Integral Recovery of Bisphenol from Poly-Bisphenol A Carbonate via Hydrolysis

Giulia Bozzano, Mario Dente, Renato Del Rosso

Politecnico di Milano, Dept. CMIC, p.zza L. Da Vinci, 32, Milano, Italy,
Giulia.Bozzano@polimi.it

Abstract

Recycling of polymer wastes is an important way for recovering valuable chemicals and reducing the environmental impact of human activities. The increasing amount of poly-bisphenol A carbonate (PC) wastes generated by electronical and electrical equipments, automobiles, optical materials, CDs and other products, makes it necessary to develop processes for their recycling. This paper investigates a recovery process based on the hydrolysis of PC with subcritical liquid water. It has been performed at 260-300 °C in a high pressure batch reactor. The effect of temperature, pressure, residence time and reaction conditions has been investigated. The products have been quantitatively and qualitatively characterized showing a substantially integral recovery of the monomer.

Keywords: Bisphenol-A, hydrolysis, polycarbonate-recycling, monomer recovery

1. Introduction

According to the research `Polycarbonate: 2009 World Market Outlook and Forecast` (http://mcgroup.co.uk/researches/P/08900/Polycarbonate) the polycarbonate (PC) market has been in bad conditions in recent times. The optical media segment shrank as a result of lower demand for CDs and DVDs. A health concern raised over Bisphenol A has influenced the demand as well. PC could disappear from the food and beverage container market in the future. However, thanks to its main strength - versatility as a manufacturing material for many different products - PC market would eventually recover. Demand would catch up with production capacity and the market would strengthen. Automotive glazing applications for PC are expected to widen. Global demand is predicted to grow of 6-7% annually over the next five years, driven by Asia and China in particular (2.4 Mt were produced in 2004). This fact suggest that PC recycling is covering more and more importance.

Thermal degradation (= pyrolysis) is a possible alternative for polymer recycling (Bok Nam Jang and C. A. Wilkie, 2004), however its low selectivity and the condensation reactions among aromatic rings makes it a not particularly suitable process, because it reduces the yield of the monomer and rises that of by-products. Several studies of possible depolymerization processes have been reported in the literature for decomposing PC to its original monomer, Bisphenol A (BPA). They are based, for instance, on the use of solvent systems (Z.Y. Pan *et al.*, 2006, R. Piñero et al., 2005) such as methylene chloride with ammonia, phenol in combination with an alkali catalyst, or via trans-esterification (alcoholysis) in supercritical or near critical conditions. These processes can require a complicate product separation in addition with environmental safety problems related to the use of highly toxic organic solvents. Other methods can be mechanical recycling (reducing however the quality of material at each cycle) or blending with other materials (Elmaghor et al.,2004). Also decomposition of

PC in subcritical and supercritical water has been taken into account by Tagaya et al. (1999) in the temperature range 230-400 °C. At 230°C a yield in BPA of 67% were obtained with the addition of Na_2CO_3 and a residence time of 24 h.

This paper investigates a recovery process based on the hydrolysis of PC with subcritical liquid water. Both pure PC and CDs wastes have been used (the driving idea for this process came from an analogy with the fats hydrolysis producing fatty acids and glycerol in the soap production field). A concerted path de-polymerization mechanism is proposed and the process kinetics is characterized and compared with the lab-scale experimental data. The residence time was in the order of 30 minutes at about 270°C. The results show that this process can become a valuable alternative for BPA recovery (it is substantially integrally recovered) mainly for its relative simplicity and absence of toxic agents or non-desired by-products. Similar hydrolysis mechanisms can take place in other fields, like for instance the production of bio-oils from biomasses.

2. Experimental

2.1 Materials

The reagent grade Poly(bisphenol A carbonate) (average M_W 64000) [CAS 25037-45-0], Bisphenol A (>99%) [CAS 80-05-7], Phenol (>99%) [CAS108-95-2] and other chemical used were all purchased from Sigma-Aldrich. Water was twice distilled. Commercial Recordable Compact Discs crashed for de-polymerization test were Verbatim Datalife CD-R.

2.2 Apparatus and methods

Depolymerisation tests were carried out in a 316 stainless steel tubular batch reactor (ID 7.8 mm L 150 mm) internal volume 7.167 mL. The typical de-polymerization test was carried out weighting the empty reactor, then charging it with 1g of PBAC and 1g of water. The reactor, exactly weighted after charge, is put into a laboratory fan assisted furnace (Heraeus M110 electronic), preheated at the temperature set point, over a support long the symmetry axis of the oven. A thermocouple fixed on the external wall of the reactor was used for measuring temperature level. Both horizontal and vertical disposition were studied. At the end of the experiment the reactor recovered from the oven, rapidly cooled and newly weighted at T_{amb}. Not more then ±0.2mg of weight difference from initial and final total weight was accepted as a probe of no spill and good capping for the test, otherwise the test is repeated. The reactor was unhead and the evolved gas collected and GC analyzed. The degassed reactor was finally weighted and the difference was assumed as CO_2 produced. The yield was evaluated from this latter information. A Mettler Toledo analytical precision balance (model B154-S) was employed. The liquid and solid internal material was discharged, identified and quantified by FTIR and GC.

3. Some analytical aspects

The solid sample was characterized by Shimadzu FTIR: IRAffinity-1 in the 500 cm^{-1} 4000 cm^{-1} range (KBr disc).

Gas chromatographic analyses of solid products were carried out (methanol as solvent) on a GC- FID HP 6890plus with SLB-5ms capillary SUPELCO column (30m length, 1μL injection volume, split ratio 1:10, Helium carrier 1.4 $mLmin^{-1}$ constant flow), the temperature was held at 65°C for the first 2 min, then increased at 255°C at a heating rate of 10°$Cmin^{-1}$, and kept at this temperature for 15 min. The main products were identified and quantified by comparing the retention time with standard compound.
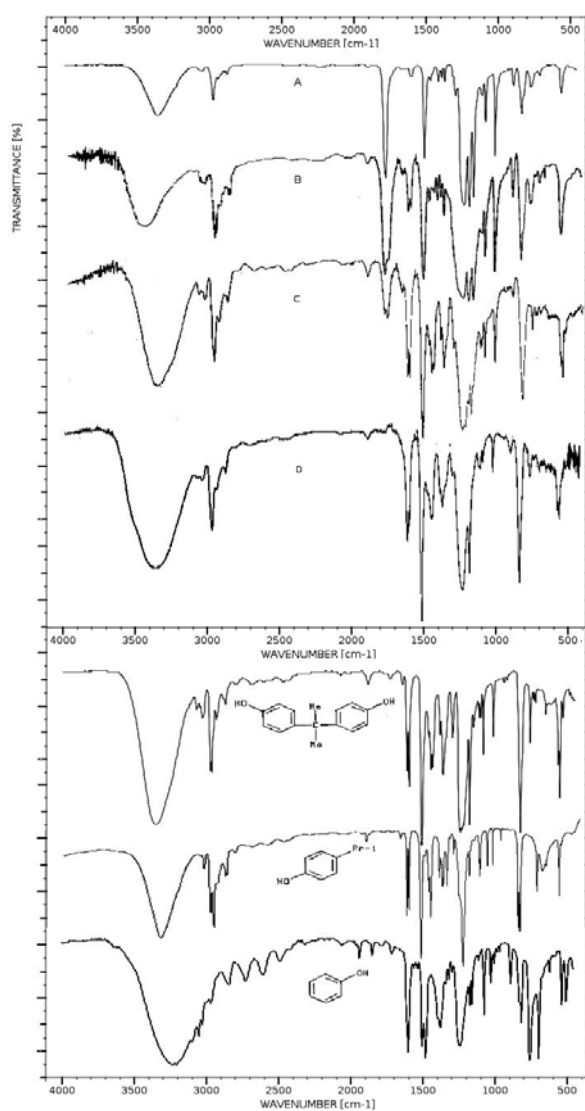
Figure 1: FT transmittance plot showing the FTIR analyses of commercial Poly(bisphenol A carbonate) (A) starting material obtained from a CD crash, a low conversion depolymerization solid discharged material (B), a medium conversion material (C), and a solid material discharged at 100% depolymerization conversion (D). The FTIR spectra of reagent grade Bisphenol A, iso-Propyl-Phenol, and Phenol are also reported.

Table 1. Retention time for products in Fig. 2

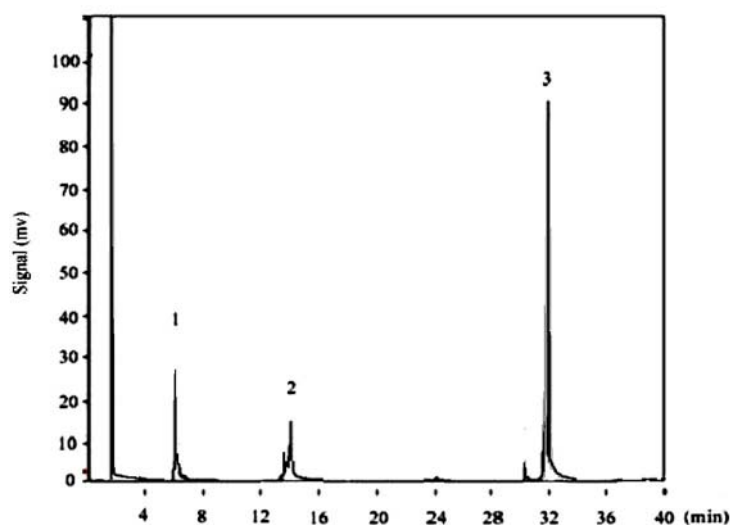| Peak number	Compound	Retention time (min)
1	Phenol	6.08
2	4-*iso*Propenylphenol	14.02
3	Bisphenol A	31.48

Figure 2: Typical gas chromatographic analysis of a solid product discharged at the end of a test at high conversion (Bisphenol A is the main product)

Gas chromatographic analyses of evolved gas were performed with a column 0.53 mm ID, Molecular Sieve 5A as stationary phase, GC-T as detector, isothermal at T=25°C. Permanent gases and CO_2 were detected.

4. Results and discussions

The experimental tests on hydrolysis have been performed at reactor temperature ranging from about 240°C to 290°C. Different residence times have been adopted. The hydrolysis yields have been evaluated by means both of the amount of CO_2 released and of the weight loss of the sample (of course the last one can be comprehensive of all the gaseous matter produced). In the preliminary tests it has been verified that the horizontal position of the reactor inside the furnace is largely preferred for enhancing the surface of contact between the molten polymer and the water. Fig.3 reports the conversions to CO2 for the different times and temperatures that have been tested; at the lowest temperature, one test has been performed for a time as long as about 6 h. There is an apparent induction time in the curves but can be essentially attributed to the heating time of the reactor. However in order to understand whether some autocatalytic effect due to the products, also some tests have been performed by adding to the feed small amounts of BPA: practically no effect has been observed.

All the reported data can be used for deducing the reaction rate and as a consequence, the activation energy of the reaction. Fig.4 reports the experimental reaction rate constant as a function of 1/T. As usually this diagram allows to deduce the activation energy of the reaction. The experimental data gives an activation energy $E_{act} \cong 80000$ kJ/kmole, a value that is very close to the theory prediction reported in the following. The gases produced (by hydrolysis) at the end of the process was evaluated not only in terms of their quantitative amount but, periodically, they were analyzed by GC for

evaluating their composition. All the analyses was clearly showing that N_2/O_2 ratio, in the effluent gases, was practically the same of the original air inclusion.
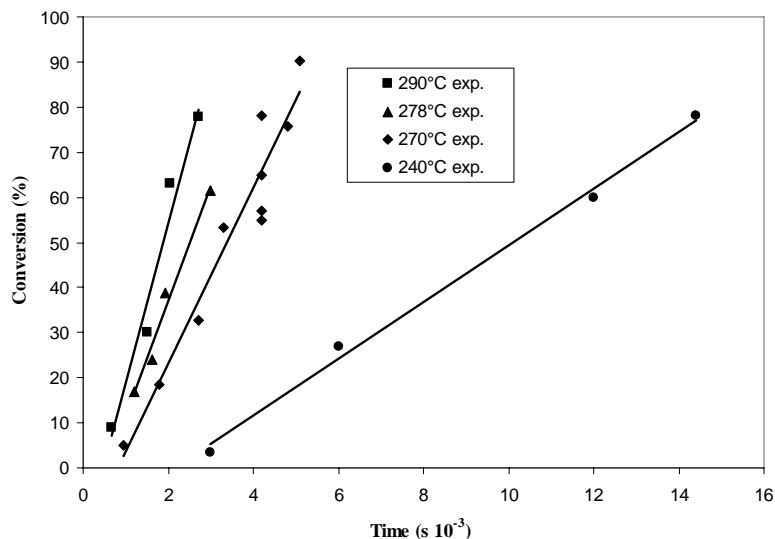

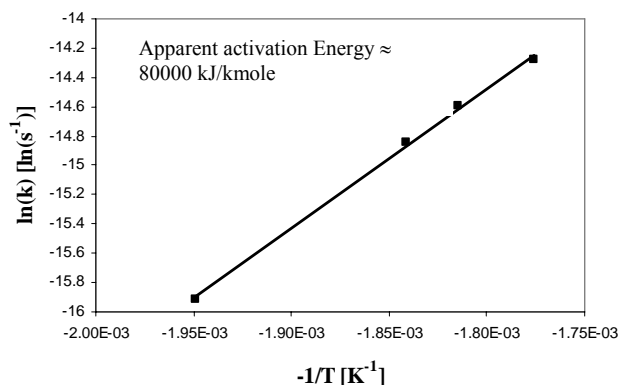Figure 3: Yield of BPA versus time at different temperatures

Figure 4: Exp. kinetic constant (s^{-1}) as a function of $-1/T$ (K^{-1}).

It simply means that no oxidative degradation is taking place during hydrolysis. Indirectly, the last considerations are also showing that practically no extra components are entering during the periodical filling and emptying of the reactor. In the tests with the highest de-polymerization degree (for instance that at 290°C with a conversion of about 80%) the products obtained after cooling at the end of the experiments have been analyzed after mixing with methanol. The results show always a high purity in Bisphenol A with respect to phenol, propyl and propylidenphenols. This means that the polymer decomposition through breakage of the bond C aromatic-C isopropylic is negligible in practice and that this kind of mechanism doesn't occurs in the selected operating conditions also in obtaining the bisphenol A. The FTIR analysis on the reaction residue after evaporation of the water shows a spectrum that is coherent with a

progressive hydrolysis of the carbonate bond, without appreciable evidences of de-alkylations with formation of terminal propylidene and phenol. The IR spectra at higher conversion are practically superposable with those of standard Bisphenol A. On the basis of these evidences, it can be affirmed that the de-polymerization reaction proceeds in the condensed phase and at the assigned conditions regenerate the monomer. It has to be pointed out that after the tests at highest conversion level the melting point of the product resulted over 145-150 °C: this is another good indication of the quasi-purity of the obtained raw monomer.

Regarding the use of pure PC or coming from the crashing of CD, essentially no difference has been observed.

5. Kinetic Mechanism

PC hydrolysis is mainly dominated by a six center concerted path reaction mechanism. By analogy with other reactions following the same mechanism involving hydrolysis of esters the kinetic constant suggested for each reaction act is:

$$k = 10^9 \exp\left(-\frac{10000}{T}\right) \qquad \text{l/mole/s} \tag{1}$$

Of course the radical path becomes more important at higher temperatures (and in practical absence of water). In our conditions water becomes a powerful reactant. It has also to be pointed out that water at nearly critical conditions has an increased solubility. Formally the reaction can be schematized as shown in equation (2):

$$PC + H_2O \Rightarrow BPA + CO_2 \tag{2}$$

6. Conclusions

In this paper the experimental results of the hydrolysis of poly-bisphenol A carbonate (PC) with water, performed both with pure and waste polymer, have been presented. The adopted temperature levels are below the critical point. The tests have shown the feasibility of the process that in these conditions in mainly based on real molecular concerted path reactions. Secondary reactions, requiring a radical path and leading to parasitic by-products, are characterized by quite higher apparent activation energy and require higher temperature levels than those here adopted. For this reason in the presented process they are negligible.

References

Bok Nam Jang and C. A. Wilkie, 2004, *A TGA/FTIR and mass spectral study on the thermal degradation of bisphenol A polycarbonate*, Polymer Degradation and Stability, 86, 419-430

Elmaghor F., Zhang L., Fan R., LI H.. 2004, *Recycling of polycarbonate by blending with maliec anhydride grafted* ABS, Polymer; 45; 6719-24

Pan Zhi Yan; Zhen Bao, Ying Xu Chen, 2006, *Depolymerization of Poly(bisphenol A carbonate) in Subcritical and Supercritical Toluene*, Chinese Chemical Lewtters, 17, 545-548

Piñero R. Juan Garcia and Maria Josè Cocero, 2005, *Chemical recycling of polycarbonate in a semi-continuous lab-plant. A green route with methanol and methanol-water mixtures*. Green Chem., 7, 380-387

Tagaya H., Katoh K., Kadokawa J., Chiba K., 1999, *Decomposition of polycarbonate in subcritical and supercritical water,* Polymer Degradation and Stability, 64, 289-292

Separation of butanol from acetone-butanol-ethanol fermentation by a hybrid extraction-distillation process

Korbinian Kraemer, Andreas Harwardt, Rob Bronneberg, Wolfgang Marquardt

Aachener Verfahrenstechnik, RWTH Aachen University, 52056 Aachen , Germany

Abstract

The alternative fuel butanol can be produced via acetone-butanol-ethanol (ABE) fermentation from renewable resources, i.e. biomass. Expensive feedstocks and the high costs for the separation of ABE from the dilute fermentation broth in the downstream processing have so far prohibited the industrial-scale production of bio-butanol. The low productivities and butanol yields of ABE batch fermentation can be increased by continuous fermentation with cell recycle and integrated product removal. In order to facilitate an effective and energy-efficient product removal, we suggest to apply a hybrid extraction-distillation process with ABE extraction in an external column. The removal of ABE outside the fermenter in an extraction column is favored from an operational point of view. By means of computer-aided molecular design (CAMD), mesitylene has been identified as a new solvent for ABE extraction from the fermentation broth. The solvent properties of mesitylene are compared to those of oleyl alcohol, which is the most common solvent for ABE extraction. Subsequently, we propose a hybrid extraction-distillation downstream process for product removal and purification. It is shown that the specific energy demand of this process is significantly lower when mesitylene is used as extraction solvent instead of oleyl alcohol.

Keywords: butanol, extraction, solvent, biofuels, CAMD

1. Introduction

Butanol has been identified as a possible fuel from renewable resources. Compared to ethanol, butanol offers several advantages as a bio-fuel such as higher energy content, lower vapor pressure, and lower hygroscopy [1]. Bio-butanol is produced via ABE fermentation from renewable feedstocks using *Clostridium acetobutylicum* or *C. beijerinckii* in anaerobic conditions. ABE fermentation ranked second only to ethanol fermentation in the first part of the 20th century, but disappeared in the second part due to the rise of the petrochemical industry. With the depletion of fossil fuels ABE fermentation becomes interesting again. The main challenges which need to be tackled in order to make ABE fermentation economically viable are [1,2] (i) expensive feedstocks, (ii) high product inhibition especially by butanol (typically 20 g/L ABE with a mass ratio of 3:6:1 is achievable), (iii) low productivity (up to 0.6 g L^{-1} h^{-1}) and ABE yields (0.3) in batch fermentation and (iv) expensive downstream processing.

Efforts are being made to use agricultural residues and energy crops such a switchgrass to reduce the cost of feedstock (i), since the butanol-producing cultures are able to catabolize a wide variety of carbohydrates [1]. To address product inhibition (ii), hyper-butanol-producing strains were developed, including *C. beijerinckii* BA101 which produces ABE up to 33 g/L with a 50% productivity threshold at about 12.5 g/L butanol [3]. Genetic engineering of butanol producing strains should allow for even lower

product inhibition, enhanced productivity and butanol yield (iii) in the future [4]. Aside from advanced butanol producing strains, productivity and yield has also been improved by continuous fermentation processes with cell recycle membrane reactors, immobilized cells reactors or packed bed reactors [5,6,7]. Continuous fermentation processes enable the use of concentrated sugar solutions, decrease product inhibition by integrated product removal, and lower the cost of waste water treatment. ABE productivities of up to 15.8 g L^{-1} h^{-1} have been achieved in immobilized cell reactors [6]. Various authors also report improved productivities by staged fermentation processes in effect cascades where the fermentation conditions are adapted to the respective cell stadium [8,9].

Despite the accomplished advances of the ABE fermentation, the expensive product removal from the dilute fermentation broth (iv) still prohibits the industrial production of bio-butanol. Since butanol has a higher boiling point than water, pure distillation processes suffer from a high energy demand. A variety of alternative hybrid separation processes have therefore been proposed: These are based on gas stripping, liquid-liquid extraction, pervaporation, perstraction, and adsorption [5,10]. Besides reduced energy demands for butanol separation, these methods also offer the advantage that they can be applied inside the fermenter to decrease product inhibition. [10] suggest that adsorption and extraction combined with distillation are the most energy efficient alternatives. According to [5], hybrid processes with pervaporation or extraction are most attractive for integrated product removal. [9] prefer extraction in combination with distillation since these techniques are conventional unit operations. Although these innovative hybrid processes constitute a leap in energy efficiency from the pure distillation process, the specific energy demand is still considerably higher than 10% of the energy content of butanol, which has been stated as the target for energy efficiency [11].

Since the high energy cost for ABE removal is still the bottleneck in industrial bio-butanol production, it is the scope of this work to explore possibilities to further lower the energy demand. Considering the above mentioned promising experiences of other authors and the preference in industry for established unit operations, we chose to study the energy savings potential of hybrid extraction-distillation processes.

2. Solvent screening for extraction of solvents in external column

Most publications on ABE removal via liquid-liquid extraction study extractive batch fermentation. Here, the fermentation products are removed in situ, i.e. inside the fermenter, into an organic solvent phase. Various authors conduct extensive solvent screenings [11,12,13,14,15]. Suitable solvents are selected based on the following criteria: non-toxicity to cells, immiscibility with water, high distribution coefficient towards butanol, low viscosity and different density as water, commercially available at low cost. Two main groups of solvents were identified: Alcohols and alkanes. While alcohols exhibit high distribution coefficients (D>5) towards butanol, they have relatively low selectivities ($D_{butanol}/D_{water}$ < 350). Alkanes, on the other hand, offer large selectivities (2500-4300) but suffer from low distribution coefficients (D<0.5).

Many authors choose oleyl alcohol as extracting agent due to its non-toxicity towards the microorganisms and its relatively high distribution coefficient for butanol (D=3.8). Oleyl alcohol has therefore become the solvent of choice for extractive fermentation and many authors report enhanced cell productivity and butanol yields for extractive fermentations with the help of oleyl alcohol [12,17,19,20]. However, most studies were carried out as batch fermentations on a lab-scale level. Under these circumstances, some disadvantages of oleyl alcohol for continuous large-scale production have little effect: The high boiling point (360°C) hinders a separation of the product from oleyl alcohol

Separation of butanol from acetone-butanol-ethanol fermentation by a hybrid extraction-distillation process

via distillation in a large-scale process. In addition, the low distribution coefficient for acetone (D=0.34) requires that a large amount of solvent is used in order to prevent an accumulation of acetone in the fermentation. The required amount of solvent is therefore determined by the removal rate of byproducts rather than butanol itself [16].

Some authors also indicate that extractive fermentation with in situ product removal may not be suitable for large-scale production due to various reasons:

· difficult process control [11]

· slow mass transfer into solvent phase (slower than butanol production) [17,18]

· formation of emulsions through agitation [5,14,18,19]

· cell inhibition by solvent (interface toxicity) and loss of cells at interface [19]

· physical shielding by attraction of cells to interface: real distribution coefficients in experiments lower than in experiments without cells [20]

· precipitates carry water into the solvent phase [14]

For these reasons, external product removal in an extraction column with recycle of solvent-lean broth seems to be better suited for large-scale production of bio-butanol [11,17,18]. When the cells are retained in the fermenter by immobilization or ultrafiltration, powerful but toxic solvents can be used in an external extraction column as long as their solubility in water is low. Hence, we performed a solvent screening where we did not exclude toxic solvents but emphasized a low solubility in water and paid attention to operational constraints like a boiling point which allows for an economic distillation to remove the products from the solvent. In addition, we emphasized high distribution coefficients not only for butanol but also for acetone and favored an optimal balance between distribution coefficient and selectivity.

The solvent screening was performed with the help of the software package ICAS [21], which uses a generate and test approach to screen molecules. First, thresholds for the desired properties are specified by the user. Then, meaningful molecules are generated by Computer-Aided Molecular design (CAMD) and tested for the desired properties based on thermodynamic group contribution methods, i.e. UNIFAC. These tested molecules can then be ranked and checked against a database to exclude non-existing molecules. The best solvent properties were predicted for methylbenzenes with more than three methyl groups, i.e. tri-, tetra-, and pentamethylbenzene. From our knowledge of the literature, these solvents were never considered in solvent screenings for ABE removal from fermentation broth before. This is probably due to the expected toxicity to the cells when applied in situ and the relatively low distribution coefficient at room temperature compared to fatty alcohols. We excluded pentamethylbenzene because of its melting point at around 50°C. Tri- and tetramethylbenzene exhibit similar properties as solvents. We chose to study 1,3,5-trimethylbenzene (mesitylene) in more detail, since it is most commonly used as a solvent in industry and research.

To validate the properties predicted in ICAS by the UNIFAC group contribution method, we measured the distribution coefficients for acetone, butanol, and ethanol in systems of water and mesitylene. We also determined the solubility of mesitylene in water experimentally. Table 1 lists the solvent properties of mesitylene (UNIFAC and measured) and gives a comparison to oleyl alcohol, which is the common solvent choice in literature. The first column contains the properties for the new solvent mesitylene predicted by UNIFAC, which led to the selection in the solving screening procedure. Note that the distribution coefficients D for butanol and acetone are predicted to be very similar. This is beneficial since not only butanol needs to be removed from the broth but also a considerable amount of acetone. Mesitylene is also predicted to exhibit a very large selectivity ($D_{butanol} / D_{water}$) and low solubilities for solvent in water and vice versa. The distribution coefficients which we measured at 25°C are considerably lower than

the predicted coefficients, particularly for acetone and ethanol. However, at 80°C we measured significantly higher distribution coefficients. Note that ethanol is the least inhibitory product and, therefore, the relatively low distribution coefficient for ethanol should not be detrimental. Table 1 shows that the UNIFAC predictions are rather inaccurate. We have therefore used the measured data in the simulation in Section 3.

Table 1. Comparison of solvent properties.

	mesitylene			oleyl alcohol
	UNIFAC 25°C	measured 25°C	measured 80°C	[15] 30°C
D butanol (kg/kg)	1.3	0.76	2.2	3.8
D acetone (kg/kg)	1.4	0.43	0.83	0.34
D ethanol (kg/kg)	0.14	0.03	0.1	0.28
selectivity	7620	1650	4760	330
solubility water in solvent (wt%)	0.017	0.046	0.112	1.14
solubility solvent in water (wt%)	0.0027	5.2e^{-3} [22]		0.0019
viscosity (mPa s)		0.66		26
melting / boiling point (°C)	-45 / 165			13-19 / 330-360

Oleyl alcohol offers an even higher distribution coefficient for butanol than mesitylene. Nevertheless, a larger amount of oleyl alcohol needs to be used for extraction compared to mesitylene at 25°C and at 80°C, since the distribution coefficient for acetone is considerably lower. Furthermore, oleyl alcohol exhibits a substantially lower selectivity than mesitylene due to the higher solubility for water. This results in noticeable amounts of water in the organic phase, which raises the cost for the downstream purification. In the following, additional advantages of the solvent mesitylene are noted.

Whereas oleyl alcohol removes the valuable intermediates butyric acid (D=3.7) and acetic acid (D=0.35) from the broth [16], mesitylene leaves these intermediates (D = 0.58 and 0.06, respectively) in the broth such that they can be catabolized in the fermenter. Groot et al. [5,14] report fouling inside the extraction column when they use oleyl alcohol as solvent due to its non-toxicity. The anticipated toxicity of mesitylene, however, will presumably reduce the issues with fouling. Both solvents have a density that allows for an efficient phase separation (0.85 g/cm^3), but the higher viscosity of oleyl alcohol results in a diffusion coefficient of only 1.1e-10 m^2/s [5] which will lead to a large height of the extraction stages. The melting and boiling points also favor mesitylene as solvent. The high boiling point of oleyl alcohol prohibits a separation of the products in a simple distillation column at normal pressure. The melting point just below room temperature can complicate large-scale production as well.

3. Simulation of hybrid extraction-distillation downstream process

As a consequence of the above mentioned favorable solvent properties of mesitylene, it is expected that the use of mesitylene as solvent in hybrid extraction-distillation downstream processes can significantly reduce the separation costs. In order to quantify the energy savings, we have modeled the entire downstream processes for the solvents mesitylene and oleyl alcohol and a pure distillation process in ASPEN PLUS.

We assume a broth flowrate of 1 m^3/h with a butanol concentration of 8 g/L. This concentration is below the threshold for butanol inhibition and has been reached in

Separation of butanol from acetone-butanol-ethanol fermentation by a hybrid extraction-distillation process

continuous fermentations in the literature [6,18]. The concentration of acetone in the broth (first column in Table 2) is determined from a mass balance around the extraction column assuming that the mass ratio of butanol and acetone in the saturated solvent stream is 2:1, which is consistent with the ratio they are produced by the cells in the fermentation [24]. For oleyl alcohol, the total concentration of ABE would then exceed 25 g/L at the minimal solvent flowrate for butanol removal. Therefore, oleyl alcohol demands a higher solvent flow than necessary for butanol removal.
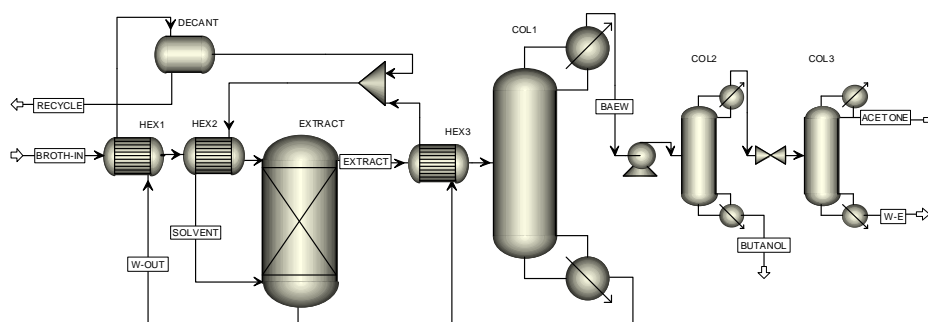

Fig.1. Process flowsheet for hybrid extraction-distillation process.

The flowsheet of the process with the new solvent mesitylene is shown in Figure 1. The fermentation broth is passed through a filter (not shown), heated by the recycles to 80°C, and given into the extraction column. The extraction is modeled with the measured distribution coefficients and solubilities at 80°C. The extraction column is assumed to consist of 10 equilibrium stages. 87.5 % of the butanol is extracted such that the product-lean fermentation broth leaves the column with a butanol content of 1 g/L. The heat of the product-lean fermentation broth is recuperated in heat exchanger Hex1. We assume a temperature difference of 2°C for this heat exchanger. While this may seem low, it results in a reasonable heat exchange area of 29 m^2. Before the cool broth is recycled back to the fermenter it is sent into a decanter where remains of mesitylene are recovered at lower temperatures. The saturated solvent stream is preheated by the solvent recycle in heat exchanger Hex3 and purified from the fermentation products in distillation column Col1. After passing through heat exchangers Hex3 and Hex2, the solvent recycle is fed into the extraction column at 80°C. The distillate product of Col1, which contains ABE and remains of water, is further split up into its pure components in columns Col2 and Col3. Note that Col3 operates at a pressure of 0.7 bar where this separation can be performed more efficiently. The vapor-liquid-equilibrium in the distillation columns is modeled by the UNIFAC (Col1) or the NRTL model (Col2 and Col3) with parameters from ASPEN. The resulting energy demands for the solvents mesitylene and oleyl alcohol and the pure distillation process are shown in Table 2.

The process with the new solvent mesitylene demands significantly less energy than both the process with solvent oleyl alcohol and the pure distillation process. The main reasons for the relatively large energy demands for oleyl alcohol are the higher solvent flowrate due to a lower distribution coefficient for acetone and the large content of water in the distillate of Col1 (45 wt%). It still needs to be determined in further experiments, whether the nutrients in the broth are extracted into the solvent in considerable amounts. In addition, possible inhibition of the cells by traces of mesitylene in the fermentation broth needs to be tested in experiments.

Table 2. Comparison of energy demands (energy content of butanol: 36 MJ/kg).

	conc. in broth (g/L ABE)	solvent flow	energy demand			
			Col1	Col 2	Col 3	total
solvent mesitylene	10/8/5	383 kg/h	9.1 kW	0.4 kW	1 kW	5.7 MJ/kg butanol
solvent oleyl alcohol	12/8/5	738 kg/h	22 kW	remaining columns 6.1 kW		15 MJ/kg butanol
pure distillation	4/8/2		4 columns 38.2 kW			19.4 MJ/kg butanol

4. Conclusions

The new solvent mesitylene for the removal of fermentation products of continuous ABE fermentation compares favorably to the solvent oleyl alcohol, which is commonly used for ABE extraction. We have therefore proposed a hybrid extraction-distillation downstream process, where the fermentation products are removed from the broth with the help of mesitylene in an external extraction column. The entire downstream process including product purification exhibits a specific energy demand of 5.7 MJ/kg butanol produced, which is 16% of the energy content of butanol. This is a significant reduction compared to the extraction with oleyl alcohol (15 MJ/kg butanol), compared to the pure distillation process (19.4 MJ/kg), and compared to the most energy efficient process reported in the literature (8.2 MJ/kg via adsorption-distillation [10,11]). Note that pure acetone, which is retrieved in a weight ratio of 1:2 (A:B), is a valuable product as well. Future research will be directed towards a rigorous optimization of the hybrid process, possibly bringing further down the energy demand and taking into account capital costs. Financial support by the cluster of excellence 'Tailor-Made Fuels from Biomass' is gratefully acknowledged.

References

[1] N. Qureshi and T.C. Ezeji, Biofuels Bioprod. Bior., 2 (2008) 319
[2] P. Dürre, Ann. N.Y. Acad. Sci., 1125 (2008) 353
[3] N. Qureshi and H.P. Blaschek, J. Ind. Microbiol. Biot., 27 (2001) 287
[4] D.R. Woods, Trends Biotechnol., 13 (1995) 259
[5] W.J. Groot, R.G.J.M. van der Lans and K.Ch.A.M. Luyben, Process Biochem., 27 (1992) 61
[6] N. Qureshi, J. Schripsema, J. Lienhardt, H.P. Blaschek, World J. Microb. Biot., 16 (2000) 377
[7] W.C. Huang, D.E. Ramey and S.T. Yang, Appl. Biochem. Biotech., 113 (2004) 887
[8] A.S. Afschar, H. Biebl, K. Schaller and K. Schügerl, Appl. Microbiol. Biot., 22 (1985) 394
[9] J. Liu and L.T. Fan, P. Seib, F. Friedler and B. Bertok, Biotechnol. Progr., 20 (2004) 1518
[10] N. Qureshi, S. Hughes, I.S. Maddox and M.A. Cotta, Bioproc. Biosyst. Eng., 27 (2005) 215
[11] A. Oudshoorn, L.A.M. van der Wielen, A.J.J. Straathof, Ind Eng Chem Res, 48 (2009) 7325
[12] S. Ishii, M. Taya and T. Kobayashi, J. Chem. Eng, Jpn., 18 (1985) 125
[13] S.R. Roffler, H.W. Blanch and C.R. Wilke, Bioproces Engineering 2 (1987) 1
[14] W.J. Groot et al., Bioprocess Engineering, 5 (1990) 203
[15] M. Matsumura and H. Kataoka, Biotechnology and Bioengineering, 30 (1987) 887
[16] M. Matsumura, H. Kataoka, M. Sueki, K. Araki, Bioprocess Engineering, 3 (1988) 93
[17] S.R. Roffler, H.W. Blanch and C.R. Wilke, Bioprocess Engineering, 2 (1987) 181
[18] S.R. Roffler, H.W. Blanch and C.R. Wilke, Biotechnol. and Bioeng., 31 (1988) 135
[19] N. Qureshi, I.S. Maddox and A. Friedl, Biotechnology Progress, 8 (1992) 382
[20] B.H. Davison and J.E. Thompson, Appl. Biochem. Biotech., 39/40 (1993) 415
[21] P.M. Harper and R. Gani, Comput. Chem. Eng., 24 (2000) 677
[22] L. Zou, G. Yang, B. Han, R. Liu and H. Yan, Science in China, 42 (1999) 400
[23] D.T. Jones and D.R. Woods, Microbiological Reviews, 50 (1986) 484

Effect of the switch strategy on the performance and stability of reactor networks for methanol synthesis

Erasmo Mancusi[a], Pietro Altimari [b], Silvestro Crescitelli[c] and Lucia Russo[d]

[a]Facoltà di Ingegneria, Università del Sannio, Piazza Roma , 82100, Benevento, Italy, mancusi@unisannio.it

[b]Dipartimento di Ingegneria Chimica Alimentare Università di Salerno, Via Ponte Don Melillo, 84084, Fisciano (SA), Italy, paltimar@unina.it

[c]dDipartimento d'Ingegneria Chimica Università "Federico II" Piazzale Tecchio 80, I-80125 Napoli, Italy, crescit@unina.it

[d]CNR, Istituto delle Ricerche sulla Combustione, P.le Tecchio 80, Napoli, 80125, Italy, Italy, lucrusso@unina.it

Abstract

A network of catalytic reactors with a periodic switching of the feed and discharge position (also called loop reactor) is studied for reversible exothermic reactions like methanol synthesis. The aim of the study is the comparison of two different switching strategies in overcoming the conversion limits imposed by the thermodynamic equilibrium. The first strategy, that is the only considered in previous works, consists of changing at each switch time the feed/discharge positions, so that the first reactor of the NTW is moved to the last place. The second strategy consists in moving the last reactor in the first place each switch time. We show through numerical simulations that the second forcing strategy is more convenient in terms of yield and methanol conversion, because it is able to create a conversion/temperature path closer to the optimal one.

Keywords: Periodically forced reactors, process intensification, equilibrium limited reactions, spatio-temporal patterns , loop reactors.

1. Introduction

As result of the increasing demand for highly efficient and size-contained process facilities, several efforts have been carried out towards the development of process intensification technologies. In this framework, periodically forced reactors have been recognized to offer a powerful solution to achieve autothermal operation of exothermic catalytic processes. The most representative example of such operation-mode is provided by the so-called reverse-flow-reactor (RFR) (Matros, 1989), where a valve system allows the period inversion of the reactant flow through the catalytic reactor. This solution has found a very successfully application in the purification of industrial off gas through autothermal catalytic combustion even with mixtures with a very low adiabatic temperature rise. Extensive numerical simulations have been reported

demonstrating the feasibility of this operation-mode for both irreversible and equilibrium limited reactions (Matros and Bunimovich, 1996).

However, the RFR has some drawbacks related to the inversion of the flow. The most important is the washout effect, that this the lost of part of reactants just after the flow inversion. To overcome the limits of the RFRs, a network of catalytic reactors in series equipped with a valve system allowing to cyclically modify the feed position has been proposed (Matros, 1989; Haynes and Caram, 1994). In this system, autothermal operation is achieved by periodically switching the feed position so as to achieve a cyclic permutation of the reactor sequence while keeping constant the flow direction throughout the fixed bed. Velardi and Barresi (2002) found that a network of three catalytic reactors for methanol synthesis has a higher yield respect to the RFR, but the process is more unstable as such high performances are reached in a region of very low switch time values where the system exhibits very complex dynamics. More recently, Sheinman and Sheintuch (2009) analyzed networks with an arbitrary number of reactors for equilibrium limited reaction and they again found complex dynamics for low switch time values. All these studies have investigated the performance of reactor networks where the feed position is switched to the next reactor of the sequence. When the reactors are just two there is obviously only one way to switch the feed position, but if more reactors are considered different switching strategies may be thought (Mancusi et a. 2007). Indeed, Mancusi et al. (2007) have shown that the range of stability of a network of three catalytic combustors is doubled when the feed position is switched to the previous (instead of the next) reactor of the sequence.

Here we compare two different switch strategies in a network of catalytic reactors where equilibrium limited reactions take place. In particular, we analyze the spatial profiles of a network of four catalytic reactors, where methanol synthesis is carried out, with the objective of detecting the switch strategy and the operating conditions which maximize the methanol yield and, at the same time, guarantee stable plant operation.

2. Mathematical Model

We consider a network of four identical fixed-bed adiabatic reactors where the methanol synthesis reaction takes place. The reactors network (NTW) is forced by two different switch strategies (Fig.1). In both cases the sequence of reactors is changed periodically following a cyclic permutation. For the strategy-1 the feed and discharge positions move clockwise (Fig.1, on the left), that is after the switch the reactor fed in the previous time interval is moved in the last position of the sequence. More precisely, let τ be the switch time, the reactors are fed according to the sequence 1-2-3-4 in the time range $[0, \tau]$; after the first switch, that is in the range $[\tau, 2\tau]$, the reactors are fed according to the sequence 2-3-4-1 and so on following a cyclic permutation until to obtain again the sequence 1-2-3-4. For the strategy-2 the feed and discharge positions move counterclockwise (Fig.1 on the right), that is after the switch the reactor fed in the previous time interval is moved in the second position of the sequence. More precisely, after the first switch, that is in the range $[\tau, 2\tau[$, the reactors are fed according to the sequence 4-1-2-3 and so on following a cyclic permutation until to obtain again the sequence 1-2-3-4. It is worth to note that for both the switch strategies the network is periodically forced system, that is after a period time $T=4\tau$ the system return in the initial configuration (1-2-3-4).

Effect of the switch strategy on performance and stability of reactor networks for methanol synthesis

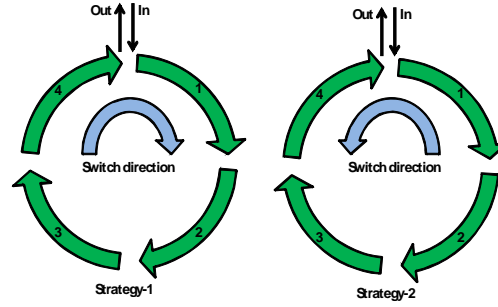

Figure1-The four catalytic reactors with two different strategies of permutation of the feed and discharge positions.

A one-dimensional pseudo-homogeneous model that takes into account mass and energy dispersive transport has been adopted for the numerical simulations. Following Sheinman and Sheintuch (2009) we assume only two reactants in the feed and that only one reaction occurs:

$$CO + 2H_2 \underset{r_{-1}}{\overset{r_1}{\square}} CH_3OH \qquad (-\Delta H) = 9.071 \cdot 10^4 \ J/mol \qquad (1)$$
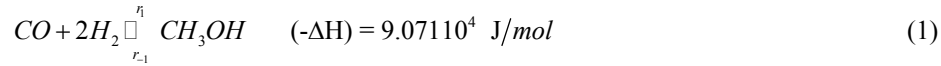
Moreover, to simplify modeling, we assume that both direct and reverse reactions are of first order with Arrhenius dependence in temperature. The transport coefficients and thermodynamic parameters are assumed to be constant. The boundary conditions are Danckwerts type at the feed and exit position. The rate equation has been reduced to a single-reactant reaction (Sheinman and Sheintuch, 2009). The dimensionless mathematical model for each reactor then reads:

$$Le \frac{\partial \vartheta_i}{\partial \bar{t}} + v \frac{\partial \vartheta_i}{\partial \xi} = \frac{1}{Pe_h} \frac{\partial^2 \vartheta_i}{\partial \xi^2} + B \ r(x_i, \vartheta_i)$$

$$\frac{\partial x_i}{\partial \bar{t}} + v \frac{\partial x_i}{\partial \xi} = \frac{1}{Pe_m} \frac{\partial^2 x_i}{\partial \xi^2} - r(x_i, \vartheta_i) \qquad (2)$$

$$r(x_i, \vartheta_i) = Da \exp\left(\frac{\vartheta_i \gamma}{\vartheta_i + \gamma} \right) \left(1 - x_i \left(1 + \psi \left(\frac{(1-\mu)\gamma^2}{\vartheta_i + \gamma} \right) \right) \right) \ (\text{ for } i=1,2,3,4)$$

In System (2), the first equation represents the heat balance while the second represents the mass balance, respectively. The index *i* identifies the reactor. The dimensionless variables and parameters are defined as:

$$\mu = \frac{E_2}{E_1}; \ \psi = \frac{A_2}{A_1}; \ \xi = \frac{z}{z_0}; \ \bar{t} = \frac{tu_0}{z_0}; \ \vartheta = \gamma \frac{T - T_0}{T}; \ x = 1 - \frac{C_{met}}{C_{met0}};$$

$$\gamma = \frac{E}{RT}; \ v_s = \frac{u}{u_0}; \ B = \frac{(-\Delta H)C_0 \gamma}{(\rho c_p)_f T_0}; \ Da = \frac{Az_0}{u_o} \exp(-\gamma); \ Le \frac{(\rho c_p)_{eff}}{(\rho c_p)_f}; \qquad (3)$$

$$Pe_h = \frac{(\rho c_p)_f z_o u_o}{k_e}; \ Pe_m = \frac{(\rho c_p)_f z_o u_o}{D_f}.$$

Table 1 Dimensionless parameters with T_{in}=100°C, T_0=200°C, u0=1 m/s, z_0=1 m (Sheinman and Sheintuch, 2009).

B=16	ϑ_{in} =-8,2	Pe_h=413	Pe_m=390	Da=0.0017
Le=29	γ =39	μ =1.4	$\psi = 3.2^8\ 29$	L=1/2

For subsequent numerical investigation, the infinite dimensional PDE system (Eqs. 5) has been reduced to a set of 400 ordinary differential equations (ODEs) by finite difference technique. A homemade software base on the robust and popular routine DLSODES libraries (Hindmarsh and Sherman, 1987) for the time integration of ODEs with sparse Jacobian matrix has been performed.

3. Results

The analysis is conducted choosing the switch time as parameter, because this is a key parameter for design, operation, and control. In fact, not only the switch time greatly affects the stability of the NTW, but it is usually used as manipulated variable in control policies.

A comparison of the temperature and methanol conversion spatial profiles along the four reactors is reported in Figure.2 for two different switch time values. As the regimes are all T-periodic, the spatial profiles vary in time and in Figure 2 are reported for the instant just before the switching.
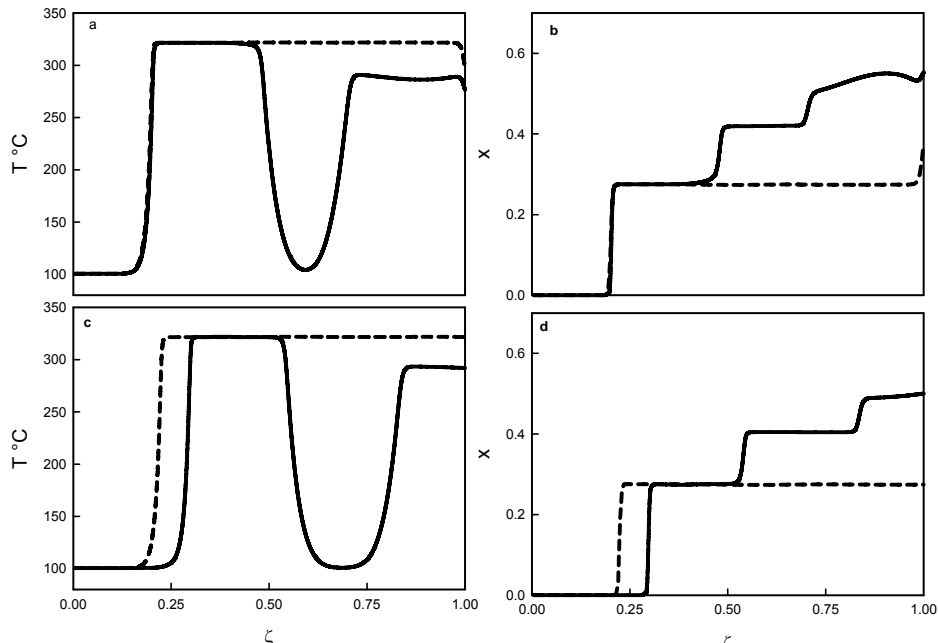

Figure-2 Temperature (a,c) and methanol conversion (b,d) spatial profiles for τ=29 (upper) and τ=34 (lower) . Dashed line: strategy-1; solid line: strategy-2.

Effect of the switch strategy on performance and stability of reactor networks for methanol synthesis

For both the switch time values, the spatial profiles for the strategy-1 exhibit just one plateau (dashed line in all Figs.2). At the entrance of the reactors sequence, the system find favourable conditions for reaction. Then, the temperature and conversion jumps reaching a plateau, where the equilibrium conversion, corresponding to the maximum temperature, is reached. For the lowest switch time value (Figs. 2(a,b)), after the plateau is reached, as the temperature decreases, the equilibrium value is shifted towards higher conversions, and more methanol is formed. For all $\tau \in$ [25, 40] the strategy-1 does not exhibit a second plateau.

For the strategy-2 the conversion jumps in three steps and it clearly shows three plateaus: in the first one methanol concentration reaches the equilibrium value, corresponding to the maximum temperature. When the temperature decreases, methanol equilibrium concentration is shifted towards higher values, but it remains constant once the temperature falls below the ignition value. Therefore, the first plateau corresponds to an equilibrium limitation, while the second one to a kinetic constraint because reaction rate is negligible below 180°C. Nevertheless, partially converted gases coming from the first three beds at a concentration below the equilibrium value, thus giving an extra push in conversion in the last reactor. Thus, thanks to these multiple plateaus, the strategy 2 allows higher conversion respect to the strategy 1.

It is worth to note that these are all T-periodic regimes corresponding to discrete travelling waves that move along the NTW with almost constant velocities and shape. This is important as efficient control policies can be applied on the basis of the front velocities (Smagina and Sheintuch, 2009).

The stability range of *T*-periodic regimes for both the strategies are determined via simulation. In Figure 3, one thousand iterates of the Poincaré map after transients that have died out, are plotted for each τ-value in the range [25, 40] (see Russo et al. 2002 for the construction of the Poincarè map of such systems). It is worth to note that, for low values of the switch time, it is observed a much richer dynamics such as multi-periodic and quasi-periodic regimes that we did not report as here we are mainly interested to regimes corresponding to discrete traveling waves.

The stability range of *T*-periodic regimes for the strategy-1 is larger than that for the strategy-2. Moreover, although the maximum temperature registered at the exit of the NTW is quite similar for both the strategies., the corresponding exit methanol conversion is bigger in the strategy -2.
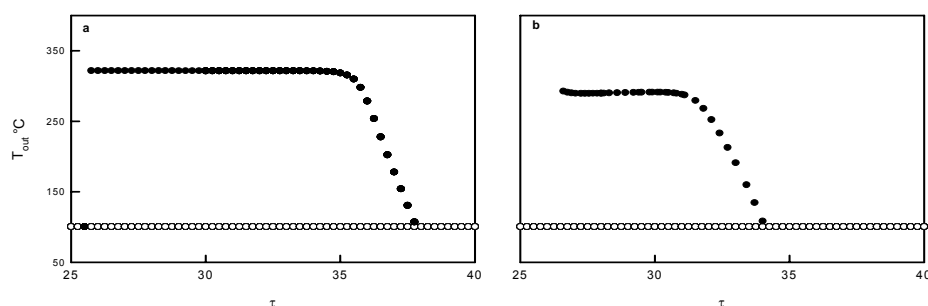

Figure-3 The asymptotic behavior for $\tau \in$ [25, 40]. The state is represented by the gas temperature at the exit of the NTW. (a) strategy-1; (b) strategy-2. Filled circle represents ignited regimes, empty one non-ignited.

This is clearly shown in Figure 4, where a comparison at the same switch time values of the time series of methanol conversion at exit is reported for both strategies.
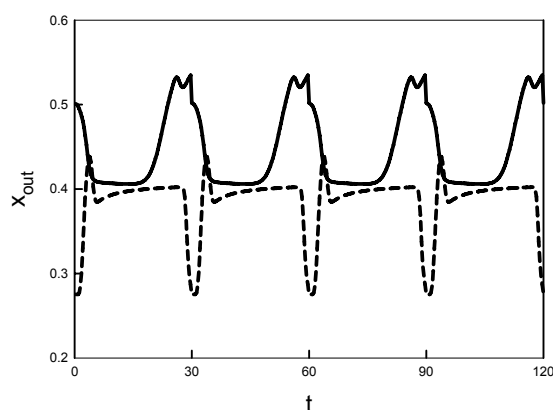

Figure-4 Methanol conversion vs. time $\tau=30$. Dashed line: strategy-1; solid line: strategy-2.

4. Conclusions

A network of reactors with periodically switching of feed and discharge positions may be forced with different switching strategies when the reactors are more than two. Here we analyzed a network of four reactors and we found for exothermic equilibrium limited reaction a new strategy more efficient respect to that previously analyzed in literature. The analysis is conducted through numerical simulation of the methanol synthesis process comparing the temperature and conversion spatial profiles of the new strategy with the old one. The new strategy gives higher methanol conversion as temperature profiles present first an increasing and then a decreasing front. This leads to temperature/conversion paths closer to the optimal one. The stability range of the T-periodic regimes for the new switching strategy is smaller respect to the old one. However, as such regimes have spatial temperature profiles which travel along the reactors with almost constant velocities, this aspect can be easily managed through a control policy based on the estimation of front velocities.

References

T.N. Haynes, H.S. Caram, 1994, *Chem. Eng. Sci.*, 49, 5465.
A. C. Hindmarsh, A.H. Sherman, A. H., 1987. LSODE, Lawrence Livermore National Laboratory, Livermore,California (http://www.netlib.org/odepack/opkd-sum)
E. Mancusi, L. Russo, P. Altimari, PL. Maffettone, S. Crescitelli S.2007, *Ind. Eng. Chem. Res.*, 46, 6510-6521.
Yu. Sh. Matros, 1989, *Catalytic process under unsteady state conditions* , Elsevier, Amsterdam.
Yu. Sh. Matros, G.A. Bunimovich, 1996, *Catal. Rev Sci. Eng.*, 38, 196.
R. Sheinman, M.Sheintuch, 2009, *Ind. Eng. Chem. Res.,*48,11, 5185.
Y. Smagina, M. Sheintuch , 2009, *J. of Proc. Cont.*, 19, 6, 954.
L.Russo,E. Mancusi , P.L.Maffettone , S. Crescitelli, 2002, *Chem. Eng. Sci.*, 57, 5065.
S.A. Velardi , A.A. Barresi, 2002, *Chem. Eng. Sci.*, 15, 2995.

Effects of fouling on performance of retrofitted heat exchanger networks; a thermo- hydraulic based analysis

Francesco Coletti[a] Sandro Macchietto[a] and Graham T. Polley[b]

[a]*Department of Chemical Engineering, Imperial College London, South Kensington Campus, SW7 2AZ U.K. s.macchietto@imperial.ac.uk*
[b]*Department of Chemical Engineering, Universidad de Guanajuato, México.*

Abstract

In refineries, fouling in crude pre-heat trains (PHTs) causes several thermal-hydraulic inefficiencies which lead to increased operating costs (from reduction in throughput and extra fuel burnt at the furnace), carbon emissions, and maintenance issues. The energy recovery performance of PHT can be severely affected over time. Such time varying effects are normally not considered in the design or retrofit of heat exchangers networks. In this paper, an existing PHT network is simulated including its fouling behaviour of over ca. two years. For this purpose, a dynamic, distributed mathematical model for shell-and-tube heat exchangers undergoing crude oil fouling (developed and validated against refinery data in previous work) is used. Three retrofit options aimed at maximizing overall heat recovery are proposed. Simulation results show that networks designs that maximize energy recovery in clean conditions (following traditional pinch rules) may not be best when fouling occurs and that a proper retrofit design must include consideration of time varying fouling effects.

Keywords: Fouling, crude oil, refinery, efficiency, network.

1. Introduction

Two thirds of the energy necessary for primary fractionation in oil refineries are recovered in the pre-heat train (PHT), an extensive network of heat exchangers (Yeap *et al.* 2004). However, PHT's efficiency is hindered by the progressive accumulation of unwanted material (fouling) on the heat exchange surfaces. Mitigating fouling is paramount for refinery operators to avoid production loss, health and safety hazards and increased fuel consumption (with related GHG emissions) at the furnace downstream the PHT. Typically, PHTs are optimised to maximise energy recovery through pinch technology or MINLP techniques. However, these techniques have several limitations:

1. They consider steady state performance of exchangers (i.e. no fouling dynamics)
2. They seek to match the hottest process streams with the crude at its highest temperature. This leads to wall temperatures being maximised. In turn fouling, which is a function of temperature and velocity, is exacerbated.
3. They rely on simple, lumped models that are limited by a number of assumptions on physical properties and temperature profiles inside the units

It was previously noted (Wilson *et al.* 2002) that failure to consider fouling behaviour at the design stage could lead to poor network performance. This is also confirmed by practical refinery experience that highly optimised networks will not work due to fouling dynamics (personal communication). The objective of this paper is to use high fidelity thermo-hydraulic simulations that capture complex dynamic interactions in
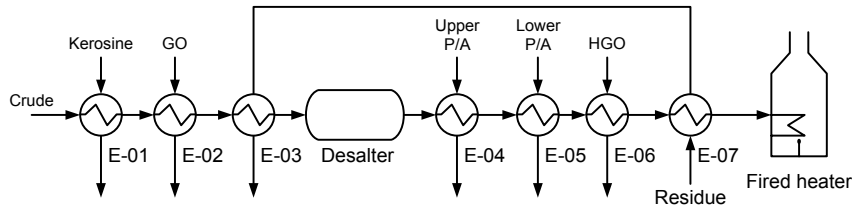
Figure 1 Pre-heat train structure in the base-case configuration C1.

the network to analyse alternative retrofit options of the network. For this purpose, a mathematical model by Coletti and Macchietto (2009), is used here to assess the fouling behaviour of an existing network. The model is dynamic, distributed and takes into account the exchanger geometry and configuration, the variation of temperature, velocity, physical properties and fouling rate along the length of each unit. It also accounts for localized fouling growth as a function of process conditions through the Ebert-Panchal model (Panchal *et al.* 1999):

$$\frac{dR_{f,n}}{dt} = \alpha \, Re^{-0.66} \, Pr^{-0.33} \exp\left(\frac{-E}{RT_{f,n}}\right) - \gamma\tau \tag{1}$$

The model also captures the interactions between the fouling layer and the fluid-dynamics by solving a moving boundary problem. Simulation results for industrial single units indicated that output temperatures are calculated within ±2% error even when the model is tested for its predictive capabilities over 1-2 years.

To illustrate the main point of the paper, first we build the network model based on an industrial case study of a small refinery. The existing network is used as base case. Parameters that characterize the fouling behaviour in each unit are estimated so as to fit this base case. Once the fouling behaviour is captured, the same values are used to assess alternative retrofit options, with the goal of increasing overall energy recovery. Three network structures are proposed. The first follows pinch rules to maximize energy recovery based on clean exchanger performance. The remaining two seek to improve the energy recovery while also minimizing the fouling behaviour. The results for the proposed retrofit designs are analyzed with respect to the coil inlet temperature (CIT) achieved and the heat load at the furnace over the entire period.

2. Case study

In order to protect proprietary information, the heat and mass balance for the existing network (Figure 1) has been adjusted and some features of the plant (reported in Table 1) have been changed. Nonetheless, these changes do not affect the validity of the conclusions. This network structure will be referred to as configuration C1 and represents the base-case for the study. The performance of the pre-heat train was monitored at start-up (after cleaning) and after 8000 hours operation (Table 2). The performance of units E1, E2 and E4 did not change significantly over the operating period whereas the other units exhibited severe fouling. The observations of the refinery operators were that fouling in units E5 and E6 mainly occurred on the tube-side (the crude oil flowed through the tubes) and this was the only side of these units cleaned. Unit E3 fouled on the shell-side (which handled column residues) but not on the tube-side. Unit E7 fouled heavily on both sides. For the following analysis it is assumed that the desalter temperature is optimally controlled and that its performance does not affect the fouling behaviour of the network. Crude is on the tube-side and hot fluids flow on the shell-side in all units. Stream properties are estimated though API relationships.

Effects of of fouling on performance of retrofitted heat exchanger networks.
A thermo- hydraulic based analysis.

Table 1 Exchangers geometry. All units have 20 ft (6.1 m) tubes of ¾ inch outside diameter.

Unit	No. Tubes	No. Tube pass	Baffle cut	Baffle spacing	No. Baffles	Shell diameter
	[-]	[-]	[%]	[mm]	[-]	[mm]
E01	152	6	22	178	32	444
E02	308	8	17	146	40	584
E03	336	2	17	130	43	584
E04	620	4	17	153	37	798
E05	1130	6	17	136	41	1100
E06	240	2	17	138	41	520
E07	968	4	17	137	41	740

3. Proposed network revamps

By analyzing the stream temperatures reported in Table 2, it is clear that more energy can be recovered from the residue stream. If pinch rules are applied, the area of E-07 would be increased. This can be done by adding an extra unit E-07x, identical to E-07 in geometry. The resulting network structure, referred to as configuration C2, is shown in Figure 2(a). In this design, the hot residue stream is matched with the crude at its highest temperature. Whilst this ensures maximum heat recovery in clean conditions, over time the fouling behaviour is expected to penalize the overall heat recovery of the network as the wall temperatures, on which fouling depends, are also maximized. To increase the performance over time, we analyse an alternative network retrofit, configuration C3 in Figure 2(b). This goes against traditional pinch rules by matching the hot residue stream with the crude at an intermediate temperature (at the exit of E-05). In his configuration, the residue stream enters the extra unit E-07x, placed between units E-05 and E06 and subsequently enters E-07 and E-03. As a result, heat recovery in clean conditions is expected to be less than that achievable with configuration C3, however, the crude at the highest temperature exchanges heat with a lower temperature residue stream. A third alternative configuration, C4 in Figure 2(c), was proposed. In this configuration the residue enters first unit E-07 as in the base-case structure and then unit E-07x which as in structure C3 is placed between E-05 and E06. The performance given by the different designs will be assessed by considering the value of the CIT achieved over time and the total energy required by the furnace to maintain a constant value of the coil outlet temperature (COT) of 360 °C. The total (cumulative) energy required for each configuration Cj, is defined as:

$$E_{Cj}(t) = \int_0^t \dot{m}c_p \left(COT - CIT \right) dt; \quad \text{for } j = 1, 2, 3, 4 \tag{2}$$

Where $\dot{m}$ is the massflowrate of the crude, c_p its specific heat capacity and COT is the coil outlet temperature. The difference in performance (extra energy recovered at the

Table 2. PHT performance in clean conditions (T_{in}^0, T_{out}^0) and after 8000 hours (T_{in}^*, T_{out}^*).

|	Shell-side (Hot streams)					Tube-side (crude oil, cold stream)				
Unit	Flow	T_{in}^0	T_{out}^0	T_{in}^*	T_{out}^*	Flow	T_{in}^0	T_{out}^0	T_{in}^*	T_{out}^*
	[kg/h]	[°C]	[°C]	[°C]	[°C]	[kg/h]	[°C]	[°C]	[°C]	[°C]
E01	10000	205	44	205	44.1	100000	32	51.3	32	51.3
E02	20000	290	83	290	83	100000	51.3	96.5	51.3	96.5
E03	50000	254.6	206.5	295.4	246.1	100000	96.5	120.5	96.5	121.2
E04	40000	220	148.2	220	147.5	100000	116	147.1	116.7	147.5
E05	60000	280	193	280	203.5	100000	147.1	199.3	147.5	193.4
E06	10000	280	208.5	280	207.2	100000	199.3	206.2	193.4	200.5
E07	50000	360	254.6	360	295.4	100000	206.2	256.4	200.5	231.3

(a)

(b)

(c)
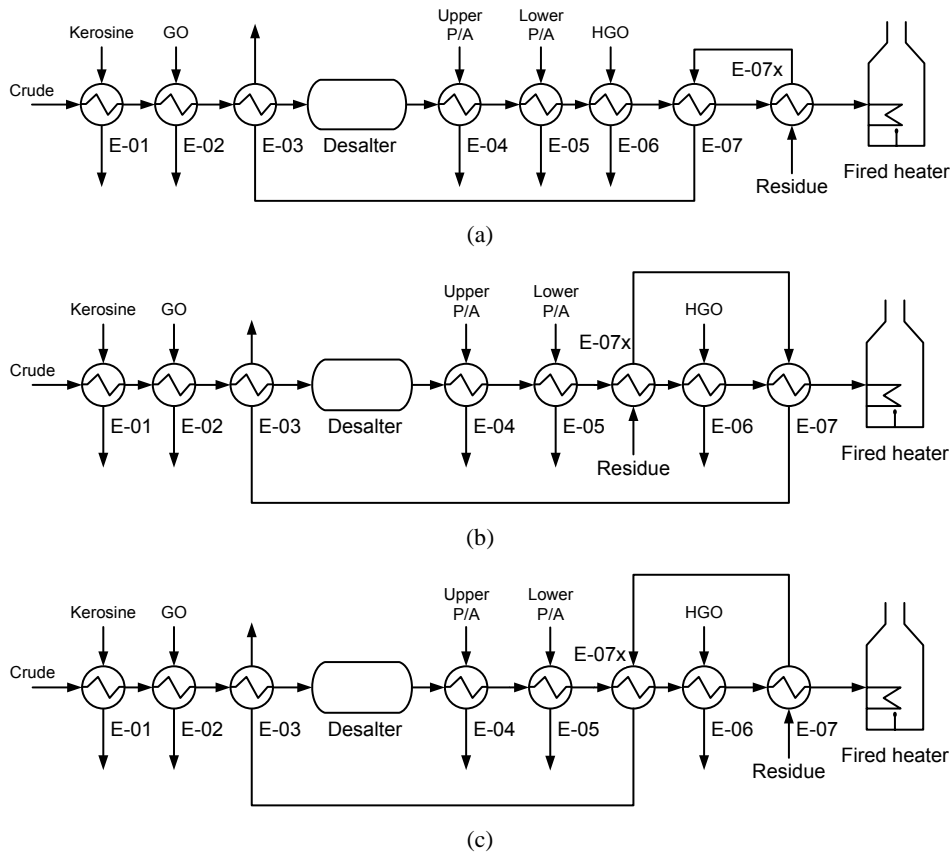
Figure 2 Network retrofit configurations C2 (a), C3 (b) and C4 (c).

furnace) of each retrofit is:

$$\Delta E_{C1-Cj} = E_{C1} - E_{Cj}; \quad \text{for } j = 2,3,4 \tag{3}$$

Equation (3) is used to assess the performance of the three retrofit configurations proposed over the base case. As the same additional heat exchanger E-07x is used in all three retrofit options, the capital costs will be the same.

4. Results and discussion

Figure 3 shows the CIT over this period for the base-case and the three retrofit options. It can be seen that in clean conditions (i.e. t=0), for configurations C2 and C4 the CIT is over 13 °C larger than that achieved in C1, indicating a good extra energy recovery. Despite the extra heat transfer area available in configuration C3 compared with C1, the initial CIT in configuration C3 is 3.5 °C lower than in the base case. This is because the performance of existing units E6 and E7 falls due to reduced temperature driving force. Over time, however, things change significantly because of fouling. Fouling rates in E-07 are highest for configurations C1 and C4, and lowest for C2 and C3 (Figure 4(a)). Unit E-07x fouls more in C2 then in C3 and C4 (Figure 4(b)). As a result, after less than a month of operations, the CIT in C3 is maintained at a higher value compared to that of the base-case C1. After 150 days C3 starts recovering more energy than C2. The structure generated according to pinch rules, C2, results in the worst performance over time. After 300 days the CIT in the retrofit structure C4 also falls below that of C3.

Effects of of fouling on performance of retrofitted heat exchanger networks.
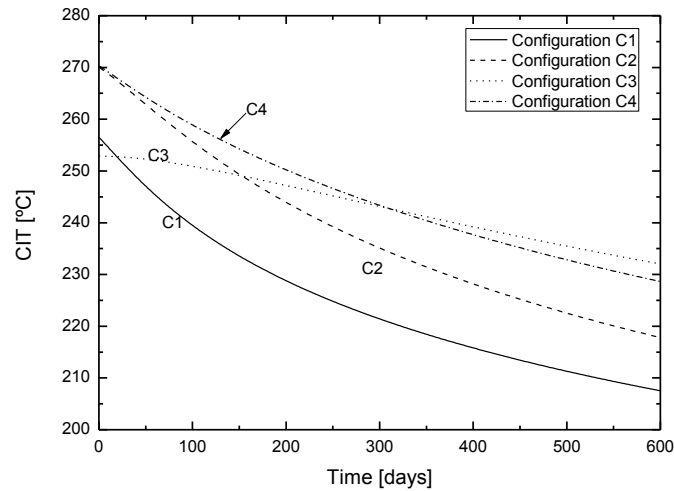*A thermo- hydraulic based analysis.*

Figure 3 CIT for the base-case and the three alternative network configurations considered.

Whilst the analysis of the CIT highlights the importance of considering fouling dynamics in the retrofit of PHT networks, it is not sufficient to asses which structure, amongst those proposed, provides the largest amount of energy recovered over time. Figure 5 shows the (cumulative) amount of extra energy required in the furnace with respect to the base-case that can be recovered trough one of the three proposed retrofits as calculated by Equation (3). Although barely appreciable in Figure 5, given the scale of the graph, it should be noted that ΔE_{C1-C3} is negative for ca. the first 50 days of operations, confirming that at clean conditions, configuration C3 has a negative impact on the overall energy recovery. However, after roughly a year of operations configuration C3 starts performing better than C2, the configuration proposed following pinch rules. By analyzing Figure 5 another important aspect can be unveiled. Whilst the CIT in configuration C3 becomes larger than that in C4 after ca. 300 days, the cumulative extra energy recovered by the latter is constantly larger than that recovered by the former. Whilst C4 increases the fouling resistance in E-07 with respect to the base-case (the tube side temperatures are higher), it results in the lowest fouling resistance in E-07x. Clearly, the placing of the extra unit and its fouling behaviour is paramount for the overall performance of the network. In fuel costs alone, savings after 600 days of over US$450,000 (assuming a cost of fuel oil of US$27/MWh) can be achieved by using configuration C2, ca. US$600,000 using C3 and ca.US$680,000 using configuration C4 and are expected to be much higher for larger refineries.
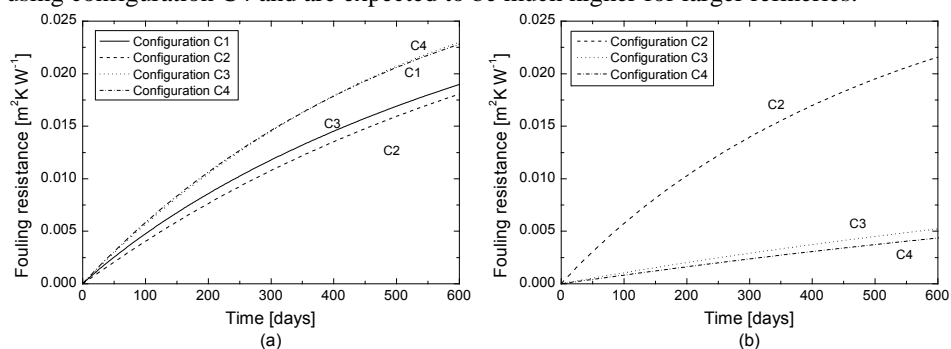

Figure 4 Average fouling resistance calculated over time for exchangers E-07 (a) and E07x (b).
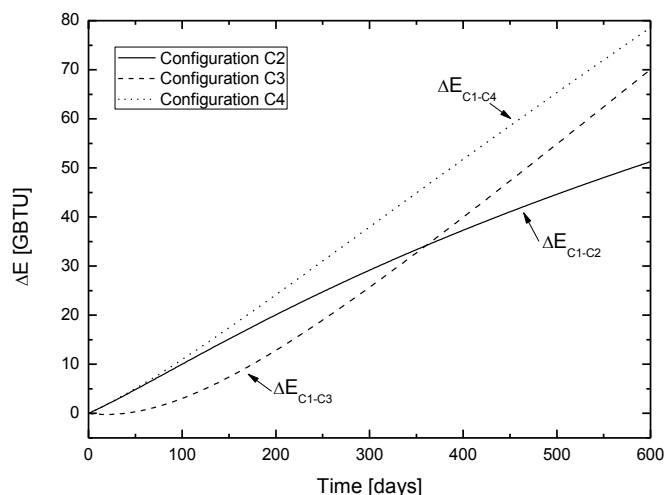
Figure 5 Extra energy recovered at the furnace with respect to the base case in the three cases considered.

5. Conclusions

A mathematical model for shell-and-tube heat exchangers undergoing crude oil fouling has been used to show rules aiming at maximize energy recovery alone may not be the best strategy to pursue in retrofitting PHTs networks. In particular it has been shown, though the case study presented, that energy recovery in three different retrofit solutions can be large. Whilst after 300 days of operation, the energy recovered in retrofit C4>C2>C3, after 600 days C4>C3>c2. In fuel costs alone, these differences translate in ca. US$ 230,000 over 600 days between network structures that follow pinch rules and those that take into account fouling.

Some key conclusion can be drawn from the study presented: (i) fouling plays a major role in the energy recovery which is not captured if steady state conditions alone are considered at the design stage; (ii) a trade-off exists between maximum energy recovery and fouling behaviour; (iii) a detailed mathematical model is required to accurately assess, energy losses and, ultimately, refinery costs related to fouling.

Of course, other factors such as reduction in throughput, which were not included in this paper, could play a crucial role in the choice of the arrangement and will be considered in future work.

References

F. Coletti and S. Macchietto (2009). "A dynamic, distributed model of shell and tube heat exchangers undergoing crude oil fouling." Submitted for publication.

C. B. Panchal, W. C. Kuru, C. F. Liao, W. A. Ebert and J. W. Palen (1999). Threshold conditions for crude oil fouling. Understanding Heat Exchanger Fouling and Its Mitigation, Castelvecchio Pascoli, Italy, Begell House Inc.

D. I. Wilson, G. T. Polley and S. J. Pugh (2002). "Mitigation of crude oil preheat train fouling by design." Heat transfer engineering 23(1): 24-37.

B. L. Yeap, D. I. Wilson, G. T. Polley and S. J. Pugh (2004). "Mitigation of crude oil refinery heat exchanger fouling through retrofits based on thermo-hydraulic fouling models." Chemical Engineering Research & Design 82(1): 53-71.

Modelling and simulation of the effect of non-condensable gases on heat transfer in the MSF desalination plants using gPROMS software

Said A. Said[a], Iqbal M. Mujtaba[a], Mansour Emtir [b]

[a] *School of Engineering, Design and Technology, University of Bradford, Bradford, West Yorkshire, BD7 1DP, United Kingdom.* Email: I.M.Mujtaba@bradford.ac.uk
[b]*National Oil Corporation, P.O. Box 2655,Tripoli, Libyan Arab Jamahiriya*

Abstract

This work describes a steady state mathematical model of Multistage Flash (MSF) desalination process incorporating correlations which take into account the effect of the presence of non-condensable gases (NCGs) and fouling factors on the overall heat transfer coefficient. First, we study the effect of different NCGs concentration on the overall heat transfer co-efficient. The simulation results showed decrease in the overall heat transfer coefficient values as NCGs concentrations increase. For a given plant configuration (fixed design) and at different seawater and steam temperatures, a 0.015 wt % of NCGs results in significantly different plant operations when compared with those obtained without the presence of NCGs. Finally, for fixed water demand and in the presence of 0.015 wt % NCGs, the performance is evaluated for different plant configurations and seawater temperature and compared with those obtained without the presence of NCGs.

Keywords: MSF, non-condensable gases, overall heat transfer coefficient, gPROMS

1. Introduction

Desalination is a technique of producing fresh water from the saline water. Industrial desalination of sea water is becoming an essential part in providing sustainable source of fresh water for a large number of countries around the world. Thermal process is the oldest and most dominating for large scale production of freshwater in today's world (El-dessouky, and Eltouney, 2002).

Multistage Flash (MSF) process produces 56% of the total fresh water produced by different desalination technologies (Hussain et al., 2003). An MSF process (Fig. 1) mainly consists of three sections: brine heater section, recovery and rejection sections each with a number of flash chambers (stages). Non-condensable gases are a serious problem in MSF. NCGs consist mainly of air (N_2 and O_2) and CO_2. The presence of NCGs is caused by the leakage of ambient air through flanges, man-holes, instrumentation nozzles, into the parts of the evaporating brine. NCG gases cause local reduction of performance, reduction of efficiency and hence a cost increase in most thermal desalination units (Khan, 1972). It was shown that even low concentrations of NCGs significantly reduce the overall heat transfer coefficient and hence the performance of desalination evaporators. CO_2 dissolves in the condensate and lowers its pH value. In presence of O_2, this may cause corrosion of the condenser tubes. The loss of freshwater production occurs due to plant shut down for maintenance which reduces the plant lifetime consequently (Oldfield, 1987)**.** Also, the release of CO_2 from the evaporating brine in seawater distillers considerably influences the concentrations of

HCO_3^-, CO_3^{2-}, H^+, and OH^- ions in the carbonate system of the brine and play an important role in alkaline scale formation. In MSF, the NCGs entering with the feed water are liberated during the evaporation process and have to be removed by adequate venting.

Modelling is very important in simulation, optimization and control of MSF desalination process (Helal et al. 1986). Most models are well developed from the basic laws of total and component mass balances and enthalpy balances coupled with heat and mass flowrate coefficients. However, many of them neglected the effects of NCGs on the overall heat transfer coefficient and scale formation in process modelling and simulations. These assumptions cause a larger discrepancy in the simulation results of the models and the actual plant data.

In this work, the effect of NGCs on the overall heat transfer coefficient is studied first. Then an MSF process model developed by Rosso et al. (1996) is refined with the new overall heat transfer co-efficient correlation featuring the effect of NCGs and is used to study the effect of the presence of non-condensable gases on the design and operation parameters of MSF processes.
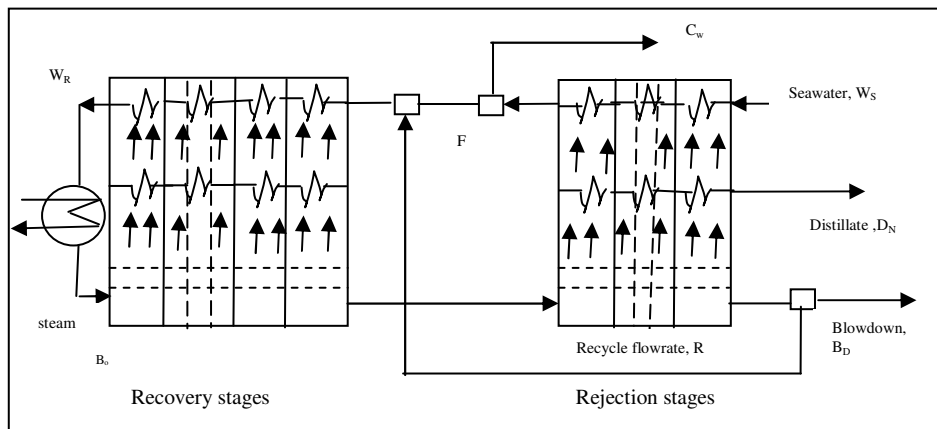

Fig.1 MSF Process description

2. Steady State MSF Process Model

There are considerable amount of work on steady state modeling and simulation of MSF desalination plants (El-dessouky, and Eltouney, 2002; Hussain et al., 2003; Tanvir and Mujtaba, 2006a,b) . The models used in these studies only involved the thermodynamic losses from stage to stage, tube velocity, tube materials, chamber geometry and the effect fouling factors on the rate of heat transfer. Study considering the effect of NCG gases on the overall heat transfer coefficient, energy consumption, plant design and production capacity of MSF process is almost nonexistent. The presence of NCGs reduces the overall heat transfer coefficient for the condensing vapor and the temperature at which it condenses at given pressure in the vapor space (El-dessouky, 1995). In this work, we used correlations reported by Wangnick (1995) which take into considerations the effect of the presence of NCGs and fouling factors on the overall heat transfer coefficient in the heat recovery section, heat rejection section, and brine heaters. These correlations calculate the inside and outside overall heat transfer coefficients

Modelling and simulation of the effect of non-condensable gases on heat transfer in the MSF desalination plants using gPROMS software

which depend on the fouling factors, flowrate, temperature and physical properties such as thermal conductivity, viscosity, density and specific heat of the condensing vapor and the brine inside the condenser.

The process model equations considered in this work are shown in Fig. 2. The other physical properties correlations presented here are reported by Helal et al. (1986), Rosso et al (1996) and Husain et al (2003). For a number of stages NS=NR+NJ, the total number of equations (TNE) is: 25NS+27. The total number of variables (TNV) is: 18NS+16. Therefore, the degrees of freedom (D.F.=TNV-TNE) is 7NS+11 which are specified before the simulation is carried out.

Stage Model

Mass Balance in the flash chamber : $B_{j-1} = B_j + V_j$

Mass Balance for the distillate tray: $D_j = D_{j-1} + V_j$

Enthalpy balance on flash brine: $B_j = (h_{Bj-1} - h_{vj})/(h_{Bj} - h_{vj})B_{j-1}$

$h_{jv} = f(T_{sj})$ $h_{Bj} = f(C_{Bj}, T_{Bj})$

Overall Enthalpy :

$W_R S_{Rj}(T_{Fj} - T_{Fj+1}) = D_{j-1}S_{Dj-1}(T_{Dj-1} - T^*)$
$+ B_{j-1}S_{Bj-1}(T_{Bj-1} - T^*) - D_j S_{Dj}(T_{Dj} -) - B_j S_{Bj}(T_{Bj} - T^*)$

Heat Transfer equation:

$W_R S_{Rj}(T_{Fj} - T_{Fj+1}) = U_j A_j X$ (replace W_R for W_S rejection stage)

$X = \{(T_{Dj-1} - T_{Fj+1}) - (T_{Dj} - T_F)\}/\ln\{(T_{Dj} - T_{Fj+1})/(T_{Dj} - T_{Fj})\}$

(replace WR for Ws rejection stage)
$SR_j = f(T_{Fj}, T_{Fj+1}, C_R)$
(replace WR for Ws rejection stage)
$SD_j = f(T_{Dj})$ $SB_j = f(T_{Bj}, , CB_j)$

Distillate and flashing brine temperature correlation:
$$T_{Bj} = T_{Dj} + TE_j + EX_j + \Delta_j$$

Distillate and flashing steam correlation:
$$T_{Bj} = T_{Dj} + \Delta_j$$
$\Delta_j = f(T_{Dj})$
$TE_j = f(T_{Dj}, C_{Bj})$ $EX_j = f(H_j, w_j, T_{Bj})$

Brine Heater Model

$B_o = W_R$ $C_{Bo} = C_R$ $B_o S_{RH}(T_{Bo} - T_{F1}) = W_{steam}\lambda_S$ $\lambda_S = f(T_{steam})$

$W_R S_{RH}(T_{Bo} - T_{F1}) = U_H A_H Y$

$Y = \{(T_{steam-}T_{F1}) - (T_{steam} - T_{Bo})\}/\ln\{(T_{steam} - T_{F1})/(T_{steam} - T_{Bo})\}$

$U_H = f(T_{steam}, T_{Bo}, T_{F1}, T_{steam}, D_H^i, D_H^o, f_H^i)$

$S_{RH} = f(T_{Bo}, T_{F1})$

Spiltters Models

     $B_D = B_{NS} - R$
$C_W = W_S - F$

Mixers Model

$W_R = R + F$ $RC_{BNS} +$
$FC_S = W_R C_R$ $W_R h_W = Rh_R + Fh_F$ $h_W = f(T_{Fm}, C_R)$ $h_F = f(T_{FNR+1}, C_F) h_R = f(T_{BNS}, C_{BNS})$

Steam Side Heat Transfer Coefficient

$h_o = .725(\dfrac{K_l^3\rho_l(\rho_l - \rho_v)g\rho\lambda_v}{\sigma_o\mu\Delta T})^{.25}C_1 C_2$

$C_1 = 1.23795 + .353808N_1 - .0017035N_1^2$,
$C_2 = 1 - 34.313X_{nc} + 1226.8X_{nc}^2 - 14923X_{nc}^3$

$N_1 = .564\sqrt{N_t}$, $N_t = \dfrac{4M_f}{\pi\sigma_i^2\rho_f V_f}$

Water Side Heat Transfer Coefficient

$h_i = (3293.5 + T(84.24 - .1714T) - x(8.471 + .1161x + .2716T))/((\dfrac{\sigma_i}{.017272})^{.2}(.656V)^{.8}(\sigma_i/\sigma_o))$

Overall heat transfer coefficients

$\dfrac{1}{U_o} = \dfrac{1}{h_i\left(\dfrac{d_i}{d_o}\right)} + \dfrac{1}{h_o} + \dfrac{1}{h_i} + r_t + r_{fi} + r_{fo}$

Fig. 2 MSF Stage Model Equations

3. Effect of Non-condensable Gases on Heat Transfer Co-efficient

First, we examine the effect of different NCGs concentration on the overall heat transfer co-efficient. The NCG concentrations were varied from 0.015 - 0.06 (wt %). The simulation results showed decrease in the overall heat transfer coefficient (Figs. 3 and 4). The corresponding decreases were up to 6% in the heat recovery section and 13% in the heat rejection section.
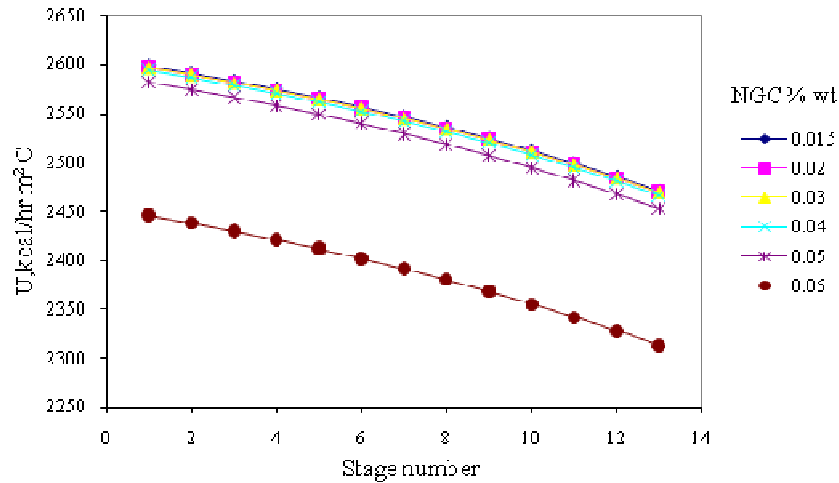
Fig. 3 Overall heat transfer coefficient through recovery stages at different NCGs wt %
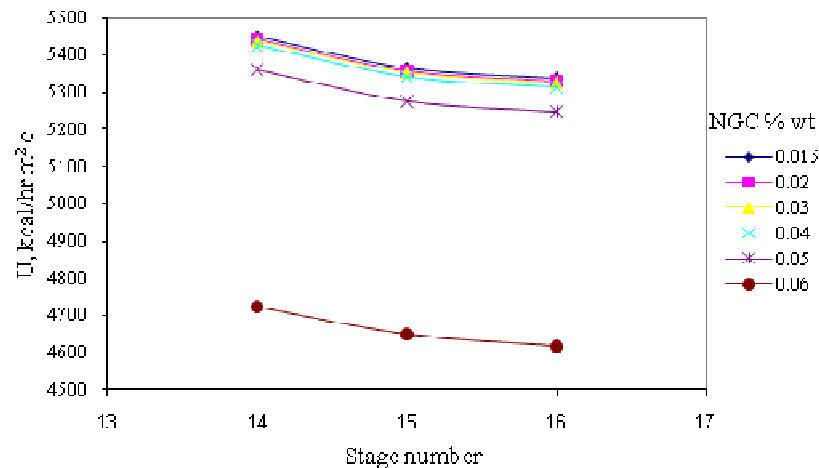

Fig. 4 Overall heat transfer coefficient through rejection stages at different NCGs wt %

4. Effect of Non-condensable Gases on the Design and Operation

In this work we have considered the case reported by Rosso et al. (1996). There are total of 16 stages with 13 recovery and 3 rejection stages. The specifications (satisfying the degrees of freedom) are same as those used in Tanvir and Mujtaba (2006a).

For a given NCGs (0.015 wt %) and steam temperature (T_{steam}) (97°C) simulations are carried out at different seawater temperature ($T_{seawater}$). The results are presented in Table 1 with the total amount of fresh water produced (D_{NS}), GOR (Gained Output Ratio), TBT (Top Brine Temperature) and final bottom brine temperature (BBT). Comparison of the results at $T_{seawater}$ = 35 C with those of Rosso et al. (1996) and Tanvir

Modelling and simulation of the effect of non-condensable gases on heat transfer in the MSF desalination plants using gPROMS software

and Mujtaba (2006a) clearly show the effect of NCGs. The amount of steam consumption goes very high due to change in overall heat transfer co-efficient. Although the freshwater production rate improves the value of GOR goes down. The results also show decrease in plant production, performance ratio, steam flowrate and increase in the top brine temperature and brine blowdown temperature as seawater temperature increases from 22°C to 45°C (similar pattern was observed by Tanvir and Mujtaba, 2006a). Further simulations are carried out for a given seawater temperature (45°C) with steam temperature varying from 110°C -120°C. The results showed increases in all the operation variables (again a similar pattern was observed but with different values by Tanvir and Mujtaba, 2006a).

Finally for 3 different sets of plant configuration (different values of number of stages in the recovery section), a series of simulation is carried out to study the effect of NCGs on the operation parameters at different seawater temperature ($T_{sewater}$) but with fixed water demand (D_{NS}) of 277.77kg/s (10^6 kg/hr). The seawater temperature is assumed to vary between 20 and 35°C and number of stages (N) between 16 and 20. The results are summarized in Table 2 and are compared with those obtained by Tanvir and Mujtaba (2006b) who did not consider the effect of NCGs in their model.

Table 1. Effect of $T_{seawater}$ and T_{steam} on D_{NS} , GOR, TBT , and BBT at 0.015 % wt of NCG

$T_{seawater}$,°C	D_{NS},kg/s	W_{steam},kg/s	GOR	TBT,°C	BBT,°C
22	320	42.8	7.48	89	29.24
35	270	35.7	7.57	90	40.87
45	230	30.4	7.56	91	50.00
T_{steam},°C					
110	293	36.3	8.07	103	50.90
115	317	38.5	8.24	108	51.26
120	342	40.7	8.40	112	51.64

Table 2. Effect of N and $T_{seawater}$ on W_{Steam} ,T_{steam} , R, and BBT at 0.015 % wt of NCG

N	$T_{seawater}$,°C	W_{steam},kg/s	T_{steam},°C	BBT,°C	R,kg/s
16	35	34.07	96.27	40.59	1660.18
	30	32.82	95.90	35.45	1370.00
	25	31.97	95.62	30.48	1153.30
	20	31.45	95.41	25.72	969.21
18	35	31.58	95.87	41.75	1732.03
	30	29.46	95.30	35.14	1364.62
	25	29.27	95.18	31.07	1178.59
	20	28.26	94.87	25.50	962.04
20	35	27.84	95.12	39.97	1623.91
	30	28.06	95.04	35.02	1358.67
	25	27.99	94.96	31.35	1190.65
	20	26.94	94.64	25.42	959.24

The results show increases in recycle flowrate (R) up 70% at seawater temperature 35°C compared to that at 20°C. Also, 57% increase in steam flowrate is obtained for the same variations of the seawater temperature. Note, although the observations (pattern) are very similar to those of Tanvir and Mujtaba (2006b) the actual values of the parameters

are quite different which are due to the consideration of the effect of NCGs in the model considered in this work.

5. Conclusions

The effect of non-condensable gases on the overall heat transfer coefficient correlations have been modelled and incorporated within an MSF process model using gPROMS modelling tool. The simulation results show decrease in overall heat transfer coefficient up to 6% in heat recovery section and 13% in heat rejection section. The model is then used to study how the process and design parameters affect due to the changes in overall heat transfer coefficient. The results show increase in steam flowrate and plant productin the presence of 0.015 wt % of NCGs compared with the same results obtained by Rosso et al. (1996) and Tanvir and Mujtaba (2006). Also, in the presence of 0.015 wt % of NCGs and to maintain a fixed water demand , a 70% increases in recycle flowrate and 57% increases in steam flowrate obtained when we change the number of stages from 16 to 20 and at seawater temperature between 20°C and 35°C.

References

El-dessouky, H., Shaban, H., Al-Ramadan, H., 1995 , Steady state analysis of multi-stage flash desalination process, *Desalination*, 103(1-3), 271.

El-dessouky, H., Eltouney, H., 2002, fundamentals of salt water desalination, Amsterdam, *Elsevier Science ltd*.

gPROMS, 2004, Introductory User Guide, Process System Enterprise Ltd (PSE).

Helal, A., Medani, M., Soliman, M., Flower, J., 1986, Tridiagonal matrix model for multistage flash desalination plants, *Computer & chemical engineering*, **10**(4), 327-342.

Hussain, A., Hassan, A., Al-Gobaisi, D., Al-Radif, A., Woldai, A., Sommariva, C.,1993, Modeling, simulation, optimization and control of multistage flash (MSF) desalination plants Part2: modeling and simulation, *Desalination ,* 92(1-3),4-41.

Khan, R.,1972, Effect of non-condensable in seawater evaporators, *Chem. Eng. Prog.,* 68, 79-80.

Oldfield, J., 1987, Vapour side corrosion in MSF plants*, Desalination*, 66,171-187.

Rosso, M., Beltramini, A., Mazzotti, M., Morbidelli, M., 1996, Modellng of multistage flash desalination plants, *Desalination*, 108, 365-374.

Tanvir, M., Mujtaba, I., 2006a, Modelling and Simulation of MSF Desalination Process using gPROMS and Neural Network based Physical Property Correlation, *Computer Aided Process Engineering*, Vol21B, 2006b, 315,ed. Marquardt and Pantelides, Elsevier.

Tanvir, M., Mujtaba, I., 2006b, Simulation of MSF Desalination Process for Fixed Water Demand using gPROMS and Neural Network based Temperature Elevation Correlation, In Proceeding of IWC2006 Conference, 300, 12-14 June, Portugal.

Wangnick, K., 1995, How incorrectly determined physical and constructional properties the seawater and brine regimes influence the design and size of an MSF desalination plant-simulus for further thought, Proceeding of the IDA world Congress on Desalination and Water science, Abu Dhabi, Vol.2, 201-218.

Simulation of Free Surface Viscoelastic Fluid Flow Using the viscoelasticInterFoam Solver

Jovani L. Favero,[a] Nilo S. M. Cardozo,[a] Argimiro R. Secchi,[b] Hrvoje Jasak[c]

[a]*Chemical Engineering Depart. – EE/UFRGS - Universidade Federal do Rio Grande do Sul, R. Eng. Luiz Englert, s/n Centro – 90040-140 – Porto Alegre, RS – Brazil, favero@enq.ufrgs.br, nilo@enq.ufrgs.br*
[b]*Chemical Engineering Program – PEQ/COPPE – Universidade Federal do Rio de Janeiro, Av. Horácio Macedo, 2030 – CT - G116 – Ilha do Fundão – 21941-972 – Rio de Janeiro, RJ – Brazil, arge@peq.coppe.ufrj.br*
[c]*FSB/Wikki – Faculty of Mechanical Engineering and Naval Architecture, University of Zagreb, Croatia, and Wikki Ltd., London W8 7PU, United Kingdom, h.jasak@wikki.co.uk*

Abstract

The understanding of polymeric fluids flows is of essential importance for several industry sectors, including plastic and food processing industries. The rheological response of viscoelastic fluids is quite complex, including combination of viscous and elastic effects and highly non-linear viscous and elastic phenomena. In a previous work we presented a solver for internal viscoelastic fluid flows, called viscoelasticFluidFoam, which was developed with the OpenFOAM computational fluid dynamics package and validated with experimental and numerical data from the literature for the analysis of a planar 4:1 contraction flow. In the present work will be tested the extension of the viscoelasticFluidFoam solver to the viscoelasticInterFoam solver used for analysis of free surface viscoelastic fluid flows using the VOF methodology. A classical die swell flow phenomena used in the rheology literature to present the concept of viscoelastic effects was simulated. The results obtained using Giesekus model showed the great potential of the developed formulation, once all phenomena observed experimentally were reproduced in the simulations.

Keywords: Viscoelastic Fluids, VOF, OpenFOAM, Die Swell, CFD.

1. Introduction

The understanding of viscoelastic fluid flow characteristics is of great interest for polymer and food processing industries. Unlike the Newtonian flow that is well understood, the flow of viscoelastic fluids is still a subject of great discussion. Most of the works on viscoelastic fluids found in literature treats internal flows using the geometry known as abrupt planar contraction. This is due mainly because this geometry was adopted as standard for viscoelastic fluids flow studies in 1987 during the fifth international workshop on numeric methods for non-Newtonians fluids (Hassager, 1988). On the other hand, simulations using free surface viscoelastic flow are receiving a special attention mainly in the last two decades (Tomé et al., 2008).

Flows considering free surface is particularly useful for industry that works with polymeric fluids, once processes as extrusion and injection molding of plastics, metallic leagues or fluids used in food industry depend always more on the knowledge of parameters that guarantee a high quality for their products (Khismatullin et al., 2006).

Regarding the modeling, the study of free surface flow presents the same bases of internal flows plus an appropriate methodology for treatment of free surface motion, such as the MAC (Marker-and-Cell) (Harlow and Welch, 1965) and VOF (Volume-of-Fluid) (Hirt and Nicols, 1981). An overview of literature works about viscoelastic free surface simulation can be found in Harvie et al. (2008). There are in the literature several works treating viscoelastic fluid flow with free surface, and the most of them use finite element method to solve simplified cases. In this work we overcome that limitation by showing results of a generic and flexible tool developed for treatment of free surface viscoelastic fluid flow using finite volume method and VOF methodology. This tool can be used to simulate complex geometries encountered in real problems of polymer and food processing industries, using non-structured and moving meshes, and allowing to choose among several different interpolation schemes, solvers for linear algebra, possibility of data processing parallelization and several constitutive equations for treatment of viscoelasticity.

This work is an extension of our previous development of a solver, called viscoelasticFluidFoam (Favero et al., 2009), for treatment of internal viscoelastic flow using the OpenFOAM software (Weller et al., 1998). Such extension consists of increasing the capability of the viscoelasticFluidFoam solver, already available to the user in OpenFOAM, for the treatment of free surfaces flow using the VOF methodology, creating a new solver called viscoelasticInterFoam. The discretization of the flow governing equations in OpenFOAM is based on the finite volume method formulated with collocated variable arrangement, with pressure and velocity solved by segregated methods. The software offers a wide range of interpolation schemes, solvers and preconditioners for the discretized algebraic equation system.

The results, studying a classic experiment in viscoelastic fluids flow with free surface: the die swell in the exit of a channel, show the success of the applied methodology. The rod-climbing experiment was also reproduced in the simulations but is not shown here due to lack of space.

2. Methodology

In this section, the mathematical formulation, the strategy used to solve free surface viscoelastic fluid flow, and a description of the test problems are presented.

2.1. Mathematical Formulation

The governing equations of laminar, incompressible, and isothermal flow of viscoelastic fluids are the mass conservation (continuity equation):

$$\nabla \cdot \mathbf{u} = 0 \tag{1}$$

and momentum conservation:

$$\frac{\partial(\rho\mathbf{u})}{\partial t} + \nabla \cdot (\rho\mathbf{u}\mathbf{u}) = -\nabla p + \nabla \cdot \boldsymbol{\tau}_S + \nabla \cdot \boldsymbol{\tau}_P + \rho \mathbf{g} \tag{2}$$

together with a mechanical constitutive equation that describes the relation between the stress and the deformation rate for a given fluid. In the above equations ρ is the density, $\mathbf{u}$ the velocity vector, p the pressure, $\boldsymbol{\tau}_S$ the solvent Newtonian contribution, and $\boldsymbol{\tau}_P$ the polymeric contribution.

The solvent contribution is given by:

$$\boldsymbol{\tau}_S = 2\eta_S \mathbf{D} \tag{3}$$

where η_S is the solvent viscosity and **D** is the deformation rate tensor, given by:

$$\mathbf{D} = \frac{1}{2}\left(\nabla\mathbf{u} + \left[\nabla\mathbf{u}\right]^T\right) \tag{4}$$

The extra elastic contribution, corresponding to the polymeric part, is obtained from the solution of an appropriate constitutive differential equation. Various constitutive equations are found in literature, as for example, Maxwell, Oldroyd-B, Phan-Thien-Tanner (PTT), Finitely Extensible Nonlinear Elastic and so on. A well known and widely used is the Giesekus constitutive model described as (Giesekus, 1982):

$$\boldsymbol{\tau}_P + \lambda\overset{\nabla}{\boldsymbol{\tau}}_P + \alpha\frac{\lambda}{\eta_p}(\boldsymbol{\tau}_P \cdot \boldsymbol{\tau}_P) = 2\eta_p\mathbf{D} \tag{5}$$

where λ and η_P are the relaxation time and the polymer viscosity at zero shear rate, respectively, α is the dimensionless 'mobility factor', and $\overset{\nabla}{\boldsymbol{\tau}}_P$ is the upper convected derivative of the elastic stress tensor defined as (Bird et al., 1987):

$$\overset{\nabla}{\boldsymbol{\tau}}_P = \frac{\partial}{\partial t}\boldsymbol{\tau}_P + \mathbf{u}\cdot\nabla\boldsymbol{\tau}_P - \left[(\nabla\mathbf{u})^T \cdot \boldsymbol{\tau}_P\right] - \left[\boldsymbol{\tau}_P \cdot \nabla\mathbf{u}\right] \tag{6}$$

Another important subject is the treatment of the free surface. In the developed solver it was used the VOF methodology (Hirt and Nicols, 1981). In this method, the volume of fluid in a cell is computed as $F_{vol} = \gamma V_{cell}$ where V_{cell} is the computational cell volume and γ is the liquid fraction in this cell. The value of γ in a cell varies between 0 and 1. A cell totally filled with one fluid presents $\gamma = 1$ and another totally filled with one different fluid have $\gamma = 0$. In the interface the value of γ is an intermediate value between 0 and 1, satisfying the following equation:

$$\frac{\partial\gamma}{\partial t} + \nabla\cdot\left(\gamma\mathbf{u}\right) + \nabla\cdot\left(\gamma(1-\gamma)\mathbf{u}_r\right) = 0 \tag{7}$$

where $\mathbf{u}_r$ is the velocity field subject to interface compression (Ubbink and Issa, 1999; Rusche, 2003). The last artificial term is only activated in the region of the interface due to the properties of the term $\gamma(1-\gamma)$.

The VOF methodology allows to solve the flow of two different fluids (two-phase flow) solving as for one fluid only, in other words, the same equations are solved for both fluids, the only difference is regarding to the physical properties. To make this possible, the physical properties in any control volume of the domain are given by the corresponding fraction of the fluid in that volume. It is mathematically obtained, considering a generic property θ, as:

$$\theta = \gamma\theta_{fluid1} + (1-\gamma)\theta_{fluid\,2} \tag{8}$$

2.2. Test Geometry

To test the implemented methodology, the reproduction of a classic example used to demonstrate viscoelasticity effects are carried out. The test consists in reproducing the die swell phenomena.

Figure 1 shows the effect of die swell due to the fluid elasticity (Bird et al., 1987). An effect due to viscoelasticity is the presence of normal stress differences in shear flow. In addition to shear stress, viscoelastic fluids present an extra stress due to stretching and

alignment of polymeric chains. In this experiment, it is considered a fluid flowing out of a capillary with diameter *d*. For Newtonian fluids the swell can vary 13% above or below of the diameter *d*, whereas for polymeric fluids, an increase in more than 300% relative to *d* can be found, depending on the fluid elasticity. To simulate this experiment a shear rate of 60 s^{-1} was used in the flowing channel.
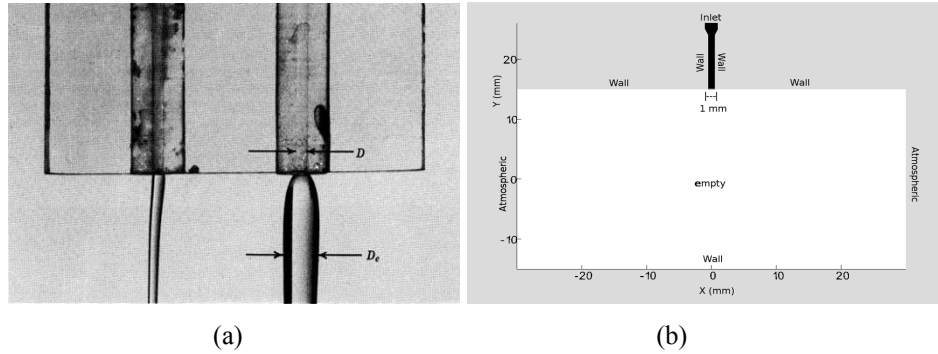

(a)   (b)

Figure 1: (a) Representation of the die swell effect for a Newtonian fluid (left) and a polymeric fluid (right) (Source: Bird et al., 1987). (b) 2D geometry used to simulate the experiment.

In Table 1, the physical properties of the fluids used in the simulations are listed. In the analyzed case, a Newtonian and two polymeric fluids are considered flowing to an atmosphere containing air. The gravitational acceleration used was -9.8 m/s^2 in Y direction. To the Newtonian fluid was used the viscosity value equal to the sum of the polymer and solvent viscosities ($\eta_P + \eta_S$). The difference between the two polymeric fluids analyzed is only on the value of λ parameter. The value of λ is responsible to the deviation of the Newtonian like fluids and then an increase of this value results in an increase of the elasticity phenomena for the same flow conditions. To Newtonian's like fluids the value of λ is assumed as being null.

Table 1: Physical properties of fluids in simulations.

Fluid	Parameters	*Die Swell*
Polymer	$\rho[kg/m^3]$	803.87
	$\eta_S[Pa.s]$	0.002
	$\eta_P[Pa.s]$	1.20
	$\lambda[s]$	0.03 and 3.0
	$\alpha[--]$	0.15
Air	$\rho[kg/m^3]$	1.0
	$\eta_S[Pa.s]$	0.0
	$\eta_P[Pa.s]$	1.48×10^{-5}
	$\lambda[s]$	0.0
	$\alpha[--]$	0.0

3. Results and Discussion

Analyzing Figure 2, we can observe three main differences comparing Newtonian and viscoelastic behavior:

1) The occurrence of jet ruptures for the viscoelastic case with high λ value. This could be attributed to the fluid accumulation at the nozzle due to elastic behavior. When there is enough accumulation of polymer a sudden acceleration going down is observed, causing a local increase of elongation rate and producing, in this way, the rupture. When jet break, recoil of fluid is also observed due to elasticity.

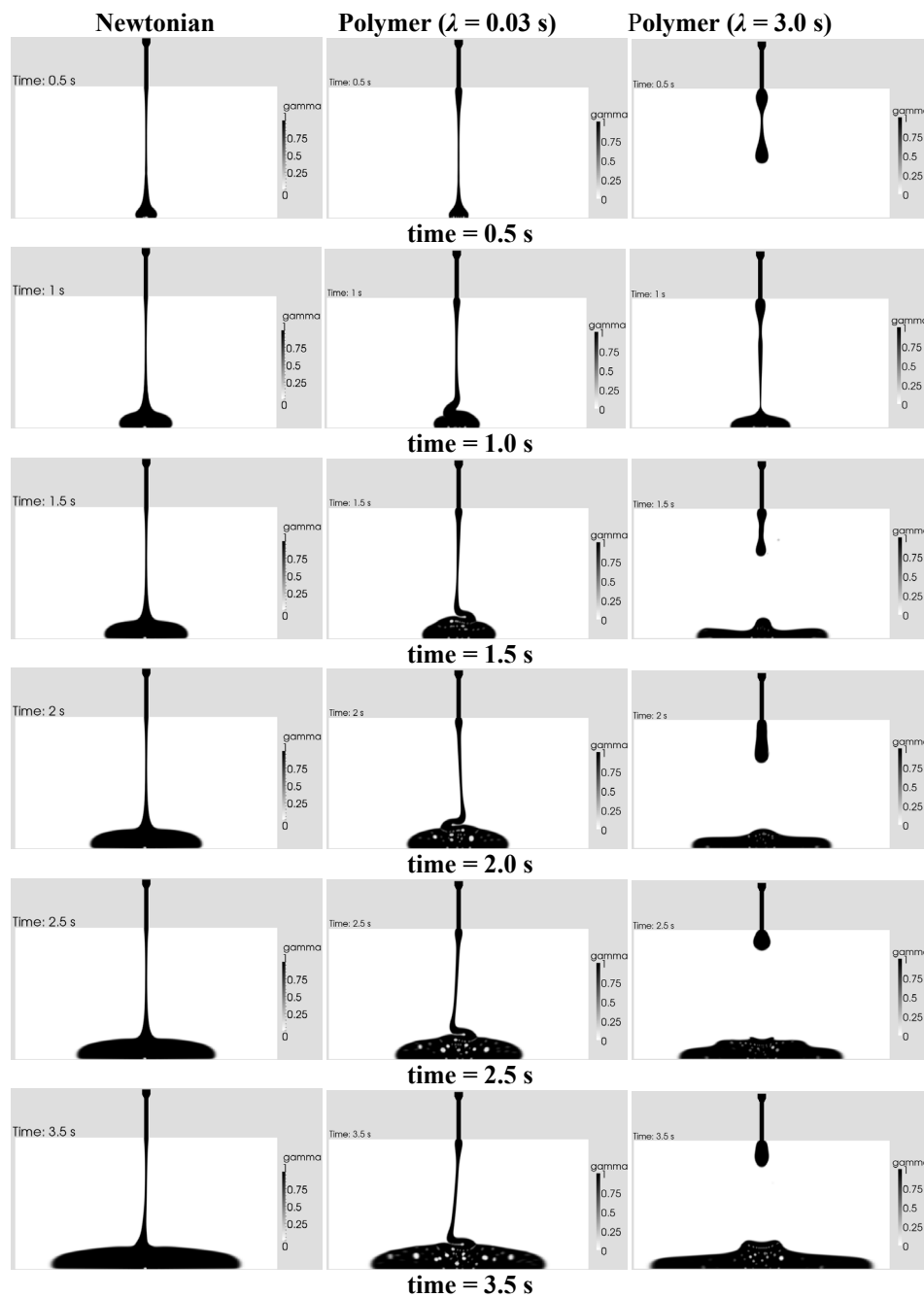*Simulation of Free Surface Viscoelastic Fluid Flow Using the Software OpenFOAM*

Newtonian	**Polymer ($\lambda = 0.03$ s)**	**Polymer ($\lambda = 3.0$ s)**

time = 0.5 s

time = 1.0 s

time = 1.5 s

time = 2.0 s

time = 2.5 s

time = 3.5 s

Figure 2: Results for die swell simulation.

2) The viscoelastic fluid presents a jet that oscillates in the X direction, and due to this effect air bubbles are occluded in the mass of fluid.

3) The die swell only happens for the polymeric fluid and it is also noticed that the die swell happens mainly next to the exit of the channel, where exists the relaxation of the polymeric chains.

In Figure 3, a zoom on the exit of the channel region is shown for a better visualization of the die swell, together with the respective stress field for the time instant 3.5 s. The larger values of stress happen inside the channel and soon at the channel exit, where the die swell exists.
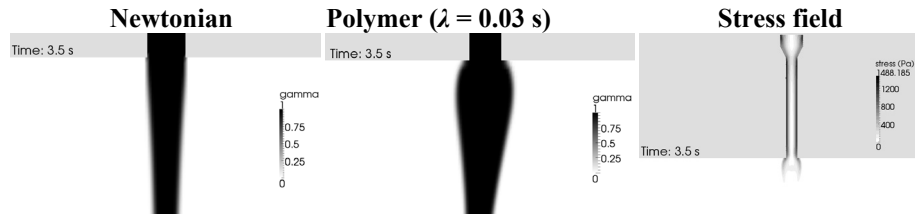

Figure 3: Die swell and stress field at time = 3.5 s.

For all the presented simulations the Crank-Nicholson method was used for time derivatives, with zero initial condition being specified for all variables unless the γ field where the value of 1.0 was specified on the channel. The resulting linear discretized systems were solved by conjugated gradient method (CG), using CG with AMG preconditioning for pressure and BiCGstab with an Incomplete Lower-Upper (ILU) preconditioning for γ, velocity and stress. The absolute tolerance for pressure was 1.0×10^{-7} and for γ, velocity, and stress was 1.0×10^{-6}.

4. Conclusion

Simulation results of viscoelastic fluid flow with free surface were presented. The analyzed cases show results with qualitative agreement with experimentally observed. Die swell was observed in the polymer jet at the channel exit, and jet rupture and air bubbles occlusion in the fluid were verified. The obtained results were satisfactory, justifying the usage of the methodology for free-surface viscoelastic fluid flow using the software OpenFOAM.

5. Acknowledgements

References

C. Macosko. Rheology: Principles, Measurements and Applications. [S.l.]: VHC, 1994.

C. W. Hirt and B. D. Nicols. J. Comp. Phys. 1981, 39, 201-225.

D. A. H. Jacobs. Central Electricity Research Laboratories RD/L/N193, 1980, 62.

D. Khismatullin; Y. Renardy; M. Renardy. J. Non-Newtonian Fluid Mech. 2006, 140, 120-131.

D.J.E. Harvie, J.J. Cooper-White, M.R. Davidson, J. Non-Newtonian Fluid Mech. 2008, 155, Issues 1-2, 67-79.

F. H. Harlow and J. E. Welch. The mac method. Physics of Fluids, 1965, 8, 2182–2189.

H. Rusche. Computational fluid dynamics of dispersed two-phase flows at high phase fractions, Phd thesis, Imperial College, University of London, 2003.

H.G. Weller, G. Tabor, H. Jasak, C. Fureby, Computers in Physics, 1998, 12, 6, 620-631.

J. L. Favero, A. R. Secchi, N. S. C. Cardozo, H. Jasak. Viscoelastic flow simulation: development of a methodology of analysis using the software OpenFOAM and differential constitutive equations. In: 10th International Symposium on Process Systems Engineering, 2009.

M.F. Tomé; A. Castelo; V.G. Ferreira; S. McKee. J. Non-Newtonian Fluid Mech. 2008, 154, 179–206.

O. Hassager. in: Working group on numerical techniques, in: Proceedings of the fifth workshop on numerical methods in non-newtonian flow. J. Non-Newton. Fluid Mech. 1988, 29, 2-5.

O. Ubbink and R.I. Issa, J. Comp. Physics, 1999, 153, 26-50.

R. B. Bird; R. Armstrong; O. Hassager. Dynamics of Polymeric Liquids. John Wiley & Sons, Inc., New York, 1987; 2nd. ed., Vol. 1.

Shape Optimization of Microchannels Using CFD and Adjoint Method

Osamu Tonomura, Manabu Kano, Shinji Hasebe

Dept.of Chemical Engineering,Kyoto University, Kyoto 615-8510, Japan
E-mail: tonomura@cheme.kyoto-u.ac.jp

Abstract

The shape of microchannels is an important design variable to achieve the desired performance. Since most microchannels are, at present, designed by trial and error, a systematic optimal shape design method needs to be established. Computational fluid dynamics (CFD) is often used to rigorously examine the influence of the shape of microchannels on heat and mass transport phenomena in the flow field. However, the rash combination of CFD and the optimization technique based on gradients of the cost function requires enormous computation time when the number of design variables is large. Recently, the adjoint variable method has attracted the attention as an efficient sensitivity analysis method, particularly for aeronautical shape design, since it allows one to successfully obtain the shape gradient functions independently of the number of design variables. In this research, an automatic shape optimization system based on the adjoint variable method is developed using C language on a Windows platform. To validate the effectiveness of the developed system, pressure drop minimization problems of a U-shaped microchannel and a branched microchannel in incompressible flows under constant volume conditions are solved. These design examples illustrate that the pressure drops of the optimally designed microchannels are decreased by 20 % ~ 40 % as compared with those of the initial shape.

Keywords: Microchannels, Shape optimization, CFD, Adjoint method

1. Introduction

In recent years, micro chemical process technology has attracted considerable industrial and academic attention in various fields [1-3]. The main characteristic of micro chemical processes is the small diameter of the channels ensuring short radial diffusion time. This leads to a narrow residence time distribution, high heat and mass transfers. In addition, micro chemical processes have a high surface to volume ratio allowing efficient heat removal and high molar flux. Through the R&D activities on micro chemical processes, the necessity of developing a systematic design method of micro devices has been recognized. The design problems of micro devices are different from those of conventional devices. In a conventional design problem, the unit operations are modeled by using terms such as perfect mixing, piston flow, and overall heat transfer coefficient. In other words, each unit operation is modeled as a lumped parameter system. However, the performance of micro devices depends on the temperature distribution, the residence time distribution and/or the degree of mixing. Therefore, each micro device should be modeled as a distributed parameter system, and the shape of the device must be included in the design variables in addition to the size of the device. Namely, the design problems of micro devices are regarded as shape optimization problems, in which a cost function defined on a flow domain and/or on its boundary is minimized or maximized under several constraints.

With the advances in computational resources and algorithms, Computational Fluid Dynamics (CFD)-based optimal shape design is an interesting field for industrial applications such as aerospace, car, train, and shipbuilding [4]. In such design, the computation of the cost function gradient, i.e., the sensitivity of some performance measure, is the heart of the optimization. Recently, the adjoint variable method [5] has attracted the attention as an efficient sensitivity analysis method, since it allows successfully obtaining the shape gradient functions independently of the number of design variables. In this research, an automatic shape optimization system based on the adjoint variable method is developed using C language on a Windows platform. In addition, in order to validate the effectiveness of the developed system, the optimal shape design problems of the pressure-driven microchannels are solved in the following sections.

2. General Formulation of the Adjoint-based Shape Optimization

The optimization technique based on gradients of the cost function is the easiest way. For each design variable, its value is varied by a small amount, the cost function is recomputed, and the gradients with respect to it are measured. In this case, the number of CFD solutions required for N design variables is $N+1$. Consequently, the gradient-based method requires enormous computation time when the number of design variables is large. In this study, the adjoint variable method is adopted to obtain gradients in a more expeditious manner.

In a fluid dynamic design optimization problem, the cost function depends on design variables and the flow variables due to them. The cost function can be written as

$$I = I(W(\beta), \beta) \tag{1}$$

where I is the cost function, W is the flow variable vector, and β is the design variable vector that represents the surface shape of channels. The cost function I is minimized or maximized subject to partial differential equation (PDE) constraints, geometric constraints, and physical constraints. Examples for the cost function I are drag or pressure drop, for PDE constraints $R(W, \beta) = 0$ the Euler/Navier-Stokes equations, for geometric constraints $g(\beta) \leq 0$ the volume or cross sectional area, and for physical constraints $h(W) \leq 0$ a minimal pressure to prevent cavitation.

The principles of the evaluation of gradients based on adjoint variables are given here [6]. A total differential in the cost function I and the PDE constraint R results in:

$$dI = \left(\frac{\partial I}{\partial W}\right)dW + \left(\frac{\partial I}{\partial \beta}\right)d\beta, \tag{2}$$

$$dR = \left(\frac{\partial R}{\partial W}\right)dW + \left(\frac{\partial R}{\partial \beta}\right)d\beta = 0. \tag{3}$$

Next, a Lagrange multiplier λ is introduced to add the flow equation to the cost function:

$$dI = \left\{ \left(\frac{\partial I}{\partial \boldsymbol{W}} \right) - \boldsymbol{\lambda}^T \left(\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{W}} \right) \right\} d\boldsymbol{W} + \left\{ \left(\frac{\partial I}{\partial \boldsymbol{\beta}} \right) - \boldsymbol{\lambda}^T \left(\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{\beta}} \right) \right\} d\boldsymbol{\beta} \tag{4}$$

This implies that if we can solve:

$$\boldsymbol{\lambda}^T \left(\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{W}} \right) = \left(\frac{\partial I}{\partial \boldsymbol{W}} \right), \tag{5}$$

the variation of I is given by:

$$dI = \left\{ \left(\frac{\partial I}{\partial \boldsymbol{\beta}} \right) - \boldsymbol{\lambda}^T \left(\frac{\partial \boldsymbol{R}}{\partial \boldsymbol{\beta}} \right) \right\} d\boldsymbol{\beta} = \boldsymbol{G} d\boldsymbol{\beta} \; . \tag{6}$$

Equation (6) means that the variation of *I* exhibits only derivatives with respect to $\boldsymbol{\beta}$, and that the shape gradient function $\boldsymbol{G}$ is independent of the number of design variables. In the case that $\boldsymbol{R}$ is PDE, the adjoint equation (5) is also PDE, and the appropriate boundary conditions must be determined. The effectiveness of the adjoint-based shape optimization is emphasized along with the increase in design variables.

3. A Shape Optimization System Development

In this research, an automatic shape optimization system based on the adjoint variable method is developed using C language on a Windows platform. The procedures for building the system are shown in Fig. 1. In principle, after a new shape is obtained, a new grid is generated, and the solution is restarted. For every design cycle, the following steps are required:

1) assume an initial shape,
2) generate computational grids,
3) solve the flow equations, viz. the Navier-Stokes equations and the continuity equations, for deriving the flow velocity and the pressure,
4) solve the adjoint equations to obtain the set of Lagrange multipliers,
5) calculate the shape gradient functions,
6) obtain a new shape by moving each point on the boundary,
7) go to step 2 unless the change in the cost function is smaller than a desired convergence parameter.

Each design cycle requires a numerical solution of both the flow and the adjoint equations, whose computational time is roughly twice that required to obtain the flow solution.

4. Case Studies

Flow in microchannels is driven by the pressure difference or the electric potential between inlet and outlet. For pressure-driven flow, an important issue is how to reduce the pressure drop required to realize a desired flow rate in a microchannel. Curved microchannels and/or branched microchannels are often used to provide long flow
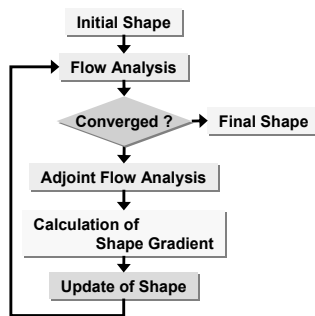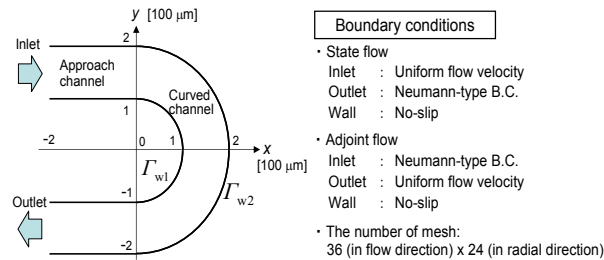
Figure 1. Flow chart of shape optimization.

Figure 2. Analytical domain and boundary condition for a U-shaped microchannel.

passage in a compact device. Modification of the shape of curved channels and/or branched channels may decrease the pressure drop. In this study, two design examples are presented to demonstrate the effectiveness of the developed system for microchannel shape optimization problems. The adjoint variable method is applied to all gradient computations. For convenience, the physical coordinates system is transformed to computational coordinates in the flow and adjoint flow analysis. The two-dimensional computations in these case studies are performed on Windows Intel® 3.0 GHz Pentium 4 processors.

4.1. U-shaped Microchannel

The first design example is a shape optimization problem of U-shaped microchannels in incompressible flows. The goal is to minimize pressure drop for inlet Reynolds numbers: $Re = 10$. The initial shape of the U-shaped microchannel and the main design conditions are shown in Fig. 2. The width of the initial shape is 100 μm. The curved channel is connected with inlet and outlet straight channels. The total number of mesh is 864. The design boundaries are assigned to Γ_{w1} and Γ_{w2}, and the design variables are associated with the grid points on both design boundaries. For pressure-driven liquid flow in a microchannel, the no-slip boundary condition is usually valid. The streamwise velocity component at the entrance is specified, and the transverse velocity component at the entrance is assumed to be zero. The prescribed pressure $p = 0$ is assumed at the exit boundary.

Design results at $Re = 10$ are presented here. Figure 3 shows the pressure distributions for the initial shape, the final shape under no volume constraint, and the converged shape under a constant volume constraint. Under no volume constraint, the width of the curved channel is widened, and the shape of the channel is significantly modified. The wider channel makes the flow velocity lower, and a large reduction of pressure drop can be achieved. On the other hand, under a constant volume constraint, both the inside and outside surfaces of the curved channel are moved toward the direction of the negative X axis, and the flow passage is shortened.

Under a constant volume constraint, the cost function is converged in 92 design iterations and pressure drop is reduced by 27.6 %, as compared with that of the initial
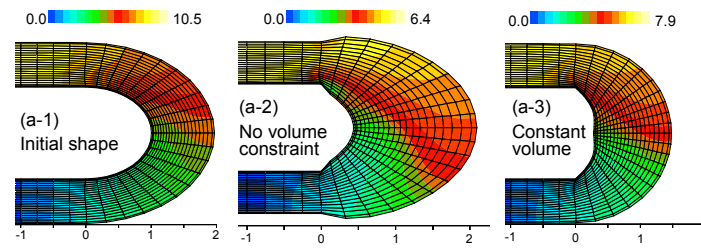
Figure 3. Pressure distributions: (a-1) initial shape, (a-2) final shape under no volume constraint, (a-3) optimal shape under a constant volume constraint. Reference frame is prepared below each shape.

curved shape. Each design iteration requires approximately 10 seconds. On the other hand, without volume constraint, the design cycles are stopped in 30 design iterations due to the fluctuation of the cost function, and pressure drop is decreased by 39.3 %, as compared with that of initial curved shape.

4.2. Branched Microchannel

The second design example targets the shape optimization of a branched microchannel shown in Fig. 4. The cost function is to minimize the pressure drop. Fluid is fed into the branched channel through a boundary Γ_i and comes out from two boundaries Γ_{o1} and Γ_{o2}. Uniform flow u_{in} is assumed at the inlet boundary, and prescribed pressure $p = 0$ is assumed at the exit boundaries. A no-slip condition is imposed on the bounding wall. The design boundaries are assigned to Γ_{w1}, Γ_{w2}, and Γ_{w3}, and the design variables are associated with the grid points on the design boundaries. The Reynolds number is defined by the uniform flow velocity u_{in} and the width of inlet boundary Γ_i. The mesh consists of 1000 elements. The numerical results of this design problem at Re = 1 and Re = 100 are presented in Fig. 5. In case of Re = 1, although 70 rounds of design cycles are required till convergence for 101 design variables, which are assigned on the channel surface, after 50 design cycles the decrease in pressure drop is small. The pressure drop of the optimally designed branched channel is decreased by about 33.1 % as compared with that of the initial shape. The design result at Re = 100 shows a 38.0 % reduction of pressure drop. As Re becomes large, the T-branched channel is transformed into Y-branched channel, in which fluid flows more smoothly without collision with wall. According to the results of two design examples of U-shaped microchannels and branched microchannels, the validity of the developed shape optimization system is confirmed.

5. Conclusions

In this work, an automatic shape optimization system based on the adjoint variable method is developed by using C language on a Windows platform. Since the pressure drop in microchannels is an important characteristic related to the energy demand for process optimization, the developed system is applied to the pressure drop minimization problems of microchannels. The last section demonstrates by representative examples that the adjoint variable method can be used to formulate computationally feasible procedures for the shape design of pressure-driven microchannels. The computational
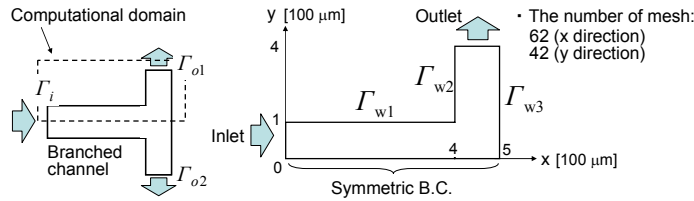
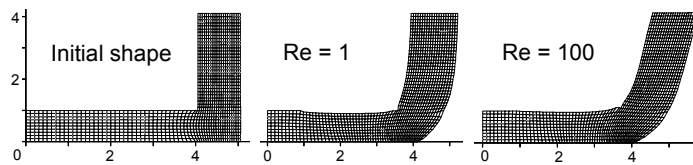Figure 4. Design conditions and assumptions of a branched microchannel.

Figure 5. Initial shape (left) and final shapes at Re = 1 (middle) and 100 (right) under the constant volume constraint.

time of each design cycle is of the same order as two flow solutions, since the adjoint equation is of comparable complexity to the flow equation. The developed system is quite general and is not limited to particular choice of cost function. Our future work will focus on the extension of the developed system to shape optimization problems of thermo-fluidic microdevices.

6. Acknowledgements

References

[1] V. Hessel, S. Hardt, H. Lowe, 2005, Chemical Micro Process Engineering: Fundamentals, Modelling and Reactions, Wiley, VCH, Weinheim.

[2] N. Kockmann, 2006, Micro Process Engineering: Fundamentals, Devices, Fabrication, and Applications, Wiley, VCH, Weinheim.

[3] M. Kano, T. Fujioka, O. Tonomura, S. Hasebe, M. Noda, 2007, Data-based and Model-based Blockage Diagnosis for Stacked Microchemical Processes, Chem. Eng. Sci., 62, 1073-1080.

[4] W. K. Anderson, V. Venkatakrishnan, 1999, Aerodynamic Design Optimization on Unstructured grids with a Continuous Adjoint Formulation, Computer & Fluids, 28, 443-480.

[5] O. Soto, R. Löhner, C. Yang, 2004, An Adjoint-based Design Methodology for CFD Problems, International Journal of Numerical Methods for Heat & Fluid Flow, 14, 6, 734-759.

[6] S. Kim, J. Alonso, A. Jameson, 2004, Multi-element High-lift Configuration Design Optimization Using Viscous Continuous Adjoint Method, Journal of Aircraft, 41, 1082-1097.

Modeling and Simulation of Particle Size Distribution in Slurry-Phase Olefin Catalytic Polymerization Industrial Loop Reactors

Vassileios Touloupides[α,b], Vassileios Kanellopoulos[α], Apostolos Krallis[α], Prokopis Pladis[α], and Costas Kiparissides[α,b]

[a]Chemical Process Engineering Research Institute, 6th km Charilaou-Thermi Road, Thessaloniki 57001, Greece
[b]Department of Chemical Engineering, Aristotle University of Thessaloniki, Thessaloniki 54006, Greece

Abstract

In the present study, a multi-scale, multi-phase, dynamic model is developed for the determination of the distributed properties (i.e., particle size distribution (PSD), molecular weight distribution (MWD)) of polyolefins produced in industrial catalytic gas and slurry phase olefin polymerization reactors (Scheme 1). The polymer MWD is determined by employing a generalized multi-site, Ziegler-Natta (Z-N) kinetic scheme (including site activation, propagation, site deactivation and site transfer reactions) in conjunction with the well-known method of moments. All the thermodynamic calculations are carried out using the Sanchez-Lacombe Equation of State (S-L EOS) (Kanellopoulos et al., 2006). A detailed population balance approach is employed to predict the PSD. The population balance model is combined with a single particle model and the comprehensive kinetic model to predict the properties of the final product in the reactors. Numerical simulations are carried out to investigate the effect of mass transfer limitations on the molecular and morphological properties of the produced polymer.

Keywords: Molecular Weight Distribution, Particle Size Distribution, Slurry Loop Reactors, Olefin Catalytic Polymerization

1. Introduction

High and low density polymers are commercially manufactured in gas and slurry phase olefin polymerization reactors using high activity transition metal catalysts such as Z-N catalysts, Phillips-Chromium oxide catalysts and supported metallocene catalysts. Gas phase processes provide an attractive environment for olefin polymerization with no liquid inventory. They process good versatility to produce a broad range of molecular weights and densities. To be competitive with slurry phase processes, gas phase reactors must operate close to the dew point of the monomer mixture in order to obtain high monomer concentration and high yields. Due to the poor heat transfer characteristics in the gas phase, particle overheating and melting must be avoided. Continuous slurry polymerization, using heterogeneous Z-N catalysts, is one of the most commonly employed processes for the production of polyolefins (Fontes and Mendes 2005). The main reasons for the wide use of slurry phase loop reactors are: (i) their simple design and operation, (ii) their well-defined mixing conditions, (iii) their excellent heat transfer capabilities, (iv) the low power requirements and (v) the high conversion rates (Touloupides et al., 2010).

Scheme 1. The multi-scale modeling approach.
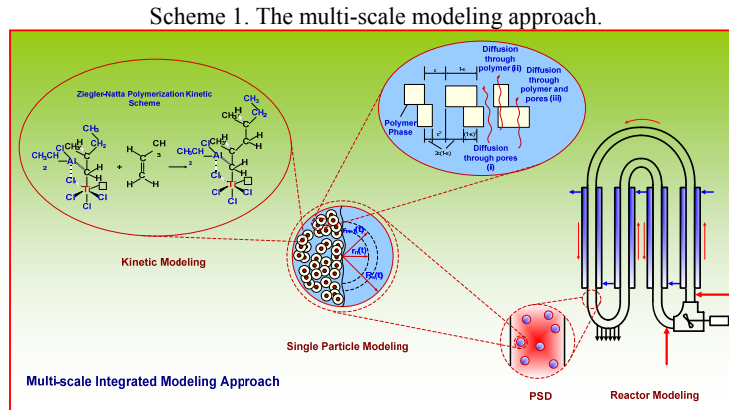
The present multi-scale model utilizes linking models at four different length and time scales, namely, a kinetic model, a single particle model, a population balance model and a reactor model (Dompazis, et al., 2008).

2. Kinetic Modeling at the Kinetic Level

To calculate the monomer(s) consumption rate(s) and molecular weight developments of polyolefins produced over a heterogeneous Z-N catalyst, a generalized two-site kinetic model is employed (Table 1). The kinetic mechanism comprises a series of elementary reactions, including site activation, propagation, site deactivation and site transfer reactions. P_0^k and D_n^k denote the concentrations of the activated vacant catalyst sites of type "k" and "dead" copolymer chains of length "n" produced at the "k" catalyst active site, respectively. The symbol $P_{n,i}^k$ denotes the concentration of "live" copolymer chains of total length "n" ending in an "i" monomer unit, produced at the "k" catalyst active site.

Table 1. Kinetic mechanism of ethylene-propylene copolymerization over a Z-N catalyst.

Activation by aluminum alkyl:	$S_p^k + A \xrightarrow{k_{aA}^k} P_0^k$
Chain initiation:	$P_0^k + M_i \xrightarrow{k_{0,i}^k} P_{1,i}^k$
Propagation:	$P_{n,i}^k + M_j \xrightarrow{k_{p,ij}^k} P_{n+1,j}^k$
Spontaneous deactivation:	$P_*^k \xrightarrow{k_{dsp}^k} C_d^k + D_n^k$
Chain transfer by hydrogen (H$_2$):	$P_{n,i}^k + H_2 \xrightarrow{k_{tH,i}^k} P_0^k + D_n^k$

The average molecular properties of interest (i.e., number- and weight- average molecular weights) and the polydispersity index can be calculated in terms of the overall "bulk" moments, ξ_ν, of "live" and "dead" polymer chains produced over the N_s catalyst active sites:

$$M_n = (\xi_1/\xi_0)\overline{MW} = \left(\sum_{k=1}^{N_s} \xi_1^k \Big/ \sum_{k=1}^{N_s} \xi_0^k \right)\overline{MW} \qquad \text{(Eqn. 1.)}$$

Modeling and Simulation of Particle Size Distribution in Slurry-Phase Olefin Catalytic Polymerization Industrial Loop Reactors

Weight-average molecular weight:

$$M_w = \left(\xi_2/\xi_1\right)\overline{MW} = \left(\sum_{k=1}^{N_s} \xi_2^k \Big/ \sum_{k=1}^{N_s} \xi_1^k\right)\overline{MW} \qquad \text{(Eqn. 2.)}$$

where $\overline{MW}$ is the average molecular weight of the repeating unit of the copolymer chains (i.e., $\overline{MW} = \sum_{i=1}^{N_m} \Phi_i MW_i$). To reconstruct the MWD of the polyolefin produced at the "k" catalyst site, the two-parameter Schultz-Flory distribution was employed:

$$W^k(x) = \frac{y^k \left(xy^k\right)^{z^k} e^{-xy^k}}{e^{\ln\left[\Gamma\left(z^k+1\right)\right]}} \quad ; \quad k = 1,2,\dots N_s \qquad \text{(Eqn. 3.)}$$

where $W^k(x)$ denotes the mass fraction of the polymer chains with a degree of polymerization x, produced at the "k" catalyst active site. Note that the Schultz-Flory parameters (i.e., z, y) can be expressed in terms of the "bulk" moments ($\xi_0^k, \xi_1^k, \xi_2^k$). The overall MWD will be given by the weighted sum of all polymer fractions produced over the different catalyst active sites.

$$W_t(x) = \sum_{k=1}^{N_s} W^k(x)\left(\xi_1^k/\xi_1\right) \qquad \text{(Eqn. 4.)}$$

where ξ_1 $\left(= \sum_{k=1}^{N_s} \xi_1^k\right)$ is the total polymer mass produced over the catalyst active sites.

3. Modeling at the Particle Level: The Random Pore Polymeric Flow Model

To simulate the growth of a single polymer particle, the random pore polymeric flow model (RPPFM) of (Kanellopoulos, et al., 2004) was employed. In the RPPFM, the growing catalyst/polymer particle is assumed to be spherical. Accordingly, the pseudo-homogeneous catalyst/polymer phase is approximated by a series of polymer shells, each one exhibiting evenly distributed properties. It is also assumed that monomer diffusion and heat conduction occur only in the radial direction, while diffusion of all the other species (e.g., polymer chains) is negligible. Transfer of the various molecular species (e.g., monomer(s), hydrogen) from the bulk gaseous phase to the catalyst active sites is assumed to occur via a dual diffusion mechanism (i.e., diffusion through the open catalyst/particle pores and diffusion through the amorphous polymer phase of the semi-crystalline polymer). As a result, the overall monomer transport rate will largely depend on the catalyst/particle morphology, which continuously changes with the polymerization time.

4. Calculation of PSD

To calculate the dynamic evolution of PSD in a catalytic gas or/and slurry phase reactor a dynamic population balance model needs to be solved together with the system of differential equations describing the radial monomer(s) concentration and temperature profiles in a single particle (Kanellopoulos, et al., 2004).
According to the developments of (Dompazis, et al., 2008) the following dynamic population balance equation, accounting for particle growth and particle agglomeration can be derived:

$$\frac{\partial n_p(D,t)}{\partial t} + \frac{\partial\left[G(D)n_p(D,t)\right]}{\partial D} = B(D,t) - D(D,t) + \frac{1}{W}\left[F_c n_c(D,t) - F_p n_p(D,t)\right] \qquad \text{(Eqn. 5.)}$$

where $n_p(D,t)$ and $n_c(D,t)$, expressed in (#/g/cm), are the corresponding number density function of the particles in the reactor and in the feed stream, respectively. The term $n_p(D,t)dD$ denotes the number of particles in the size range $(D, D+dD)$ per mass of polymer particles. $K_{ag}(D,D')$ is a temperature and particle size dependent functional, governing the agglomeration rate of particles of sizes D and D'. To calculate the individual particle growth rate and temperature profile, one has to solve the monomer and energy balances for each discrete class of particles (Kanellopoulos, et al., 2004). The particle growth rate, $G(D)$, can be expressed in terms of the overall particle polymerization rate, $R_{pp}(D)$, as follows:

$$G(D) = 2R_{pp}(D)/\rho_p \pi D^2 \qquad \text{(Eqn. 6.)}$$

Moreover, one can easily show that the steady-state mass balance in the reactor will be given by the following equation:

$$F_p = F_c + W \int_{D_{\min}}^{\infty} G(D)n_p(D,t)d(\frac{\rho_p \pi D^3}{6}) \qquad \text{(Eqn. 7.)}$$

Notice that the second term on the right hand side of eqn.7. accounts for the total polymer production rate in the reactor. The number- and weight- average molecular weights for all polymer particles present in the reactor, at time t, will be given by the respective weighted sum (i.e., $M_n(D)p_p(D,t)dD$ and $M_w(D)p_p(D,t)dD$), calculated over the total variation of D (Dompazis, et al., 2008). Similarly, the MWD for all polymer particles in the reactor, at time t, will be given by the weighted sum of $W_{tot}(x,D)p_p(D,t)dD$ term, calculated over the total variation of D (Dompazis, et al., 2008).

5. Results and Discussion

The multi-scale integrated model described above was subsequently employed to assess the effect of reactor media on the distributed molecular and morphological polymer properties (i.e., MWD, PSD), in catalytic gas phase and slurry phase, ethylene-propylene copolymerization industrial scale reactors.

In gas phase processes gaseous monomer(s) are introduced in the reactor at certain conditions (i.e., pressure, temperature). The transport of monomer(s) from the bulk gas phase to the catalyst active sites strongly depends on the internal and external mass transfer resistances. In slurry phase processes the transport of gaseous monomer(s) over a gas-liquid boundary is of profound interest, since controls the rate of monomer(s) transport from the reactor media to the catalyst active sites, where the polymerization takes place (Kittilsen, et al., 2001).

Mass transfer resistances can play an important role in the slurry polymerization processes. In particular, when highly active catalysts are employed, mass transfer can become the rate limiting step (Kittilsen et al., 2001). In the present study it was assumed that gaseous monomer (i.e., ethylene), under similar to the gas phase process conditions is introduced into the slurry olefin polymerization reactor. Accordingly, in such a process, the monomer partially dissolves in the solvent (i.e., n-hexane), and then diffuses towards the growing catalyst/polymer particles through the solvent media.

Modeling and Simulation of Particle Size Distribution in Slurry-Phase Olefin Catalytic Polymerization Industrial Loop Reactors
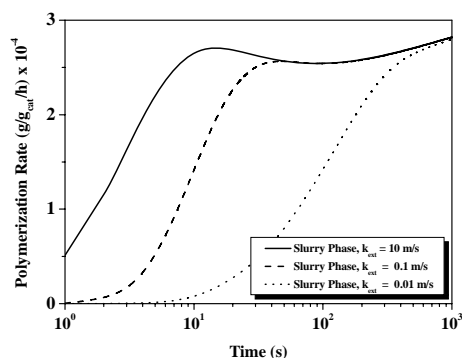

Figure 1: Effect of the external mass transfer limitations on the particle polymerization rate.
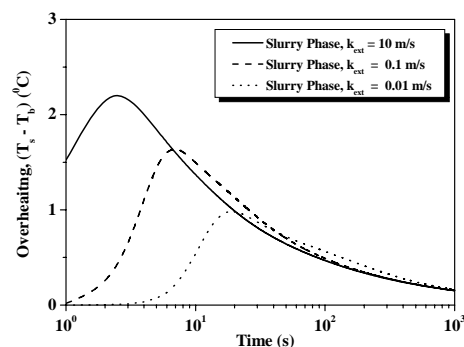

Figure 2: Effect of the external mass transfer limitations on the particle overheating.

In Figures 1 and 2, the effect of the external mass transfer limitations on the polymerization rate and particle overheating, in a growing catalyst/polymer particle in a slurry phase reactor is depicted. It is apparent that the transport of gaseous monomer (e.g., ethylene) in the diluent significantly affects the rate of monomer transfer from the bulk liquid phase to the active catalyst sites, especially during the first few seconds of the polymerization. Thus, as the value of the K_{ext} increases (i.e., negligible external mass transfer limitations) the monomer concentration at the external particle's surface reaches its final thermodynamic value more quickly. As a result, the time needed for the polymerization rate to attain its maximum value decreases as the value of K_{ext} increases (see Figure 1). In fact, catalyst/polymer particles of the same size fed into a slurry reactor exhibit different polymerization rates due to differences in the external mass transfer resistances. On the other hand, due to the sufficient heat removal, particle overheating is not affected by the presence of the external mass transfer limitations in the liquid phase (see Figure 2). The evolution of the polymerization rate and temperature in a single growing polymer particle strongly depends on the reactor media (i.e., slurry or gas). In Figure 3, the effect of the reactor media on the polymerization rate-time histories of a single particle is shown. As can be seen, in gas phase, particle polymerization rate attains higher values than the corresponding slurry phase process due to the lower monomer mass transfer limitations at the external particle boundary layer. In Figure 4, the time evolution of particle overheating is depicted for the studied polymerization processes. It is apparent that in slurry phase polymerization, the growing polymer particle exhibits lower particle overheating due to better heat removal.
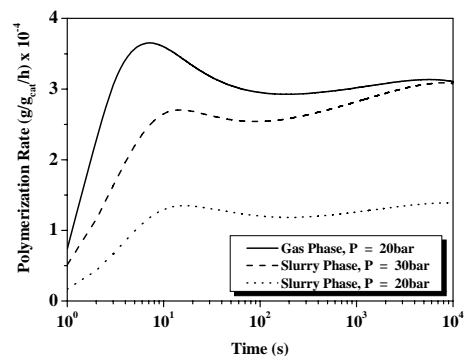

Figure 3: Effect of the reactor type on particle polymerization rate.
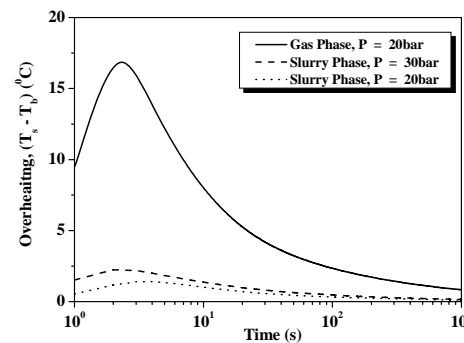

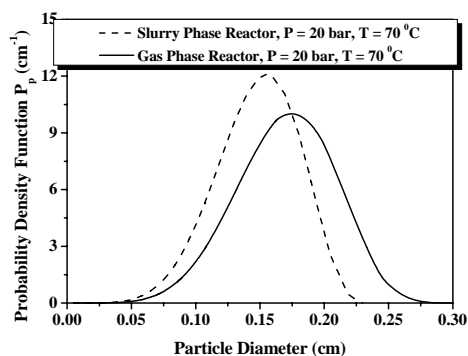Figure: 4. Effect of the reactor type on particle overheating.
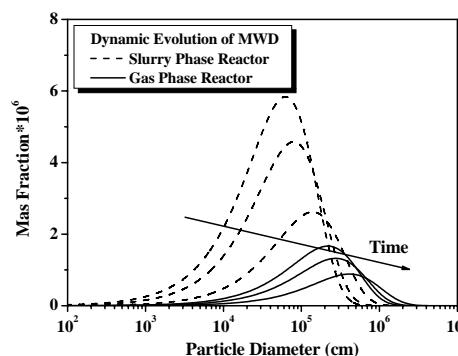
Figure 5: Effect of the reactor type on PSD.

Figure 6: Effect of the reactor type on MWD.

Figure 5, illustrates the effect of the polymerization process on the PSD developments in the reactors. As can be seen, PSD of polymer particles produced in the gas phase reactor becomes broader while, its peak value is shifted to higher particle diameters. This can be explained by the fact that the individual particle growth in gas phase is significantly higher than the corresponding particle growth in the slurry phase. In Figure 6, the dynamic evolution of the MWDs of polymer particles produced in gas phase and slurry phase reactors is demonstrated. It should be pointed out that in gas phase polymerization process, the time required for the MWD to reach its final steady state value is less than the corresponding time of slurry reactor.

6. Conclusions

In the present study, a comprehensive multi-scale mathematical model is developed to predict the dynamic evolution of morphological (i.e., particle size distribution (PSD)) and molecular (i.e., molecular weight distribution (MWD)) distributed polymer properties in catalytic Ziegler-Natta (Z-N) gas and slurry phase olefin polymerization reactors. The multi-scale description of the continuous olefin polymerization reactors utilizes models at three different levels, namely, a kinetic model, a single particle model and a population balance model. The effect of reactor media (e.g., slurry vs gas phase) on the MWD and PSD of the produced polymer particles is thoroughly analyzed.

References

V. Touloupides, V. Kanellopoulos, P. Pladis and C. Kiparissides (2010). Modeling and Simulation of an Industrial Slurry-phase Catalytic Olefin Polymerization Reactor Series. *Submitted to Chem. Eng. Sci.*

G. Dompazis, V. Kanellopoulos. V. Touloupides and C. Kiparissides (2008). Development of a Multi-scale, Multi-phase, Multi-zone Dynamic Model for the Prediction of Particle Segregation in Catalytic Olefin Polymerization FBRs'. *Chem. Eng. Sci.,* 63, 4735-4753.

V. Kanellopoulos, D. Mouratides, P. Pladis and C. Kiparissides (2006). Prediction of Solubility of α-olefins in Polyilefins Using a Combined Equation of State-Molecular Dynamics Approach. *Ind. Eng. & Chem. Res.*, 45, 5870-5878.

C.H. Fontes and M.J. Mendes (2005). Analysis of an Industrial Continuous Slurry Reactor for Ethylene–butene Copolymerization. *Polymer*, 46, 2922–2932.

V. Kanellopoulos, G. Dompazis, B. Gustafsson and C. Kiparissides (2004). Comprehensive Analysis of Single-particle Growth in Heterogeneous Olefin Polymerization: The Random-Pore Polymeric Flow Model. *Ind. Eng. & Chem. Res.*, 43, 5166-5180.

P. Kittilsen, R. Tøgersen, E. Rytter and H. Svendsen (2001). Determination of Gas-liquid Mass-transfer Limitations in Slurry Olefin Polymerization. *Ind. Eng. & Chem. Res.*, 40, 1329.

Application of Different Turbulence Models to Study the Effect of Local Anisotropy for a Non-Premixed Piloted Methane Flame

Claudia Pfeiler, Harald Raupenstrauch

Chair of Thermal Processing Technology, Department Metallurgy, University of Leoben, Franz-Josef-Str. 18 8700 Leoben, Austria, claudia.pfeiler@unileoben.ac.at

Abstract

Turbulent non-premixed combustion of gaseous fuels is of importance for many technical applications, especially for the steel and refractory industry. Accurate turbulent flow and temperature fields are of major importance in order to predict details on the concentration fields. The performances of the realizable k-ε and the RSM turbulence model are compared. Detailed chemistry is included with the GRI-Mech 3.0 mechanism in combination with the laminar flamelet combustion model. The combustion system selected for this comparison is a piloted non-premixed methane flame surrounded by co-flowing air. The simulation results are compared with experimental data of the "Sandia Flame D" published by the international TNF workshops on turbulent flames. For simplification a lot of steady-state flame simulations are performed axisymmetrical in 2D. Simple RANS models do not account for the local anisotropy in turbulent flows. To consider this effect a 3D calculation and the application of the RSM turbulence model, which accounts for these anisotropy, is necessary. In axially symmetric 2D flame simulation the realizable k-ε and in 3D the RSM give unexpected similar results. But still the predicted turbulence and temperature field shows some differences to the experimental data. A modification of a single empirical model constant for the turbulence helped to get better results in both, 2D and 3D.

Keywords: burner, flamelet, turbulence, combustion, non-premixed, flame D

1. Introduction

In the past computational capabilities have been low. In industrial combustion devices, such as furnaces, simpler models have to be used because of computer capability limitations. Therefore, computational models have to be applied in a resource friendly way. One of them, the eddy dissipation model of Magnussen, based on chemical equilibrium and fast chemistry, has been employed e. g. to model the temperature distribution in industrial kilns or pollutant concentration fields in turbulent flames [1, 2]. Further improvements of CFD software and computer hardware have focused the modeling potential on detailed chemical kinetics including non-equilibrium chemistry or detailed pollutants studies in technical flames and furnaces. The direction is to get not only tendencies but also realistic values of temperature and pollutants concentrations. Composition PDF Transport Models give nowadays the most accurate predictions especially for kinetically controlled species such as NO_x or CO. PDF is mainly used for detailed studies on flames including ignition or extinction phenomena. Even if an ISAT algorithm is used, the model is computationally expensive [3, 4]. A method to combine detailed chemistry and turbulence within a moderate computation time, suitable for the prediction of industrial flames and also kilns, is the Flamelet model which is adapted for non-equilibrium effects [5, 6]. The purpose of the present work is to evaluate the possible combustion and turbulence models, which can be used within a moderate computation time

also for big geometries, such as industrial kilns. Numerical simulation results for turbulent flames and comparisons with experimental data are performed.

2. Model Description

Mass and momentum conservation in the computational domain were achieved by solving the continuity equation and the Navier Stokes equations for Newtonian fluids. The solution yields the pressure and velocity components at every point in the 2D and 3D domain. The solution of the flow and the mixing field is performed in ANSYS Fluent, a Computational Fluid Dynamics (CFD) code, with the stationary laminar flamelet model (SLFM) used for chemistry [6].

2.1. Flamelet Model

The Flamelet model is a method to combine detailed chemical reactions with turbulent flow within moderate computational time. It is based on the assumption, that the turbulent non-premixed flame is composed of an ensemble of one-dimensional discrete laminar counter-flow diffusion flames called flamelets. As the velocity of the counter flowing jets is increased, the flame is strained and increasingly departs from chemical equilibrium. The governing equations can be simplified to one dimension along the axis of the fuel and oxidizer jets. There, complex chemistry calculations can be performed. The width of these flamelets is assumed to be smaller than the Kolmogorov scale, which separates combustion and turbulence at different scales [7]. The flamelet equations are derived applying a coordinate transformation with the mixture fraction as an independent coordinate to the governing equations for the temperature, T, and the species mass fractions, Y_i, [8],

$$\rho\frac{\partial T}{\partial t} = \frac{1}{2}\rho\chi\frac{\partial^2 T}{\partial Z^2} - \frac{1}{c_p}\sum_i H_i S_i + \frac{1}{2c_p}\rho\chi\left[\frac{\partial c_p}{\partial Z} + \sum_i c_{p,i}\frac{\partial Y_i}{\partial Z}\right]\frac{\partial T}{\partial Z} \tag{1}$$

$$\rho\frac{\partial Y_i}{\partial t} = \frac{1}{2}\rho\chi\frac{\partial^2 Y_i}{\partial Z^2} + S_i \tag{2}$$

Here, ρ is the density, $c_{p,i}$ and c_p are the ith species specific heat capacity and mixture-averaged specific heat, respectively. t is the time, Z the mixture fraction, S_i the species reaction rate, H_i the specific enthalpy and χ is the instantaneous scalar dissipation rate defined as $\chi = 2D_Z|\nabla Z|^2$, where D_Z is an adequate diffusion coefficient. It is assumed that the pressure is constant and the Lewis number for all the species is unity. To include the effect of density variation, the modeling of the scalar dissipation is based on [9],

$$\chi(Z) = \frac{a_s}{4\pi}\frac{3\left(\sqrt{\rho_\infty/\rho}+1\right)^2}{2\sqrt{\rho_\infty/\rho}+1}\exp\left(-2\left[\mathrm{erfc}^{-1}(2Z)\right]^2\right) \tag{3}$$

where ρ_∞ is the density of the oxidizer stream, a_s is the flamelet strain rate and erfc[-1] is the inverse complementary error function.

The advantage of the reduction of the complex chemistry to two variables allows the flamelet calculations to be preprocessed, which makes the simulation faster. It is assumed that the flame respond immediately to the aerodynamic strain. Therefore the model cannot capture deep non-equilibrium effects e.g. slow chemistry or ignition.

The conservation equations for the mean mixture fraction, $\bar{z}$, and its variance, $\bar{z}'^2$, are solved during flow field calculation. Z is assumed to follow the presumed β-function PDF, p, [10]. The mean values of mass fractions of species, temperature and density, presented as $\bar{\phi}$ and can be calculated as,

$$\bar{\phi} = \iint \phi(Z,\chi_{st})p(Z,\chi_{st})dZd\chi_{st} \tag{4}$$

where, χ_{st} is the scalar dissipation rate at stoichiometric mixture. The β-PDF is defined by its first two moments. The first moment, the mean scalar dissipation, is

$$\overline{\chi}_{st} = \frac{2\overline{\varepsilon Z}'^2}{k} \tag{5}$$

Here, k is the turbulent kinetic energy and ε is the turbulent dissipation rate.

2.2. Realizable k-ε Model

For the prediction of an accurately spreading rate of round jets the realizable k-ε model is a good opportunity. The Boussinesq approach of this model assumes the turbulent viscosity as an isotropic scalar quantity. The advantage of this approach is the relative low computational cost, although the isotropic assumption is not valid for most of the turbulent flows. The model includes an eddy-viscosity formula and a model equation for dissipation [11].The equation for the turbulent kinetic energy is the same as that in the standard k-ε model. The dissipation equation is based on the dynamic equation of the mean-square vorticity fluctuation. C_μ is computed as a function of k and ε. The model constants C_2, σ_k, and σ_ε have been established to ensure that the model performs well for certain canonical flows [10]. The standard value of the model constant $C_2 = 1.9$ was changed to $C_2 = 1.8$ as recommended in literature [12].

2.3. Reynolds Stress Model (RSM)

Measurements on the Sandia flame D show anisotropic turbulent fluctuations [13]. Compared to the k-ε model the Reynolds stress model is able to account for anisotropic turbulent flows. The Reynolds stress model solves a transport equation for each of the stress terms in the Reynolds stress tensor. RSM is better for situations in which the anisotropy of turbulence has a dominant effect on the mean flow. Since in 3D seven additional transport equations has to be solved, this turbulence model is more computational intensive as the realizable k-ε model.

3. Experimental Data

The flame simulations were performed and compared with experimental data from literature [12, 13]. The flame is a turbulent non-premixed piloted methane flame. A coflowing air was placed around the flame to avoid the influence of air flow in radial direction. The boundary conditions of the measurements are shown in Table 1.

3.1. Sandia Flame D

The Sandia flame D was measured at the Sandia National Laboratories in Livermore, California and at the Technical University of Darmstadt in Germany [12, 13]. This flame is partial premixed with air to prevent almost completely soot formation and to provide a stable flow field. However this flame burns as a non-premixed flame with a single reaction zone near the stoichiometric mixture. In fuel rich regions a significant premixed reaction was not observed [13].

Table 1. Experimental and Simulation Conditions

jet diameter [m]	0.0072 [12]
pilot diameter [m]	0.0182 [12]
jet Reynolds number [-]	22,400 [12]
jet composition CH_4/air [vol.-%]	25/75 [12]
pilot mixture fraction [-]	0.27 [12]
co flowing air velocity [m/s]	0.9 [12]
jet velocity [m/s]	profile [13]
pilot velocity [m/s]	profile [13]
fuel temperature [K]	294 [12]
co flowing air temperature [K]	291 [12]

The Reynolds number of the jet exit is 22,400 with a low probability of localized flame extinction. The pilot flame burns a mixture of gases having the same composition and enthalpy as a CH_4/air mixture at 0.27 mixture fraction. The coflowing air parallel to the flame was about 0.9 m/s. The experiment was performed at room temperature.

4. Simulation Details

Assuming axisymmetrical flow, the calculation domain consists of a two-dimensional axisymmetric plane. The computational domain is descretized into structured elements which are refined around the inlet of the burner nozzle. The grids consist of 2300 elements. The inlet velocity profile for the pilot and burner is positioned at the exit plane of the burner. At the outlet a constant pressure were applied as boundary condition. Inlet turbulence parameters at the inlet were fixed at average values for k and ε, based on measurements [13]. An additional simulation was done in a three dimensional domain for the Sandia flame D. The grid was performed symmetrically with structured elements. In order to get an equivalent comparison, the grid elements are distributed in the same way as for the axisymmetrical grid. Therefore, the grid consists of 228,800 elements. No mesh sensitivity analysis was performed. Material and process data which are used for the simulations are summarized in Table 1. The equations were solved using the ANSYS-FLUENT CFD package. The thermo physical data base of FLUENT was used. To account for detailed chemical kinetic the Gri-Mech 3.0 mechanism was implemented, considering 53 species and 325 reactions [14]. For all simulations presented in this paper, a second order upwind scheme was used for the conservation equation of momentum, turbulent kinetic energy, turbulent dissipation rate, mean mixture fraction and mean mixture fraction variance. The Presto scheme [10] was used for interpolation methods for pressure. Simple [10] was chosen for the coupling between the velocity and the pressure.

5. Results and Discussions

In the present study, the stationary laminar flamelet approach is applied and the turbulence flow field is predicted with the realizable k-ε model in 2D case and RSM in the 3D case. The effect of the assumption of isotropic turbulence of k-ε and the possibility of anisotropic calculation with RSM as well as the change of the empirical turbulence constant C_2 are studied. The results are compared to experimental data [12, 13].

Comparison of the axial temperature is shown in Fig. 1. Radiation is not considered in these studies, since former works have shown less influence of radiation on the temperature distribution for unconfined flames [15].

Both, the 3D RSM and 2D realizable k-ε model predict a temperature maximum which is shifted upstream. Realizable k-ε produces a slightly lower maximum temperature as RSM. Measurements have shown anisotropic turbulence in this flame [13].The estimation that the consideration of the known anisotropic turbulence in this flame with the RSM model can maybe reproduce the measured temperatures, was not achieved. Both models, realizable k-ε and RSM with standard turbulence constants, were not able to reproduce the measurements. The change of the proposed turbulence constant C_2 for both turbulence models to a value of $C_2 = 1.8$, produces a big change as can be seen in Fig. 1. The value of C_2 was recommended for the standard k-ε model [12]. But also for realizable k-ε and RSM it seems to have an important effect for an accurate prediction of turbulent non- premixed jet flames. The change of the constant leads to a later increase of temperature. Realizable k-ε and RSM with $C_2 = 1.8$ predict very well the measurements till the position x/d = 50. After that the decline of temperature is lower than in the experiment. The highest temperature is predicted by RSM with $C_2 = 1.8$. These results demonstrate that realizable k-ε in 2D and RSM model in 3D with $C_2 = 1.8$ are able to predict the flame temperature better as with standard turbulence constants.

*Application of Different Turbulence Models to Study the Effect of Local Anisotropy for
a Non-Premixed Piloted Methane Flame*

In these figures, *x* and *r* are the distances in axial and radial directions and *d* is the diameter of the burner nozzle.
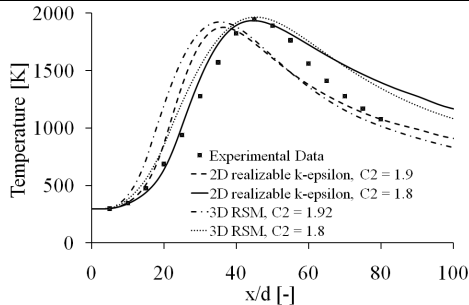

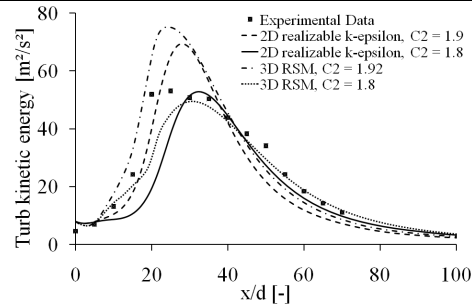Figure 1. Comparison of the axial temperature profiles with experimental data [12].

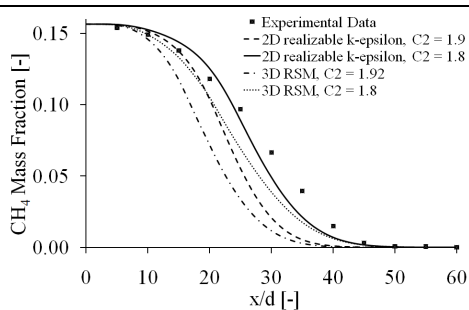Figure 2. Axial turbulent kinetic energy profiles compared with experimental data [13].

Figure 3. Axial methane mass fraction compared with experimental data [12].
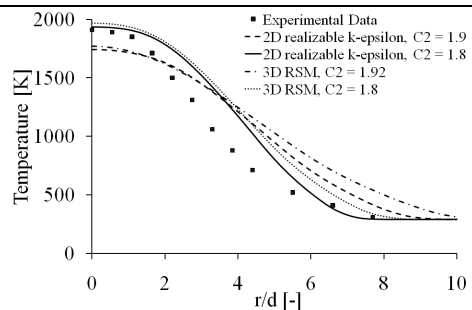
Figure 4. Radial temperature profiles at $x/d = 45$ compared with measurements [12].

The 2D realizable *k-ε* calculation, which is computational less intensive, is able to produce similar results as the more complex 3D RSM model, which can be useful for calculation of bigger simulation domains, e. g. for industrial kilns. The reason for the effect of the change of the turbulence constant C_2 can be shown in Fig. 2. The simulations using the standard turbulence parameters over predict the turbulent kinetic energy in the flame center. A decrease of C_2 in both cases decreases the maximum of the curve. Also a slight shift downwards can be observed. C_2 is a factor for a sink in the dissipation equation. A decrease of that factor leads to a lower reduction of dissipation and a faster reduction of turbulent kinetic energy. Thus, C_2 is the amount of lost rotational energy due to friction. Chemistry is strongly coupled with turbulence. The lower turbulence decreases the mixing of methane and air. Therefore, the consumption of methane is slower. Representatively, the mass fractions of methane are shown in Fig. 3. All of the calculations show a faster decrease in the methane mass fractions in comparison to the measurements. The fastest consumption of methane occures with the predictions by the use of the standard values for $C_2 = 1.9$ and $C_2 = 1.92$ for realizable *k-ε* or RSM, respectively. The complete methane concentration is consumed at the position $x/d = 40$. The experimental data and the simulations with $C_2 = 1.8$ show the total consumption at the position of $x/d = 50$.

Radial profiles of temperature give the same tendency as for the axial profiles above. By changing the turbulence constant the predicted curves are agreeing better with the measurement data (Fig. 4). But here, the accordance is less than for the axial profiles. The calculations predict a later temperature decay along the radial position $x/d = 45$. The decline is the same in case of $C_2 = 1.8$. By contrast, the predicted curves with standard values for the turbulence constants, the decrease in temperature is lower. Due to the shifted temperature

peak at the axis (Fig. 4), the predicted maximum temperature is lower than that in experiment at this position.

6. Conclusions

In the present paper the applicability of two different turbulence models on the predictions on a piloted non-premixed turbulent flame to account for an accurate prediction of the flow and temperature field of a jet flame was investigated. The results of realizable *k-ε* and RSM, together with the Steady Laminar Flamelet Model are compared with measurement data from literature [12, 13]. A better agreement with measurements with the RSM model, which accounts for the anisotropic Reynolds stresses, could not be observed. Discrepancies between predictions and measurements are observed using the standard turbulence constants for both studied turbulence models. These discrepancies are related to an over prediction of turbulent kinetic energy. This leads to a higher mixing and earlier consumption of methane. By changing the turbulence constant C_2 to a value of $C_2 = 1.8$, a general better agreement with the measurements is observed. The changed constant C_2 leads to a lower turbulent kinetic energy, a higher maximum temperature with a temperature peak shifted downstream and slower methane consumption due to a lower turbulent mixing. The previous results indicate that the distribution of temperature and concentration is sensitive to turbulence modeling. Interesting is, that the 2D realizable and the 3D RSM model give similar results. For the industry, which is interested in the simulation of industrial kilns, the combination of SLFM and realizable *k-ε* using a modified constant C_2 is an attractive and accurate method. For the prediction of NO_x an accurate temperature field is of great importance. With a simple change of this turbulence constant, a satisfying temperature, velocity and concentration field can be achieved, even by application of a simpler turbulence model. The advantage is the lower computation time. The change of the turbulence constant C_2 has an effect on the results in case of the turbulent non-premixed Sandia flame D. However, further investigations and comparisons with measurements on further flame and burner configurations should be done to evaluate our present approach.

7. Acknowledgements

References

[1]   B. F. Magnussen and B. H. Hjertager, 16th Symp. on Comb., PA, Pittsburg, 1976.
[2]   S. M. Kaustubh and V. V. Ranade, Asia-Pac. J. Chem. Eng,.3 (2008) 106.
[3]   S. B. Pope, Combustion Theory and Modeling, 1 (1997) 41.
[4]   S. B. Pope, Progress Energy Combustion Science, 11(1985) 119.
[5]   N. Peters, Prog. Energy Combust. Sci., 10 (1984) 319.
[6]   N. Peters, 21st Symp. (Int.) on Combustion, PA, Pittsburgh, 1986.
[7]   N. Peters, Turbulent Combustion. Cambridge University Press, Cambridge, 2000.
[8]   H. Pitsch et al., SAE Paper 962057, SAE, 1996.
[9]   J. S. Kim and F. A. Williams, Eng. Math, 31 (1997) 101.
[10]   FLUENT 6.3 User's Guide, Fluent Inc. (2006).
[11]   T.-H. Shih et al., Computers Fluids, 24 (3) (1995) 227.
[12]   R. S. Barlow, J.-Y. Chen (eds.), 3rd TNF Conf. Proc., Colorado, Boulder, 1998.
[13]   C. Schneider et al., Combustion and Flame, 135 (2003) 185.
[14]   G. P. Smith et al., http://www.me.berkeley.edu/gri_mech/
[15]   M. Baburic, PhD Thesis, Univ. of Zagreb, Fac. Mech. Eng. and Naval Arch., 2005.

Mechanism of drop coalescence at the interface of two immiscible liquids

Giulia Bozzano[a], Mario Dente[a]

[a]*Politecnico di Milano, CMIC Dept., P.zza L. Da Vinci, 32, 20133 Milano, Italy,*
giulia.bozzano@polimi.it

Abstract

Several processes involve two immiscible liquid phases with dispersed drops surrounded by a continuous phase. The drops coalesce for joining their mother phase. This phenomenon occurs in many industrial mixing, separation, as well as in environmental processes. This paper is focused on the problem of the coalescence of a single drop at the planar interface of two immiscible liquids. An original model for coalescence mechanism is proposed. It has been compared with some experimental data of the literature. The comparisons are quite satisfactory.

Keywords: Drop shape, film drainage, drop coalescence mechanism

1. Introduction

In many industrial processes often it is required to produce a dispersion of two immiscible liquids one inside an other to generate a large interfacial area. Of course it can be necessary mainly for enhancing heat and mass transfer among the phases. However, these dispersions need to be separated into their original phases in order to proceed to subsequent process steps. Frequently gravitational settlers are used to separate dispersed drops from the surrounding continuous phase. The dispersed drops move to the interface and, after resting for some time, will coalesce in their reservoir. In order to improve the knowledge on the fluid-dynamics of the separation process, it is convenient to have a better understanding of the mechanism of coalescence. The study of the coalescence of single drop is the first elementary step for the description of the process. Then the mechanism can be extended to systems of drops and foams. Several authors have studied the problem (among others Charles and Mason (1960), Zulfaa and Longmire (2004), Blanchette and Bigioni, (2006), Chen *et al.* (2006)) and mainly regression models are proposed. Here a mechanistic model, based on the analysis of the occurring phenomena, is proposed. The model considers only single step coalescence and the problem of the origin of the so-called coalescence cascade here is only marginally taken into account. The experimental data available for the model validation are relatively few and characterized by a modest level of reproducibility. This facts allowed to introduce some approximations in the mathematical approach to the problem.

2. The mechanism of drop –interface coalescence

Let assume a drop of a phase A rising or falling through an immiscible phase B separated from its reservoir by an interface. The terminal velocity of approach to the interface is dictated by the drag coefficient that depends on the deformation. A convenient theory for evaluating that velocity is reported in Bozzano and Dente (2001 and 2009). When the single drop gets close to the reservoir interface its kinetic energy is quickly dissipated (by the viscous resistance) and its velocity is quite immediately

reduced from several centimeters per second to few microns per second. Capillary waves radially expanding are formed contributing to the energy dissipation. At this stage the film thickness is evaluated by means of an energy balance between variation of kinetic energy and dissipation due to viscosity (the other terms being negligible). The simplified balance is:

$$\frac{d}{dt}\left(m_{app}v^2/2\right) \cong = -E_{diss} \tag{1}$$

(v= actual drop velocity, E_{diss}= energy dissipation per unit time, m_{app} = total mass of the drop included the virtual adhering mass). Due to the thin thickness of the remaining film (when the drop velocity becomes about zero), it is possible to estimate the power dissipated as:

$$E_{diss} = 3/2 \cdot \pi\mu_c \cdot v_\infty^2 R_0^4 / h_f^3 \tag{2}$$

(μ_c=viscosity of the continuous phase, h_f=film thickness, v_∞ = terminal drop approach velocity). The adhering volume is about 0.5 times that of the drop. Substituting in equation 1) the film thickness remaining (after the dissipation of kinetic energy and before the final drainage) is obtained:

$$h_f = \frac{3}{4}\sqrt[3]{\frac{\mu_c R_0^4}{(0.5\rho_c + \rho_d)R_S^2 v_\infty}} \tag{3}$$

(ρ= density, c,d= indexes for continuous and dispersed phase). The drop and of the quasi-planar interface become deformed (for supporting the net weight of the drop). The shape of this latter presents an inflection circle (see fig.1 showing a lateral picture of a water drop at an oil-water interface) and can be symmetrical or not. A thin film of the continuous phase is contained in between; the drop rests there the time necessary for its drainage. The final drainage time can be relatively large, so that the process is apparently showing a static drop shape.

2.1. Shape of the drop

Along the film drainage time, the usual drop shape is similar to an oblate spheroid. Fig.2 shows a sketch with reference to minor and major axis and to the radius at the inflection point of contact with the planar interface. Taking into account the difference of capillary pressure between the top of the drop and the equatorial position, to be balanced by buoyancy force, the following expressions can be derived for the drop and the beneath meniscus:

$$\text{drop: } \sigma\left[\left(\frac{1}{a}+\frac{a}{b^2}\right)-\frac{2b}{a^2}\right]=\Delta\rho gb \qquad \text{meniscus: } \sigma\left[\left(\frac{1}{R_1}+\frac{1}{R_2}\right)\right]=\Delta\rho gy \tag{4}$$
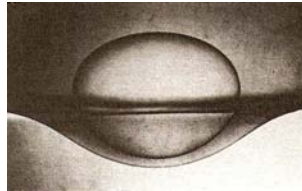

Fig. 1: Drop of water at water-oil interface

Mechanism of drop coalescence at the interface of two immiscible liquids

(g = acceleration due to gravity, $\Delta\rho$ = density difference, σ = interfacial tension, y = vertical coordinate). The meniscus of the quasi-planar interface supporting the drop has two main radius of curvature R_1 and R_2. By linearizing the curvature, it is possible to obtain:

$$\frac{1}{R_1} \cong \frac{1}{r}\frac{dy}{dr} \qquad\qquad \frac{1}{R_2} \cong \frac{d^2 y}{dr^2} \tag{5}$$

By introducing the Eötvös number ($\Delta\rho g D_s^2/\sigma$, D_s = diameter of the equivalent sphere) and defining z=b/a, the following expression is obtained from the first of eqs. (4):

$$Eo = 4\cdot\left[(1-z)(1+z+2z^2)\right]/z^{7/3} \tag{6}$$

from which z can be evaluated. The parameters of the shape of the quasi-static drop gently moving against the interface is then given by: $a = R_s \cdot z^{-1/3}$ $b = R_s \cdot z^{2/3}$. By using the second expression of eq.(4) and substituting the equations (6):

$$\frac{d^2 y}{dr^2} + \frac{1}{r}\frac{dy}{dr} - \frac{\Delta\rho g y}{\sigma} \cong 0 \qquad \Rightarrow \qquad y = y_0 \frac{H_0(i\xi)}{H_0(i\xi_0)} \tag{7}$$

with: $\dfrac{y_0}{a} = (2\cdot z^{4/3}u)/\left[Eo^{1/2}(1-u^2)^{1/2}\right]\cdot\left\{\left[-H_0(i\xi_0)\right]/\left[iH_1(i\xi_0)\right]\right\}$ $\xi = (\Delta\rho g/\sigma)^{1/2} R$

$$\xi_0 = \xi\cdot R_0 = \frac{1}{2}Eo^{1/2}z^{-1/3}u \qquad\qquad u = R_0/a = 1\bigg/\sqrt{1+6z^{1/3}/Eo}$$

H_0 and H_1 are Hankel functions of 0 and 1 order. Other parameters of interest are the radius of curvature R_c, the angle θ and the fraction of drop volume α.

$$\frac{1}{R_c} \cong -\frac{d^2 y}{dx^2}\bigg|_{x=0} = \frac{a^2}{b}\left[1 - \frac{R_0^2}{a^2}\left(1 - \frac{b^2}{a^2}\right)\right]^{3/2} \qquad \cos\theta = \sqrt{\frac{(1-u^2)}{1-(1-z^2)u^2}} \tag{8}$$

$\alpha = V_{sup}/V_{tot}$ $V_{sup} = \pi/3\cdot R_s^3\left[2 + \sqrt{1-u^2}\,(2+u^2)\right]$ (R_s = radius of the equivalent sphere, V_{sup} = drop volume standing over the interface, V_{tot} = drop volume)
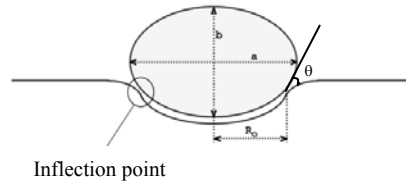

Inflection point

Fig. 2: Drop shape. a = major axis, b = minor axis, R_0= distance from central axis and inflection point, θ = contact angle between drop and interface

2.2. Film drainage and coalescence

As said before viscous forces control film drainage.It is meaningful to observe that as soon as the drainage proceeds, the constant force due to the "net weight" of the drop cannot overcome the viscous resistance (that increases when the film becoms thinner). Then at this final stage the molecular attractive energies, essentially dispersion forces connected to Lennard-Jones potential energy, will constitute the final driving force. By applying a balance among net weight, attractive and dissipative forces, the following expression arises (for describing the final velocity of film drainage):

$$\alpha \frac{4}{3}\pi R_s^3 \Delta\rho g + \frac{\pi^3}{12}\sqrt{2}A\frac{R_0 R_c^{0.5}}{h_f^{5/2}\cos\theta^{3/2}} + A\frac{R_0^2}{6h_f^3\cos\theta^2} = -6\sqrt{2}\pi^2\mu_c\frac{R_0 R_c^{1.5}\cos\theta^{1/2}}{h_f^{3/2}}\frac{dh_f}{dt} \qquad (9)$$

(A = Hamaker constant, related to the refractive index n). When a medium is interposed, A can be evaluated following Mahanty and Ninham (1976):

$$A = \frac{3\pi^2}{2}\left\{\frac{\mu_1^2}{2I_1} + \frac{\mu_0^2}{2I_0} - \frac{2\mu_0\mu_1}{I_0+I_1}\right\} \qquad (10)$$

($\mu_i = (\varepsilon_i - 1)/4\pi$, $I_i = h/(2\pi)\omega_0^{(i)}$, $\varepsilon = n^2$, h=Planck constant, $\omega_0 = 2\pi c/\lambda$, λ=wave length, c = light velocity). A simplified approach to the solution of eq.(10) leads to the following expression for the drainage time:

$$t_{drain} = 63.5\mu_c\frac{R_0^{2/3}R_c^{3/2}(\cos\theta)^{5/12}}{A^{1/6}(\alpha\Delta\rho g R_s^3)^{5/6}}\cdot\frac{1}{1+1.5\left(\dfrac{\cos\theta^{0.5}AR_c^3}{\alpha\Delta\rho g R_s^3 R_0^4}\right)^{1/30}} \qquad (11)$$

Expression (11) shows that the drainage time depends on several physical properties, mainly on the viscosity of the continuous phase. In the case of axi-symmetrical coalescence, the critical boundary for the coalescence is constituted by the line of inflection points. Fig. 3 shows a drop of colored water in seed oil viewed from the bottom (a), the same drop 0.02 s after film drainage (b), and the curls generated by the wake (c). In figure 3a) it is also evident underlying shape of the interface, while in 3b) can be observed the imprint of the drop contact area (= inside inflection circle).

3. Results and comparisons

A first comparison with experimental data is related to the evaluation of the drop shape. They are taken from Hartland (1967) for Golden Syrup(GSy)-Paraffin and Glycerol(Gly)-Water mixture and Paraffin systems, while for water and seed oil has been obtained by the authors. The comparison (Table 1) is very satisfactory.
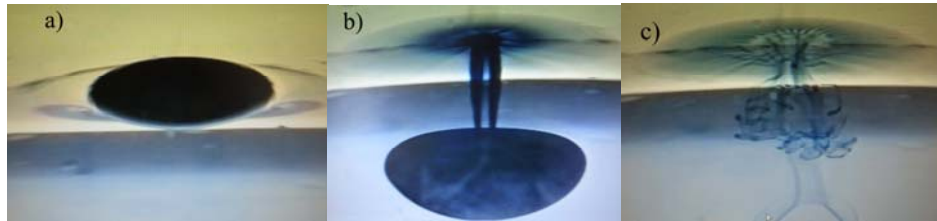

Fig. 3: drop (0.5 ml) coalescence

Mechanism of drop coalescence at the interface of two immiscible liquids

Table 1: Comparisons for the shape of the drop (a e b in cm)

	Volume	Eo/4	b calc.	b exp.	a calc.	a exp.
GSy-Paraffin	0.5	3.15	0.35	0.36	0.59	0.58
	0.25	1.98	0.30	0.31	0.44	0.45
	0.1	1.08	0.24	0.25	0.31	0.33
Gly – Paraffin	0.5	0.43	0.34	0.35	0.59	0.59
	0.25	2.16	0.29	0.31	0.45	0.44
	0.1	1.17	0.24	0.25	0.31	0.33
Water-seed oil	0.7	1.19	0.46	0.43	0.60	0.61
	1.0	1.6	0.50	0.48	0.69	0.71

Table 2: Comparisons of exp. drainage time (Davis et al. 1971) with the calculated one

Drop volume (ml)	Exp. t_{drain} (s)	Calc. t_{drain} (s)	Coalescence Stages
Water-diethyl carbonate			(4-7 exp.)
0.089	2.38÷6.72	8.93	4
0.049	5.35÷12.04	6.75	4
0.204	6.6÷15.42	10.1	4
0.112	0.93÷3.49	9.24	4
0.107	2.32÷2.84	9.17	4
0.09	4.27÷6.81	8.93	4
0.094	1.08÷5.4	9.00	4
0.203	0.82÷11.26	10.1	4
0.275	1.21÷9.87	12.0	5
Water-methyl isobutyl ketone			(several exp.)
0.038	1.03÷2.73	2.32	4
0.026	1.29÷4.05	2.20	4
0.011	2.45÷6.01	1.59	3
0.020	1.95÷6.43	1.74	3
0.0043	1.55÷4.39	1.01	3
0.0052	0.55÷0.69	1.04	2
0.021	0.74÷1.36	1.75	2
0.090	0.52÷2.02	2.55	4
Water-cyclohexanone			(1-2 exp.)
0.0306	4.30÷11.72	8.96	2
0.0952	1.52÷5.32	10.3	2
0.0152	6.87÷17.65	8.08	2
0.0111	1.49÷3.67	4.78	1
Water-amyl alcohol			(1 exp.)
0.023	1.77÷2.49	1.75	1
0.0136	1.06÷1.30	1.64	1
0.0105	0.91÷1.15	1.57	1
0.0059	0.85÷0.97	1.41	1
Water-anisole			(6-7 exp.)
0.441	29.2	25.44	6
0.398	20.47	25.04	6
0.50	10.67÷14.33	25.94	6
0.243	9.62÷15.12	20.99	5
0.40	23.95÷33.59	25.06	6
Water-kerosene			(1-6 exp.)
0.336	6.01÷14.33	8.00	7
0.12	2.89÷8.89	6.36	6
0.0038	1.63÷4.27	4.78	5

A further validation of the model has been made by comparing the calculated drainage times with those coming from the experiments of Davis et al. (1971). The properties of organic liquids used for the experiments are reported in Table 3, while the comparisons are reported in Table 2. The data are related both to single and multiple coalescence stages. For multiple stages the reduction of the daughter droplet diameter with respect to the mother droplet has been evaluated taking into account the surface energy resulting after dissipation and ratio among mother and daughter drop given in Aryafar and Kavehpour (2006). A more detailed model for the coalescence cascade is under development and will be reported a future paper. It has to be pointed out that data related to the drainage time of the interposed films are rarely reported into the literature while it is possible to find the coalescence time evaluated taking into account the period starting from film breakage till daughter drop formation. The data of Davis, despite the use of a careful technique for obtaining the maximum of purity and the less of disturbances for measuring the drainage time of a large number of drops (about 100 for each test) show a large variation of values so that a range of times is reported.

Table 3: Physical properties of phases 25°C

System	Water (n= 1.333)		Organic Phase			
	ρ g/cm^3	μ cP	ρ g/cm^3	μ cP	n	σ g/s^2
Diethyl carbonate-water	0.9995	1.06	0.9756	0.821	1.3845	13.1
Methyl isobutyl ketone-water	0.9949	1.25	0.8114	0.623	1.3939	12.8
Cyclohexanone-water	0.9984	1.35	0.9512	2.28	1.451	4
Amyl alcohol- water	0.9922	1.15	0.8254	4.21	1.41	4.88
Anisole-water	0.9982	1.00	0.9941	1.09	1.5174	16.1
Kerosene-water	0.9982	1.01	0.7880	1.93	1.40	52.5

4. Conclusions

The presented model for liquid droplets coalescence is a mechanistic one. The obtained results show its reliability both for predicting shape and film drainage time. It can be a reasonable basis also for the study of the cascade of coalescences and foams.

References

Aryafar H., Kavehpour H.P., 2006, *Drop coalescence through planar surfaces*, Physics of Fluids, 18, 072105-1÷072105-5.

Blanchette F., T.P. Bigioni, 2006, *Partial coalescence of drop at liquid interfaces*, Nature Physics, 2, 2006, letters, 254-257

Bozzano G., M. Dente, 2001. *Shape and terminal velocity of single bubble motion: a novel approach*, Comp. & Chem. Engng, 25, 571-576

Bozzano G., M. Dente, 2009, *Single bubble and drop motion modeling*, Chemical Engineering Transactions,17, 567-572

Charles G.E., S.G. Mason, 1960, *The coalescence of liquid drops with flat liquid/liquid interfaces*, Journal of Colloid Science,15, 236-267

Chen X., S. Mandre, J.J. Feng, 2006, *Partial coalescence between a drop and a liquid-liquid interface*, Phisics of Fluids, 18, 051705-1, 051705-4

Davis G.A., G.V. Jeffreys, D.V. Smith, *Coalescence of liquid droplet-correlation of coalescence times*, Proceedings of the International Solvent Extraction Conference, ISEC 71, The Hague, 19-23 April, 1971, vol.1, 385-399, paper 35, session 3A, Coalescence

Hartland S., 1967, *The coalescence of a liquid drop at a liquid-liquid interface, part I: drop shape*, Trans. Instn. Chem. Engrs., 45, T97-T100.

Mahanty J., B.W. Ninham, 1976, *Dispersion Forces*, Academic Press, London

Princen H. M., 1963, *Shape of a fluid drop at a liquid-liquid interface*, J. of Colloid Sci., 18, 178-195

Zulfaa M., E.K. Longmire, 2004, *Drop coalescence through a liquid/liquid interface*, Phisics of Fluids, 16, n°7, 2170-2181

Combining Statistical and Physical Considerations in Deriving Targeted QSPRs Using Very Large Molecular Descriptor Databases

Inga Paster,[a] Greta Tovarovski,[b] Mordechai Shacham,[a] Neima Brauner,[b]

[a]*Department of Chemical Engineering, Ben-Gurion University of the Negev, Beer-Sheva 84105, Israel*
[b]*School of Engineering, Tel-Aviv University, Tel-Aviv 69978, Israel*

Abstract

The use of 3-D molecular descriptors in Quantitative Structure Property Relationships (QSPR) is considered. Such descriptors are known to have high level of uncertainty, as different minimization algorithms tend to yield different 3-D structures, which in turn yield different descriptor values.

An algorithm has been developed in which the uncertainty in a 3-D descriptor is determined by the difference between the values obtained when using minimized structures coming from different sources for the same compound. This uncertainty can be used as an estimate in the descriptor "noise" level for determining its "signal to noise" ratio. A descriptor of a low signal-to-noise ratio should not be included in QSPRs even if it is highly correlated with the property values of the training set.

Keywords: property prediction, QSPR, 3-D structure, molecular descriptor

1. Introduction

Current methods used to predict physical and thermodynamic properties can be classified into "group contribution" methods, methods based on the "corresponding-states principle", "asymptotic behavior" correlations and Quantitative Structure Property Relationships (QSPRs). The QSPRs are correlations derived from limited amount of available structural information in terms of molecular descriptors and experimental property data.

Unlike in the traditional QSPR methods, the Targeted QSPR (TQSPR) method [1] is targeted to one particular compound or a group of compounds, and enables predicting properties within experimental uncertainty, depending on the level of similarity between the target compound and the predictive compounds selected to the training set. The similarity level is determined based on the corresponding molecular descriptors values. The accuracy and the consistency of the molecular-descriptor database are therefore critical in the stage of the selection of the predictive compounds. Erroneous, inaccurate and inconsistent values of molecular descriptors introduce a noise, which may prevent the selection of the most similar compounds available in the database to the training set.

In recent years computer programs that can calculate several thousands of molecular descriptors have emerged. It is practically impossible to check the accuracy and consistency of the individual-descriptor values, because of the large number of descriptors and compounds involved.

In principle, descriptor values are calculated by mathematical formulas, and as such they are considered accurate up to the last decimal digit reported. However, for 3-D descriptors this assumption may not be valid due to the uncertainty associated with the

minimization of the 3-D structure. This raises concerns regarding the accuracy and consistency of the molecular descriptors used and the probability of obtaining "chance" correlations, while applying stepwise regression procedures to large descriptor databases in order to obtain a QSPR (or TQSPR) for representing a particular property. Using the stepwise regression algorithm we developed [2], the noise (uncertainty) in both the dependent (property value) and the independent variables (descriptor values) is considered. The signal-to-noise ratio tolerances embedded in the algorithm prevent variables associated with high level of uncertainty entering the regression model. Thus, when selecting the members of the training sets using molecular descriptors and statistical methods it is essential to have realistic estimate of the noise level in the 3-D descriptors.

The generation of the 3-D structure requires minimization of the total energy of the molecule. The computation of the minimized 3-D structure may involve the use of several different programs in order to increase the probability of convergence to a global minimum, and most programs allow use of different models for representing the energy and various convergence tolerances. For a more detailed discussion of this subject, including references, see for example the "Computed 3-D Structures" section in the NIST[3] database.

Some preliminary results (Paster et al. [4]) have shown that minimized 3-D structure files generated by different algorithms may yield completely different 3-D molecular descriptor values when fed into a program that generates such data. To check the consistency and the reliability of the descriptor database the following studies were carried out:

1. Plotting the vectors of descriptors of two immediate neighbors in a homologous series, one versus another.
2. Plotting molecular descriptors versus the number of carbon atoms for homologous series
3. Using subsets of the descriptors to identify training sets that belong to the homologous series of the target compound.
4. Using physical considerations for selecting the training set in prediction of melting point temperature
5. Comparing 3-D descriptors obtained from 3-D structure files minimized by different algorithms.

The results of the first four studies were described in detail by Paster et al. [4]. This paper is dedicated to the description of the results of study No. 5.

When comparing 3-D descriptors obtained from various structure files, the difference between the descriptor values can be used to obtain an estimate of their "noise level". The noise level is used by the SROV stepwise regression algorithm [2] to determine the signal-to-noise ratio in a regression model. To avoid over fitting and obtain a stable model (QSPR), the reduction in the model variance upon the inclusion of the descriptor considered should significantly exceed the noise level (i.e., the signal-to-noise ratio indicators of the descriptor selected to the QSPR at each stage of the stepwise regression must be larger then 1). This sets a limit on the number of descriptors that can be included in a QSPR and descriptors with large noise level are prevented from entering the QSPR.

2. Methodology

For this study 31 compounds for which 3-D MOL files are available in both the NIST [3] database (source 1, henceforward) and a library associated with the Dragon 5.5

Combining Statistical and Physical Considerations in Deriving Targeted QSPRs

program [5] (source 2) were used (see Table 1). These MOL files were fed into the Dragon program to calculate 1664 descriptors for the compounds included in the study. As 0-D, 1-D and 2-D descriptors are not affected by the variation of the 3-D structure only the 721 3-D descriptors (out of 1556) were considered.

Table 1. Compounds for which 3-D MOL files are available from NIST and Dragon

No.	Compound	No.	Compound	No.	Compound
1	methane	12	pyrrole	22	trans-2-butene
2	ethanol	13	furan	23	cyclopentane
3	ethane	14	isobutane	24	n-pentane
4	2-propanone	15	cyclohexane	25	cis-2-butene
5	benzene	16	cyclopropane	26	cyclohexanone
6	2-propylamine	17	n-butane	27	anthracene
7	2-propanol	18	toluene	28	1-propanol
8	propane	19	thiophene	29	dibenzofuran
9	naphthalene	20	2-butyne	30	2-methylpentane
10	phenol	21	neopentane	31	n-hexane
11	cyclobutane				

Dragon	NIST
Ethanol	
n-hexane	
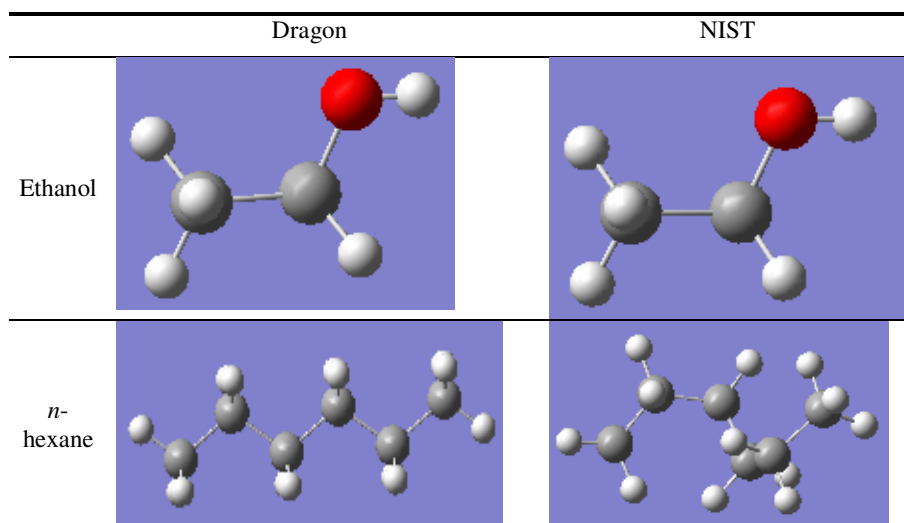
Figure 1. Visual comparison of Dragon and NIST 3-D structure files using Gaussian 3

The Gaussian [6] program was used for graphical display of the 3-D molecular structures and the JMP (http://www.jmp.com/) statistical software was used to analyze the differences between the descriptor values calculated from different 3-D mole files.
The 3-D structure representations of the 31 compounds were visually inspected using the Gaussian [6] program. For most of the compounds the differences were small, hardly distinguishable. One example of such a case is shown in Figure 1 (for ethanol).

For three compounds: n-hexane, 2-methylpentane and 1-propanol, the H-depleted molecular structure appears planar when the MOL file from the Dragon library is used as input to the Gaussian, while the NIST mole files yielded 3-D structures (see demonstration in Figure 1 for n-hexane). Based on this comparison, it was decided to study separately the group of compounds with apparent major differences (n-hexane, 2-methylpentane and 1-propanol) and the compounds with minor differences (the rest of the compounds in Table 1).

Table 2. Groups of 3-D descriptors calculated by the Dragon [2] program

No.	Descriptor Group	Description
1	3D-MoRSE	Descriptors calculated by summing atom weights viewed by different angular scattering function
2	Geometrical	Different kinds of conformationally dependent descriptors based on the molecular geometry.
3	GETAWAY	Descriptors calculated from the leverage matrix obtained by the centred atomic coordinates (molecular influence matrix, MIM)
4	Randic molecular profiles	Descriptors derived from the distance distribution moments of the geometry matrix defined as the average row sum of its entries raised to the k-th power, normalized by the factor k!.
5	RDF	descriptors obtained by radial basis functions centred on different interatomic distances (from 0.5A to 15.5A)
6	WHIM	Descriptors obtained as statistical indices of the atoms projected onto the 3 principal components obtained from weighted covariance matrices of atomic coordinates.

Table 3. Numbers of 3-D descriptors in the various groups

No.	Descriptor Group	No. of Descriptors	Average No. Of Nonzero Descriptors
1	3D-MoRSE	160	158
2	Geometrical	74	33
3	GETAWAY	197	124
4	Randic molecular profiles	41	25
5	RDF	150	45
6	WHIM	99	92

The 3-D descriptors that can be calculated by Dragon are divided into 6 groups of descriptors: 3D-MoRSE, geometrical, GETAWAY, RDF, WHIM and Randic molecular profiles. Brief descriptions of the various groups are provided in Table 2. The numbers of the descriptors in the various groups are summarized in Table 3. Non-zero and zero value descriptors are counted separately, as the zero value usually indicates that the descriptor is either undefined or cannot be calculated for a particular compound, and it cannot be considered as a calculated zero value. The fraction of the zero value

descriptors varies from group to group. For the 3D-MoRSE group close to 99% of the descriptors are calculated non-zero values, while for the RTF group only 30% of the descriptors are non-zero. Only non-zero descriptors were included in the analyses that follow.

3. Results and Discussion

The sensitivity of the descriptors to differences in the 3-D structures is measured by the difference in the values of the descriptors calculated using the MOL files from sources 1 and 2.

Table 4. Percent differences between 3-D descriptors based on NIST and Dragon library MOL files in the various groups (28 compounds)

No.	Descriptor Group	$\leq 0.2\%$	$\leq 3\%$	$\leq 10\%$	$\leq 50\%$	$> 50\%$
1	3D-MoRSE	4.51	15.09	30.00	67.03	32.97
2	Geometrical	12.97	81.38	92.89	97.18	2.82
3	GETAWAY	33.17	85.06	96.94	99.43	0.57
4	Randic molecular profiles	16.95	58.47	81.92	99.15	0.85
5	RDF	1.03	4.61	22.67	70.01	29.99
6	WHIM	30.13	72.51	92.40	97.75	2.25

Table 4 summarizes the percent differences obtained for the group of 28 compounds for which the MOL files from both sources yielded similar structures (no visually noticeable differences). The differences are grouped according to the upper difference limit: ≤ 0.2 %; ≤ 3 %; ≤ 10 % and ≤ 50 %. Descriptors with difference > 50 % are grouped separately. The meaning of these categories is that descriptors associated with an upper difference limit of say 0.2 %, the "noise" level can be set at 0.2%. Observe that the great majority of the descriptors of the geometrical, GETAWAY and WHIM groups are on the ≤ 3 % difference level. For the Randic Molecular Profile group the great majority is on the ≤ 10 % level. Thus, most of the descriptors included in these four groups can be used for predicting properties, as their noise level will not exclude them a-priory from consideration to be included in a QSPR. As for the 3D-MoRSE and RDF group of descriptors, only a small percentage of them can be included in the ≤ 10 % difference category. Thus, very few descriptors of these groups can be included in TQSPRs. Furthermore, the noise of the descriptors included in the QSPR is propagated to the predicted property values. Consequently, the prediction error increases beyond the minimal level set by the experimental errors in the property values of the predictive compounds (the dependent variables). Thus, inclusion of descriptors from the > 10% difference category in the QSPR will result in excessive prediction errors.

In Table 5 the percent differences are summarized for n-hexane, 2-methylpentane and 1-propanol where the MOL files from the Dragon library yield (apparently) planar structures. For these 3 compounds, the number of descriptors for which the difference level is low enough to enable their use in QSPR models is much smaller compared to Table 4.

Table 5. Percent differences between 3-D descriptors based on NIST and Dragon library MOL files in the various groups (n-hexane, 2-methylpentane and 1-propanol)

No.	Descriptor Group	≤ 0.2%	≤ 3%	≤ 10%	≤ 50%	> 50%
1	3D-MoRSE	3.96	10.00	20.83	58.75	41.25
2	Geometrical	0.98	15.69	39.22	92.16	7.84
3	GETAWAY	6.25	15.73	43.10	93.10	6.90
4	Randic molecular profiles	0.00	3.81	12.38	36.19	63.81
5	RDF	0.00	6.01	20.22	55.19	44.81
6	WHIM	9.09	15.15	31.99	87.54	12.46

4. Conclusions

It has been demonstrated that in spite of the uncertainty associated with the minimization of 3-D structure files, it is possible to use part of the 3-D descriptors in QSPRs (or TQSPRs). The maximal uncertainty in the descriptor, as determined by the difference between the values of the same descriptor when calculated from two minimized structure files coming from different sources for the same compound, can be used as the "noise" level. The latter is used for determining "signal to noise" ratio in the descriptor value. A large noise level (low signal-to-noise ratio) will prevent the inclusion of such a descriptor in the QSPR, even if it is highly correlated with the property values of the training set.

The ranking of the descriptor groups according to the percentage of the descriptors with low sensitivity to the 3-D structure file source is the following: GETAWAY, geometrical, WHIM, Randic molecular profiles, 3D-MoRSE and RDF. Structure files that were not minimized properly yield much smaller number of usable (thus with low level of uncertainty) 3-D.

To further enhance and extend the usability of 3-D descriptors in QSPRs it is important to make reliable and consistent 3-D structures available in the property databases.

References

[1] Brauner, N., Stateva, R. P., Cholakov, G. St. and Shacham, M. "A Structurally "Targeted" QSPR Method for Property Prediction", Ind. Eng. Chem. Res. 45, 8430-8437 (2006)

[2] Shacham, M.; Brauner, N. The SROV Program for Data Analysis and Regression Model Identification. Computers Chem. Engng. 2003, 27, 701.

[3] National Institute of Standards and Technology (NIST). In: Linstrom PJ, Mallard WG, eds. Chemistry WebBook, NIST Standard Reference Database Number 69. Gaithersburg, MD: NIST; June 2005 (http://webbook.nist.gov).

[4] I. Paster, M. Shacham and N. Brauner, " Investigation of the Relationships between Molecular Structure, Molecular Descriptors and Physical Properties", Ind. Eng. Chem. Res., 2009, 48, 9723.

[5] Todeschini, R.; Consonni, V.; Mauri, A.; Pavan, M. DRAGON user manual, Talete srl, Milano, Italy, 2006.

[6] The Official Gaussian Website Home Page: (http://www.gaussian.com/).

Simulated preparation of supported porous catalyst and evaluation of its reaction-transport properties

V. Novák,[a], F. Štěpánek[b], P. Kočí[a], M. Marek[a], M. Kubíček[c]*
[a]*Department of Chemical Engineering*
[b]*Chemical Robotics Laboratory*
[c]*Department of Mathematics*
Institute of Chemical Technology, Prague, Technická 5, 166 28 Praha, Czech Republic
Corresponding author, E-mail: milan.kubicek@vscht.cz, Tel.: +420 220 443 104

Abstract

In this contribution, mathematical model for the description of solvent evaporation and noble metal crystallization in a porous medium is presented. The methodology is based on the volume-of-fluid method and the aim is validation of the model by comparing the numerical simulation results with analytical solutions for evaporation from a single pore and for particle growth and Ostwald ripening of two freely suspended particles in a saturated solution.

Keywords: crystallization, drying, catalyst, volume-of-fluid method

1. Introduction

The preparation process of a supported metal porous catalyst can be formally divided into three stages: (i) impregnation of the metals or their precursors into the porous structure, where nucleation and crystallization take place; (ii) drying of liquid solution from the porous structure and (iii) after drying, reduction and calcination of the catalyst [1]. Mathematical modeling of such processes based on continuum-level equations can predict the resulting properties of porous catalysts on the macroscopic scale (level of entire particle), c.f. e.g. [2]. The aim of this contribution is to present a methodology based on the volume-of-fluid method which describes crystallization and drying processes on the nano-scale level, where the heterogeneous porous structure and spatial arrangement of individual crystallites are explicitly described and typical pore radii are in the order of 10 nm (e.g., in catalytic converters used in automobile industry [3]).
The methodology presented here follows from our previous work which dealt with the characterization and prediction of porous structure [4]–[7] and modeling of reaction and transport inside the porous medium [8]–[11] and thus forms a new segment to the virtual platform for porous media simulations.

2. Studied system

The catalytic washcoat used in automobile converters is a bimodal porous structure with deposited noble metal crystallites acting as active catalytic sites, Fig. (1a). Simulation of crystallization and drying can be carried out in the porous medium representing the supporting material, reconstructed, e.g., by a particle packing method [6]. The digital porous medium is represented by a volume phase function [4].
At the beginning of our simulation the pores are filled with a liquid solution of the active metals. Nucleation centers are randomly generated on the surface of the solid phase and then liquid evaporation and crystallization are simulated. Final situation in

the system is depicted in Fig. (1b).Afterwards the influence of process conditions on the final noble metal distribution can be studied *in silico*.

At this moment the following assumptions are considered:

1. mass transport in the liquid phase is governed by Fick's diffusion;
2. migration caused by electrical charge is not considered;
3. capillary flow is not taken into account;
4. vapor pressure in the bulk is constant during drying;
5. isothermal conditions apply over the studied system;
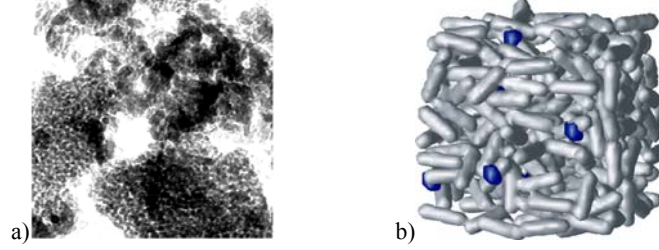6. absence of adsorption/desorption effects.

Fig. 1. a) SEM picture of meso-porous structure of Pt/γ-Al₂O₃ (white pores, grey γ-alumina, and black Pt); b) 3D reconstructed structure with dispersed Pt particles (blue)

3. Model

3.1 Transport of dissolved species

Spatially 3D diffusion is described by mass balances in the form of the following partial differential equations for individual components $i = 1..I$:

$$\frac{\partial c_i}{\partial t} = \nabla \left(D_i \nabla c_i \right) \tag{1}$$

Here $c_i = c_i(\mathbf{x})$ is the local concentration of component i, and $D_i = D_i(\mathbf{x})$ is the local value of its diffusion coefficient. Diffusion in liquid can be described by Wilke-Chang equation [12]. Eq. (1) is solved only in liquid phase. Different types of initial and boundary conditions can be defined in the studied section of porous catalyst to account for different modes of crystallization and drying. In this paper the following set of initial and boundary conditions is used:

$$c_i \left(\mathbf{x} \right)\Big|_{t=0} = c^{\text{init}}, \forall \mathbf{x} : 0 < f_1(\mathbf{x}) \leq 1 \tag{2}$$

$$\frac{\partial c_i}{\partial x}\bigg|_{x=x_0,x_1} = 0, \frac{\partial c_i}{\partial y}\bigg|_{y=y_0,y_1} = 0, \frac{\partial c_i}{\partial z}\bigg|_{z=z_0,z_1} = 0 \tag{3}$$

$$-D_i \nabla c_i = -c_i r_v \frac{M_w}{\rho_w}, \forall \mathbf{x} : 0 < f_1(\mathbf{x}) < 1, \; -D_i \nabla c_i = r_d, \forall \mathbf{x} : 0 < f_s(\mathbf{x}) < 1 \tag{4}$$

These boundary conditions correspond to a batch system with initial concentration c^{init} in time $t = 0$ s, where the rates of solvent evaporation r_v and deposition of noble metal (or precursor) crystallites r_d are computed only at the gas/liquid and noble solid/liquid interfaces, respectively.

Preparation of a porous catalyst and evaluation of reaction-transport properties by the computer simulations

Eq. (1) is solved numerically within the spatially 3D section of porous catalyst. After equidistant spatial discretisation of the domain with a step h, Eq. (1) can be rewritten as (shown only for x direction and subscript i is omitted for sake of brevity):

$$\frac{c^{t+\Delta t} - c^t}{\Delta t} = \overline{D}_{<m-1,m>,n,o} \frac{c^t_{m-1,n,o} - c^t_{m,n,o}}{h^2} + \overline{D}_{<m+1,m>,n,o} \frac{c^t_{m+1,n,o} - c^t_{m,n,o}}{h^2} \tag{5}$$

Here Δt is time step, the subscripts m, n, o are discretisation indices of the spatial coordinates x, y, z, respectively. The $\overline{D}_{<m-1,m>,n,o}$ denotes mean diffusion coefficient of the component between two adjacent volume elements with the coordinates $(m-1,n,o)$ and (m,n,o):

$$\overline{D}_{<m-1,m>,n,o} = \frac{2 D_{<m-1,m>,n,o} D_{m,n,o}}{D_{<m-1,m>,n,o} + D_{m,n,o}} \tag{6}$$

The resulting set of equations is then solved iteratively by the Successive over-relaxation Gauss-Seidel method [13].

3.2 Solvent evaporation

The rate of liquid evaporation $r_v(\mathbf{x})$ into its vapor surrounding during drying is computed from Hertz-Knudsen equation [14]:

$$r_v = \left(p^*(\mathbf{x}) - p_v \right) \varepsilon \left(\frac{M}{2\pi R^g T} \right)^{1/2} \tag{7}$$

Here ε is evaporation coefficient defined as the experimental rate of evaporation divided by the theoretical maximum rate for the same conditions and $p^*(\mathbf{x})$ is the equilibrium vapour pressure above the curved interface in a surface point $(\mathbf{x})$ computed from the Kelvin equation:

$$p^*(\mathbf{x}) = p_s \exp\left[\frac{\kappa_1(\mathbf{x}) V_1^M \gamma_1}{R^g T} \right] \tag{8}$$

The local curvature of the liquid-vapour interface $\kappa_l(\mathbf{x})$ is evaluated from Eq. (14). The shift of the liquid-vapour interface l_1 is then changed according to the following equation:

$$\frac{dl_1}{dt} = \frac{r_v M_1}{\rho_1} \tag{9}$$

3.3 Formation of crystallites

The rate of crystal growth r_d per unit surface area A is considered as a pseudo m-th order reaction A $\rightarrow$ B where the order of reaction can be different for deposition and dissolution. The rate is computed from:

$$r_d = k \left(c(\mathbf{x}) - c^*(\mathbf{x}) \right)^m \tag{10}$$

Here k is the crystallisation/dissolution rate constant obtained from the Arrhenius relation. The rate of crystallization r_d depends on the local concentration of liquid solution $c_i(\mathbf{x})$ above the solid-liquid interface and on the equilibrium concentration $c_i^*(\mathbf{x})$ which is evaluated from the Ostwald-Freundlich equation, Eq. (11). As it can be seen Ostwald-Freundlich equation is analogous to the Kelvin eqution, Eq. (8).

$$c^*(\mathbf{x}) = c_s \exp\left[\frac{\kappa_s(\mathbf{x})V_s^M\gamma_s}{R^gT}\right] \tag{11}$$

The local curvature of the solid-liquid interface $\kappa_s(\mathbf{x})$ is evaluated from Eq. (14). Then the shift of the solid-liquid interface l_s is computed from the following equation:

$$\frac{dl_s}{dt} = \frac{r_d M_s}{\rho_s} \tag{12}$$

3.4 Local interface curvature

The local curvature, $\kappa(\mathbf{x})$ required for calculating $p^*(\mathbf{x})$ and $c^*(\mathbf{x})$ appearing in Eq. (8) and Eq. (11), is evaluated numerically [15] from the phase function of i-th phase f_i in every iteration (where i is liquid or solid phase). First, the unit normal vectors, oriented from the i-th phase outwards, are constructed at all interface points of phase i:

$$\mathbf{n}_i(\mathbf{x}) = -\frac{\nabla\widehat{f}_i(\mathbf{x})}{\left\|\nabla\widehat{f}_i(\mathbf{x})\right\|}, \forall\mathbf{x}: 0 < f_i(\mathbf{x}) < 1 \tag{13}$$

Here $\widehat{f}_i(\mathbf{x})$ is the "mollified" phase function, obtained from f_i by the application of the 1-6-1 smoothing kernel. Once the interface normal vectors are known, the radius of curvature at point $(\mathbf{x})$ can be calculated from:

$$\kappa_i(\mathbf{x}) = -\nabla\mathbf{n}_i(\mathbf{x}), \forall\mathbf{x}: 0 < f_i(\mathbf{x}) < 1 \tag{14}$$

The second-order, symmetric finite difference approximations of the partial derivatives have been used in Eq. (13) and Eq. (14), e.g.,

$$\left.\frac{\partial f_L}{\partial x}\right|_i \approx \left(f_L^{i+1} - f_L^{i-1}\right)/2h \tag{15}$$

In a triple phase point s-l-g the normal vector of the liquid is obtained from $\mathbf{n}_l\mathbf{n}_s = \cos\theta$, where $\mathbf{n}_s$ is computed from Eq. (13) and θ is the contact angle.

4. Validation of the model

In this section the validation of the presented methodology for liquid drying and solids deposition modeling is shown. The numerical simulation results will be compared with results obtained from analytical solution in simple geometries, namely evaporation from a cylindrical pore and dissolution/deposition (Ostwald ripening) of two spherical particles.

An example of a volume-of-fluid numerical simulation of evaporation from a single cylindrical pore is shown in Fig (2a). The simulation starts from a completely filed pore and a curved meniscus gradually forms during evaporation as a consequence of Eqs. (5) and (6). The height of the meniscus H for varying equilibrium contact angle θ between the liquid and the solid is depicted in Fig. (2b). Results from numerical simulation are compared with analytical calculation of the meniscus height calculated according to:

$$H_{anal} = abs\left(\frac{R}{\cos\theta}(1 - \sin\theta)\right) \tag{16}$$

The simulation verifies that the liquid is able to "adapt" to a new condition and thus creates the correct shape of meniscus. In a general porous medium the local meniscus results from the local shape, size and wettability of a pore during evaporation. It can be seen in Fig. (2b) there is quantitative agreement between the height of meniscus H obtained from numerical simulation and that obtained analytically over most contact

Preparation of a porous catalyst and evaluation of reaction-transport properties by the computer simulations

angles. However, for contact angles below approximately 30 deg a deviation is observed, most likely caused by a discretisation error when evaluating interface curvature near the three-phase contact line.
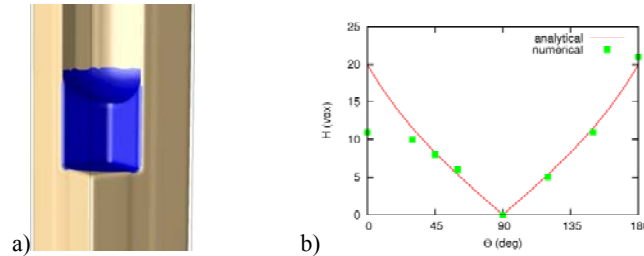

a)   b)

Fig. 2. a) Example of liquid meniscus (blue) in a 3D reconstructed cylindrical pore with r=20 vox; b) height of meniscus in dependence on contact angle θ.

The crystallization/dissolution module was separately validated by simulating Ostwald coarsening. The simulation was carried out in a batch system (boundary condition Eq. (3)) for two spherical particles of different initial size (r_1=16 vox and r_2=8 vox), as shown in Fig. (3a). Results obtained from the volume-of-fluid simulation are compared with a model based on material balances with a presumption of spherical shape and an ideal mixing of the batch according to equations:

$$\frac{\mathrm{d}n_1}{\mathrm{d}t} = A_1 k \left(c_L - c_1^*\right)^m ; \quad \frac{\mathrm{d}n_2}{\mathrm{d}t} = A_2 k \left(c_L - c_2^*\right)^m \tag{17}$$

$$V \frac{\mathrm{d}c_L}{\mathrm{d}t} = -\left[A_1 k \left(c_L - c_1^*\right)^m + A_2 k \left(c_L - c_2^*\right)^m\right] \tag{18}$$

The subscript 1 refers to the initially larger particle and 2 to the smaller one. The quantity of the solid in each particle is n [mol], A is the particle surface area and V the volume of the liquid batch. Equilibrium concentrations c_1^* and c_2^* are evaluated from Eq. (11) where $\kappa_s = 2/R_s$, and R_s is the radius of each particle. It can be seen in Fig. (3b) that the small particle dissolves while the bigger one grows – the bigger particle is energetically more stable than the smaller one. The agreement between numerical simulations and analytical solution is quantitative in this case.
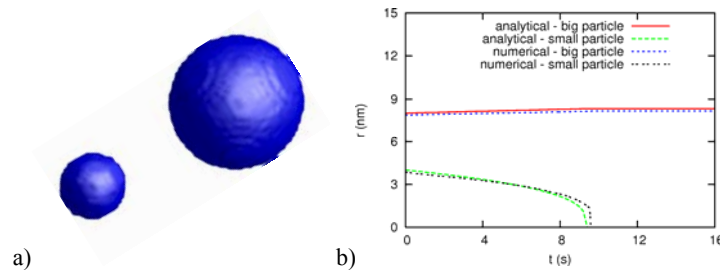

a)   b)

Fig. 3. a) Example of two spherical particles of active metal, reconstructed in 3D; b) simulation of Ostwald coarsening during the crystallisation - dependence of particle radius on time.

An example of the evaporation and metal particles deposition in a general porous medium is depicted in Fig. (4). The porous medium filled with a liquid solution of the active metal is shown in Fig. (4a). Nucleation centers are randomly generated on the surface of the solid phase and then the evaporation and crystallization are simulated dynamically. The final distribution of metal particles after the evaporation is shown in

Fig. (4b). It can be seen that a couple of larger particles developed while several smaller particles dissolved.
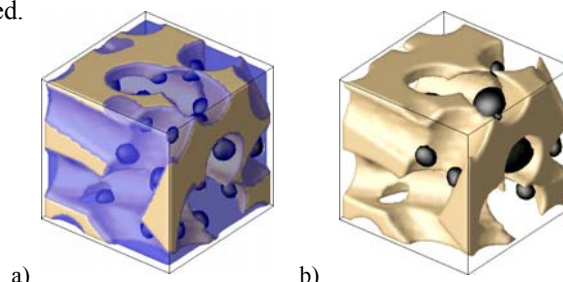

a)   b)

Fig. 4. Example of evaporation and crystallization in a general porous medium with the discretization the 60x60x60 voxels and space step h = 0.25 nm

5. Conclusions

Novel models were presented for description of liquid evaporation and crystallization of particles in a general 3D porous medium. These processes are strongly relevant to practical application during the preparation of porous supported catalysts (impregnation of active metals dissolved in a liquid, followed by drying of the solvent). These models extend an existing range of tools for the 3D simulations of transformation, reaction and transport in porous media [4]–[11].

The model was validated separately for evaporation of liquid from a cylindrical pore and for crystallization of two spherical particles of different size. An application of the model in a general porous medium was also demonstrated. The following research will be aimed on a parametric study of catalyst preparation and description of unknown parameters from experiments as well as comparison of simulation results with measurements.

Acknowledgements

References

[1]   Nijhuis T.A., Beers A.E.W., Vergunst T., Hoek I., Kapteijn F., Moulijn J.A. Catal. Rev. Sci. Eng. 43 (2001) 345-380.
[2]   Liu X., Khinast J.G., Glasser B.J. Chem. Eng. Sci. 63 (2008) 4517-4530.
[3]   Starý T., Šolcová O., Schneider P., Marek M. Chem. Eng. Sci. 278 (2006) 5934.
[4]   Kosek J., Štěpánek F., Marek M. Adv. Chem. Eng. 30 (2005) 137.
[5]   Štěpánek F., Marek M., Adler P.M. AIChE J. 45 (1999) 1901.
[6]   Štěpánek F., Ansari M.A. Chem. Eng. Sci. 60 (2005) 4019.
[7]   Štěpánek F., Soos M., Rajniak P. Coll. Surf. A 300 (2007) 11.
[8]   Kočí P., Štěpánek F., Kubíček M., Marek M. Chem. Eng. Sci. 61 (2006) 3240.
[9]   Kočí P., Štěpánek F., Kubíček M., Marek M. Molecular Simulation 33 (2007) 369-377.
[10] Kočí P., Novák V., Štěpánek F., Marek M., Kubíček M. Chem. Eng. Sci. 65 (2010) 412.
[11] Novák V., Štěpánek F., Kočí P., Marek M., Kubíček M. Chem. Eng. Sci. (2010) in press, doi:10.1016/j.ces.2009.09.009
[12] Wilke C.R., Chang P. AIChE J. 1 (1955) 264-269.
[13] Carré B.A. The Comp. J. 4 (1961) 73.
[14] Eames I.W., Marr N.J., Sabir H. Int. J. Heat Mass Transfer 40 (1997) 2963-2973.
[15] Bullard J.W., Garboczi E.J., Carter W.C., Fuller E.R. Comp. Mater. Sci. 4 (1995) 103-116.

Fluid-Dynamics Study of Multiphase Flow in a Sieve Tray of a Distillation Column

Joel G. Teleken, Leandro O. Werle, Iaçanã G. B. Parisotto, Cintia Marangoni,

Ana Paula Meneguelo, Ricardo, A. F. Machado.

Federal University of Santa Catarina - Department of Chemical Engineering - Laboratory of Control of Processes LCP/CTC - Technological Center - University Campus - Trindade. P.O. Box 476. ZIP Code: 88010-970. Florianópolis, Santa Catarina, Brazil, ricardo@enq.ufsc.br.

Abstract

Many studies on distillation columns are based on macroscopic models of mass and energy conservation. This paper deals with the complex hydrodynamics of sieve trays in pilot plant distillation columns based on a distributed control system with heating action at intermediate points, through electrical resistance heaters placed on the surfaces of sieve trays, using computational fluid dynamics. In this context, the objective of this study was to evaluate the influence of these electrical resistance heaters on the hydrodynamics. A three-dimensional mathematical homogeneous biphasic model was implemented in the commercial code of computational fluid dynamics (CFD) for a numerical experimentation study. The results showed that electrical resistance heaters placed on the surfaces of the sieve trays influenced the flow patterns, without affecting the hydrodynamics as a whole, and helped to mix and homogenize the region with active bubbles. Thus, the use of electrical resistance heaters could be applied in a distributed control system approach.

Keywords: distillation columns, fluid dynamics model, CFX.

1. Introduction

In recent years, there has been considerable academic and industrial interest in the use of computational fluid dynamics (CFD) to model two-phase flows in some chemical engineering systems. Many attempts have been made to simulate sieve tray hydrodynamics using CFD. Liu et al. (2000) neglected the variations in the direction of gas flow along the height of the dispersion to simulate the two-phase flow behavior, and only the hydrodynamics of the liquid flow was obtained. The inter-phase momentum exchange (drag) coefficient is required to model the hydrodynamics of multiphase flow on a sieve tray. This drag coefficient is not appropriate for the description of the hydrodynamics of sieve trays operating in either the froth or spray regimes. Krishna et al. (1999) and van Batten and Krishna (2000) described the hydrodynamics of sieve trays by estimating a new drag coefficient correlation for a swarm of large bubbles on the basis of the correlation of Bennett et al. (1983) for the liquid hold-up. There are several studies in the literature using fluid dynamics modeling for distillation columns, notably the studies by Li et al. (2009), Nikou and Ehsani, (2008), and Noriler et al. (2007), who, in most cases, studied the behavior of the speed vectors and the distribution of the kinetic energy of the liquid-gas flow in sieve trays. The model was solved using the finite volume method with variables located in a system of generalized coordinates (Maliska, 2004).

The objective of this study is to propose the development, implementation and application of a microscopic model for momentum conservation subjected to turbulent flow of the vapor phase, to represent the fluid dynamics of vapor-liquid flow in a perforated plate distillation pilot plant and then examine the influence of electrical resistance used to control the heating of a distributed distillation pilot plant, placed on the surface of the trays, on the fluid flow dynamics (Marangoni and Machado, 2007). To this aim, we performed a preliminary study on the flow of water only. The flow with the injection of air into the holes of the distillation plate was then investigated and, finally, the presence of electrical resistance on the plates.

2. Materials and Methods

The model considers the gas and liquid flows in a Eulerian-Eulerian framework, where the phases are treated with transport equations. To solve this problem mass continuity and momentum equations were used, and in order to solve these it was necessary to add and apply the Momentum Flow equation (Liu et al., 2000). It was considered that the fluctuations (turbulence) consist of small swarms of bubbles being formed and dispersed, and that the Reynolds stresses can be linearly related to the mean velocity gradients (eddy viscosity hypothesis), similarly to the relationship between the stress and the strain tensors in laminar Newtonian flow. Also, the standard k-ε turbulence model for multiphase flow was assumed (Nikou and Ehsani 2008).

3. Boundary and Initial Conditions

Physical space is mapped to a cylindrical computational space and the boundary conditions of all boundaries of the physical domain are required: at the inlet, uniform profiles of the velocities and turbulent properties are imposed, no-slip conditions on the wall are assumed for both phases and pressure conditions at the outlet were also applied for the two phases. Figure 1 shows the configuration of the systems which were simulated: A system with air injection in the absence of electrical resistance (Figure 1a) and a system with air injection and the presence of electrical resistance (Figure 1b). This order of study was followed to observe the influence of each variable on the flow and thus to understand its behavior. The diameter of the tray is 0.21 m with a height of 0.15 m. The length of the weir is 0.16 m and the diameter of the resistor is 0.01 m. The liquid enters the tray through a stocking moon opening.
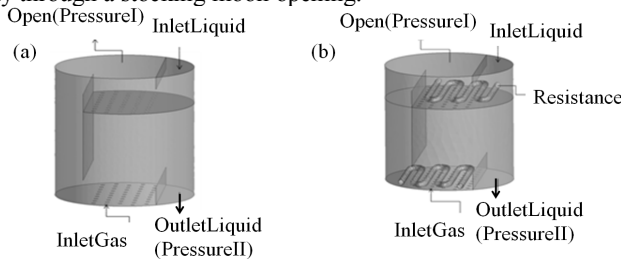
Figure 1 - 3-D Multiphase physical domain: (a) no air injection, (b) without electrical resistance, and (c) with electrical resistance.

Air, at ambient pressure, and water were used as the gas and liquid phases, respectively. At the beginning of the simulation, the conditions consisted of filling up with liquid up to the weir height, and air up to the weir height at a homogeneous temperature equal to T0. The velocity fields and the turbulent properties were also considered as initial conditions for each simulation and will be shown below. The time increment used in the

simulations is 0.01 s. During the simulation the volume fraction of the liquid phase in the gas–liquid dispersion in the system is monitored and quasi-steady state is assumed to prevail if the value of the hold-up remains constant for a period sufficient to determine the time-averaged values of the various parameters.

4. Results and Discussions

In this section, the results of simulations performed using the three physical domains studied and analyzed are presented. The hydrodynamics of the liquid-gas flow on the tray in the two flow areas can be observed in Figure 2.

We performed several simulations using different values for the liquid feed flow rate and increasing the air power in order to investigate the behavior of the two phases. Also, the main characteristics of each flow were analyzed, and the supply of air and liquid tracked. The main results for a liquid feed rate of 0.1 (m / s) and vapor phase of 0.3 (m / s) are given below.

Figures 2a and 2c show the distribution profile of the volume fraction of water in the XY plane and at three different heights (1.5, 9, 16.5 cm) in the XZ plane, respectively, without the presence of electrical resistance on the surface of the plates. Figures 2b and 3d show the behavior of the velocities of the air (in the XY plane) and liquid (in the XZ plane at 1.5 cm from the surface of the plate) vectors, respectively.

(a) Volume Fraction of Liquid phase in the XY plane

(b) Velocity vector of gaseous phase in the XY Plane

(c) Volume Fraction of liquid phase in the XZ plane at three different heights

(d) Velocity vector of liquid phase in the XZ plane at a height of 1.5 cm above the surface of the plate
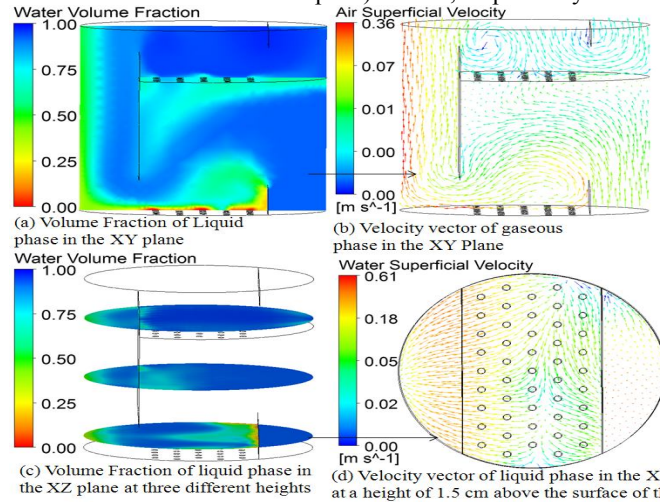
Figure 2 - Distribution of the liquid volume fraction and velocity vectors for multiphase flow without electrical resistance.

It can be observed in Figure 2a that the liquid layer retained (hold-up), represented by the dark blue color (volume fraction of liquid equal to one), undergoes a small increase; with mixing of the two phases and foam formation occurring, because of the high gas flow. The liquid elevation is more pronounced in the region near the wall of the exit from the lower plate, since the entrance of air through the holes is a control condition with a constant velocity and normal boundary, which forces the emergence of some route preferences, such as that observed, and making a part of this air rise through the liquid fall tube.

Figure 2b shows the behavior of the velocity vectors of the gas phase in the XY plane. It can be observed that there are regions of large recirculation and turbulence near the walls. Also, some route preferences appear with greater velocity of the gas phase, as in

the region near the fall barrier of the hold-up and also near the wall of the liquid fall. This verifies the results from the observation of the volume fraction distribution of the liquid in this plane.

Figure 2c confirms the distribution of the liquid volume fraction on three different levels in the XZ plane. It can be observed that the liquid phase distribution at 1.5 cm below the plate has a less homogeneous mixture of the liquid and vapor phases, with a region of greater concentration of air close to the fall barrier of the hold-up and near the wall of liquid fall represented by the color green. This probably occurred due to the initial condition and the boundary imposed on the gas entrance flow, which is constant and normal at the boundary.

When observing the distribution of the liquid volume fraction at 1.5 cm from the top plate, there is greater homogeneity of the mixture, due to the behavior of the vapor flow being closer to reality, that is, in the holes of the top plate an air entry condition was not imposed because the air that passes through the liquid retained in the lower plate is the same as that which enters the holes of the top plate. Figure 2d shows the behavior of the velocity vector of the liquid phase at 1.5 cm from the surface of the plate.

The presence of some circulation regions near the walls of the plate and also near the liquid fall tube can be noted. This distillation plate behavior presents some operational and efficiency problems, however, it should be remembered that this study is focused on the hydrodynamics of the plate in order to study these phenomena in particular.

In order to evaluate the influence of the electrical resistors placed on the surface of the plate on its hydrodynamics, new simulations were performed using the physical domain in Figure 3. The results of the simulation performed using the same air and liquid flows as the previous case as the new physical domain are represented in the same figure.

(a) Volume Fraction of Liquid phase in the XY plane in the presence of electrical resistence

(b) Velocity vector of gaseous phase in the XY plane in the presence of electrical resistence

(c) Volume Fraction of liquid phase in the XZ plane at three different heights in the presence of electrical resistence

(d) Velocity vector of liquid phase in the XZ plane at a height of 1.5 cm above the surface of the plate in the presence of electrical resistence
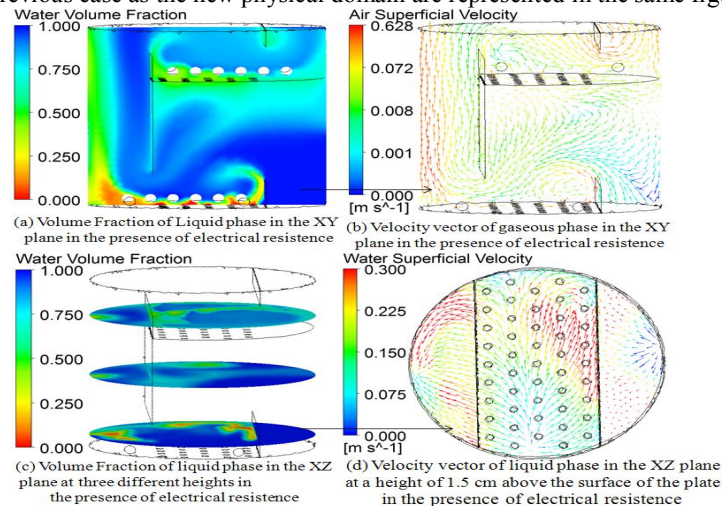
Figure 3 - Distribution of the liquid volume fraction and velocity vectors for multiphase flow with the presence of the electrical resistance.

Figure 3a gives the distribution of the liquid volume fraction and again shows the formation of a preferential route for the gas movement for the whole of the liquid drop, however, in the active bubbling area the mixture occurred in a more homogeneous way with a lower elevation of the hold-up. In Figure 3b, the behavior of the velocity vectors of the gas phase is observed, in the same plane for which the profile of the liquid volume fraction was demonstrated, showing circulation zones near the walls. Again it is

possible to observe a preferential route for the gas phase drainage through the liquid fall tube.

In Figure 3c, it can be observed that at 1.5cm below the surface of the lower plate there is no homogeneous mixture of phases for the plate with electrical resistance because there are points with different concentrations of water (or gas), as noted by the appearance of regions with higher concentrations of water (strong blue) and regions where water is almost absent (green tending toward red). In this situation, mass transfer does not occur in a maximized way in the bubbling region, and there are concentration gradients in this region. However, if compared to Figures 2c and 3c, the mixture occurring on the surface of the plates with the presence of resistance is more effective than in the case without resistance. Figure 3d shows the behavior of the velocity vectors of the liquid phase at 1.5 cm from the surface of the lower plate, where it is possible to observe that circulation occurs near the curved walls of the plate in the liquid fall region and some small areas near the electrical resistor wall.

With the aim of analyzing the liquid volume fraction distribution on the surface of the plates, the distribution of this fraction in the central region of plates 2 cm from the surface (bottom plate) was plotted across the entire range, as seen in Figure 5. To demonstrate the liquid volume fraction distribution in the active bubbling region (-0.05 to 0.05 in X[m]), three lines were drawn on it (one in the center of the plate and two 5 cm away to the right and to the left of the central line) for both the lower plate (red lines) and the top plate (blue lines).
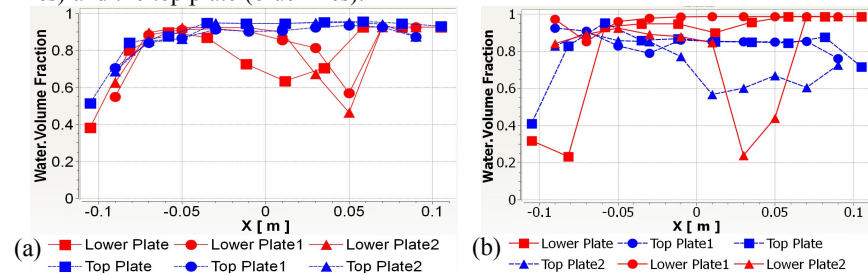

Figure 4 - Distribution of liquid volume fraction on the plate surface with (a) no air injection, (b) air injection, and (c) air injection with electrical resistance.

It can be observed in Figure 4a (without air injection) that the distribution of the liquid volume fraction in the lower plate (represented by the red lines) presents a great variation in the bubbling region, presenting points of minimum fluid concentration nearer to the wall than the hold-up (close to 0.05 m), which is an undesirable factor in terms of the operation of the plate, because the presence of concentration and temperature gradients will occur on the surface of the plate, reducing its efficiency. In the top plate the mixture of the phases occurs in a considerably more homogeneous and satisfactory manner, in terms of the operation of perforated plate distillation.

In Figure 4b (with air injection) the distribution of the liquid volume fraction on the lower plate (represented by red lines) shows only a small region where there is a considerable variation in the volume fraction in the liquid phase in the bubbling region nearer to the wall than the hold-up (close to 0.05m). On the top plate the mixture between the phases occurred in a considerably more homogeneous and satisfactory way in terms of the operation of perforated plate distillation.

Comparing the results obtained with and without the presence of electrical resistance on the surface of the plates, it can be observed that the distribution of the liquid volume fraction, that is, the hydrodynamic flow on the plate, is not adversely affected by its

presence. It can be verified in Figures 4a and 4b that the presence of the electrical resistance even favors the mixture since there is a more homogeneous distribution of the liquid volume fraction on the plate. From the separation process point of view this is beneficial because it minimizes the possibility for concentration and temperature gradients in this region.

5. Conclusions

The main results of this study confirmed that understanding the inter-phase transfer is key to predicting gas-liquid flow in a distillation tray. The CFD tools presented and discussed herein enabled a better understanding of the turbulent gas-liquid flow in a distillation column sieve tray and they could be applied to optimize the design and operating conditions of such processes.

It is concluded from the results obtained in this study that the presence of electrical resistance on the surface of the plates influenced the patterns of the liquid flows on it, but this influence does not adversely affect the hydrodynamics, in fact, it favors the homogenization of phases, highlighting the benefits of this approach to distributed control. It is thus concluded that the use of electrical resistance on the surface of the plates is a feasible option for use in control systems with distributed heating. Moreover, it improves the homogenization of the phase mixture in the active bubbling region, which is very important in terms of maximizing the mass transfer. Based on these results, in a subsequent stage, a study will be carried out with ethanol and water mixtures in order to analyze the efficiency of the separation process in the two domains, as well as to evaluate the effect of electrical resistance on the mass transfer process in this unit.

6. Acknowledgements

References

D.L., Bennett, R. Agrawal, P.J. Cook, P.J, 1983, New pressure drop correlation for sieve tray distillation columns, AIChE J., v. 29, p. 434–442.

R. Krishna, J.M. van Baten, J. Ellenberger, A.P. Higler, R. Taylor, 1999, CFD simulations of sieve tray hydrodynamics, Chem. Eng. Res. Des, v. 77, p. 639–646.

X.G. Li, D. X. Liu, S. M.; Xu, H. Li, 2009, CFD simulation of hydrodynamics of valve tray. Chemical Engineering and Processing: Process Intensification, v. 48, p. 145-151.

C.J.X. Liu, G. Yuan, K.T. Yu, X.J. Zhu, 2000, A fluid-dynamic model for flow pattern on a distillation tray, Chemical Engineering Science, v. 55, p. 2287-2294.

C.R. Maliska, 2004, Transferência de calor e mecânica dos fluidos computacional. Rio de Janeiro: Livros Técnicos e Científicos.

C. Marangoni, R.A.F. Machado, 2007, Distillation tower with distributed control strategy: Feed temperature loads. Chemical Engineering an Technoly, v.30, p.1292-1297.

M. R. K. Nikou, M. R. Ehsani, 2008, Turbulence models application on CFD simulation of hydrodynamics, heat and mass transfer in a structured packing, International Communications in Heat and Mass Transfer, v. 35, p. 1211-1219.

D. Noriler, H. F. Meier, A. A. C. Barros, M. R. Wolf Maciel, 2007, Thermal fluid dynamics analysis of gas-liquid flow on a distillation sieve tray, Chem. Eng. J., doi:10.1016/j.cej.2007.03.023.

J.M. van Baten, R. Krishna, 2000, Modelling sieve tray hydraulics using computational fluid dynamics, Chem. Eng. J., v. 77, p. 143–151.

Robust algorithms for the calculation of phase equilibrium

F.E. Pereira, G. Jackson, A. Galindo , C.S. Adjiman[a]

a Department of Chemical Engineering, Centre for Process Systems Engineering, Imperial College London, London SW7 2AZ, U.K.. c.adjiman@imperial.ac.uk.

Abstract

An algorithm for the solution of the phase equilibrium problem at constant pressure and temperature (*P,T* flash) is presented. The approach is based on the formulation of the phase stability problem as a dual optimisation problem that results from the minimisation of the Helmholtz free energy. The problem is solved in composition and volume space, which is especially suited to complex equations of state (EOS) that are higher than cubic functions in volume, and formulated in the Helmholtz free energy, such as SAFT (statistical associating fluid theory) [Chapman, W.G., Gubbins, K.E., Jackson, G. and Radosz, M. Ind. Eng. Chem. Res., 29 (1990), pp. 1709.]. With the proposed method, one is mathematically guaranteed to find all stable phases for multi-component systems with any number of fluid phases. Results are presented for the case of liquid-liquid equilibria in an asymmetric ternary system, modelled with a SAFT EOS.

Keywords: Multicomponent phase equilibrium, Global optimisation, Duality.

1. Introduction

The solution of the phase equilibrium (PE) problem at constant pressure and temperature (*P,T* flash) is an important step in process or product performance and hence in model-based design activities. Even for binary mixtures, this problem can be very challenging [1; 2; 3] due to the high degree of nonlinearity and the presence of discontinuities in the derivatives of the Gibbs free energy function.

The solution of this PE problem corresponds to the global minimum of the total system Gibbs free energy [4]. However, locating this equilibrium state is complicated because the number of equilibrium phases (*np*) is generally unknown *a priori* [5]. The alternating flash/stability test approach first introduced by Michelsen[6; 7] is the ubiquitous framework for addressing this problem. In this approach, *np* is postulated, and is an input in the construction of the objective function. The correctness of this assumption is tested *a posteriori* through tangent plane analysis. If the specification of *np* is proven to be inaccurate, the entire system must be resolved for a different number of phases. This can slow down calculations and, unless global optimisation is employed[8], the equilibrium state obtained may be influenced by poor initial guesses or the presence of metastable solutions.

Process engineering applications could therefore benefit from PE algorithms that are more robust and require less *a priori* knowledge of the system in question. In particular, applications in polymer systems and solvent design would profit from greater numerical stability, since inbuilt assumptions are often a cause of failure in these systems.

An additional hurdle to PE calculations exists if the thermodynamic properties are obtained from one of the many equations of state (EOS) that are formulated in the Helmholtz free energy, such as Soave-Redlich-Kwong (SRK)[9], Peng-Robinson[10] and SAFT[11; 12]. Unless the volume roots of the[13] EOS can be obtained

analytically, as is the case for cubic equations of state, calculations at specified pressure and temperature become time consuming. This is because the calculation of $G(\underline{x},T,P)$ (the Gibbs free energy at specified composition, temperature and pressure) at each point in composition requires the use of a nonlinear solver that can reliably identify the appropriate volume root for use in the calculation. Nagarajan *et al.*[14; 15] have developed a stability test and flash based on the Michelsen framework, which takes the molar densities of the components as the independent variables. This essentially adds volume as a variable to the optimisation problem, and removes the need to identify the correct volume root for a given $\underline{x},T,P$ combination.

The link between phase stability and Lagrangian duality [13] provides some interesting avenues for research [16]. Mitsos and Barton [17] have recently proposed an alternative approach to determining whether a phase is stable. They showed that the solutions of a dual problem stemming from the minimisation of the Gibbs free energy, where a single phase is assumed, and mass balance constraints are imposed, are stable equilibrium phases. When posed as a dual problem, the convergence may be largely decoupled from any initial guesses for the properties of the stable phases, even when global optimisation is not employed. In addition, *np* is no longer present in the formulation, and consequently the problem dimensions are reduced to *nc*-1 (with *nc* the number of components), from *np**(*nc*-1) dimensions present in a 'traditional' formulation.

In this paper, we present an alternative formulation of the PE problem, extending the method of Mitsos and Barton[17], and propose an efficient algorithm for its solution. Our approach is particularly well-suited to complex equations of state formulated in the Helmholtz free energy, such as SAFT because, following Nagarajan *et al.*[14], it does not require the explicit identification of volume roots in the calculation of the system properties at specified *T* and *P*. Unlike [14] however, our translation of the PE problem into the volume space does not introduce bilinearity into the mass balance constraint. Subsequently, we propose a deterministic algorithm for the guaranteed identification of all stable phases, which makes minimal use of global optimisation solvers. We describe our implementation and present computational results with an EOS that is a non cubic function of volume: the SAFT-HS (Statistical Associating Fluid Theory for Hard Spheres) EOS[18; 19]. We examine polymer-like phase behaviour in a ternary system, and show the effectiveness of the proposed approach in dealing with this asymmetric and numerically challenging example.

2. Problem formulation

Consider a mixture at fixed temperature T^0 and pressure P^0 with total composition $\underline{x}^0$, where $\underline{x}^0$ is an (*nc*-1)-dimensional vector of feed mole fractions, and *nc* is the number of components. A trivial minimisation problem may be formulated in which the system is treated as if in a single phase:

$$\min_{\underline{x},V} \quad A(\underline{x},V,T^0) + P^0 V$$
$$\text{s.t.} \quad x_i^0 - x_i = 0 \quad i = 1,\ldots,nc-1$$
$$\underline{x} \in X \subset \mathbf{R}^{nc-1}$$
$$V > 0 \tag{1}$$

where $A(\underline{x},V,T^0)$ is the Helmholtz free energy, $\underline{x}$ is the vector of component mole fractions, and V is volume. It can be shown that solution of the *P,T* flash problem may

be formulated as finding all the global solutions of a dual problem arising from problem (1):

$$\max_{\lambda \in \mathbb{R}^{nc-1}} \theta^V(\lambda)$$

$$s.t. \quad \theta^V(\lambda) = \min_{x \in X, V > 0} L^V(x, V, \lambda)$$

$$s.t. \quad L^V(x, V, \lambda) = A(x, V, T^0) + P^0 V + \sum_{i=1}^{nc-1} \lambda_i (x_i^0 - x_i)$$

$$\tag{2}$$

where $\theta^V(\lambda)$ is the dual objective function, and λ is the component-wise vector of Lagrange multipliers on the mass balance. In this context, the solution of the PE problem corresponds to the highest supporting hyperplane on the $G(x, P^0, T^0)$ surface at x^0, which is the tangent plane corresponding to the phase stability criterion [20].

Problem (2) is a bilevel program. The outer problem is solved though a maximisation of the concave function $\theta^V(\lambda)$. In order to evaluate $\theta^V(\lambda)$ for a given λ, one must solve the inner problem by minimising the Lagrangian function L^V with respect to x and V, for fixed λ. The inner problem is solved through an unconstrained minimisation of the nonconvex Lagrangian, defined as:

$$\min_{x \in X, V > 0} L^V(x, V, \lambda^k) = \min_{x \in X, V > 0} \left\{ A(x, V, T^0) + P^0 V + \sum_{i=1}^{nc-1} \lambda_i^k (x_i^0 - x_i) \right\}$$

$$\tag{3}$$

A proof that the global solution of this dual optimisation problem corresponds to a stable phase is presented in *[21]*, by showing the equivalence of (2) with a dual problem formulated in the x space only, as discussed in*[17]*. At equilibrium, $\lambda = \lambda^*$, and is related to the equilibrium chemical potential vector. $\theta^V(\lambda^*)$ corresponds to the system Gibbs free energy over all phases. The stable phases are joined by a supporting plane for both the Gibbs and Helmholtz free energy surfaces *[21]*. The global solutions (stable phases) are described by a unique λ^* and different values of x and V.

3. Phase equilibrium algorithm

The outer problem is solved with a cutting-plane algorithm [22; 23] summarised in figure 1. Local minimisations of the inner problem are used to generate cutting planes, which provide progressively tighter constraints on the feasible region of λ, and therefore the upper bound on the outer problem. The PE problem is solved through alternating solution of the inner and outer problems. The outer problem is a linear maximisation, subject to the cutting planes created by the solutions of the inner problem:

$$UBD^V = \max_{v, \lambda \in \mathbb{R}^{nc-1}} v$$

$$s.t. \quad v \leq A(x^m, V^m, T^0) + P^0 V^m + \sum_{i=1}^{nc-1} \lambda_i (x_i^0 - x_i^m) \ \forall [x^m, V^m] \in \quad (4)$$

$$v \leq G^p$$

where G^P is the solution of (1). Each solution of the inner problem creates a new linear constraint. This is added to the outer problem through the set ⬚. If the inner problem is solved to global optimality, then the constraint created is tangential to the function $\theta^V(\underline{\lambda})$, and hence the tightest possible cutting plane for the given $\underline{\lambda}$. Local solutions of (3) provide looser bounds on the outer problem. Local solvers are employed to the greatest extent possible while retaining a deterministic guarantee on the final solution. The algorithm resorts to the global solver only when no suitable solution is obtained for the inner or outer problem, and at convergence, as a final check of global optimality. The thick arrows in figure 1 denote the most common route taken during an iteration of the algorithm. The multipliers $\underline{\lambda}$ are initialised as $(\partial G(\underline{x}, P^0, T^0)/\partial x_i)|\underline{x}=\underline{x}^0$, the gradients of the Gibbs free energy at the feed conditions. This allows the first iteration to double as a stability test. The constraint set on the outer problem is initialised so as to bound $\underline{\lambda}$ effectively.

All the stable phases present at equilibrium are global solutions of the dual problem. Therefore, at the equilibrium value of $\underline{\lambda} = \underline{\lambda}^*$, each stable phase is a global solution of the inner problem. If a spatial branch-and-bound algorithm (sBB) is used to solve the inner problem, it can be used to find all global solutions (all stable coexisting phases) to the inner problem, at little extra computational cost. This step is denoted "Find all solutions" in Figure 1.
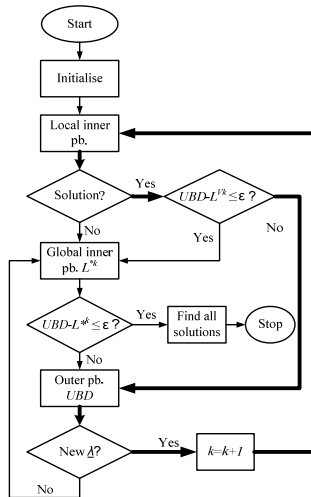

Figure 1. The PE algorithm. The local inner problem is problem (3), *UBD* is the best solution of the outer problem (4), L^{Vk} is the best local solution of the inner problem at iteration k, and L^{*k} is the global solution of the inner problem at iteration k.
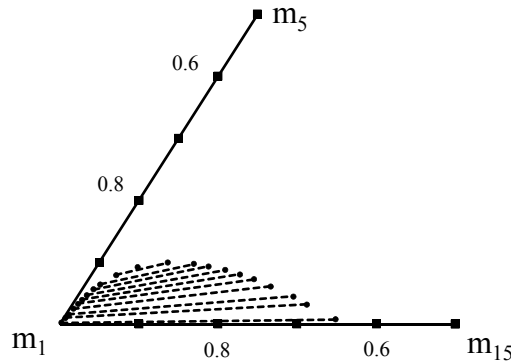
Figure 2. LLE in the example SAFT-HS system, at $T^*=0.091$ and $P^*=0.0037$. The hollow circle indicates the feed composition, $\underline{x}^0$, of the point calculation. Note that the plot is in weight fraction and does not extend to the pure component apexes of m_5 and m_{15}.

4. Implementation

The algorithm described in section 3 has been implemented in GAMS 23.2 [24]. The inner problem is solved globally with BARON V8.1.5 [25] and locally with MINOS 5.51 [26]. The linear outer problem is solved with CPLEX 12.1.0 [27]. We have found

that the most effective way to identify all stable phases in a GAMS implementation of this type is through a final, 'multi-solution' BARON call. Thus, the inner problem for the optimal λ^* vector is solved a second time. In our implementation, we use BARON to return the best n solutions, with a minimum separation between these solutions of $1x10^{-4}$ ('isoltol=$1x10^{-4}$'), so that the same phase is not identified twice. We use BARON to return four solutions during a final iteration for a ternary mixture, and three for a binary. One may subsequently examine these solutions to ascertain how many phases, or global solutions are present.

5. Example calculations

To illustrate the effectiveness of our proposed algorithm we carry out calculations for a ternary system modelled with a simplified version of the SAFT EOS in which the molecules are composed of chains of m equal-size spherical segments [18; 19; 28]. The ratio of molecular chain lengths is chosen as (component 1:component 2:component 3) 1:15:5. This corresponds to an asymmetric system, which exhibits liquid –liquid equilibrium terminating in a plait (critical) point, at the specified conditions of T^*=0.091 and P^*=0.0037 (where T^* and P^* are both reduced by the (symmetrical) attractive interaction of the molecules; $T^*=kTb/a$, and $P^*=Pb^2/a$. b represents the volume of the spheres comprising the molecule and a the attractive interaction). This solvent + bi-disperse polymer type phase behaviour is highly non-ideal and presents a difficult challenge to any PE solution method.

A section of the phase diagram of the system at these conditions, generated using the proposed algorithm, is shown in figure 2, plotted in weight fraction. In generating the diagrams, no initial guesses were used, and no numerical difficulties were encountered. To illustrate computational performance, a point calculation was carried out at $\underline{x}^0$= $[0.98,0.01,0.01]^T$, shown by a hollow circle on figure 2. Ten local minimisations were performed for the solution of each inner problem. The problem required 13 major iterations (solutions of the inner and outer problems), and took 20.1s to converge to the optimal chemical potential, on an HP Workstation (Intel Xenon, 3GHz). The global solver was called only once, at convergence. The absolute optimality tolerance used for the dual (stability) problem was $\varepsilon=1x10^{-6}$. The optimality tolerance for BARON, MINOS and CPLEX was $\varepsilon=1x10^{-8}$. The two stable phases at this feed mole fraction have the compositions, $\underline{x}^\alpha$=[0.962,0.022,0.016] and $\underline{x}^\beta$= [0.995,0.0003,0.00047].

6. Conclusions

In this contribution we have presented a formulation of the P,T flash as a dual optimisation problem in the volume-composition space, using the Helmholtz free energy. This space is more convenient for calculations with complex EOS written in the Helmholtz free energy. It avoids the hurdles encountered when working directly in the Gibbs free energy, such as the time consuming process of finding the appropriate volume for a specified pressure. In addition, we have proposed an algorithm for the efficient solution of the problem. It is based on a cutting plane approach, in which cutting planes are generated at low computational cost, via local optimisation. Global optimisation is used to ensure the correct solution has been found and that all stable phases can be identified. The low dimensionality of the dual formulation (the number of variables is independent of the number of phases) and the absence of any guess as to the number of stable phases, or their properties, are important advantages.

The algorithm has been implemented in GAMS with the SAFT-HS EOS. An example calculation has been shown for the challenging phase behaviour in an asymmetric, polymer + polymer + solvent type system.

7. Acknowledgements

References

1. Gau, C.Y., Brennecke, J.F. and Stadtherr, M.A. 2000, Fluid Phase Equilibria, Vol. 168, p. 1.
2. Giovanoglou, A., Galindo, A., Jackson, G. and Adjiman, C.S. 2009, Fluid Phase Equilibria, Vol. 275, p. 79.
3. Giovanoglou, A., Adjiman, C.S., Jackson, G. and Galindo, A. 2009, Fluid Phase Equilibria, Vol. 275, p. 95.
4. White, W. B., Johnson, S. M. and Dantzig, G. B. 1958, The Journal of Chemical Physics, Vol. 28(5), p. 751.
5. Castillo, J. and Grossmann, I. E. 1981, Computers & Chemical Engineering, Vol. 5, p. 99.
6. Michelsen, M.L. 1982, Fluid Phase Equilibria, Vol. 9, p. 1.
7. Michelsen, M.L. 1982, Fluid Phase Equilibria, Vol. 9, p. 21.
8. Floudas, C.A. and Gounaris, C.E. 2009, J. Glob. Optim., Vol. 45(1), p. 3.
9. Soave, G. 1972, Chemical Engineering Science, Vol. 27, p. 1197.
10. Peng, D.Y. and Robinson, D.B. (1), 1976, Ind. Eng. Chem. Fundam., Vol. 15, p. 59.
11. Chapman, W.G., Gubbins, K.E., Jackson, G. and Radosz, M. 1989, Fluid Phase Equilibria, Vol. 52, p. 31.
12. Chapman, W.G., Gubbins, K.E., Jackson, G. and Radosz, M. 1990, Ind. Eng. Chem. Res., Vol. 29, p. 1709.
13. Modell, M. and Reid, R. C. *Thermodynamics and Its Applications in Chemical Engineering.* NJ : Prentice Hall, 1983.
14. Nagarajan, N. R., Cullick, A. S. and Griewank, A. 1991, Fluid Phase Equilibria, Vol. 62, p. 191.
15. Nagarajan, N. R., Cullick, A. S. and Griewank, A. 1991, Fluid Phase Equilibria, Vol. 62, p. 211.
16. McDonald, C. M. and Floudas, C. A. 1995, AIChE Journal, Vol. 41, p. 1798.
17. Mitsos, A. and Barton, P.I. 2007, AIChE Journal, Vol. 53(8), p. 2131.
18. Galindo, A., Whitehead, P. J., Jackson, G. and Burgess, A. N. 1996, J. Phys. Chem., Vol. 100, p. 6781.
19. Galindo, A., Whitehead, P. J., Jackson, G. and Burgess, A. N. 1997, J. Phys. Chem. B, Vol. 101, p. 2082.
20. Baker, L.E., Pierce, A.C. and Luks, K.D. 1981, SPE/DOE Second Joint Symposium on Enhanced Oil Recovery, Tulsa, Oklahoma.
21. Pereira, F.E., Jackson, G., Galindo, A. and Adjiman, C.S. 2009, In preparation.
22. Blankenship, J.W. and Falk, J.E. 1976, J. Optim Theory Appl., Vol. 19(2), p. 261.
23. Falk, J.E. and Hoffman, K. 1977, Naval Res Logistics, Vol. 24, p. 441.
24. *General Algebraic Modelling System (GAMS).* 2009, www.gams.com.
25. *BARON with GAMS user guide.* Sahinidis, N. and Tawarmalani, M. 2009.
26. *MINOS with GAMS user guide.* Murtagh, B.A., Saunders, M.A. and Gill, P.E. 2009.
27. *CPLEX 12 with GAMS user guide.* Corporation/Ilog, GAMS Development. 2007.
28. Paricaud, P., Galindo, A. and Jackson, G. 2003, Molecular Physics, Vol. 101(16), p. 2575.

Prediction of Interfacial Tension of Binary Mixtures

Oscar G. Nino-Amezquita, Sabine Enders*

*TU Berlin, TK7, Institute of Process Engineering, Department of Thermodynamics and Thermal Engineering, Straße des 17. Juni, 10623 Berlin, Germany, *sabine.enders@tu-berlin.de, gabriel.nino.a@mailbox.tu-berlin.de*

Abstract

The PCP-SAFT is applied in the density gradient theory for prediction of the surface tension of mixture containing a component having a dipole or a quadrupole moment. The surface tension of mixture can be predicted very close to experimental data taken from the literature.

Keywords: Interfacial tension, density gradient theory, dipolar and quadrupole substances, PCP-SAFT

1. Introduction

Besides the phase behavior the interfacial properties play an important role in process design. This is especially true in the field of oil- and gas gathering. In order to reduce the global warming effect carbon dioxide (CO_2) can pressed into the oil- or/and gas reservoir. From the economical point of view the interfacial tension between the carbon dioxide and the oil- or/and gas phase, and hence the necessary pressure is an essential information. Recently [1], a method was suggested to predict the surface tension of non-polar mixture based on the density gradient theory [2] in very good agreement with experimental data. However, the most important feature of CO_2 is the quadrupole moment. The present paper aims to the incorporation of molecules having a dipole or a quadrupole moment in the theoretical framework [1] without the use of an additional adjustable parameter. One of the last extension of the well-known Perturbed - Chain - Statistical Association Fluid theory [3] (PC-SAFT) is the application to molecules having a dipole [4] or a quadrupole [5] moment. This new equation of state is called PCP-SAFT. The basic idea of this paper is the combination of the density gradient theory with the new PCP-SAFT in order to predict the interfacial properties of mixtures containing molecules with dipole or quadrupole moment.

2. Theoretical Approach

The Cahn - Hilliard Theory [2] describes the thermodynamic properties of a system where an interface exists between two fluid phases. At thermodynamic equilibrium the pressure, the temperature and the composition in the bulk-phases are fixed by the thermodynamics equilibrium conditions. For pure components only the density changes within the interface. For binary mixtures the density as well as the concentration will change throughout the interface. In order to take both effects into account we define the partial densities ρ_i.

$$\rho_i = X_i \rho \tag{1}$$

where X_i is the mole fraction of component i and ρ the density of the mixture. In general, the thermodynamic quantities of binary systems made from the components A and B at constant temperature are functions of density and composition, expressed by mole fraction (e.g. Helmholtz energy $F(\rho, X_i)$). Applying an equation of state the thermodynamic functions can be rewritten as functions of both partial densities (e.g. $F(\rho_A, \rho_B)$). Using a similar procedure like Cahn and Hillard [2] to calculate the interfacial tension of a planar interface, Poser and Sanchez [6] worked out a method that considers the change of two variables (ρ_A, ρ_B) within the interface. The interfacial tension between two fluid phases in equilibrium reads:

$$
\begin{aligned}
\sigma &= 2^{1/2} \int_{\rho_B^I}^{\rho_B^{II}} \left[\kappa_B + 2\kappa_{AB}\left(\frac{d\rho_A}{d\rho_B}\right) + \kappa_A\left(\frac{d\rho_A}{d\rho_B}\right)^2 \right]^{1/2} \Delta\omega^{1/2} \, d\rho_B \\
&= 2^{1/2} \int_{\rho_A^I}^{\rho_A^{II}} \sqrt{\kappa' \Delta\omega} \, d\rho_B
\end{aligned}
\tag{2}
$$

where κ' results from the so-called influence parameters of the pure components, κ_i, and $\Delta\omega$ is the grand thermodynamic potential. The κ_i-parameter with i=A or B in Eq. (2) can be adjusted to one experimental surface tension of the pure component at one temperature. The parameter κ_{AB} can be calculated using geometrical average of the pure component parameters:

$$\kappa_{AB} = \sqrt{\kappa_A \kappa_B} \tag{3}$$

This situation allows the prediction of the surface tension of binary mixtures as a function of temperature, pressure and composition because only information about the bulk phase via the EOS and one surface tension of pure components, which can be found in a data base [7] enter the theoretical framework.

The limits of integration ρ_B^I and ρ_B^{II} are the partial densities in the coexisting bulk phases and they can be obtained by the calculation of the phase equilibrium. The grand thermodynamic potential $\Delta\omega$ for binary mixtures results in:

$$\Delta\omega = \left(F(T,V,X_B) - X_A \mu_A^V - X_B \mu_B^V + PV \right) \rho \tag{4}$$

The suggested procedure permits the calculation of the interfacial tension, if a suitable equation of state is applied. The goal of this contribution is the description of the interfacial properties of mixtures, where a dipolar or a quadrupole component is present. Gross [5] as well Gross and Vrabec [4] developed a version of the Perturbed Chain Statistical Association Fluid Theory (PC-SAFT) [3], where polar and quadrupole contributions are incorporated (PCP-SAFT). Using the PCP-SAFT-approach the

residual Helmholtz free energy (F^{res}) consists of the hard-chain reference (F^{hc}), the dispersion (F^{disp}), the dipole (F^{dipol}) and the quadrupole (F^{quad}) contributions:

$$F^{res} = F^{hc} + F^{disp} + F^{dipol} + F^{quad} \qquad (5)$$

In principle an association term can also be used, however the molecules considered in this paper have no associative forces and therefore this term is not used. Depending from the chemical nature of the molecules different contributions from Eq. 5 were applied. i.e. carbon dioxide has a quadrupole moment and the term F^{quad} is used. In order to model F^{dipol} or F^{quad} the equations given in the literature [4,5] were used, because no additional adjustable parameter is required.

3. Results and Discussion

3.1. Pure components

Based on the suggested theoretical framework first the phase equilibria and the surface tensions of pure components in order to estimate the influence parameter of the pure components must be calculated. The most critical quantity for the calculation of surface tension is the liquid volume in equilibrium with the corresponding vapor phase. In Figure 1 the calculated liquid volumes of CO_2 using PC-SAFT and PCP-SAFT are compared with experimental data taken from the literature [8]. The pure component parameters were taken from Gross [5]. It can be clearly recognized the improvement of the calculation results, if the quadrupole of CO_2 is included in the theoretical framework, especially in the critical region.
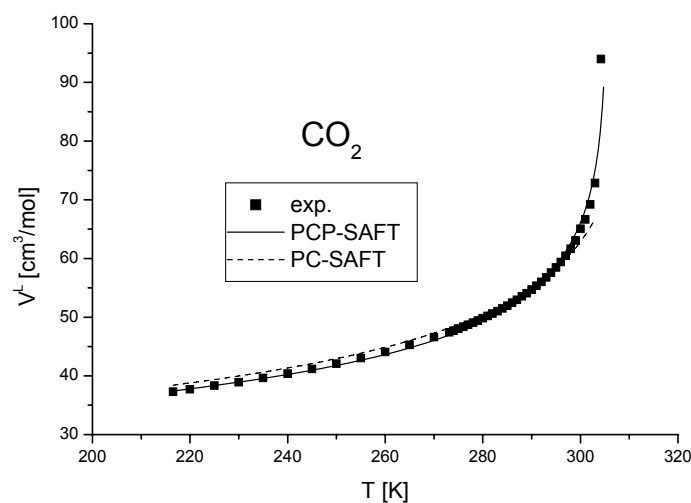
Figure 1: Comparison of calculated liquid volumes of CO_2 using PCP-SAFT and PC-SAFT with experimental data [8].

The improvement of the calculated liquid volumes leads to an improvement in the calculated surface tension (Figure 2). Additionally the calculated and experimental surface tensions of dipolar molecules, like acetone and methyl chloride, are plotted in Figure 2. The pure-component parameters are also taken from the literature [4]. The surface tension at one temperature is used to estimate the influence parameter κ. The

obtained parameters are listed in Table 1. Analyzing the data given in Figure 2 it can be concluded the density gradient theory in combination with the PCP-SAFT EOS is able to calculated the temperature dependency of the surface tension in an excellent agreement with experimental data.

Figure 2: Comparison of calculated surface tension of CO_2, acetone and methyl chloride using PCP-SAFT (solid line) and PC-SAFT (broken line) with experimental data [7,8,,9,10].

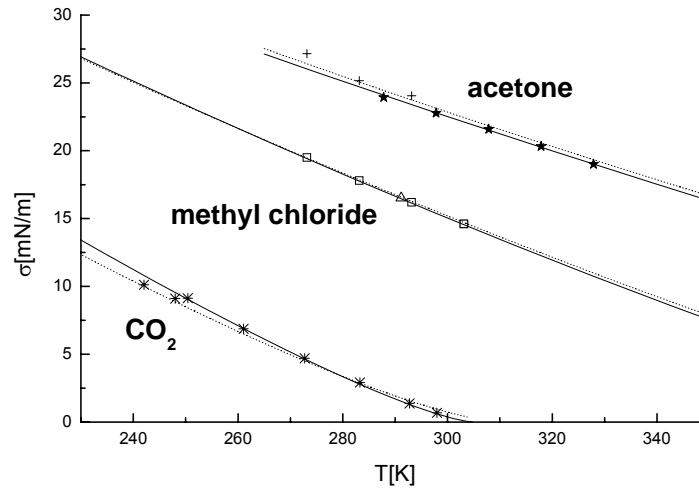Table 1: Influence parameter of the pure components, κ, used in PC-SAFT and PCP-SAFT.

EOS	component	CO_2	methyl chloride	acetone	butane
PC-SAFT	$\kappa\ 10^{-20}\ Jm^5/mol^2$	2.312	5.65	11.57	-
PCP-SAFT	$\kappa\ 10^{-20}\ Jm^5/mol^2$	2.327	5.67	11.49	17.81

3.2. Systems Carbon Dioxide + n-butane

In order to study the possibility to predict the surface tension of mixtures containing one non-polar and one component having a polar or quadrupole moment we choose the system CO_2 + n- butane, because for this system experimental data are available in the literature. We focus our attention to the data measured by Hsu et al. [11]. The phase equilibria of this system at other temperatures were also calculated by Gross [5]. In this paper [5] a temperature-independent binary interaction parameter k_{ij}=0.036 was found. Using this binary interaction parameter the resulting phase equilibria at temperatures where the surface tension was measured are shown in Figure 3. The vapor as well the liquid volumes at equilibrium condition could be predicted with a high accuracy. The statement holds true also in the critical region.

After phase equilibria calculations the surface tension can be predicted using Eq. 2 in combination with Eq. 3 . In Figure 4 the predicted surface tensions are compared with experimental data taken from the literature [11]. At 344.3K and 377.6K the predicted surface tensions are within the experimental error. Similar calculations were carried out by Cornelisse [12] using the Peng-Robinson EOS; however, for the influence parameter of both components, κ_A and κ_B, and for the binary interaction parameter, k_{ij}, a linear temperature dependency was assumed. Taking the quadrupole moment of CO_2 into

account via the PCP-SAFT-EOS leads to an improvement of the prediction results and the temperature dependency of the adjustable parameters can be neglected.
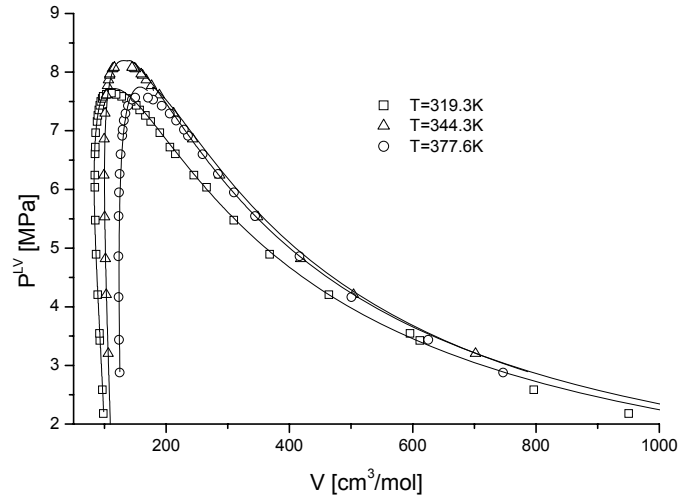
Figure 3: Experimental [11] and predicted phase equilibria of the system CO_2 and n-butane.
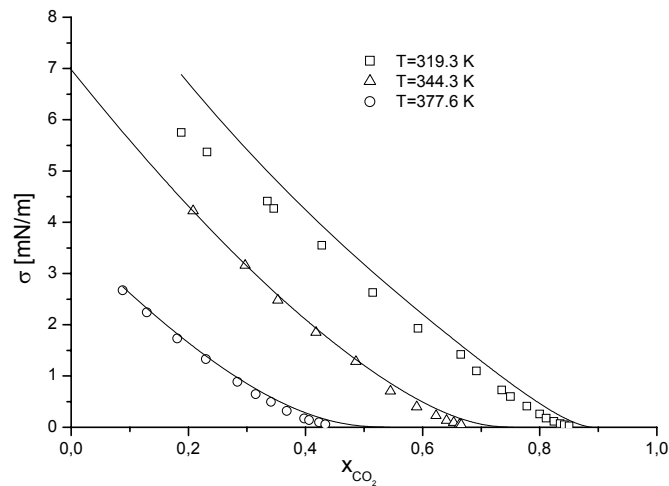
Figure 4: Experimental [11] and predicted surface tension of the system CO_2 and n-butane.

The density gradient theory allows also the calculation of the surface profiles (Figure 5). These profiles for CO_2 run through a maximum and hence CO_2 will be enriched in the surface and can lead to a barrier for mass transport through the interface, especially at low CO_2 mole fractions in the liquid mixture. The profiles for n- butane have the usually tanh-shape without a maximum.
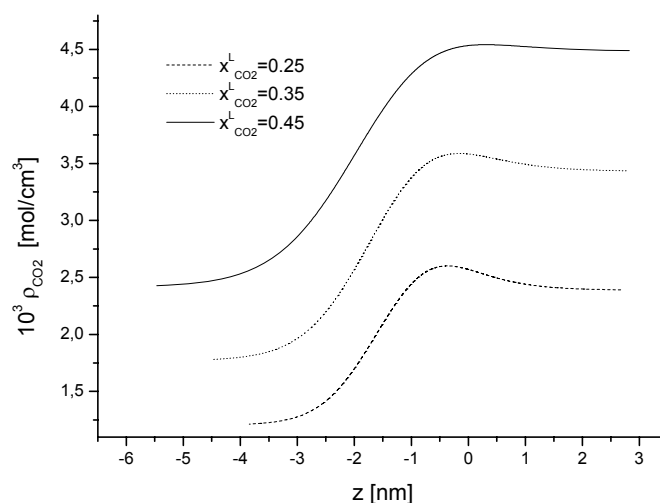
Figure 5: Surface profiles of CO_2 at different CO_2 levels in the liquid phase and at 344.3K.

4. Summary

The combination of the PCP-SAFT with the density gradient theory leads to a powerful tool for the prediction of surface tension of binary mixtures, if molecules with dipole or quadrupole moments are present.

5. Acknowledgements

References

[1] O.G. Nino-Amezquita, S. Enders, P.T. Jaeger , R. Eggers, Ind. Eng. Chem. Res. 2009 in press.
[2] J.W. Cahn and J.E. Hilliard, J. Chem. Phys. 28 (1958) 258.
[3] J. Gross and G. Sadowski, Ind. Eng. Chem. Res. 40 (2001) 1244.
[4] J. Gross and J. Vrabec, AIChE J. 52 (2006) 1194.
[5] J. Gross, AIChE J. 51 (2005) 2556.
[6] C.I. Poser and L.C. Sanchez, Macromolecules 14 (1981) 361.
[7] Ch. Wohlfarth, B. Wohlfarth, in M.D. Lechner (Ed.) Numerical Data and Functional Relationships in Science and Technology, Vol. 16 Surface Tension of Pure Liquids and Binary Liquid Mixtures, in Landoldt-Börnstein, New Series Group IV Physical Chemistry, Springer Verlag Berlin, Heidelberg, 1997.
[8] J.M. Soerensen and W. Arlt, Vapour-Liquid Equilibrium Data Collection, Bd 1, Binary Systems. DECHEMA Frankfurt/Main, 1979.
[9] H. Kahl, T. Wadewitz, J. Winkelmann, J. Chem. Eng. Data 48 (2003) 1500.
[10] H.Li, W. Chen, Phys. Chem. Liq. 43 (2005) 343.
[11] J.J.C. Hsu, N. Nagarajan, R.L. Robinson, J. Chem. Eng. Data 30 (1985) 485.
[12] P. Cornelisse, thesis, Technical University Delft, 1997.

Transport Phenomena Modeling During Drying of Shrinking Materials

Maria Aversa, Stefano Curcio, Vincenza Calabrò, Gabriele Iorio

Department of Engineering Modeling, Ponte P.Bucci - Cubo 39/c , University of Calabria, Rende (CS) 87030, ITALY
maria.aversa@unical.it; stefano.curcio@unical.it; vincenza.calabro@unical.it; gabriele.iorio@unical.it

Abstract

Different kind of materials are usually submitted to drying. In some cases the aim is to decrease their transportation costs in some others to preserve materials from deterioration. The latter is the case of foods drying. Foods usually exhibit changes in shape and dimensions (known as shrinkage) during drying caused by water loss. The aim of the present work is the modeling of the transport phenomena involved in drying process accounting for also the shrinkage effects. The simultaneous transfer of momentum, heat and mass occurring in a convective drier where hot dry air flows about the food sample have been modeled. The system of non-linear unsteady-state partial differential equations modeling the process has been solved by means of Finite Elements Method coupled to the ALE (Arbitrary Lagrangian Eulerian) procedure that, by a proper modification of integration domain, accounts for shrinkage effects. The model proposed is suitable for industrial equipment optimization.

Keywords: Food drying, Transport phenomena, Finite Elements Method, Process Modeling, Shrinkage

1. Introduction

In a previous paper (Curcio, Aversa, Calabrò, Iorio 2008) the authors of the present work formulated a theoretical model describing the transport phenomena involved in food drying process. The attention was focused on the simultaneous transfer of momentum, heat and mass occurring in a convective drier where dry and hot air flowed, in turbulent conditions, around a wet and cold food sample with a low porosity. Moisture transport inside food with low porosity, where inner evaporation can be neglected, (May & Perré 2002), was modelled by an effective diffusion coefficient of water in the food, thus not distinguishing between the actual transport of both liquid water and vapour within the food structure. Porous foods are hygroscopic materials that contain both bound and free water (Datta 2007 I). Actually during drying of food with high porosity, water evaporation takes place inside the food as well as at the food external surface. Moreover, although evaporation takes place inside the food, the transfer rates occurring at air/food interface are strongly dependent on the drying air velocity field existing in the drying chamber and, particularly, in the boundary layers developing close to the food surfaces exposed to air. For this reason, to improve the precision of the model, in particular close to the solid surfaces, the k-ω model (Wilcox 1998) has been used in the present paper to calculate drying air velocity field and to describe the momentum transport in turbulent conditions. The aim of the present work is to adopt a conservation-based approach to develop a multiphase transport model so to describe convective drying process. The model is based on conservation of liquid water,

vapour and energy in both air and food domain. Moreover, the transfer of momentum in air, in turbulent conditions, is also modelled by k-ω model. To properly describe shrinkage phenomenon (food volume variation during drying), the above-mentioned transport equations have been coupled with a structural mechanics analysis performed on the food sample.The system of non-linear unsteady-state partial differential equations were solved by means of the Finite elements method.

2. Theory

The food under study was a potato sample, dried in a convective oven . The mass balance referred to the liquid water and vapour in the food sample leads, respectively, to the unsteady-state mass transfer equations (Bird, Stewart, Lightfoot, 1960, Welty, Wicks, Wilson, Rorrer, 2001):

$$\partial C_w / \partial t + \underline{\nabla} \cdot \left(-D_w \underline{\nabla} C_w \right) + \dot{I} = 0 \tag{1}$$

$$\partial C_v / \partial t + \underline{\nabla} \cdot \left(-D_v \underline{\nabla} C_v \right) - \dot{I} = 0 \tag{2}$$

where C_w is the water concentration in food, C_v is the vapour concentration in food, D_w is the capillary diffusivity in food and D_v is the diffusion coefficient of vapour in food and $\dot{I}$ is the evaporation rate. The energy balance in the food material leads, according to the Fourier's law, to the unsteady-state heat transfer equation (Bird, Stewart, Lightfoot, 1960, Welty, Wicks, Wilson, Rorrer, 2001).

$$\rho_s C_{p_s} \partial T / \partial t - \underline{\nabla} \cdot \left(k_{eff} \underline{\nabla} T \right) + \lambda \cdot \dot{I} = 0 \tag{3}$$

where T is the food temperature, ρ_s is the density of food sample, C_{p_s} its specific heat, k_{eff} is the food effective thermal conductivity accounting for a combination of different transport mechanisms, λ is the water latent heat of evaporation. The non-isothermal turbulent flow of air within the drying chamber was modeled by means of the well-known k-ω model (Curcio, Aversa, Calabrò, Iorio 2008), that is based on two additional semi-empirical transport equations for the variables *k* and *ω* i.e. the turbulent kinetic energy and the dissipation for unit of turbulent kinetic energy respectively. The unsteady-state momentum balance coupled to the continuity equation written for the drying air leaded to (Bird, Stewart, Lightfoot, 1960, Verboven, Scheerlinck, De Baerdemaeker, Nicolaï, 2001):

$$\partial \rho_a / \partial t + \underline{\nabla} \cdot \rho_a \underline{u} = 0 \tag{4}$$

$$\rho_a \partial \underline{u} / \partial t + \rho_a \underline{u} \cdot \underline{\nabla} \underline{u} = \underline{\nabla} \cdot \left[-p \underline{I} + \left(\eta_a + \eta_t \right) \left(\underline{\nabla} \underline{u} + \left(\underline{\nabla} \underline{u} \right)^T - \left(\tfrac{2}{3} \right) \left(\underline{\nabla} \cdot \underline{u} \right) \underline{I} \right) - \left(\tfrac{2}{3} \right) \rho_a k \underline{I} \right] \tag{5}$$

$$\rho_a \partial k / \partial t + \rho_a \underline{u} \cdot \underline{\nabla} k = \underline{\nabla} \cdot \left[\left(\eta_a + \sigma_k \eta_t \right) \left(\underline{\nabla} k \right) \right] + \eta_t P(\underline{u}) - \left(\tfrac{2 \rho_a k}{3} \right) \left(\underline{\nabla} \cdot \underline{u} \right) - \beta_k \rho_a k \omega \tag{6}$$

$$\rho_a \partial \omega / \partial t + \rho_a \underline{u} \cdot \underline{\nabla} \omega = \underline{\nabla} \cdot \left[\left(\eta_a + \sigma_\omega \eta_t \right) \left(\underline{\nabla} \omega \right) \right] + \left(\tfrac{\alpha \omega}{k} \right) \left[\eta_t P(\underline{u}) - \left(\tfrac{2 \rho_a k}{3} \right) \left(\underline{\nabla} \cdot \underline{u} \right) \right] - \beta \rho_a \omega^2 / k \tag{7}$$

Transport Phenomena Modeling During Drying of Shrinking Materials

where ρ_a is the air density, η_a is its viscosity, both expressed in terms of the local values of temperature and of water content, p is the pressure within the drying chamber, u is the velocity vector. β_k, σ_k, σ_ω, α, β are constants (Wilcox 1998). The term, P(u), contains the contribution of the shear stresses and η_t is the turbulent viscosity.

The energy balance in the drying air, accounting for both convective and conductive contributions, leaded to (Bird, Stewart, Lightfoot, 1960, Welty, Wicks, Wilson, Rorrer, 2001):

$$\rho_a C_{pa} \frac{\partial T_2}{\partial t} - \underline{\nabla} \cdot (k_a \underline{\nabla} T_2) + \rho_a C_{pa} \underline{u} \cdot \underline{\nabla} T_2 = 0 \tag{8}$$

where T_2 is the air temperature, is its specific heat and k_a its thermal conductivity.

The mass balance in the drying air, referred to the water vapour and accounting for both convective and diffusive contributions, leaded to (Bird, Stewart, Lightfoot, 1960, Welty, Wicks, Wilson, Rorrer, 2001):

$$\frac{\partial C_2}{\partial t} + \underline{\nabla} \cdot (-D_a \underline{\nabla} C_2) + \underline{u} \bullet \underline{\nabla} C_2 = 0 \tag{9}$$

where C_2 is the water concentration in the air and D_a is the diffusion coefficient of water in air. The transport and physical properties of potato has been obtained by Srikiatden, Roberts (2008), Zhang, Datta (2004) and Maroulis, Saravacos, Krokida, Panagiotou (2002). As far as the boundary conditions were concerned, the continuity of both heat and mass fluxes and of temperature was formulated at the food/air interfaces. Moreover an equilibrium relationship between the liquid water and vapour inside food was used (Smith, Van Ness, Abbot, 1987). At the oven outlet section, conduction and diffusion phenomena were neglected with respect to convection (Danckwerts conditions). As far as the boundary conditions referred to the momentum balance were concerned, a two-velocity scale logarithmic wall function was used on the solid surfaces, (Lacasse, Turgeon and Pelletier, 2004).

Food shrinkage was modelled defining the local total strains {dε} as a function of changes in mechanical strains *{dεₛ}* (constrained deformation due to food mechanical properties) and in shrinkage strains *{dε₀}* (the sum of a free deformation due to moisture loss):

$$\{d\varepsilon\} = \{d\varepsilon_s\} + \{d\varepsilon_0\} \tag{10}$$

Total strain *{dε}* is actually a function of total displacement *{dU}*

$$\{d\varepsilon\} = [A]\{dU\} \tag{11}$$

Changes in stresses *{dσ}* are function of changes in mechanical strains *{dεₛ}*

$$\{d\sigma\} = [D]\{d\varepsilon_s\} \tag{12}$$

Where [D] is the stress-strain matrix containing the Young Modulus reported by Yang and Sakai (2001) in the case of potatoes drying.

To express the free drying shrinkage strains *{dε₀}*, it was assumed that the free deformation due to moisture loss was proportional to the water content variation, through a constant (the hydrous compressibility factor). The constant was estimated from the experimental data showing drying shrinkage vs. weight loss.

Virtual work principle was formulated to obtain the equilibrium equation. By assuming that zero body and surface forces are applied to food, it can be written:

$$\int_V \delta \{d\varepsilon\}^T \{d\sigma\} dV = 0 \tag{13}$$

As far as the boundary conditions are concerned, one side of the food rests on the drier net (fixed position) whereas the other three are free to move.

The transport equations, together with the virtual work principle represent a system of unsteady-state, non-linear, partial differential equations that can be solved only by means of a numerical method. Equations were written referring to a time dependent deformed mesh that accounted for food volume variation due to water transport.

3. Mathematical Modeling

An Arbitrary Lagrangian-Eulerian (ALE) description, implemented by Comsol Multiphysics, was adopted. It is worthwhile to remark that the ALE method can be considered as "intermediate" between the Lagrangian and Eulerian approaches since it combines the best features of both of them and allows describing moving boundaries without the need for the mesh movement to follow the material.

The boundary conditions control the displacement of the moving mesh with respect to the initial geometry. At the boundaries of food sample, this displacement is actually that calculated by solving the structural mechanic problem.

The system of unsteady, non-linear PDEs resulting from the present study was solved by the Finite Elements Method using Comsol Multiphysics 3.4. Both food and air domains were discretized into a total number of 9195 triangular finite elements leading to about 93000 degrees of freedom. In particular, the mesh consisted of 3151 elements and 6044 elements within food and air domains, respectively. The considered mesh provided a satisfactory spatial resolution for the system under study. Lagrange finite elements of order two were chosen for the components of air velocity vector u, for the turbulent kinetic energy, the dissipation for unit of turbulent kinetic energy and for the pressure distribution within the drying chamber but, also for water concentration and temperature in, both, air and food. The time-dependent problem was solved by an implicit time-stepping scheme, leading to a non-linear system of equations for each time step. Newton's method was used to solve each non-linear system of equations, whereas a direct linear system solver was adopted to solve the resulting systems of linear equations. On a dual-core computer running under Linux, a typical drying time of 5 hours was typically simulated in about 50 minutes.

4. Experimental

Potato samples were air-dried by air in a convective oven. The potatoes were cut in slab shape. Each sample was dried in lab-scale convective oven (Memmert Universal Oven model UFP 400) monitoring, with respect to time, food weight, by a precision balance (Mettler AE 160, accuracy of ±10-4g) and food dimensions by a vernier caliper.

The lab-scale oven allowed monitoring, by a Dostmann electronic Precision Measuring Instrument P 655, air temperature and its humidity (by a rh 071073 probe) and air velocity, by a H 113828 probe. Air velocity was equal to 2.2 m/s and air temperature, T_a, was chosen equal to 50°C and to 70°C. Experimental data were used also to obtain some important parameters regarding food shrinkage modelling. Food weight and food dimensions were, indeed, used to evaluate food moisture content and food volume: a linear fitting procedures of the aforementioned experimental data allowed estimating the hydrous compressibility factor that were equal to 7.44 10-4 and to 8.24 10-4, in the case of a drying temperature of 50°C and 70°C respectively.

Transport Phenomena Modeling During Drying of Shrinking Materials

5. Results

The proposed model was able to provide worthwhile information about water, temperature and velocity profiles in both the considered domains. This is mainly interesting in the case of food matter where, the local value of water content and temperature is an index of the conditions that give rise to deterioration reaction due to micro-organism activity. Some of the most representative results are shown in the following Fig. 1, where the time evolution of potato moisture content (on a wet basis) is presented for a potato sample whose shape and dimension change with respect to time because of the shrinkage.
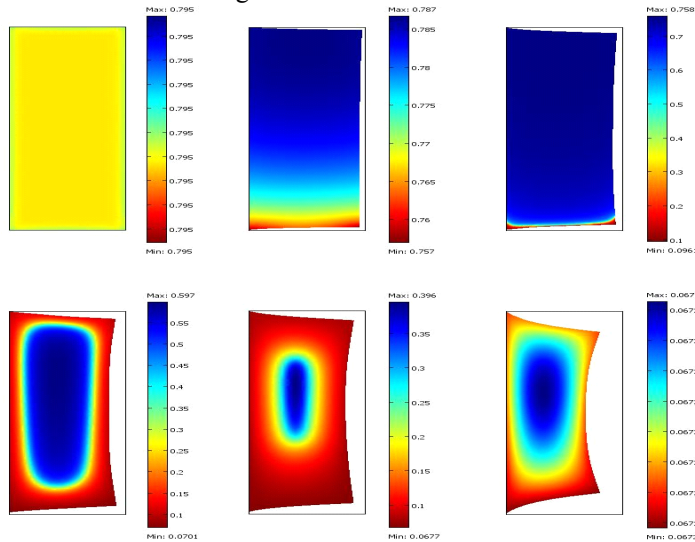
Figure 1. Time evolution of potato moisture content (on a wet basis) during drying accounting for shrinkage effect. (Air temperature of 50°C, air velocity of 2.2m/s)

Comparison between measured and calculated food dimensions in terms of total volume is shown in figure 2, showing a remarkable agreement between model predictions and experimental data.
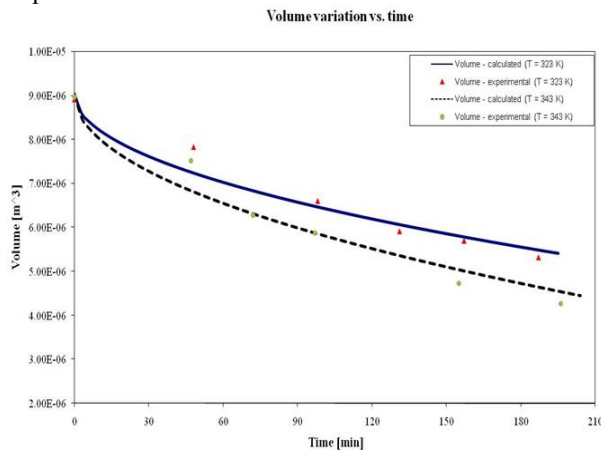
Figure 2. Comparison between experimental and model results in terms of total volume evolution of potato samples during drying.

6. Conclusions

The transport phenomena involved in food drying process have been analyzed. A general predictive model, i.e. not based on any semi-empirical correlation for estimating heat and mass fluxes at food-air interface, has been formulated.

The proposed model also predicted the spatial moisture profiles at all times, thus allowing detecting the regions within the food core, where high values of moisture content can promote microbial spoilage.

The model has been also improved to predict food shrinkage by coupling transport equations to virtual work principle, written with reference to a deformed mesh whose movement has been described by ALE method. The obtained results are promising. It is intended to improve the transport model by considering also the influence of convection inside the food. Also shrinkage description needs to be improved, for instance by formulating a different assumption relating free deformation due to moisture loss to water content variation, or by taking into account the influence of body and surface forces applied to food.

References

R.B. Bird, W.E. Stewart, E.N.Lightfoot, 1960, Transport Phenomena. John Wiley & Sons, London, UK.

S.Curcio, M. Aversa, V.Calabro', G. Iorio, 2008, Simulation of food drying: FEM analysis and experimental validation, Journal of Food Engineering, Vol 87 (4), pages 541-553.

A.K. Datta, 2007, Porous media approaches to studying simultaneous heat and mass transfer in food processes. I: Problem formulations. Journal of Food Engineering, Vol 80 (1), pages 80-95.

D. Lacasse, É. Turgeon and D.Pelletier, 2004, On the judicious use of the k-e psilon model, wall functions and adaptivity, International Journal of Thermal Sciences, Vol 43 (10), pages 925-938.

Z.B., Maroulis, G.D. Saravacos, M.K. Krokida, N.M. Panagiotou, 2002, Thermal conductivity prediction for foodstuffs: effect of moisture content and temperature, Int J Food Propert , Vol ,5 pages231–45.

B.K. May, P. Perrè, 2002, The importance of considering exchange surface area reduction to exhibit a constant drying flux period in foodstuffs, Journal of Food Engineering, Vol 54, pages 271-282.

J.M.Smith, H.C.Van Ness, M.M. Abbott, 1987, Chemical Engineering Thermodynamics (IV edition). New York, McGraw-Hill.

J.Srikiatden, J.S.Roberts, 2008, Predicting moisture profiles in potato and carrot during convective hot air drying using isothermally measured effective diffusivity, Journal of Food Engineering Vol 84, pages 516–525.

P.Verboven, N. Scheerlinck, J. De Baerdemaeker, B. M. Nicolaï, 2001, Sensitivity of the Food Centre Temperature with Respect to the Air Velocity and the Turbulence Kinetic Energy, Journal of Food Engineering Vol 48, pages 53-60.

J.Welty, C.Wicks, R.Wilson, G.Rorrer, 2001, Fundamentals of Momentum, Heat, and Mass Transfer. New York: John Wiley and Sons.

D. C.Wilcox, 1998, Turbulence Modeling for CFD, DCW Industries, Inc., La Canada, CA.

H.Yang, N. Sakai, and M.Watanabe, 2001, Drying model with non-isotropic shrinkage deformation undergoing simultaneous heat and mass transfer, Drying Technology, Vol 19 (7), pages 1441-1460.

J. Zhang, and A. K. Datta, 2004, Some considerations in modeling of moisture transport in heating of hygroscopic materials, Drying Technology, Vol 22 (8), pages 1983- 2008.

Capacity planning and financial optimization of the bioethanol supply chain under price uncertainty

Matteo Dal Mas, Sara Giarola, Andrea Zamboni and Fabrizio Bezzo[a]

[a]*CAPE-Lab − Computer-Aided Process Engineering Laboratory,*
Dipartimento di Principi e Impianti di Ingegneria Chimica "I. Sorgato",
Università di Padova, via Marzolo 9, I-35131, Padova, Italy, fabrizio.bezzo@unipd.it

Abstract

This work addresses the development of a dynamic spatially explicit MILP (Mixed Integer Linear Programming) modeling framework devised to optimize the design and planning of biomass-based fuel supply networks according to financial criteria and accounting for uncertainty on market conditions. The model capabilities in steering strategic decisions are assessed through a real-world case study related to the emerging corn-based bioethanol production system in Northern Italy. Two optimization criteria are considered, based on a risk-seeking or, alternatively, on a risk-adverse-approach.

Keywords: supply chain optimization, financial analysis, bioethanol, design under uncertainty.

1. Introduction

In the last few years, uncertain perspectives related to energy demand and supply as well as concerns about environmental health have endangered the stability of traditional hydrocarbon-driven energy systems. These issues have been pushing the EU policies toward the promotion of renewable sources as viable alternatives in energy generation. In this context, biomass has been incurring a special interest because of its flexibility and its further possibility to be converted into liquid fuels. As a consequence, the European Commission has outlined an ambitious program (Directive 2003/30/CE) to support biofuels penetration within the conventional fuel market through setting as mandatory a minimum blending quota of 5.75%$_E$, recently lifted up to 10%$_E$ by 2020. Among the other options, bioethanol has been assuming a leader position although some doubts are present on the effective economic feasibility of such a business, which is still tightly related to the geographical location and the technological background on which the system is built as well as to the market conditions volatility relating to feedstock purchase costs and products prices. Thus, it arises the need for specific decision-making tools capable to manage such uncertainties and to adopt a wider approach beyond the limited company-centric view of the business [1].

The objective of this work is to develop a Mixed Integer Linear Program (MILP) modeling framework capable of representing the dynamic evolution of a bioethanol supply chain (SC) in the presence of price uncertainty. The mathematical pattern builds on the spatially-explicit multi-echelon static model devised by Zamboni et al. [2] and adopts a stochastic formulation to handle the effect of uncertainty on biomass and bioethanol prices [3]. A sensitivity analysis is carried out to take the DDGS market price volatility into account. A case study is used to assess the proposed framework by considering Northern Italy as the geographical reference and the dry grind process as the benchmark technology for ethanol production. The financial criteria according to which the supply network is optimized are captured by two financial indicators: the

Opportunity Value (OV), suitable for risk-seeking investors, and the Value-at-Risk (VaR), oriented to the risk-adverse ones. The economics have been assessed by means of SC Analysis (SCA) techniques and the model is used first to maximize the most optimistic expected profit (i.e., the OV) and then to minimize the risk on investment (VaR) over a ten years time horizon by identifying the best network topology in terms of biomass cultivation site locations, ethanol production plant capacities and locations as well as transport logistics.

2. Supply Chain Analysis key issues

The financial assessment of a production system by means of SCA techniques requires the characterization of the supply chain components as well as of the logistics nodes. Fig. 1 depicts the general structure of the corn-based ethanol system under assessment. Two main SC substructures are identified: the fuel upstream production network (involving biomass cultivations, biomass delivery and fuel production sites) and the downstream product distribution to the demand centers.
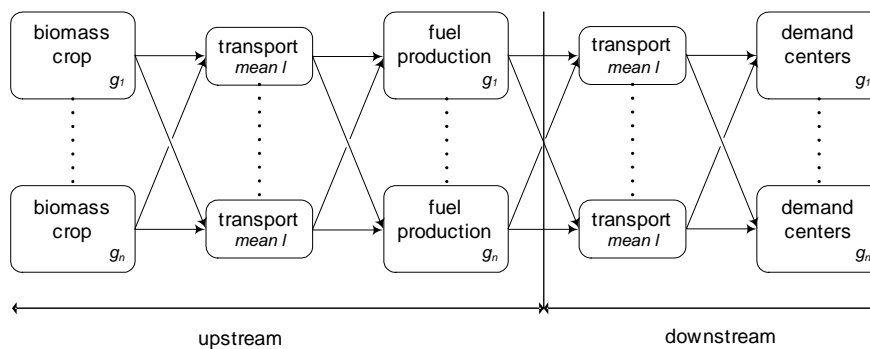
Figure 1. Bioethanol supply chain superstructure

2.1 Biomass cultivation

First generation production technology is assumed as the most convenient solution over a short-term perspective within the Italian industrial context, and corn as the most suitable biomass for ethanol production. Spatially specific parameters for corn cropping such as yield and land availability are taken from Zamboni et al. [2].

In order to consider the conflict between "biomass-for-food" and "biomass-for-fuel", a sustainability factor is introduced and set equal to 14% of the total potential corn crops [2]. It represents the maximum percentage of corn suitable for ethanol production.

2.2 Transport system

The Northern Italy infrastructure includes a full-scale range of transport options available for industrial purposes. The whole set has been considered including trucks, rail, barges and ships. Goods delivery is assumed to be an additional service provided by other actors operating within the existing infrastructure, thus transport costs have been evaluated by multiplying the freight loads by a unit transport cost [€/(t·km)]. The transport infrastructure is defined through feasibility constraints capturing the actual availability of each transport option and a tortuosity factor accounts for the different routes that each specific transport means has to go through.

Optimization of the bioethanol supply chain under price uncertainty

2.3 Ethanol production
The dry grind process has been considered in characterizing the production facilities. The economy of scale effect on ethanol production capital and operating costs has been taken into account by considering four different discrete capacity ranges. For each range specific production and capital costs have been calculated trough a purpose-designed model [4].

2.4 Demand centers
Bioethanol is assumed to be sent to blending terminals existing at given locations. Their gasoline delivery rates (satisfying the regional demand centers) are supposed to be constant all over the time horizon. Location and number of units are fixed whilst the actual gasoline delivery rates of each terminal are defined by means of a secondary distribution optimization model [2,5]. In accordance to the EU Directive, the bioethanol quota is set equal to $3\%_E$ for 2009, $5.75\%_E$ by 2010 and from 2010 to 2019 minimum increments of bioethanol percentages are provided in order to achieve 2020 EU target of $10\%_E$. Five time periods t are considered (t represents a 2 years' time interval).
The problem is conceived so as to keep the ethanol demand as a free variable at which the problem solution aims by optimizing the financial performances of the system. This means that is not compulsory for the production system to fulfill the market demand of ethanol whenever it results not to be profitable from an economic point of view. The actual market demand represents indeed an upper bound parameter defined by imposing the minimum blending quota to the gasoline market demand values.

3. Modeling framework

The mathematical formulation is based on the modeling approaches adopted in the strategic design of a multi-echelon SC encompassing features to address the location of spatially explicit facilities [6], capacity planning of strategic fuel systems [7], as well as the uncertainty on costs and products prices [3,8].
The objective is to determine the optimal system configuration in terms of supply chain profitability and financial risk on a ten years' investment. Therefore, the key variables to be optimized over the planning time horizon are: (i) geographical location of biomass production sites, (ii) biomass production for each site, (iii) biofuel production facilities location and scale, (iv) biofuel market demand satisfaction rate, (v) transport logistics, (vi) supply chain profit and (vii) financial risk under uncertainty.
The uncertainty dealing with the biomass purchase costs and ethanol market price is handled through a stochastic optimization based on a set of possible average scenarios over the time horizon.

3.1. Definition of price scenarios
Each scenario is a particular combination of ethanol market price and corn purchase cost. The analysis of historical price trends has been performed by fitting the following functions for biomass and ethanol prices respectively:

$$pdf(UPCc) = 18.4240 \cdot \Gamma[UPCc, 87.6592, 1.4964] + 10.5057 \cdot [UPCc, 57.0740, 3.0429] \qquad (Eqn.1)$$

$$pdf(MPe) = 83.8326 \cdot \Gamma[MPe, 362.6378, 2.0258] \qquad (Eqn.2)$$

where UPC_c [€/t] is the corn purchase cost and MP_e [€/t] is the ethanol market price. The probability functions are discretized into the 4 elements-vectors of scenarios S and R respectively related to corn purchase costs and ethanol market prices. The two vectors are combined into a 4 x 4 matrix (s,r) of 16 scenarios. Each scenario (s,r) is assumed to

represent an average cost or price for all the 10 years' period. The method is a simplification of a more rigorous approach recombining all the 16 realizations along the time horizon. However, we preferred to retain a high complexity in the SC description.

3.2. SCA optimization criteria

The first financial objective function adopted as optimization criterion represents the mean zero absolute OV [€] indicator; it stands for the expected profit for a quantile of (*1-p*), with *p* set to 10%. As indicated in [9], the formulation should be consistent to the VaR formulation (Eq. 5). Accordingly:

$$Obj_{OV} = -\frac{\sum_{s,r} NPV_{s,r} \cdot \pi_{s,r}}{\sum_{s,r} \pi_{s,r}} \quad \forall \, s,r \in best_{s,r} \qquad (Eqn.3)$$

where $best_{s,r} \subseteq S \cup R$ is the combination set representing the most profitable market scenarios occurring with a 90% confidence level; $NPV_{s,r}$ is the Net Present Value and $\pi_{s,r}$ is the event probability. Both $NPV_{s,r}$ and $\pi_{s,r}$ are related to each scenario deriving from the combination of corn purchase cost *s* and ethanol market price *r*.

$NPV_{s,r}$ for each scenario is defined as the discounted profits minus the discounted costs:

$$NPV_{s,r} = \sum_t \left(PBT_{t,s,r} \cdot \varepsilon Cap_t - FCC_t \cdot \varepsilon FCC_t \right) \quad \forall \, s,r \qquad (Eqn.4)$$

where $PBT_{t,s,r}$ [€/y] represents the profit before taxes, FCC_t, are the facility capital costs [€/y] for time period *t* and scenario (*s,r*). Both terms are discounted through factors, depending on the MARR interest rate and collected in two arrays εCap_t and εFCC_t. Capital costs are allocated at the beginning of each year and are yearly discounted, while revenues are received at the end of the year and discounted every two-year period. The second objective function is given by the negative value of the mean zero absolute VaR [€] as defined by [10], representing the expected loss for a certain confidence level *p*, with *p* set to 10%:

$$Obj_{VaR} = -\frac{\sum_{s,r} NPV_{s,r} \cdot \pi_{s,r}}{\sum_{s,r} \pi_{s,r}} \quad \forall \, s,r \in worst_{s,r} \qquad (Eqn.5)$$

where $worst_{s,r} \subseteq S \cup R$ is the combination set representing the most unprofitable market scenarios occurring with a 10% cumulative probability.

The NPV-related variables depend on design variables, subjected to logical constraints, such as products (ethanol, corn) demand as well as production and mass-flows between the grid cells. Additional information about the major modeling assumptions and parameters can be found in [2].

4. Case study

The proposed framework has been applied to the design and planning of a representative case study, namely the ethanol production network in Northern Italy. The stochastic optimization problem is addressed by formulating two alternative case studies corresponding to the maximization of the OV at 90% quantile (Case A) and maximization of the VaR at 10% quantile (Case B). Moreover, a sensitivity analysis on three different options for DDGS price MP_d (€/t) is assessed for each case study through three optimization instances: $MP_d = 300$ €/t (instance 1); $MP_d = 200$ €/t (instance 2); and finally considering a progressive DDGS devaluation over time according to which MP_d is supposed to vary as follows: 300 €/t from year 2009 to 2012, 200 €/t from 2013 to 2016 and 100 €/t from 2017 to 2019 (instance 3).

Optimization of the bioethanol supply chain under price uncertainty

4.1. Results and discussion

The optimization on the two financial criteria (OV and VaR) results in two different best optimum configurations. The OV optimization criterion (Case A) is scarcely sensitive to the amount of revenues coming from the side production of DDGS. In fact, there is always a reasonable chance to obtain profitable results even if DDGS market value is low: even for instance A.3, the ROI index averaged on the business period is about 36%, thus demonstrating the good financial performance of the business. A higher value for the DDGS market price (instance A.1) leads to a higher ROI value; short payback times allow for facility constructions even in 2015. The high profitability also reduces the importance of the economy of scale in planning the ethanol production capacity: the optimization results propose a rather distributed supply chain configuration in which even small production plants are profitable. The optimal supply chain configuration as planned at the end of the network lifetime for instance A.1 is shown in Fig. 2. Plant sites are located in Milan ($g = 27$), Venice harbor ($g = 32$), Genova ($g = 46$) and Porto Viro ($g = 43$). The network configuration over time is nearly the same for instances A.1, A.2 and A.3. Fig.2 also shows that biomass importation is usually preferred to domestic production and accordingly there is a convenience in locating the production plants along the coastline.
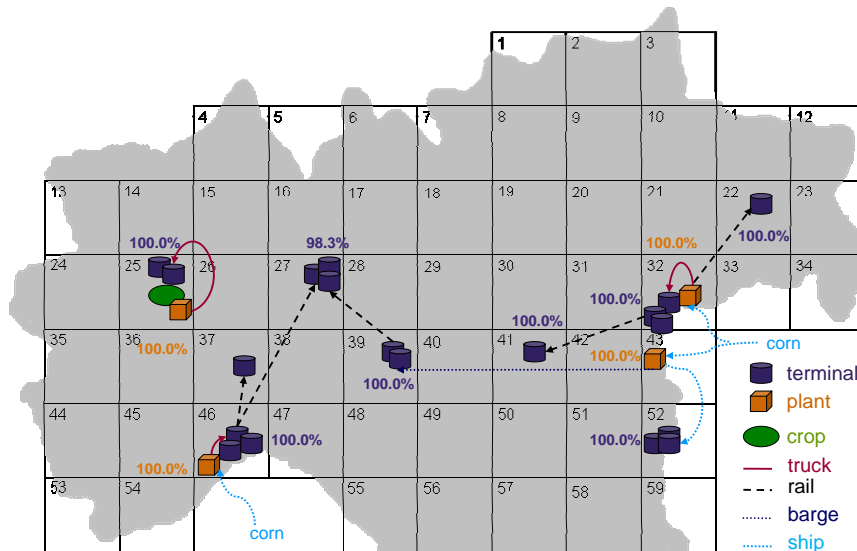
Figure 2: Optimal SC configuration for instance A.1

Truck delivery is selected only for local transport of high density products (ethanol from plants to blending terminals), whereas for long distances bulkier transport means, such as barges or train, are preferred. Finally, the optimal network solution for instance A.1 proposes the complete fulfillment of the available quota for biofuel blending.

If the VaR optimization (Case B) is taken into account, only a high DDGS market price (instance B.1) would allow a well-performing configuration; in the other instances the best solution is not to enter the business as the discounted cumulative cash position always results in a negative value at the end of the time horizon. With reference to instance B.1 and comparing it to instance A.1, a few plants of bigger size are planned according to an economy of scale rule. The average ROI index shows a return on the capital expenditures of 19.95% while the high payback times cannot justify plant establishment after 2011.

The biofuel demand fulfillment quota for all instances are collected in Table 1. As clearly shown, the investment profitability of Case A always allows for a full provision of bioethanol. On the other hand, in Instance B.1 the demand is never fulfilled since 2015 onwards. Instances B.2 and B.3 never suggest to start up the production business. These results are clearly representative of a situation in which risk prevention is considered a better solution than that of meeting the market needs.

Table 1. Percentages of bioethanol demand fulfillment: comparison among all the instances

period	Instance A.1	Instance A.2	Instance A.3	Instance B1	Instance B.2	Instance B.3
1	100%	100%	100%	74%	-	-
2	100%	100%	100%	100%	-	-
3	100%	100%	100%	90.6%	-	-
4	100%	100%	100%	81.4%	-	-
5	100%	100%	100%	74%	-	-

5. Final remarks

A dynamic, stochastic and spatially-explicit modeling framework for the design and planning of multi-echelon biofuels SC under uncertainty has been presented. The bioethanol production system of Northern Italy has been chosen as a case study in order to demonstrate the model pertinence and capabilities.

The design of the bioethanol SC by considering the best market scenarios (OV criterion) suggests that there is always a reasonable probability for investments on distributed networks to be profitable. On the contrary, if only the worst market scenarios are taken into account (VaR criterion), the results show that only a high DDGS selling price allows for a profitable SC configuration.

6. Acknowledgements

References

[1]   L.G. Papageorgiou, Comput. Chem. Eng. 33 (2009) 1931
[2]   A. Zamboni, N. Shah. F. Bezzo, Energy Fuels 23 (2009) 5121.
[3]   K.H. Tsang, N.I. Samsatli, N. Shah, Food and Bioproducts Processing 85 (2007) 120.
[4]   G. Franceschin, A. Zamboni, F. Bezzo, A. Bertucco, Chemical Engineering Research and Design 86 (2008) 488.
[5]   M.T. Kong, Downstream oil products supply chain optimization. PhD Thesis,: University of London; London, 2002.
[6]   A. Almansoori, N. Shah, Chemical Engineering Research and Design 84, (2006) 423.
[7]   A.Hugo, E.N. Pistikopoulos, Journal of Cleaner Production 13, (2005), 1471.
[8]   G. Guillén-Gonsalbez, I. E. Grossmann, AIChe J 55 (2009), 99.
[9]   K.H. Tsang, N.I. Samsatli, N. Shah, Food and Bioproducts Processing 85 (2007) 129.
[10]   G.J. Alexander, A.M. Baptista, Journal of Economic Dynamics & Control 26 (2002) 1159.

Integrating pricing policies in the strategic planning of supply chains: a case study of the sugar cane industry in Argentina

Andrey M. Kostin[a], Gonzalo Guillén-Gosálbez[a*], Fernando D. Mele[b], Miguel J. Bagajewicz[c], Laureano Jiménez[a]

[a]Universitat Rovira i Virgili, Av. Països Catalans 26, Tarragona 43007, Spain, gonzalo.guillen@urv.cat, laureano.jimenez@urv.cat
[b]Universidad Nacional de Tucumán, Av. Independencia 1800, S. M. de Tucumán T4002BLR, Argentina, fmele@herrera.unt.edu.ar
[c]School of Chemical, Biological and Materials Engineering, University of Oklahoma, Norman, Oklahoma, 73019 , USA , bagajewicz@ou.edu

Abstract

In this paper, we address the strategic planning of supply chains for ethanol and sugar production under uncertainty in the market trends. The design task is formulated as a mixed-integer linear programming (MILP) problem that decides on the capacity expansions of production and storage facilities over time along with the associated sales decisions. The main novelty of the work lies in the incorporation of a price-demand relationship explicitly at the modeling stage. Thus, prices are regarded in our work as decision variables, whereas the demand is determined based on the prices calculated by the model. To incorporate the trade-off between the expected profit and risk that naturally exists in an uncertain environmental of this type, we employ the sample average approximation (SAA) algorithm. The capabilities and possibilities of this approach are demonstrated thought case studies based on Argentinean sugar cane industry.

Keywords: Optimization; pricing policies, bioethanol, supply chain, sugar cane; financial risk

1. Introduction

The production of plant-based fuels has recently gained wider interest given their potential to reduce both the consumption of crude oil and the associated environmental impact. Specifically, bioethanol is currently among the most important biofuels in terms of its production rate, which has risen from 13,100 millions of gallons in 2007 to 17,300 millions of gallons in 2008 [1]. This rapid development has been motivated by several advantages of the bioethanol compared to other liquid fuels. Particularly, the octane number (*i.e.,* 99) is higher than the average octane number of gasoline. On the other hand, pure ethanol shows a lower heating value: 21 MJ/L for ethanol versus 30 MJ/L for gasoline. However, by blending it in low proportion with gasoline, it becomes suitable for common internal combustion engines without any significant changes in engine performance and construction [2].

Argentina has great potential to produce bioethanol. It has abundant natural resources, a very efficient agricultural production sector, and the domestic sugar industry has a strong structure, having increased efficiency and production significantly in the last years. In 2007 the Argentine Government published Law 26,093 on biofuels. It forces the use of biofuels by 2010 with an obligatory mix of 5 percent of ethanol in gasoline

and 5 percent of biodiesel in diesel. To comply with it, private analysts forecast that a volume of 250 million L/year of ethanol will be needed [3].

The circumstances open a big challenge for the sugar cane industry: the opportunity to divert part of the cane from sugar into ethanol production, providing mills more flexibility in their production mix, and increasing overall operational efficiency. This flexibility is intricately related to internal and external products prices and demand, and subsidy measures.

The aim of this paper is to provide a decision-support tool for the optimal strategic planning of supply chains (SCs) for ethanol production. The main novelty of our work is the development of a mathematical formulation that integrates all the components of the ethanol SC into a single framework in order to optimize strategic SC decisions. The model presented in this paper allows to determine in a systematic way the capacity expansions of the production and storage facilities of the network over time as well as the associated planning decisions.

2. Problem statement

The proposed model has to determine the structure of the three-echelon SC (production-storage-market). This network includes a set of plants and a set of storage facilities, where products are stored before being delivered to the final customers. The facilities can be installed in a set of regions or grids that correspond to the potential locations in which the overall regions of interest is divided. The problem addressed in this paper can be formally stated as follows:

Given are a fixed time horizon (T), a set of possible product prices (PR_{igkte}) and corresponding demands (D_{ikte}), cost parameters for production (UPC_{ipgt}), storage (USC_{isgt}) and transportation of materials (FP_{lt}, FE_l, DW_{lt}, GE_{lt}, $TCap_l$, $EL_{gg'}$, SP_l), taxes rate (φ), minimum and maximum capacity of plants (τ, $PCap^{min}_p$, $PCap^{max}_p$, $PCapE^{min}_p$, $PCapE^{max}_p$), storages ($SCap^{min}_s$, $SCap^{max}_s$, $SCapE^{min}_s$, $SCapE^{max}_s$) and transportation links (Q^{max}_l, Q^{min}_l), capital cost data (α^{PL}_{pgt}, β^{PL}_{pgt}, α^S_{sgt}, β^S_{sgt}), interest rate for NPV calculations (ir), landfill tax (LT_{ig}) and upper limit for capital investment (UCI).

The goal is to determine the configuration of the bioethanol network and the associated planning decisions with the goal of maximizing the expected net present value ($E[NPV]$). The model determines the number, location, capacity of production plants and warehouses to be set up in each grid, their capacity expansion policy, the transportation links and transportation modes that need to be established in the network, the production rates and flows of feed stocks, wastes and final products, as well as the prices of final products and demands over the planning horizon.

3. Pricing model

In microeconomics, one way to obtain the demand function is to solve a consumer utility-maximization problem limited by a consumer budget (Y) and a maximum possible demand D. Taking into account the two products with demands d_1 (for our product) and d_2 (for competitor's product), we maximize the utility function $U(d_1,d_2)$ subject to a budget limitation and a total demand limitation:

$$\begin{aligned}
\max \quad & U(d_1,d_2) \\
s.t \quad & \\
& p_1 d_1 + p_2 d_2 \leq Y \leq D
\end{aligned} \tag{1}$$

where p_1 is our price, and p_2 is a competitor's price. In this work, we use a constant elasticity of substitution (*CES*) utility function:

Integrating pricing policies in the strategic planning of supply chains: a case study of the sugar cane industry in Argentina

$$U(d_1, d_2) = (x_1^\rho + x_2^\rho)^{1/\rho} \tag{2}$$

In Eq. (2), x is a satisfaction function, which has the following forms:

$$x_1 = \alpha d_1, \qquad x_2 = \beta d_2 \tag{3}$$

where β is a measure of how much a consumer prefers our product to product 2, and α is a measure of how much the consumer population is aware of the quality of our product. When the consumer budget is binding, the solution of the problem 1 [4, 5] is:

$$p_1 d_1^{1-\rho} = (\frac{\alpha}{\beta})^\rho p_2 \left[\frac{Y - p_1 d_1}{p_2} \right] \tag{4}$$

Several parameters of the previous equation, namely α, β and ρ, were determined based on historical data on sugar and ethanol sales. The competitor's price is assumed to be constant and equal to the current market price. Regarding our price, the model is free to choose from a set of k possible prices (PR_{igkte}). This condition is valid only for sugar, due to the fact, that ethanol price in Argentina is regulated by government and in a medium-range planning problem can be considered as fixed. The uncertainty is introduced in the consumer budget (Y_e). Thereby, the solution of Eq. (4) provides product demand (d_{lietk}) for each randomly generated consumer budget and corresponding price PR_{igkte}. Due to the nonlinear nature of the price-demand relationship, Eq. (4) was solved separately in Matlab®, and the obtained demands for the discrete sizes were then regarded as parameters in the optimization model.

4. Stochastic formulation

As previously mentioned, the uncertainty is introduced through the demand and sugar prices. Two models were developed: a "general" planning problem without price-demand relationship and an "advanced" planning problem with the incorporation of the price-demand relationship. For both formulations, the stochastic results are obtained using a sample average approximation (SAA) algorithm [6]. Following this approach, a full deterministic model is solved for each scenario, and then the design variables (the number of production and storage facilities, their expansion policies within planning horizon and the number of transportation units) are fixed and the model is run again for the remaining scenarios. After that, non-dominated risk curves are selected and analyzed.

5. Mathematical model

The MILP formulation is based on that introduced by Almansoori and Shah [7] and Guillén-Gosálbez *et al.* [8] for the case of hydrogen SCs and Guillén-Gosálbez and Grossmann for the case of petrochemical SCs [9]. It includes integer variables representing the number of plants (NP_{pgt}) and storage facilities (NS_{sgt}) established in each potential location and time period, as well as transportation units (NT_{lt}) circulating between the SC entities. On the other hand, binary variables are used to represent the establishment of transportation links between grids in each time period ($X_{lgg'te}$) and to choose one optimal price from the set of possible prices (YP_{ikte}), whereas continuous ones denote the transportation flows, capacity expansions, storage inventories and production rates.

The model includes three main blocks of equations: mass balances, capacity constraints and objective function. A brief outline of each of these sets of equations is next given.

5.1. Mass balance equations

Overall mass balance for each grid zone in scenario *e* is represented by Eq.(5). In accordance with it, for every material form *i*, the initial inventory kept in grid *g* ($ST_{isgt-1e}$) plus the amount produced *(PT_{igte})* the amount of raw materials purchased *(PU_{igte})* and the input flow rate *($Q_{ilg'gte}$)* must equal the final inventory *(ST_{isgte})* plus the amount delivered to customers *(DTS_{igte})* plus the output flow *($Q_{ilgg'te}$)* and the amount of waste generated *(W_{igte})*:

$$\sum_{s \in SI(i,s)} ST_{isgt-1e} + PT_{igte} + PU_{igte} + \sum_l \sum_{g' \neq g} Q_{ilg'gte} =$$
$$\sum_{s \in SI(i,s)} ST_{isgte} + DTS_{igte} + \sum_l \sum_{g \neq g'} Q_{ilgg'te} + W_{igte} \qquad \forall i,g,t,e \tag{5}$$

5.2. Capacity constraints

Production and storage capacity expansions are bounded between upper and lower limits:

$$NP_{pgt} PCap_p^{\min} \leq PCapE_{pgt} \leq NP_{pgt} PCap_p^{\max} \qquad \forall p,g,t \tag{6}$$

$$NS_{sgt} SCap_s^{\min} \leq SCapE_{sgt} \leq NS_{sgt} SCap_s^{\max} \qquad \forall s,g,t \tag{7}$$

The material flow in scenario *e* is constrained by the minimum and maximum allowable transportation capacity:

$$Q_l^{\min} X_{lg'gte} \leq \sum_{i \in IL(i,l)} Q_{ilg'gte} \leq Q_l^{\max} X_{lg'gte} \qquad \forall l,g,g',t,e \tag{8}$$

The production rate of sugar cane *($PT_{1,gte}$)* in each scenario *e* is limited by the available capacities of sugar cane plantations situated in grid *g* in time *t* *($CapCrop_g$)*:

$$PT_{igte} \leq CapCrop_g \qquad \forall i = 1,g,t,e \tag{9}$$

5.3. Objective function

At the end of the time horizon, a different value of NPV_e is obtained for each realization of demand uncertainty for the selected price. The proposed model must account to the maximization of the expected value (*E[NPV]*) of the resulting NPV_e distribution:

$$E[NPV] = \sum_e prob_e NPV_e \tag{10}$$

The NPV_e can be determined from the discounted cash flows *(CF_{te})* generated in each of the time intervals *t* in which the total time horizon is divided:

$$NPV_e = \sum_t \frac{CF_{te}}{(1+ir)^{t-1}} \qquad \forall e \tag{11}$$

Bagajewicz [10] points out that NPV is not necessarily a good measure of profit in conditions where the capital investment is not fixed. Therefore, it is convenient to

Integrating pricing policies in the strategic planning of supply chains: a case study of the sugar cane industry in Argentina

consider the maximization of NPV subject to different maximum available capital and then check the return of investment. For brevity, we omit this step.

6. Results and discussion

All the models were written in GAMS and solved with the MILP solver CPLEX 11.0 on a HP Compaq DC5850 PC with AMD Phenom 8600B 2.29 GHz triple-core processor and 2.75 Gb of RAM.

The model contains 10 grids representing the provinces located in the north of Argentina. Only 5 of them possess sugar cane plantations, while the remaining provinces can import sugar cane, which eventually leads to an increase in the transportation operating cost. Domestic Argentinean sugar cane industry operates with 5 production technologies, 2 different storage technologies and 3 types of transportation trucks. The length of the time horizon is 4 years. The uncertainty is represented by a set of 30 scenarios that were generated from Monte Carlo sampling. The set of allowable prices contains 8 discrete values. The risk curves obtained by the "general model" are depicted in Fig. 1. The thick dashed line is the solution with Maximum E[NPV]. This curve indicates a 30% risk of loosing money. An inspection of the lower portion of the curve indicates the existence on non-dominated solutions that can be used to reduce risk, but at a cost of reducing the upside potential to that of the maximum E[NPV] or even lower.

The resulted risk curves of the "advanced" model are shown in Fig. 2. The results indicate, that this model provides the solution with higher E[NPV] and less risk than the "general" model. Particularly, the solution with Maximum E[NPV] shows a 20% risk of loosing money. Having a possibility to determine a price, the "advanced" model chooses the higher values of sugar prices than the "general" model. That leads to shrinkage in the sugar demand and construction of the SC with the less production capacity and risk of loosing money.
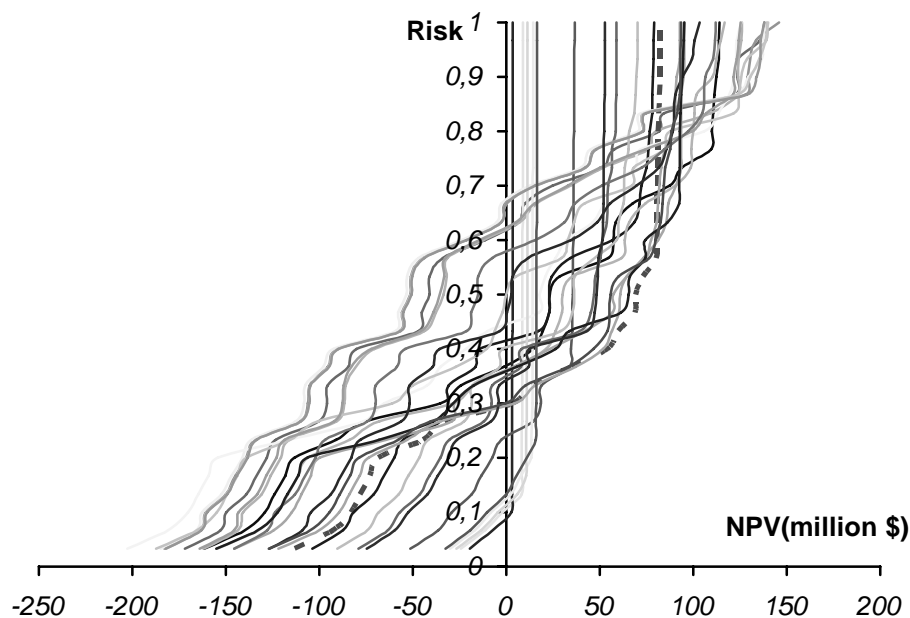

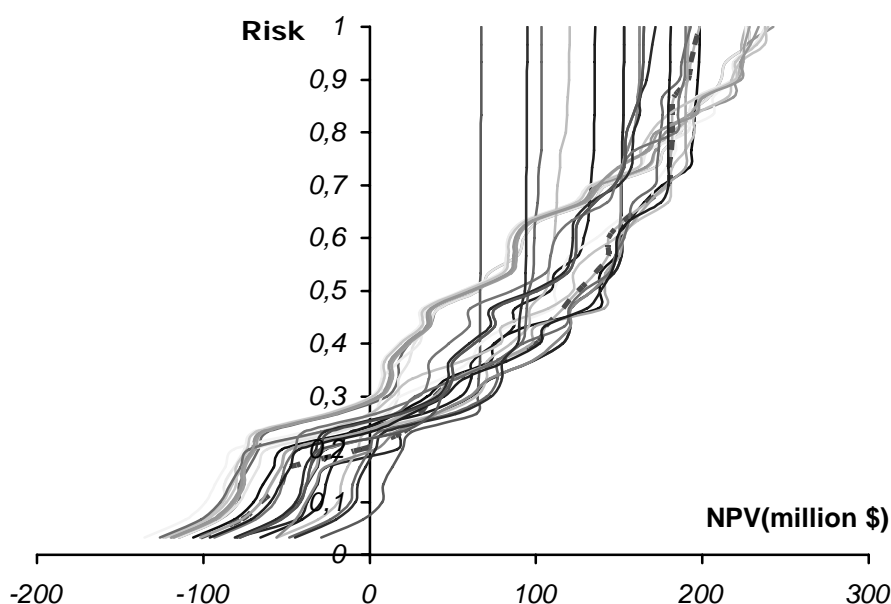Figure 1. The risk curves of the "general" planning model.

Figure 2. The risk curves of the "advanced" planning model.

7. Conclusions

In this paper, an optimization model for strategic planning of SC for ethanol production was described. The price-demand relationship was integrated into proposed model to determine the optimal price and corresponding demand for final products. The SAA algorithm was employed to consider the uncertainty in consumer budget and generate the risk curves. After that, the results generated by the model with pricing decision and the general model with fixed prices were compared. The results shown, that the "advanced" model provides the solution with higher E[NPV] and less risk of loosing money.

8. Acknowledgements

References

[1]   Ethanol industry statistics, Renewable Fuels Association, 2007.
[2]   A. Pandey, Handbook of plant-based biofuels, CRC Press, Boca Raton, 2008.
[3]   K. Joseph, R. Hoff, GAIN Report, Argentina, 2007.
[4]   H.Lakkhanawat, M. Bagajewicz, Ind. & Eng. Chem. Res., 47(2008), 6622.
[5]   M.Bagajewicz, AIChE Journal, 53(2007), 3155.
[6]   A. Aseeri, M. Bagajewicz, Comp. & Chem. Eng., 28(2004), 2791.
[7]   A. Almansoori, N. Shah, Chem. Eng. Res. & Des., 84(2006), 423.
[8]   G. Guillen-Gosalbez, F. Mele, I. E. Grossmann, AIChE Journal, 2009 .
[9]   G. Guillen-Gosalbez, I. E. Grossmann, AIChE Journal, 55(2009), 99.
[10] M.Bagajewicz, Ind. & Eng. Chem. Res., 47(2008), 9413.

Green Supply Chain Design and Operation by Integrating LCA and Dynamic Simulation

Ei Sandi Nwe[a], Arief Adhitya[b], Iskandar Halim[b], Rajagopalan Srinivasan[a,b,*]

[a]*Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576, Singapore*
[b]*Institute of Chemical and Engineering Sciences, A*STAR (Agency for Science, Technology and Research), 1 Pesek Road, Jurong Island, Singapore 627833, Singapore*
Corresponding author: chergs@nus.edu.sg

Abstract

With sustainability increasingly becoming an important business factor, companies are now looking for methods and tools to help assess the fuller picture of the environmental impacts associated with their manufacturing and supply chain activities. Life cycle assessment (LCA) is a widely-used technique for measuring the environmental costs assignable to a product or service. However, LCA takes a high-level view and assumes a fixed supply chain structure and operation. It does not explicitly consider the effect of supply chain design and practices which can be a significant contributor to the overall environmental impacts. This paper presents an approach integrating LCA indicators and dynamic simulation for green supply chain design and operation. Environmental impact indicators are incorporated into a dynamic model of the supply chain along with profit and customer satisfaction, so that sustainability of various design and operational decisions can be assessed comprehensively. The application and benefits of the proposed approach are demonstrated using two case studies.

Keywords: sustainability, supply chain, LCA, lubricants, simulation

1. Introduction

A supply chain (SC) is the network of suppliers, manufacturing plants, warehouses and distribution channels organized to acquire raw materials, convert them to finished products and distribute these products to customers. With sustainability footprint increasingly becoming a dominant issue, companies are now beginning to re-examine their manufacturing processes and the configuration of their supply chains to become more sustainable not just in terms of their economic viability but also their environmental and social impacts. In recent years, much attention has been given to the concept of green SC to help companies improve their environmental performance along the value chain. The basic idea involves integrating into the existing SC structure processes, metrics, and best practices that not only improve the environmental performance but also add value to the business. While much literature exists on various facets of green SC, less attention is given to the modeling of SC network related aspects (Srivasta, 2007). In this paper, we propose a framework comprising dynamic simulation and life cycle assessment (LCA) for integrated economic and environmentally-sound design and operation of a SC.

LCA is a tool to evaluate the environmental burdens over the entire life-cycle of the product, starting from raw material extraction to production process, distribution, point of use, and final disposal (Sonnemann et al., 2004). Throughout all these stages, extraction and consumption of resources and releases into air, water and soil are

identified and quantified. LCA is thus a useful tool for comparing the environmental impacts of products with the same functionality (for example, paper cup vs. plastic cup). While LCA can be used for measuring the environmental costs associated with bringing a product to market, its approach is limited to taking a high-level view of the entire life cycle of a product from raw material extraction, processing through final disposal. The detailed issues related to environmental impacts along the manufacturer-distributor-customer chain are generally overlooked. For instance, LCA usually does not consider the impacts of various logistics options from manufacturing plant to customer. Further, it does not consider the different SC policies although they can be a significant contributor to the overall environmental impacts. In this paper, we examine such issues through integration of SC dynamic model and LCA. We use a case study involving manufacturing of metal-working lubricants.

2. Dynamic Model of Metal-Working Lubricant Supply Chain

The SC involves suppliers, the lubricant manufacturer company, and customers. The company has a centralized sales department and a number of production plants in different locations. A customer places an order with the sales department, which then assigns the order to one of the plants. Each plant has different departments performing the different SC functions: scheduling, procurement, storage, operations, and packaging. The SC dynamics are complex due to the multi-tiered structure and numerous interactions among the entities. The impact of an entity's action on the overall SC performance, both economic and environmental, may not be immediately obvious. An integrated analysis of the impacts of decisions on the overall system is required. These motivate the use of SC simulation models, which could capture the behavior of the entities, their interactions, the resulting dynamics, and the various uncertainties, and evaluate the overall performance of the SC. Dynamic simulation offers a good way to uncover direct and indirect effects. In this paper, we adapt the dynamic simulation model of Yang et al. (2009) for integration with LCA.

The SC operates in pull-mode, driven by customer demand. Customer orders for each product type are randomly generated by sampling from a pre-specified demand curve and order frequency. Sales receives an order and decides which plant to assign the order following an *order assignment policy*. For example, if assignment is based on earliest completion date, sales asks for projected completion date for that order from the scheduler in each plant and assigns the order to the plant which can deliver at the earliest.

Scheduler of the selected plant then inserts a new job to produce the assigned order into its schedule, based on a *job scheduling policy*, e.g. first-come-first-serve, shortest processing time, earliest processing due date, etc. Production is carried out following the job schedule, which consists of a sequence of jobs to be processed. The plant operates in batch mode, i.e. production is done batch-wise in non-continuous processing units. The series of processing units can be considered as one big reactor. Raw materials are fed into this reactor and after a certain processing time, products are collected. A unique product recipe defines the type and amount of raw materials needed to make the product. The required raw materials are transferred from storage to the reactor. When processing is done, the product is collected and sent to packaging. Scheduler removes the completed job from the schedule and continues to the next job. After packaging, the packaged product is immediately delivered to the customer. No product inventory is kept. The delivery time is calculated based on the distance between the plant and the

customer. Location of a plant or a customer is represented through a pair of coordinates *(x, y)*.

Raw material inventory management within each plant involves the storage department, the procurement department, and the suppliers. Different *procurement policies* for replenishment of raw materials can be adopted. For example, under the reorder point policy, raw material will be purchased when its inventory falls below a reorder point to bring it back to the top-up level.

Economic performance is measured through indicators such as profit and customer satisfaction. While each plant has its own local performance measures, we are also concerned with the profit and customer satisfaction of the overall enterprise. Profit is calculated as revenue from product sales minus costs:

$$Profit = Revenue - \begin{bmatrix} Raw\ Material\ Cost\ + Variable\ Operating\ Cost \\ +\ Fixed\ Operating\ Cost\ + Packaging\ Cost \\ +\ Transportation\ Cost\ + Inventory\ Cost \\ +\ Late\ Penalty\ Cost \end{bmatrix} \tag{1}$$

Customer satisfaction (CS) is measured as the percentage of non-late out of the total number of customer orders:

$$CS = (1 - \frac{number\ of\ late\ orders}{number\ of\ orders}) \times 100\% \tag{2}$$

The model uses discrete time representation, where one day is divided into 100 time ticks. It has been implemented in MATLAB/Simulink (MathWorks, 2007).
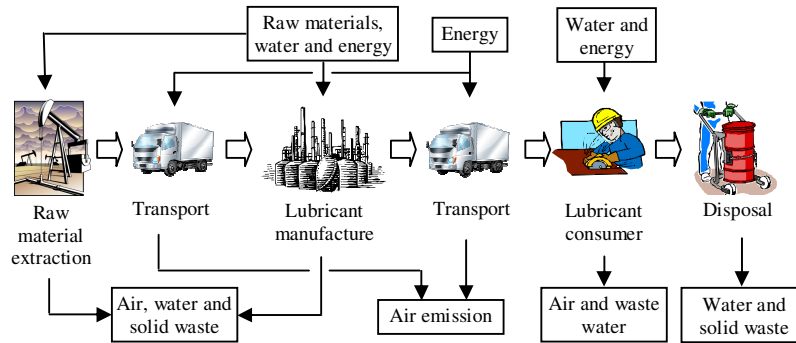
Figure 1. LCA of a metal-working lubricant supply chain

3. Incorporating Environmental Impacts into the Supply Chain Model

LCA is one suitable tool for analyzing the environmental burdens associated with the design and operation of a SC (Bojarski et al., 2009). By using the 'cradle to grave' approach of LCA, each entity of the supply chain including suppliers, manufacturers, distributors/retailers and customers can be similarly identified with products, processes and/or transport activities which require material and energy and emit pollutants. This is illustrated using Figure 1 which shows the environmental impacts of a metal-working lubricant SC involving suppliers, the manufacturer and customers. Traditionally, the SC performance is measured only in terms of economic indicators such as profit and customer satisfaction. Here, we include environmental impact indicators on the basis of

the amount of products produced. Eight indicators are used, taking into consideration raw material acquisition, processing, packaging, and transportation from the plants to customers. These are acidification, global warming potential over 100 years (GWP 100), solid waste, water use, land use, ecotoxicity, non-renewable energy consumption, and transportation. These indicators are incorporated into the SC simulation model.

This work focuses on two types of metal-working lubricants, one mineral-based (naphthenic oil) and the other bio-based (rapeseed oil). The lubricants are composed of the base oil, anionic surfactant and non-anionic surfactant. They make up about two percent of the metal-working fluid while the balance would be a carrier which can either be water, gaseous air or carbon dioxide. Since environmental concerns usually are far more significant for the lubricants rather than the carrier, we exclude the carriers in our subsequent analysis.

Table 1 lists the indicator values per kg product for the two lubricants. The production indicators cover the impact from the raw material extraction stage up to the processing stage, for the production of 1 kg of lubricant (Clarens et al., 2008). We assume that the packaging used is 200L plastic drums made of high density polyethylene. The whole life cycle of the drum is considered during the calculation of packaging impact to the environment. The indicators for packaging (per kg of products packaged) in Table 1 are calculated based on Manuilova (2003). Transportation of packaged products is assumed to be done by 28-tonne trucks and the corresponding indicator is taken from Goedkoop et al. (2000). The indicators in Table 1 have been incorporated into the SC simulation model. The total impact for each indicator is the sum of the production and packaging impacts, calculated based on the mass of products produced. This total impact is then normalized based on the total mass of product produced of both lubricant types.

4. Case Studies

In this section, two case studies are presented, one involving a policy decision and the other a design decision. The lubricant company has three plants located in Singapore, Houston, and Japan, with equal storage and production capacity. The simulation horizon is 120 days. The scheduling policy used is earliest processing due date and the procurement policy is reorder point.

Table 1. Environmental impact indicator values

Indicators	Rapeseed-based		Naphthenic-based	
	Production	Packaging	Production	Packaging
Acidification (g SO_2/kg product)	6.55	0.429	14.3	0.494
GWP 100 (kg CO_2 eq/kg product)	0.855	0.2	5.9	0.23
Solid Waste (g/kg product)	23.93	4.207	29.3	4.838
Water Use (kg/kg product)	10.96	0.968	17	1.114
Land Use (m^2/kg product)	3.89	0	0	0
Ecotoxicity (cg Pb eq/kg product)	14.2	0	0.3	0
Non-renewable Energy (MJ/kg product)	26.39	1.371	33.4	1.576
Transportation (millipoints/m^3km)	8			

4.1. Case Study 1: Order assignment policy
This case study focuses on the assignment policy of the sales department and compares the earliest completion date policy with another policy based on customer location. The

earliest completion date policy assigns an order to the plant which can deliver the product to the customer at the earliest. On the other hand, under the *customer location* policy, sales assigns the order to the plant nearest to the customer. This should minimize transportation cost and its associated environmental impact. Simulation is used to compare these two assignment policies. The results are shown in Table 2. The completion date policy leads to higher profit (29.25 M$ vs. 28.08 M$) due to more jobs completed since orders are assigned to the least busy plant, which can deliver at the earliest. This also means higher customer satisfaction (100% vs. 93.55%). The customer location policy results in a 17% reduction in transportation impact at the expense of profit and customer satisfaction. Other environmental impacts are comparable. Simulation brings out the trade-off between environmental and economic performance resulting from a policy decision. Further study can be done to find the preferred balance, for example by using a combination of the two policies.

4.2. Case Study 2: Responding to change in demand pattern

This case study considers a design decision in response to an anticipated change in demand patterns. As customers become more environmentally conscious, they would increasingly favor the more environmentally-friendly rapeseed-based lubricant. Currently, customer demand is split equally between the rapeseed-based and the naphthenic-based lubricants. The company forecasts that 25% of customer demand for naphthenic-based lubricant will be switched to rapeseed-based. In this case study, each plant is specialized – Singapore plant produces only rapeseed-based lubricant, while both Houston and Japan plant produce only naphthenic-based lubricant. In response to the expected demand pattern change, it is proposed to upgrade the Houston plant to produce rapeseed-based lubricant instead of naphthenic-based. Simulation is performed to study the economic and environmental impacts of the proposal. The results are shown in Table 3.

The environmental benefit due to the change in customer preference towards rapeseed-based lubricant is clearly reflected by the decrease in all environmental impact indicators under the forecast demand pattern, except for land use and ecotoxicity. This is because rapeseed has to be cultivated and therefore uses more land and fertilizers, which contribute to more environmental discharges. Upgrading results in higher profit under the current demand scenario (25.1 M$ vs. 22.93 M$) and more significantly under the forecast demand scenario (28.75 M$ vs. 20.44 M$), as rapeseed-based lubricant has a higher price and profit margin than naphthenic-based. With the planned upgrade, customer satisfaction would decrease temporarily under the current demand profile (77.36% to 64.15%), but would increase to 100% for the forecast demand scenario. The only drawback of upgrading is the drop in customer satisfaction while the demand is changing towards the new pattern. Profit and environmental impacts do not suffer. Under the new pattern, upgrading leads to higher profit, customer satisfaction, and less environmental impacts. Hence, it is recommended to proceed with upgrading the Houston plant.

5. Concluding Remarks

Sustainability is a critical issue for today's SCs. In this paper, we propose an approach integrating dynamic simulation and LCA which enables the evaluation of both economic and environmental impacts of the SC. Two case studies on metal-working lubricant SC demonstrate how the proposed approach provides integrated decision support for green SC design (e.g. upgrading a plant) and operation (e.g. order assignment policy) considering both economic and environmental performance. Current

work includes breaking down the economic and environmental impacts to each stage of the SC to allow for a more detailed environmental analysis.

Table 2. Simulation results for Case Study 1: Order assignment policy

Indicators	Earliest Completion	Customer Location
Profit (M$)	29.25	28.08
Customer Satisfaction (%)	100	93.55
Number of jobs completed	97	93
Acidification (g SO_2/kg product)	11.12	10.79
GWP 100 (kg CO_2 eq/kg product)	3.75	3.53
Solid Waste (g/kg product)	31.32	31.07
Water Use (kg/kg product)	15.21	14.95
Land Use (m^2/kg product)	1.83	1.99
Ecotoxicity (cg Pb eq/kg product)	6.83	7.42
Non-renewable Energy (MJ/kg product)	31.59	31.28
Transportation (millipoints/kg product)	16.95	14.12

Table 3. Simulation results for Case Study 2: Responding to change in demand pattern

Indicators	Without Upgrade		With Upgrade	
Rapeseed-based Plant	Singapore		Singapore, Houston	
Naphthenic-based Plant	Houston, Japan		Japan	
Demand ratio Rapeseed-based : Naphthenic-based	1:1	1.25:0.75	1:1	1.25:0.75
Profit (M$)	22.93	20.44	25.1	28.75
Customer Satisfaction (%)	77.36	76.84	64.15	100
Acidification (g SO_2/kg product)	11.39	10.77	10.26	9.83
GWP 100 (kg CO_2 eq/kg product)	3.92	3.52	3.19	2.91
Solid Waste (g/kg product)	31.52	31.05	30.66	30.33
Water Use (kg/kg product)	15.42	14.93	14.53	14.19
Land Use (m^2/kg product)	1.70	2.00	2.26	2.47
Ecotoxicity (cg Pb eq/kg product)	6.36	7.46	8.36	9.13
Non-renewable Energy (MJ/kg product)	31.83	31.26	30.79	30.39
Transportation (millipoints/kg product)	24.99	23.12	27.23	24.91

References

A.D. Bojarski, J.M. Laínez, A. Espuña, L. Puigjaner, 2009, Incorporating environmental impacts and regulations in a holistic supply chains modeling: An LCA approach, Computers and Chemical Engineering, 33(10), 1747-1759.

A. F. Clarens, J. B. Zimmerman, G. A. Keoleian, K. F. Hayes, S. J. Skerlos, 2008, Comparison of life cycle emissions and energy consumption for environmentally adapted metalworking fluid systems, Environmental Science and Technology, 42, 8534-8540.

M. Goedkoop, S. Effing, M. Collingnon, 2000, Eco-indicator 99: Manual for designers, <http://www.pre.nl/download/EI99_Manual.pdf>.

A. Manuilova, 2003, Life cycle assessment of industrial packaging for chemicals, <http://www.dantes.info/Publications/Publication-doc/Packaging-public.pdf>.

MathWorks, 2007, Simulink User's Guide.

G. Sonnemann, F. Castells, M. Schuhmacher, 2004, Life-cycle assessment, Integrated life-cycle and risk assessment for industrial processes, Boca Raton: CRC Press LLC, 37-73.

S. K. Srivasta, 2007, Green supply-chain management: A state-of-the-art literature review, International Journal of Management Reviews, 9(1), 53-80.

J. Yang, A. Adhitya, R. Srinivasan, 2009, Dynamic modeling of a multi-site specialty chemical manufacturing supply chain, International Conference on Infrastructure Systems and Services (NGInfra09).

Energy and Economic Assessment of Soda and Organosolv Biorefinery Processes

Araceli García, María González, Rodrigo Llano-Ponte, Jalel Labidi*

Chemical and Environmental Engineering Department, University of the Basque Country, Plaza Europa, 1, 20018, Donostia-San Sebastián, Spain
E-mail: jalel.labidi@ehu.es

Abstract

The paper industry is facing the necessity of improving process efficiency, especially in terms of energy and raw materials consumption. In most of the cases, the production of pulp and paper generates an energy surplus that is often not conveniently used. The actual trend is to convert traditional pulp and paper processes in biorefineries where all components of lignocellulosic materials can be converted in added value products such as bioethanol, lignin, cellulose whiskers and other chemicals than can replace the common petroleum based products. In the present work, the energetic and economical efficiencies of two biorefinery processes were studied using the simulation software Aspen Plus. Soda and organosolv processes were used for raw material delignification. The simulation design and treatment sequences (reaction, solid fraction washing, products recovery and liquid fraction processing) are similar for both processes. Mass and energy balances were established and the processes were compared in terms of yield, solvents/reactants recovery and energy consumption based on the same amount of processed raw material. In this way, the development of rigorous simulations allowed to determine the economical feasibility of both biorefinery models, and to establish the optimal operation conditions for both processes.

Keywords: biorefinery, organosolv, soda process, energy, economic analysis

1. Introduction

The existing pulp and paper production processes require large consumption of raw materials, energy and water and huge amounts of effluents are generated (Nurmesniemi et al., 2007). The analysis and improvement of this technology is being widely studied applying process simulation, for designing and optimizing the different parts involved in the process in order to improve energetic and economical efficiencies (Bujak, 2008). Several methods have been established to simulate pulp and paper processes. Some authors have developed mathematical models and correlations that represent the process (Bhargava et al., 2008) and they have subsequently applied them to the simulation (Mercangöz and Doyle, 2008). Commercial software (Cadsim Plus, Aspen Plus, Wimgems…), which incorporates mathematic models of different usual equipments, has been also used to simulate and optimize pulp and paper processes (González et al., 2009). Biorefinery technology, especially the 'Lignocellulose Feedstock Biorefinery', is becoming an actual alternative to petroleum based industry to produce energy, chemicals and products. The conversion of 'nature-dry' raw material (wood, straw, forest and agricultural lignocellulosic residues) into goods is getting more and more important due to the abundance and variety of available raw materials, its renewable

nature and the good position of the conversion products on the market (Kamm and Kamm, 2004). Lignocellulose, the most abundant renewable biomass produced from photosynthesis, is a complicated natural composite with three main biopolymers (cellulose, hemicellulose and lignin), whose structures and compositions vary greatly, depending on plant species, growth conditions and the method used to treat biomass (Ding and Himmel, 2006). There are several technologies to fractionate lignocellulosic feedstock components: enzymatic fractionating, hot water or acid chemical hydrolysis, steam or ammonia fiber explosion, alkaline treatment and organosolv processes (Mosier et al., 2005). In the present work two different technologies have been studied to fractionate lignocellulosic material: alkaline treatment (soda) and organosolv-ethanol processes (based on using organic alcohols and acids as reagents). Experimental data have been used to develop process simulations in order to establish mass and energy balances in a large-scale operation. The results obtained by process simulation have been used to establish the energetic and economic efficiency of both processes.

2. Biorefinery Processes Description

In the present work, soda and organosolv-ethanol processes have been applied to fractionate lignocellulosic material. Both treatments present different process conditions and yields, and, therefore, they require different equipment and unit operations. Nevertheless, the treatment sequence in both studied biorefinery processes present is similar: raw material fractionation (reaction), cellulosic solid fraction SF treatment (washing) and liquid fraction LF processing (by-products obtaining, recovery of chemicals and energy). A brief description of these stages for each process is given in the following paragraphs.

2.1. Soda Process

The soda process is one of the most commonly applied in the non-wood pulp and paper industry (Iglesias et al., 1996), since it reaches high yields at low temperature, good pulp and paper properties and also because of the obtaining of sulfur-free products. In figure 1, a general diagram of the proposed soda process is shown. The raw material (biomass) is first treated with a soda solution (NaOH concentration: 15% w/w) in an atmospheric reactor (T: 90ºC, time: 90 min, liquid/solid ratio: 6/1 w/w) obtaining two streams: a wet solid fraction, SF, mainly constituted by cellulose, and a liquid fraction, LF, containing dissolved hemicelluloses, lignin and part of the non reacted inorganic chemicals. The SF is sent to the washing stage, where water is used for cleaning fibers and removing most of the dissolved components in the SF. The resulting washing liquor WL is then mixed with the LF from the reactor and this stream is treated for energy and chemicals recovery. The first step of this stage consists of the concentration of LF in a multiple effect evaporator system, where part of the contained water is removed by using low pressure steam. The concentrated LF is then burned in the recovery boiler, where the heat released by the organic matter combustion (hemicelluloses and lignin) is used for high pressure steam generation. This steam, after temperature conditioning, is then used for heating requirements in the plant. LF combustion generates a smelt, mainly composed by Na_2CO_3. For chemicals regeneration and reuse, a causticizing step is applied, where the majority of the NaOH is regenerated, recovered and sent back to the biomass fractionation stage. Thus, in this scheme, no by-products are obtained and lignin and hemicelluloses are burned to recover energy and reactants.
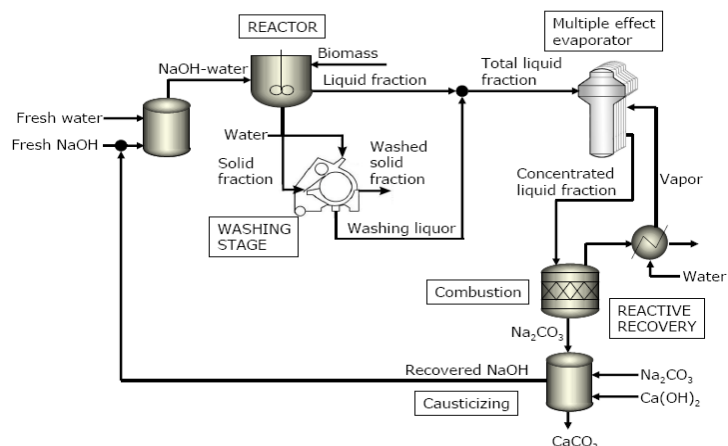
Figure 1. Process diagram of the soda biorefinery process proposed.

2.2. Organosolv-Ethanol Process

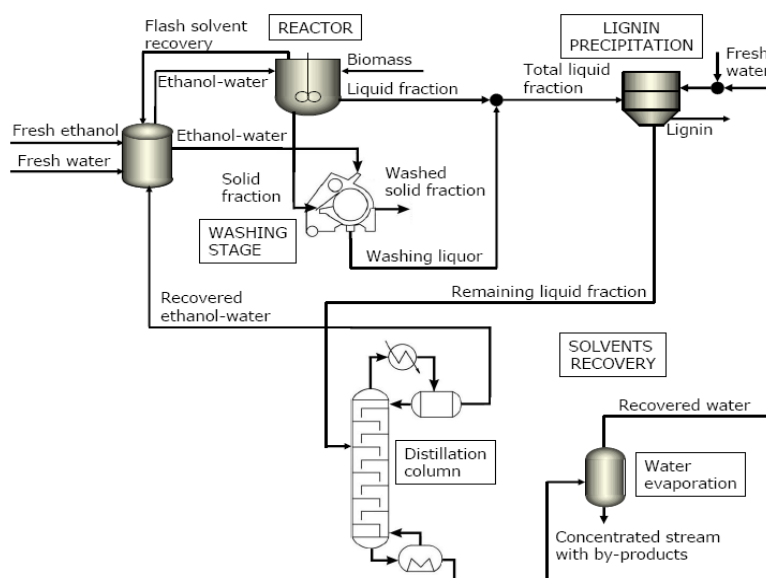Organosolv process diagram is presented in figure 2.

Figure 2. Process diagram of the organosolv-ethanol biorefinery process proposed.

Organosolv processes, based on using mixtures of water and organic alcohols or acids and particularly, ethanol organosolv process, gained new relevance for biomass pre-treatment in a biorefinery sense (Lignol process) as it allowed obtaining several high-value products (cellulose, lignin, hemicelluloses, extractives) under mild reaction conditions by a cost-effective process in which solvents were recovered and recycled at the end of the process (Pan et al., 2005). The raw material (biomass) is mixed with the solvent (ethanol-water, 60/40 w/w) in a pressurized reactor (T: 160ºC, time: 90 min, liquid/solid ratio: 6/1 w/w). Once reaction time is finished, a flash stream composed by ethanol and water, is obtained from the reactor, condensed and recycled to the reaction

step. Two other streams are obtained: a wet solid fraction, SF, mainly constituted by cellulose, a liquid fraction, LF, containing dissolved hemicelluloses, lignin and remaining solvent. The SF is sent to the washing stage, where a mixture of ethanol-water with the same concentration used in reaction stage is used for cleaning fibers. The resulting washing liquor WL is then mixed with the LF from the reactor and this stream is treated to recover by-products and solvents. Firstly, lignin content in LF is obtained by precipitation by adding of two volumes of water. Remaining LF after lignin precipitation is sent to the distillation unit, where a mixture of ethanol-water is obtained as distillate and recycled to the reaction unit. The residue, composed by water and co-products, mainly hemicellulosic sugars, is treated by heating in a flash unit to obtain a clean water stream, which is sent back to the lignin precipitation unit, and a concentrated stream with the remaining process by-products for subsequent treatment and use.

3. Simulation Approach of Biorefinery Processes by using Aspen Plus

Aspen Plus was used to design and simulate both processes on the basis of experimental results. Lignin, cellulose and hemicelluloses were defined by their chemical structure and physical properties which were obtained from the National Renewable Energy Laboratory (NREL) database (NREL/MP-425-20685; task number BF521004), whereas other conventional components were selected from the ASPEN PLUS data bank. ELECT-NRTL model (Non-Random, Two Liquids for electrolyte systems) was used to simulate the thermodynamic properties of solutions in the soda process, whereas NRTL-RK (Non-Random, Two Liquids - Redlich Kwong) model was selected in the case of organosolv one. These routes include the NRTL equation, obtained by Renon and Prausnitz, for the liquids activity coefficients calculation, Henry's law for the dissolved gases and, in the second case, RKS (Redlich-Kwong-Soave) equation of state for the vapour phase. The simulation process was developed using the following inlet streams to the reactor: 1000 kg/h of dry raw material and 6000 kg/h of solvent (soda-water 15% w/w and ethanol-water 60% w/w respectively) which corresponded to a liquid/solid ratio (w/w) of 6. Raw material composition was defined as a typical lignocellulosic non-wood material composition: 45% cellulose, 28% hemicelluloses, 25% lignin, 2% inorganic compounds (% on a dry weight basis). Reaction conditions for the simulation were: Soda process temperature: 90ºC, ethanol process temperature: 160ºC.

4. Simulation Results and Comparative Study

Obtained process solid fraction yield was 450 kg/h and 495 kg/h (on a dry weight basis) for soda and ethanol systems respectively. Furthermore, the latter scheme allowed the obtaining of 130 kg/h of organosolv lignin, which is a sulfur-free aromatic type structure with high amount of phenolic hydroxyl groups that favour their incorporation into polymer formulations and their chemical modification (Kubo and Kadla, 2004).

Hemicellulosic sugars (311 kg/h) were also obtained in an aqueous concentrated solution that could be subsequently treated for different applications as, for example, by enzymatic hydrolysis to produce bio-ethanol. Chemicals and water consumption of both processes is included in Table 1. According to energy requirements, the utilities consumptions for both processes have been calculated (Table 2 for soda process and Table 3 for ethanol-organosolv process). To determine the corresponding cost of steam and cooling water requirements, a factor of 48 €/kW·year (heating) and 4.7 €/kW·year (cooling) has been applied respectively.

Energy and Economic Assessment of Soda and Organosolv Biorefinery Processes

Table 1. Chemicals and water requirements for the two biorefinery processes analyzed (for 1000 kg/h of raw material).

REQUIREMENTS	Soda process (kg/h)			Organosolv-ethanol process (kg/h)	
	Soda	Water	Na$_2$CO$_3$	Ethanol	Water
Reaction stage	900	5100	0	3600	2400
SF washing	0	1275	0	390	260
Lignin precipitation	0	0	0	0	2500
Soda recovery	0	1340	1087	0	0
Total required	**900**	**7715**	**1087**	**3990**	**5160**
RECOVERED	**820**	**840**	**0**	**3895**	**4214**
FRESH INPUT	**80**	**6875**	**1087**	**85**	**861**

Table 2. Utilities requirements and associated costs for the soda process (for 1000 kg/h of raw material). MPS: Medium pressure steam; LPS: Low pressure steam.

PROCESS STAGE	Soda process				
	Utility requirements			Costs	
	MPS	LPS	Cold water	Heating	Cooling
	kg/h			k€/year	
Reaction stage	776	0	0	20.2	0
Soda recovery					
Multiple evaporators	0	5463	0	153.6	0
Fresh vapor	0	1340	0	37.7	0
Vapor condensation	0	0	25341	0	2.76
Recovery boiler	0	495	0	13.9	0
Total	**776**	**7298**	**25341**	**225.4**	**2.76**
Energy (MW)	**0.42**	**4.275**	**0.587**		
GENERATED	**794**	**7298**	**0**		
NET REQUIREMENT	**0**	**0**	**25341**	**0**	**2.76**
TOTAL COST				**2.76**	

Table 3. Utilities requirements and associated costs for the organosolv-ethanol process (for 1000 kg/h of raw material). MPS: Medium pressure steam; LPS: Low pressure steam.

PROCESS STAGE	Organosolv-ethanol process				
	Utility requirements			Costs	
	MPS	LPS	Cold water	Heating	Cooling
	kg/h			k€/year	
Reaction stage	3549	0	0	92.2	0
Solvents recovery					
Flash stream condensation	0	0	88500	0	9.6
Distillation condenser	0	0	59144	0	6.4
Distillation reboiler	0	768	0	21.6	0
Water evaporation	0	1571	0	44.2	0
Steam flash condensation	0	0	39717	0	4.3
Total	**3549**	**2339**	**187361**	**157.9**	**20.4**
Energy (MW)	**1.92**	**1.37**	**4.34**		
GENERATED	**0**	**0**	**0**	**0**	**0**
NET REQUIREMENT	**3549**	**2339**	**187361**		
TOTAL COST				**178.3**	

As it can be observed, soda process energetic requirements are considerably lower than organosolv-ethanol ones, as most of the heating utilities are generated within the process. On the other side no by-products are obtained and only the enriched cellulosic solid fraction is produced. However, organosolv process allowed the obtaining of high quality cellulose, lignin and profitable hemicellulosic components.

4. Conclusions

Simulation tools have been successfully applied to compare two biorefinery processes treating the same raw material, in terms of energy and chemicals consumptions. Soda process was found to be energy efficient, as energy requirement of the process was generated by the combustion of the organic fraction of the black liquor in the recovery boiler. Organosolv-ethanol process allowed an integral use of the raw material and the obtaining of high added value by-products, as well as the recovery of the solvents. Process simulation has permitted the study of biorefinery processes and the evaluation of different alternatives design of the studied processes and also different operation conditions.

5. Acknowledgements

References

R. Bhargava, S. Khanam, B. Mohanty, A.K. Ray, 2008, Simulation of Flat Falling Film Evaporator System for Concentration of Black Liquor. Comput. Chem. Eng., 32, 3213-3223.

J. Bujak, 2008, Energy Savings and Heat Efficiency in the Paper Industry: A Case Study of a Corrugated Board Machine. Energy, 33, 1597-1608.

S.Y. Ding, M.E. Himmel, 2006, The maize primary cell wall microfibril: A new model derived from direct visualization. J. Agric. Food Chem. 54 3 597-606.

M. González, A. Tejado, M. Blanco, I. Mondragon, J. Labidi, 2009, Agricultural Palm Oil Tree Residues as Raw Materia for Cellulose, Lignin and Hemicelluloses Production by Ethylene Glycol Pulping Process. Chem. Eng. J., 148, 106-114.

G. Iglesias, M. Bao, J. Lamas, A. Vega, 1996, Soda pulping of Miscanthus Sinensis. Effectcts of operational variables on pulp yield and lignin solubilization. Biores. Tech., 58, 17-23.

B. Kamm, M. Kamm, 2004, Biorefinery – Systems. Chem. Biochem. Eng. Q. 18 1 1–6.

S. Kubo, J.F. Kadla, 2004. Poly(ethylene oxide)/organosolv lignin blends: relationship between thermal properties, chemical structure, and blend behaviour, Macromolecules 37, 6904–6911.

M. Mercangöz, F.J. Doyle III, 2008, Real-Time Optimization of the Pulp Mill Benchmark Problem. Comput. Chem. Eng., 32, 789-804.

N. Mosier, C. Wyman, B. Dale, R. Elander, Y.Y. Lee, M. Holtzapple, M. Ladisch, 2005, Features of promising technologies for pretreatment of lignocellulosic biomass. Bioresour. Technol. 96 673–686.

H. Nurmesniemi, R. Pöykiö, R.L. Keiski, 2007, A Case Study of Waste Management at the Northern Finnich Pulp and Paper Mill Complex of Stora Enso Veitsiluoto Mills, Waste Manage., 27, 1939-1948.

X. Pan, J.F. Kadla, K. Ehara, N. Gilkes, J.N. Saddler, 2006, Organosolv ethanol lignin from hybrid poplar as a radical scavenger: relationship between lignin structure, extraction conditions, and antioxidant activity, J. Agric. Food Chem., 54 16 5806-5813.

Multi-objective optimization of a hydrogen supply chain for vehicle use including economic and financial risk metrics. A case study of Spain

Nagore Sabio, Mamdouh Gadalla, Laureano Jiménez, Gonzalo Guillén-Gosálbez[*].

Universitat Rovira i Virgili, Av. Països Catalans 26, Tarragona 43007, Spain,
[*]: gonzalo.guillen@urv.cat

Abstract

In this paper we present a decision-support tool to address the strategic planning of hydrogen supply chains for vehicle use under uncertainty in the operating costs of the network. Given a superstructure of alternatives that embeds a set of available technologies to produce, store and deliver hydrogen, the objective of our study is to determine the optimal design of the production-distribution network capable of fulfilling a predefined hydrogen demand. The design task is formulated as a multi-scenario mixed-integer linear programming problem (MILP) that decides on the production rates and expansions on capacity over time. The main novelty of the approach presented is that it allows controlling the variability of the economic performance of the hydrogen network at the design step in the space of uncertain parameters. This is accomplished by using a risk metric that is appended to the objective function as an additional criterion to be optimized. An efficient decomposition method is also presented in order to expedite the solution of the underlying mathematical model by exploiting its specific structure. The capabilities of the proposed modeling framework and solution strategy are illustrated through its application to a real case study based on Spain, for which valuable insights are obtained.

Keywords: Hydrogen, supply chain, mixed-integer linear programming, risk management, multi-objective optimization.

1. Introduction

The growing concern about possible disruptions in the oil supply and the need to reduce greenhouse gases (GHG) emissions have fostered in recent years the research for a more sustainable energy and transport model. In this context, hydrogen appears as a promising alternative fuel and energy carrier, since it can be produced locally, has the potential to be environmentally friendly and shows a wide range of applications [1]. Recent views suggest that the transition to the hydrogen economy will depend on two factors that must be developed in parallel: the construction of a hydrogen infrastructure has to be accompanied by policies promoting market for fuel cell technologies [1].
Focusing on the first of these factors, several mathematical programming models have been proposed that address the design of hydrogen supply chains (SCs). Most of these models are deterministic approaches assuming that all problem parameters can be perfectly known in advance. In the context of taking strategic decisions for developing a hydrogen infrastructure, this is an important limitation since the problem is in practice affected by different sources of uncertainty that can have a large impact on the final solution.

A possible way to overcome this limitation consists of applying stochastic optimization methods. Stochastic programming techniques allow assessing the performance of the problem under study in the space of uncertain parameters by optimizing the expected value of the objective function. In a more accurate way, financial risk metrics enable to take a step forward and control the variability of the objective function in the space of uncertain parameters [8]. Therefore, the introduction of a financial risk metric as an additional criterion to be optimized can be used to shape the form of the probability curves in order to reduce the possibility of having undesirable outcomes.

In the context of designing hydrogen SCs for vehicle use, only a few models have incorporated uncertainties [2, 3, 4], whereas none of them has applied financial risk management techniques. Hence, the aim of this paper is to consider the financial risk management associated with the long-term design and planning of hydrogen supply chains for vehicle use under uncertainty in production prices. The design task is posed as a multi-objective optimization problem where the expected total discounted cost of the network and a specific metric for financial risk (i.e. the worst case) are the objectives considered. The resulting large scale MILP tends to be computationally prohibitive as the number of time periods and potential locations for the SC entities increase. Hence, an efficient decomposition method is also presented to expedite the search for the Pareto solutions of the model by exploiting its mathematical structure. Finally, the capabilities of the proposed modeling framework and solution strategy are illustrated through its application to a real case study based on Spain.

2. Problem Statement

The design problem addressed in this paper has the objective of determining the configuration of a three-echelon hydrogen network for vehicle use (production-storage-market) with the goal of minimizing the expected total cost and financial risk. We consider a given region of interest (e.g., a country, a continent, etc.) that can be divided into a set of potential locations (g) for production and storage facilities. Each of these regions is characterized by a given hydrogen demand ($\overline{D_{gt}}$). This set of potential locations along with the associated geographical distribution of the demand must be provided as input data to the problem. The network design can therefore be formally stated as follows.

Given are a fixed time horizon and number of time periods (t), the set of available production (p), storage (s) and transportation technologies (l), the capacity limitations of plants ($\overline{PC_p^{PL}}$, $\underline{PC_p^{PL}}$) and storage facilities ($\overline{SC_s^{ST}}$, $\underline{SC_s^{ST}}$) the costs associated with the network operation (facility $FOC_{t,e}$ and transportation costs $TOC_{t,e}$), the investment costs (FCC_t, TCC_t) the probabilistic information that describe the uncertain parameters (i.e., type of probability distribution, mean and variance) and interest rate (ir).

The final goal is to determine: 1) the optimal SC design, including the number, type, location and capacity of plants and storage facilities (N_{gpt}^{PL}, N_{gst}^{ST}), along with the number and type of transportation units (N_{lt}^{TR}, e.g. tanker trucks, railway tube cars, etc.), and transportation links to be established between the SC facilities ($X_{gg'lt}$); and 2) the associated planning decisions, covering production rates at the plants (PR_{igpt}), inventory levels at the storage facilities (S_{igst}) and flows of hydrogen between plants, storage facilities and final markets ($Q_{igg'lt}$); in order to minimize simultaneously the total cost and financial risk.

Multi-objective optimization of a hydrogen supply chain for vehicle use including economic and financial risk metrics. 'A case study of Spain'

3. Mathematical Model

The model presented is motivated by previous formulations [3, 4, 5, 6]. Specifically, our approach relies on a multi-period optimization model that extends the formulations previously presented by the authors in order to account for more production and transportation technologies and study the evolution of the network over time. The model also considers the uncertainty of the coefficients of the objective function via a multi-scenario stochastic programming approach.

Particularly, the mathematical formulation considers the possibility of establishing different production, storage and transportation facilities, which are represented by specific integer variables ($N_{gpt}^{PL}, N_{gst}^{ST}, N_{lt}^{TR}$), in a set of potential locations with known demand and uncertain operating costs. The establishment of transportation links between the potential locations is modeled through a binary variable ($X_{gg'lt}$).

The mathematical formulation can be divided into three main blocks of equations, which include the mass balance constraints, the capacity constraints and the objective function equations.

3.1. Mass balance equations

The mass balance must be satisfied in each potential location g and time period t. Thus, for every hydrogen form i, liquid or gas, the initial inventory kept in a location S_{igst-1} plus the amount produced (PR_{igpt}) and the input flow rate ($Q_{ilg'gt}$) must equal the final inventory (S_{igst}) plus the amount delivered to the customers (D_{igt}) and the output flow rate ($Q_{ilgg't}$):

$$\sum_{s \in SI(i)} S_{igst-1} + PR_{igpt} + \sum_{l} \sum_{g' \neq g} Q_{ilg'gt} = \sum_{s \in SI(i)} S_{igst} + D_{igt} +$$
$$+ \sum_{l} \sum_{g \neq g'} Q_{ilgg't} \qquad \forall i,g,t \tag{1}$$

In this equation, *SI(i)* represents the set of technologies that can be used to store product form *i*.

3.2. Capacity constraints

Equations 2 and 3 are applied to limit the capacity expansions of production and storage technologies within lower and upper bounds.

$$NP_{gpt}^{PL} \overline{PC_{p}^{PL}} \leq CE_{gpt}^{PL} \leq NP_{gpt}^{PL} \underline{PC_{p}^{PL}} \qquad \forall g,p,t \tag{2}$$

$$NS_{gst}^{ST} \overline{SC_{s}^{ST}} \leq SE_{gst}^{ST} \leq NP_{gst}^{ST} \underline{SC_{s}^{ST}} \qquad \forall g,s,t \tag{3}$$

Regarding the transportation flows, a zero value of the binary variable $X_{gg'lt}$ prevents the flow of materials between potential locations from taking place, whereas a value of one allows for the transport of materials within some specified lower and upper bounds.

$$Q_{l}^{min} X_{lg'gte} \leq \sum_{i \in IL(i,l)} Q_{ilg'gte} \leq Q_{l}^{max} X_{lg'gte} \qquad \forall l,g,g',t,e \tag{4}$$

3.3. Objective function

The expected total cost is given by the mean value of the cost distribution:

$$E[TDC] = \sum_e prob_e TDC_e \qquad (5)$$

The total cost (TDC) is calculated as the summation of the discounted costs associated with each time period:

$$TDC_e = \sum_t \frac{TC_{t,e}}{(1+ir)^{t-1}} \qquad \forall e \qquad (6)$$

In the aforementioned expressions, e is the number of scenarios implemented to represent the uncertain parameters space, and $prob_e$ is the probability of occurrence associated to each scenario. The equations that model the establishment of gas pipelines and maritime transportation facilities are omitted here due to space limitations. In Equation 6, ir represents the interest rate and $TC_{t,e}$ is the total amount of money spent in period t and scenario e, which includes the capital (FCC_t, TCC_t) as well as operating costs ($FOC_{t,e}$, $TOC_{t,e}$) given by the production, storage and transportation facilities of the network:

$$TC_{te} = FCC_t + TCC_t + FOC_{te} + TOC_{te} \qquad \forall t,e \qquad (7)$$

4. Financial Risk

In this work, the probability of meeting unfavorable scenarios is controlled by adding the worst case cost as an additional objective to be minimized. This specific risk metric is easy to implement and leads to good numerical performance in stochastic models [9]. It can be determined from the maximum cost attained over all the scenarios as follows:

$$WC \geq TDC_e \qquad \forall e \qquad (8)$$

5. Solution Strategy: Two-step Sequential Approach

Our solution method is a two-step sequential approach that exploits the specific structure of the model. Inspired by previous bi-level decomposition methods presented so far in the literature [6, 7], the method relies on decomposing the original formulation into two hierarchical levels: a lower bound master problem (MP) and an upper bound slave problem (SP), that are solved in a sequential way. The master problem corresponds to a specific relaxation of the integer variables of the problem ($N_{gpt}^{PL}, N_{gst}^{ST}$, N_{lt}^{TR}), whereas the slave problem is obtained from the original full-space model rounding up some relaxed integer variables provided by the master problem solution.

6. Results and Discussion

All the problems were implemented in GAMS [10] and solved in an AMD Phenom ™ 8600 B Triple-Core 2300 MHz processor machine using CPLEX 9.0. We first solved several problems that differed in the level of complexity (*i.e.* number of time periods). The goal is to illustrate the performance of the proposed solution method as compared to the full-space model. The results show that the two-step sequential approach provides near optimal solutions in CPU times that are on average one order of magnitude lower than those reported by the full-space approach. The difference between the lower bound (MP) and the upper bound (SP) remains below 1% for resolutions up to eight periods.

Multi-objective optimization of a hydrogen supply chain for vehicle use including economic and financial risk metrics. 'A case study of Spain'

The model contains 19 potential locations representing the autonomous communities of Spain, whose hydrogen demand is supposed to cover 60% of the actual fossil fuels demand. Hydrogen can be obtained from four different production technologies, each one with the ability of producing liquefied or gaseous hydrogen, two different storage technologies, and six types of transportation modes including road, railway, pipelines and maritime transportation facilities. The length of the time horizon is twelve years and is divided in bi-annual periods. The uncertainty is represented by fifty scenarios, generated using Monte Carlo sampling on a set of normal distributions that characterize the uncertain operating costs. Specifically, in this particular example we aim to analyze the impact that the large variability in the natural gas price has on the SC design.
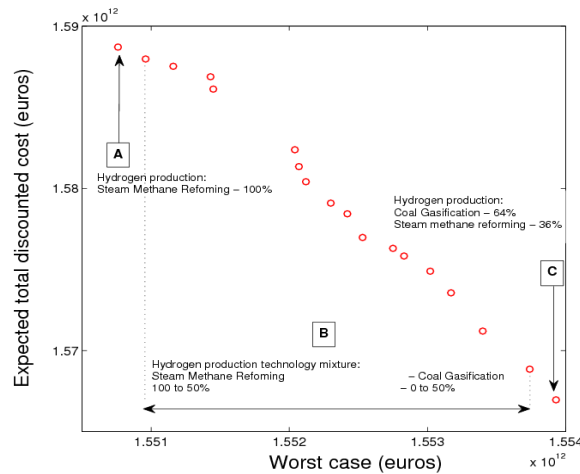

Figure 1. Pareto curve for slave problem using the 2-step sequential approach.

Figure 1 presents the Pareto frontier of the multi-objective problem. The results obtained show that in order to minimize the worst case, the model resolves to shift from steam methane reforming to coal gasification. This is because the natural gas price shows higher variability than the coal. Let us note that both technologies include carbon sequestration facilities for reducing the GHG emissions.
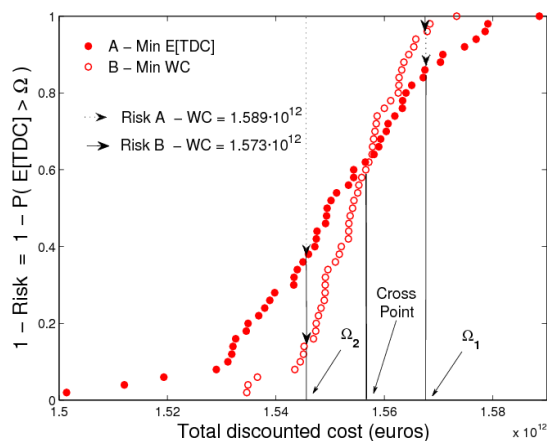

Figure 2. Cumulative probability curves of the Pareto extreme solutions.

Figure 2 depicts the cumulative probability curves of the feasible extreme of the Pareto set. The figure reveals that for low cost levels (lower than $1.557 \cdot 10^{12}$ €), the minimum cost solution shows a level of risk lower than the minimum worst case one. The former solution represents a hydrogen network composed by steam methane reforming plants, whereas the latter one involves a mixture of production plants dominated by coal gasification ones. On the other hand, for high cost levels, the coal gasification technology shows less financial risk than steam methane reforming. These extreme solutions and the ones in between reflect different possible attitudes towards.

7. Conclusions

This work has introduced a novel decision-support tool for risk management in the strategic design and planning of hydrogen supply chains under uncertainty in production cost. The problem has been posed as a bi-criteria stochastic MILP that simultaneously accounts for the minimization of the expected cost and the worst case. A new two-step sequential approach has also been presented in order to expedite the solution of such model. Numerical results have shown the convenience of replacing steam methane reforming by coal gasification to reduce the variability of the cost distribution. The approach presented is relevant to guide policy makers towards the adoption of more robust solutions in the face of uncertainty.

8. Acknowledgements

References

[1]   M. Ball, M. Wietschel, 2009, The future of hydrogen: opportunities and challenges, International Journal of Hydrogen Energy, 34:615-13

[2]   M. Melo, S. Níkel, F. Saldanha-da-Gama, 2009, Facility location and supply chain management – A review, European Journal of Operation Research, 196:401-12

[3]   J. Kim, I. Moon, 2008, Strategic design of hydrogen infrastructure considering cost and safety using multiobjective optimization, Int. J. Hydrogen Energy, 33:5887-21.

[4]   J. Kim, Y. Lee, I. Moon , 2008, Optimization of a hydrogen supply chain under demand uncertainty. International Journal of Hydrogen Energy, 33:4715-18.

[5]   A.Almansoori, N.Shah, 2006, Design and operation of a future hydrogen supply chain - Snapshot model. Chemical Engineering Research & Design, 84:423-16

[6]   G. Guillén-Gosálbez, F. Mele, I.E. Grossmann, 2009, A bi-criterion optimization approach for the design and planning of hydrogen supply chains for vehicle use with economic and environmental concerns, AIChE Journal, 56(3):650-667.

[7]   R. Iyer, I.E. Grossman, 1998, A bilevel decomposition algorithm for long-range planning of process networks, Industrial and Engineering Chemistry Research, 37:474-481

[8]   G. Guillen, F. Mele, M. Bagajewicz, A. Espuña, L. Puigjaner, 2005, Multiobjective supply chain design under uncertainty. Chemical Engineering Science, 60(6):1535-1553.

[9]   A. Bonfill,M. Bagajewicz, A. Espuña, L. Puigjaner, 2004; Risk management in scheduling of batch plants under uncertain market demand, Ind.Eng.Chem.Res, 43:741-750.

[10]   A. Brooke, D. Kendrick, A.Meeraus, R. Raman, 1998, GAMS: A User's Guide; GAMS Development Corporation.

Fuzzy Performance Indicators for Decision Making in Selection of Wastewater Treatment Methods

Yury Avramenko, Martin Kamami, Andrzej Kraslawski[a]

[a]*Faculty of Technology, Lappeenranta University of Technology, P.O. Box 20, FIN-53851 Lappeenranta, Finland, E-mail: Yury.Avramenko@lut.fi*

Abstract

The main objective of the work is to develop a decision support method for selection of sustainable wastewater treatment technologies. The method is based on use of the set of fuzzy formulated economic and environmental indicators. The environmental indicators are related to efficiency in removal of contaminants. The economic indicators correspond to energy and land requirements, capital and operational costs and reuse potential. The indicators are converted to fuzzy performance ratings which are categorized from excellent to very poor. The ratings are tabulated and used according to developed algorithm for selection of wastewater treatment technologies. The final decision grade is the weighted conjunction of all indicators. The calibration of the table and the validation for reliability of the obtained results has been done by the application of existing wastewater treatment decision support tool and previous cases.
Keywords: decision support, wastewater treatment, fuzzy notions

Notations

BOD – Biochemical oxygen demand   COD – Chemical oxygen demand
TSS – Total suspended solids   TDS – Total dissolved solids
TP – Total phosphorous   O & M – Operation and maintenance

1. Introduction

The design of wastewater treatment systems is a demanding task for the engineers. It consists in determination of the treatment levels to be achieved and sequencing of the methods to be applied in order to meet the ecological requirements. Usually there exist various options to achieve the objectives of the wastewater treatment. They should be always evaluated against many criteria of economic, social and environmental nature.

The process of evaluating and selecting appropriate wastewater treatment technology should consider the life cycle cost of such a system, including design, construction, operation, maintenance, repair and replacement (Massoud et al., 2009). Over the operational lifetime of the system the O&M costs are equally important as construction costs. In consequence, it calls for the development of appropriate methods to assist in the decision making process. The technologies would ensure the protection of the environment and public health and alleviate pressure on fresh water demand.

Simple, affordable, and efficient sewage treatment systems are urgently needed, especially in developing countries, where most of the conventional technologies currently in use in industrialized nations are too expensive and too complex (Grau, 1996). The problem associated with the current treatment technologies is a lack of sustainability. In comparison to the United States and Europe, domestic wastewater in arid areas like the Middle East are up to five times more concentrated in the amount of oxygen demand per volume of sewage. This is extremely high and may cause a large amount of sludge production (Bdour et al., 2009).

2. Problem Statement

The development of the decision supporting system for wastewater treatment often meet such obstacles as incomplete data description, uncertain characteristics, high dependences on local conditions. Therefore, the sustainable method must be able to cope with this behavior. One of the solutions could be to represent the problem by fuzzy descriptors and to propose the technique how to apply them.

The main objective of the present work is to develop a decision support method for selection of sustainable wastewater treatment technologies. The method is based on the use of several economic, technological and environmental indicators formulated as fuzzy concepts. The approach is purposed to analyze the treatment options resulting from either engineering considerations or the other design supporting tools.

3. Method Development

The method for decision-making support in the selection of wastewater treatment processes has been developed through stages of gathering and analyzing of data on practice of wastewater treatment to fuzzification of environmental and economic indicators. The steps can be described as follows:

1) Identification of significant technology performance indicators
2) Collection and tabulation of data on typical concentration ranges of wastewater contaminants and performance efficiency of the treatment technologies
3) Classification of concentration ranges of important wastewater characteristics into fuzzy classes
4) Translation of actual performance data into fuzzy notification based on analyzing of application of technologies efficiency to the given classes of concentrations
5) Rating of wastewater treatment technologies according to the level of performance on economic and environmental indicators
6) Setting the rules and operations to produce the grade for decision making based on tabulated rating of technologies performance
7) Validation of the method on cases provided by existing process selection tool

3.1. Indentification of the set of indicators

The relative importance associated with wastewater economic, environmental and societal attributes differ by community, region, and country. The performance data used was documented through the targeting regions that are classified as water scarce like the Mediterranean countries of Tunisia, Morocco, Egypt and Jordan. Wastewater treatment technologies that were investigated involved mainly the secondary biological treatment processes. These were activated sludge process, trickling filter, rotating biological contactors, waste stabilization ponds, constructed wetlands, land treatment and septic tank. These are the main technologies employed for wastewater treatment especially in developing countries (Muga, 2008; Volkman, 2003).

The selected important wastewater environmental indicators are BOD, COD, TSS, TDS, fats, oils & grease, nutrients, pathogens and heavy metals. The considered economic indicators are energy and land requirements, capital costs, operation and maintenance costs, hydraulic retention time, odor potential and sludge generation. The reuse considerations and regulatory requirements had a great impact on the importance of the specific indicators. The economic indicators determine the affordability of a particular technology to a community while environmental indicators measure resource utilization, and performance of the technology in removing conventional wastewater constituents.

3.2. Classification and fuzzy translation of indicators.

The indicators for the selection of wastewater treatment technologies were gathered from such literature sources as technical reports, journals and e-books. The obtained data were analysed and tabulated for each wastewater indicator translated into fuzzy concept.

The environmental indicators are related to efficiency of the technologies for contaminants removal. The efficiency is based on the degree of concentration reduction between influent and effluent flows. The range of contaminants contents can be from the highest (extreme) to the lowest (traces) concentration. The wastewater composition is represented using fuzzy notions to denote the class of concentration, namely, E-extreme, G-grand, H-high, M-medium, L-low, S-small and T-traces. The efficiency of a technology is allocated to one or two classes of performance ratings, - Excellent, High, Good, Moderate, Poor and Very poor; the rating is assigned based on level of reduction of concentration classes after treatment (for example, High performance corresponds to reduction of class E into H, or, G into M, etc). The table of environmental performance rating and corresponding levels of concentration class reduction in the row E-G-H-M-L-S-T are replicated in Table 1.

Table 1. Environmental performance rating and corresponding level of reduction

Performance rating	Level of reduction
Excellent	the higher concentrations reduced by 3 classes
High	the higher concentrations reduced by 2 classes
Good	the higher concentrations reduced by 1 class
Moderate	resulted in 67% reduction in concentration; reduction on 1 class is achieved when Moderate performance on same indicator is repeated twice; and 2 classes – when the performance is repeated 3 times.
Poor	resulted in 30% reduction in concentration; reduction on 1 class is achieved when Poor performance on same indicator is repeated three times.
Very poor	resulted in no reduction

The economic indicators correspond to energy and land requirements, capital costs, operation and maintenance costs, hydraulic retention time, effluent reuse potential for agriculture and sludge generation. Each indicator is converted to performance rating which is again represented as a fuzzy class – from excellent to very poor. The rating is obtained as a result of analysis of economic effects of the corresponding indicator, - the greater contribution to the cost the worse rating is. For instance, a technology that requires a lot of energy for treatment purposes or has high capital costs is rated poorly.

3.3. Deriving overall performance grade

Overall performance of treatment method that may include several technologies is derived from a consideration of weights assigned to importance of indicators and performance rating for both environmental and economic aspects.

Environmental and economic indicators were weighted depending on their relative degree of importance in determining agricultural reuse potential of the treated effluent. The weights, which are also of fuzzy nature, represent six categories from extremely

important to not important. Indicators' weights of importance were also determined by the reuse purpose of treated effluent. For instance, importance of indicators for agricultural reuse would be different from that of intended use as potable water.

The overall grade is derived according to the following procedure:

1) the environmental performance rating of a current technology from the set of wastewater treatment method is applied to fuzzy concentration of process influent; the corresponding performance fuzzy parameter is applied to every presented environmental indicator (such as TSS, BOD, COD etc).

2) the obtained fuzzy representation of process effluent (resulted level of reduction according to previous technology performance) is used as the influent for the following process from the set; and next, step 1 is repeated.

3) when all technologies from the treatment set are applied the final performance rating of the entire set is obtained based on the level of treatment: the lower final effluent the higher final rating is (effluent class T corresponds to E rating; class S – to G etc).

4) the economic performance rating is applied to every technology from the treatment set and every economic indicators.

5) the final rating is obtained by fuzzy convergence of all indicators with corresponding weights of importance over all technologies from the set

6) the overall grade the junction of environmental and economic final performance ratings; the ratings are defuzzified, divided to the sum of weights of all involved indicators converted to cardinal numbers, and then the result is represented as fuzzy concept again - from excellent (most appropriate) to very poor (unacceptable).

Based on the overall grade the decision on selection of treatment options is made. The better grade the more acceptable in local situation the treatment method is.

4. Implementation and Validation

The data on over 30 wastewater technologies and their various modifications are collected and the performance ratings for them are tabulated. The table includes both environmental and economics indicators represented in fuzzy denotation. A small part of the table is given in Table 2.

Table 2. A part of the Table of performance ratings for technologies

Technology	BOD removal	COD removal	TSS removal	Capital Cost	O & M	Sludge generation
Activated Sludge Process (ASP)	high	high	high	poor	poor	poor
Rotating Biological Contactors	good	good	good	moderate	moderate	poor
Trickling Filter + APS	high	high	high	very poor	very poor	very poor
Maturation Ponds (MP)	moderate	moderate	poor	high	high	moderate
Facultative Ponds (FP)	moderate	moderate	moderate	high	high	moderate
Intermittent Sand Filter	good	poor	good	good	moderate	moderate

4.1. Use of the table to make decision

The table and the introduced operations over fuzzy classes of performance are used for decision support in selection of sustainable wastewater treatment systems. For each indicator, the corresponding rating is retrieved for each technology from the system

sequence and then all values are united through special fuzzy conjunction. The overall criterion for the considered indicator is produced and next indicator is evaluated. The final decision grade is the weighted conjunction of all indicators. The grade has values from excellent (most appropriate) to very poor (unacceptable).

In order to illustrate the process of decision support the case of municipal wastewater treatment is further considered. The treatment technology employed is a sequence of units comprising of a screening device, a grit chamber and finally waste stabilization ponds, which in this case is comprised of anaerobic and facultative lagoons. The performance of the various units in the treatment sequence of the case is given in Table 3. The indicators are taken from Table of performance rating for corresponding units. The table 3 presents also the treatment outcome for the units for given wastewater indicators as well as the final ratings on environmental and economic indicators.

Table 3. Production of performance rating of treatment sequence

Technology\Indicator	BOD	COD	TSS	Capital Cost	O&M	Land	Energy
mg/l	705	1890	591				
influent fuzzy class	H	G	G				
Screening	-	-	poor	high	high	high	high
Grit removal	-	moderate	high	high	high	high	high
Anaerobic lagoon	moderate	moderate	moderate	high	high	very poor	excellent
Facultative ponds	moderate	moderate	moderate	high	high	poor	excellent
effluent fuzzy class	M	M	M				
Rating	Moderate	Moderate	Moderate	High	High	Moderate	Good

The overall rating of the technology sequence, taking into account weights of the importance (marked as cursive), is as follows:

 Moderate $\cap$ *Extreme important* $\cup$ Moderate $\cap$ *Very important* $\cup$ Moderate $\cap$ *Extreme important* $\cup$ High $\cap$ *Very Important* $\cup$ High $\cap$ *Extreme important* $\cup$ Moderate $\cap$ *Important* $\cup$ Good $\cap$ *Very Important* = GOOD.

The overall rating is Good; hence the conclusion is that the proposed treatment sequence is appropriate for application in the current conditions.

4.2. Validation of the method

The calibration of the table and the reliability validation of the obtained results has been done by using the wastewater treatment decision support tool, ED-WAVE, developed by the authors (Avramenko & Kraslawski, 2008). The method has been tested on the case base of wastewater treatment sequences which contains data of environmental and economical efficiency.

Validation provided the basis for verification on the accuracy of wastewater treatment results obtained through the decision support method. The fuzzy classification methods (on concentration, efficiency and economic evaluations) have been adjusted to get the results in accordance with the actual cases data.

During the verification process, actual case study results were compared with performance results obtained from application of decision support method. Table 4 presents the comparison of existing case data – municipal case study provided by ED-WAVE tool and results given by the presented decision support method (DSM).

Table 4. Comparison of actual case data with results of decision support method

Technology\Indicator	Influent actual	Influent fuzzy	Effluent actual	Effluent fuzzy	Effluent DSM	Rating from case data	Rating DSM
BOD (mg/L)	155.6	M	51.6	L	L	High	High
COD (mg/L)	397.2	M	106.2	L	L	High	High
TSS (mg/L)	154.1	M	118	M	L	Moderate	High
Nitrate (mg/L)	1.84	L	0.43	S	S	Good	Good
Capital costs ($/m3/d)			25.7-34.3	L	L	High	High
O & M ($/m3/d)			0.53-1.67	L	L	High	High
Land req. (m2/m3/d)			12.5-14	M	H	Moderate	Poor
Overall Rating						**High**	**High**

5. Summary

The goal of the work has been the development of a decision support method for the selection of sustainable wastewater treatment. In this research, the sustainability of treatment technologies was evaluated using a set of environmental and economic indicators. The performance ratings are attached to the considered indicators; they are supplemented with corresponding weights of importance representing the area of application of the treatment. This enabled a comparison of the overall performance of technologies to be made. The comparison of the existing case data provided the basis for verification on the accuracy of wastewater treatment results obtained through the decision support method. The results showed a very good agreement for these sets of data and for the cases where the variations occurred, the necessary adjustments of the method were done.

The most appropriate technology to be applied must be economically affordable, environmentally sustainable and socially acceptable. Societal attributes were not investigated during this research work and this is another important area that can be considered for further work.

References

Massoud M.A, Tarhini A. and Nasr, J.A. (2009). *Decentralized approaches to wastewater treatment and management: Applicability in developing countries.* Journal of Environmental Management 90, 652–659.

Grau, P. (1996). *Low cost wastewater treatment.* Water Science and Technology

Bdour A.N., Hamdi M.R and Tarawneh Z (2009). *Perspectives on Sustainable Wastewater Treatment Technologies and Reuse Options in the Urban Areas of the Mediterranean Region.* Desalination 237, 162–174.

Muga, H.E (2008). *Sustainability of Wastewater Treatment Technologies.* Journal of Environmental Management 88, 437-447.

Volkman, S (2003). *Sustainable Wastewater Treatment and Reuse in Urban Areas of the Developing World.* Michigan Technological University.

Avramenko, Y and Kraslawski, A. (2008). *Case Based Design. Applications in Process Engineering,* Studies in Computational Intelligence, Vol. 87, Springer-Verlag Berlin Heidelberg.

Simulation-Optimization for Business Decision Support in a Global Specialty Chemicals Enterprise

Jia Hui Tan[a], Arief Adhitya[b], Rajagopalan Srinivasan[a,b,]*

[a]*Department of Chemical and Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576, Singapore*
[b]*Institute of Chemical and Engineering Sciences, A*STAR (Agency for Science, Technology and Research), 1 Pesek Road, Jurong Island, Singapore 627833, Singapore*
Corresponding author: chergs@nus.edu.sg

Abstract

To tap on the vast business opportunities offered by globalization, companies are increasingly shifting from single-site manufacturing to multi-site enterprise operation. This results in a more complex supply chain structure, involving numerous entities in different locations with intricate dynamics. The optimization of such systems is not amenable to mathematical programming approaches. In this paper, we propose a simulation-optimization framework for business decision support in a global specialty chemicals enterprise. A dynamic simulation model is used to capture the behavior of the entities, their interactions, the various uncertainties, and the resulting dynamics. Optimization is done by coupling simulation with a non-dominated sorting genetic algorithm, implemented in a parallel computing environment for computational efficiency. The application of the proposed approach for business decision support are demonstrated in two case studies.

Keywords: dynamic simulation, optimization, NSGA-II, multi-site, specialty chemicals

1. Introduction

In today's competitive business environment, the ability of companies to reach out globally to different markets spanning across the world opens up vast business opportunities to be seized. Shifting from one-plant manufacturing facilities to multi-plant enterprise enables a company to have the flexibility of producing different products, focus on specialization activities, be close to low cost raw material sources as well as its targeted market. The supply chain (SC) of such multi-plant enterprise spans across continents, involving numerous entities with different interests and contends with various uncertainties. Enterprises consider supply chain management (SCM) to be a key factor for achieving better profitability and customer satisfaction.

Different types of decisions have to be made in SC management. Here, we focus on design and operational decisions. The former includes decisions on building a new plant, closing down an old plant, specialization, plant location, storage capacity, reactor size, etc, while the latter is mostly implemented through operational policies. Figure 1 shows a schematic of the global specialty chemicals SC for lubricant additive products. The specialty chemicals enterprise consists of a global headquarter and a number of production sites located at different regions in the world. Materials flow from suppliers to the plants and from the plants to customers (solid arrows in Figure 1) as controlled by the information exchanges (dotted arrows). Customer places an order with the global sales office, who then works with the scheduling department of each local plant to decide which plant to assign the order to, following an *order assignment policy*. The

scheduling department of the selected plant then inserts the order to its job schedule based on its *scheduling policy*. The procurement department buys raw materials based on a *procurement policy*. These policies determine how the departments operate and their combined actions result in the local plant performance and in turn the overall enterprise performance. For example, the procurement policy affects raw material availability for processing an order and the scheduling policy affects the completion date of the order, which accordingly affects the enterprise's profit and customer satisfaction.

Mathematical programming models have been proposed for managing multi-site SCs. Timpe and Kallrath (2000) presented a general mixed-integer linear programming (MILP) model for planning of production, distribution, and marketing for the multi-site SC. Dondo et al. (2008) focused on the management of logistic activities in the multi-site SC. Mathematical programming approaches generally work well for small-scale, short-term supply chain problems, but they are limited by the exponential increase in computation time for large-scale, long-term, integrated problems. At present, simulation is the predominant methodology for dealing with such problems. Optimization for business decision support can thus be done by coupling simulation with optimization techniques. Simulation enables the evaluation of a set of decision-making parameters in terms of certain SC performance indicators such as profit and customer satisfaction. Optimization guides the search for the parameters that would give the best performance. In this paper, we present the simulation-optimization-based decision support for the global specialty chemicals enterprise. We use a simulator of the lubricant additive SC, called Integrated Lubricant Additive Supply Chain (ILAS), developed by Yang et al. (2009), coupled with a non-dominated sorting genetic algorithm.

2. Proposed Simulation-Optimization Method

2.1. Simulation through ILAS

ILAS is built on the Matlab/Simulink (MathWorks, 2007) platform and designed with the intention of mimicking the entities in Figure 1, their activities and interactions, with considerations of economics and stochastics in the SC. ILAS allows the user to assess the overall impact of policies or decisions on the enterprise performance. Here, performance is measured through three indicators: profit, accepted order index, and customer satisfaction. While each plant has its own local performance measures, we are interested in the overall enterprise performance.
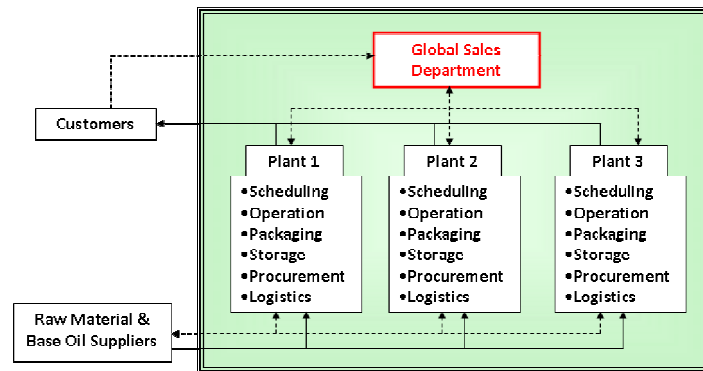

Figure 1. Global specialty chemicals supply chain

Simulation-Optimization for Business Decision Support in a Global Specialty Chemicals Enterprise

Profit is calculated as revenue from product sales minus costs:

$$Profit = Revenue - \begin{pmatrix} Raw\ Material\ Cost \\ +Variable\ Operating\ Cost \\ +Fixed\ Operating\ Cost \\ +Packaging\ Cost \\ +Transportation\ Cost \\ +Inventory\ Cost \\ +Late\ Penalty\ Cost \end{pmatrix} \qquad (1)$$

Each customer order comes to the global sales department with a due date. If none of the plants can deliver the order by the required date, the global sales department will decline the order and it will not be assigned to any plant. This is considered a missed opportunity for the enterprise and is measured through the accepted order index:

$$CS = (\frac{number\ of\ accepted\ orders}{number\ of\ orders}) \times 100\% \qquad (2)$$

Customer satisfaction is measured as the percentage of non-late order deliveries out of the total number of accepted orders:

$$CS = (1 - \frac{number\ of\ late\ orders}{number\ of\ accepted\ orders}) \times 100\% \qquad (3)$$

2.2. Optimization through NSGA-II

The above three performance indicators are the objectives for the multi-objective optimization. NSGA-II (Deb et al., 2002) is one of the most widely-used algorithms for multi-objective optimization problems due to its low computational requirements, elitist approach, and parameter-less sharing, and is chosen for this work. Considering stochastics in the simulation runs, we adopt the two-phase optimization scheme in Koo et al. (2008). In Phase 1, which is the NSGA-II run, each individual chromosome containing the decision variables is evaluated via ILAS simulation once and sorted into fronts based on the dominance of objective values. Evaluated and sorted chromosomes are then selected for breeding, where two chromosomes are selected at random from the current population and their fronts are compared. The chromosome that belongs to a better front is selected as parent for breeding. If both chromosomes belong to the same front, the chromosome with a larger crowding distance will be chosen. Breeding process continues for generations until a user-determined number of generations, *n*, has been reached. Chromosomes belonging to Front 1 in generation *n* are then evaluated more accurately in Phase 2 where they are simulated 100 times to get the representative performance indicators. Based on the mean objective values from 100 runs and considering the error band of each objective, the chromosomes are again sorted into fronts. If the error band is *e%*, chromosome A dominates chromosome B only if (100-*e*)% of the mean objective value of A is greater than (100+*e*)% of the mean objective value of B, for all objectives. This error band is determined from the convergence index from 100 runs (Koo et al., 2008).

2.3. Parallel Implementation

Since the simulation needs to be run thousands of times for each decision problem, we have used parallel computing implementation to reduce the total time. The problem is broken into discrete parts that can be executed simultaneously on different CPUs with some coordination. The time consuming step is the evaluation of objective functions, as

each run of ILAS simulation takes ~10 minutes. Also, simulation of one chromosome in a population is independent of simulation of other chromosomes, so they can be executed in parallel. Multiple compute nodes are used, with one given the role of master and the others as slaves. The master executes the NSGA-II routine – sorting, selection, and operations on initial population to generate offspring chromosomes. The slaves then pick a chromosome, evaluate it by performing ILAS simulation and attach the objective values to the chromosome. Each slave then picks the next available chromosome. Once the children population is completely evaluated, the master proceeds to generate the next population and the whole procedure is repeated. In this work, we used three slave compute nodes, which results in a 67% saving of computational time.

3. Case Studies

The global specialty chemicals enterprise considered in the case studies has three plants located in Singapore, Houston, and Japan. They are producing three types of product with five different grades for each product. Each grade is produced using a specific combination from five components and three base oils. The order assignment policy used is *projected completion date*, where the order is assigned to the plant whose projected completion date can best meet the order due date. The scheduling policy used in all plants is *earliest due date*, where the jobs are sorted according to its due date, from the earliest to the latest. Procurement follows the *reorder point* policy. Under this policy, raw material will be purchased when its inventory falls below a certain reorder point to bring it back to a certain top-up level.

3.1. Case Study 1 (Design): Optimal process reactor capacity

This case study illustrates how the proposed method supports a design decision: finding the optimal process reactor capacity. Reactor size determines the size of each production batch and each batch has a certain cycle time, therefore the production time required to process a customer order will depend on the capacity of the reactor. A large reactor enables a bigger batch size production which implies that a particular order size can be made in fewer batches and hence overall shorter production time. Hence, a large reactor is favored over a small reactor because of its ability to process larger order within fewer batches and cut down on the overall production time. With current reactor capacities of 2500 kg/batch, 1500 kg/batch, and 1500 kg/batch for Singapore plant, Houston plant, and Japan plant respectively, management considers the option of increasing reactor capacities. Although a larger reactor capacity implies the ability to handle larger order in shorter time, and thus the ability to take up more jobs and generate higher revenue, the initial capital cost of such a larger reactor has to be considered. Furthermore, the fixed operating cost, which factors in depreciation of equipment associated, will be higher given the higher capital cost of larger reactor. Therefore, while accepted order index and customer satisfaction are favored with increased reactor capacity, total profit of the enterprise does not necessarily improve or benefit from such an investment.

The possible reactor sizes and their corresponding capital costs are given in Table 1. Simulation-optimization is performed with the three plants' reactor capacities as decision variables using the NSGA-II parameters listed in Table 2. The resulting non-dominated front chromosomes are shown in Table 3. We can see that all six chromosomes give higher customer satisfaction and accepted order than the base case. It is clear that the current capacities are not enough to capitalize on the demand. Studying the results, management decides that it prefers both customer satisfaction and accepted order index to be higher than 98%. Two chromosomes satisfy this requirement: [3000;

3000; 5000] and [5000; 5000; 5000]. From this two options, [3000; 3000; 5000] gives a higher profit of 4.408 M$ and is selected.

Table 1. Range of parameters for the case studies

Case	Parameter	Range	
1 (Design)	Process reactor capacity and the corresponding capital cost	1000	$60,000
		1250	$80,000
		1500	$130,000
		1750	$170,000
		2000	$230,000
		2500	$360,000
		3000	$520,000
		3500	$700,000
		4000	$910,000
		5000	$1,440,000
2 (Policy)	Reorder point for 5 components	1000; 1500; 2250	
	Reorder point for 3 base oils	5000; 7500; 10000; 12500	

Table 2. NSGA-II parameters used

Population size	100		
Number of generations	30		
Number of crossover operations per generation (Crossover rate)			
Arithmetic: 10	Heuristic: 10		Simple: 10
Number of mutation operations per generation (Mutation rate)			
Boundary	10	Non-uniform	10
Multi-non-uniform	10	Uniform	10

Table 3. Results for Case Study 1: Optimal process reactor capacity

Process Reactor Size			Total Profit (M$)	Customer Satisfaction (%)	Accepted Order Index (%)
Singapore Plant	Houston Plant	Japan Plant			
2500 (base)	1500 (base)	1500 (base)	4.224	95.75	62.88
2500	3000	5000	4.675	96.47	89.47
5000	1000	5000	4.574	97.09	91.07
3000	3500	5000	4.417	96.74	92.19
3000	3000	5000	4.408	98.14	98.35
3500	3500	5000	4.136	96.72	93.08
5000	5000	5000	3.708	98.66	98.85

Table 4. Results for Case Study 2: Optimal reorder point for components and base oils

Reorder Point			Total Profit (M$)	Customer Satisfaction (%)	Accepted Order Index (%)
Singapore Plant	Houston Plant	Japan Plant			
1000 5000	1000 5000	1500 7500	3.864	93.92	77.53
1500 7500	1500 5000	1500 7500	3.748	97.19	78.79
1500 10000	2250 10000	1500 7500	3.328	98.33	79.29
1500 7500	2250 10000	1500 10000	3.327	98.32	79.21

3.2. Case Study 2 (Policy): Optimal raw material reorder point

This case study focuses on a policy decision: finding the optimal raw material reorder point for procurement. A higher reorder point implies that raw materials have to be procured more frequently because the duration before the reorder point is reached is shorter, incurring more fixed procurement cost which is charged for each procurement. It also leads to higher raw material inventory and consequently higher inventory cost. While a lower reorder point is favorable for greater cost saving through reduction in fixed procurement cost and inventory cost, it increases vulnerability to stock-out situations. When a plant runs short of raw materials to process the next-in-schedule job, production has to stop or slow down, resulting in more late job deliveries (lower customer satisfaction) as the plant waits for fresh supply of raw materials to resume production. Furthermore, a production delay pushes back the whole schedule and this leads to a later projected completion date for a potential order, increasing the probability of missing the potential order (lower accepted order index).

The top-up point is 3000 for each of the five components and 15000 for each of the three base oils. The various reorder points under consideration are shown in Table 1. Each plant will have two reorder points, one common for the five components and the other for the three base oils. Since there are three different plants, six decision variables form the chromosome. Simulation-optimization is performed using the NSGA-II parameters listed in Table 2 and the resulting non-dominated front chromosomes are shown in Table 4. Management desires customer satisfaction to be higher than 98% and accepted order index to be around 80%. From the two options that qualify, [1500 10000; 2250 10000; 1500 7500] gives a higher profit and is selected.

4. Concluding Remarks

The optimization of design and operational decisions in a global multi-site specialty chemicals is not amenable to traditional optimization approaches. In this paper, a simulation-optimization-based approach has been developed to support such decisions. Optimization is performed using a non-dominated sorting genetic algorithm linked to a dynamic simulator of the global specialty chemicals enterprise. Two case studies on selecting optimal reactor capacities and reorder points illustrate how the proposed method supports design and policy decisions. With the simulation-optimization framework, it is also possible to evaluate optimal strategies for dealing with disruptions. This is the direction of our current research.

References

K. Deb, A. Pratap, S. Agrawal, T. Meyarivan, 2002, A fast elitist multi-objective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, 6(2), 182-197.

R. Dondo, C. A. Mendez, J. Cerda, 2008, Optimal management of logistic activities in multi-site environments, Computers and Chemical Engineering, 32(11), 2547-2569.

L. Y. Koo, A. Adhitya, R. Srinivasan, I. A. Karimi, 2008, Decision support for integrated refinery supply chains: Part 2. Design and operation, Computers and Chemical Engineering, 32(11), 2787-2800.

MathWorks, 2007, Simulink User's Guide.

C. H. Timpe, J. Kallrath, 2000, Optimal planning in large multi-site production networks, European Journal of Operational Research, 126(2), 422-435.

J. Yang, A. Adhitya, R. Srinivasan, 2009, Dynamic modeling of a multi-site specialty chemical manufacturing supply chain, International Conference on Infrastructure Systems and Services (NGInfra09), 9-11 December, Chennai, India.

Generalized Classes for Lower Levels of Supply Chain Management: Object-Oriented Approach

Flavio Manenti,[a] Nádson M. N. Lima,[b] Lamia Zuñiga Liñan,[b] Alberto Cuoci,[a] Alessio Frassoldati[a]

[a]Politecnico di Milano, CMIC dept. Giulio Natta, Piazza Leonardo da Vinci 32, 20133 Milano, ITALY; Email: flavio.manenti@polimi.it
[b]University of Campinas (UNICAMP), Department of Chemical Processes, PO Box 6066, 13081-970, Campinas, São Paulo, BRAZIL; Email: nadson@feq.unicamp.br

Abstract

A possible and promising approach to effectively tackle the distinct optimization levels of supply chain hierarchy is to combine the object-oriented programming with the parallel computing. From this perspective, the paper proposes a generalized framework to solve the lowest levels of supply chain management paradigm by means of the *BzzMath* library as numerical kernel (optimizers and differential solvers) and *openMP* directives for exploiting shared memory machines. Also, the object-oriented approach allows the implementation of more solvers to force the same generalized class to automatically select the best one among them according to the problem and the user has not to worry about what solver and which optimizer are preferable.

Keywords: Dynamic Optimization, Supply Chain, Parallel Computing, Object-Oriented Programming

1. Introduction

This paper deals with object-oriented programming and parallel computing to effectively tackle the lower optimal control levels of supply chain management hierarchy (Biegler, 2007; Manenti and Manca, 2009), specifically the nonlinear model predictive control and the real-time dynamic optimization levels (Morari and Lee, 1999), both characterized by continuous variables only, by leaving the development of generalized classes for strategic levels (i.e., scheduling and planning) to future works.

The twofold aim of studying a generalized class for solving optimal control problems is the need of finding an efficient solution for the supply chain management problem as well as to propose and validate a freely downloadable tool to support users in settling nonlinear model predictive control and business-wide dynamic optimization especially looking at the increase in last 5 years of such industrial applications (Bauer and Craig, 2008; Qin and Badgwell, 2003).

A generalized C++ class to solve nonlinear model predictive control and dynamic optimization problems is proposed here. As optimal control problems usually involve (i) differential equation systems to foresee the plant behavior and/or process unit dynamics and (ii) constrained optimization issues to meet process specs and requirements, the freely downloadable *BzzMath* library (Buzzi-Ferraris, 2009) is adopted as kernel to handle both these tasks numerically.

The proposed class allows both FORTRAN and C++ users to easily solve predictive control and dynamic optimization problems by only defining their own convolution system and the desired objective function, without taking care of any numerical problem

that may occur in integrating differential systems, in searching for the minimum of a constrained/complex objective function, and in implementing a moving horizon methodology as these issues are usually consciously tackled by the same generalized class. Actually, thanks to the object-oriented philosophy, classes are able to automatically detect the best combinations of algorithms and numerical methods as well and to switch among them during the solution so to solve such issues at best. As an example, in the optimization of a multidimensional multimodal function subject to nonlinear constraints, narrow valleys, and even undefined regions somewhere in the domain of search, a robust optimizer is firstly adopted to check the function at the macro-scale, whereas more efficient methods are adopted to locally refine the search. The switch from the methods is automatically managed by the C++ class (Buzzi-Ferraris and Manenti, 2010a). Some specific industrial cases already discussed elsewhere by the same authors (Lima *et al.*, 2009; Manenti *et al.*, 2009; Manenti and Rovaglio, 2008) were selected to validate the proposed approach.

2. Optimizers and Differential Solvers Belonging to BzzMath Library

This research activity is based on BzzMath library, which is freely downloadable at Professor Buzzi-Ferraris's homepage. BzzMath is a numerical library entirely written in C++ by exploiting object-oriented programming. It covers several scientific fields such as linear algebra, linear/nonlinear regressions, optimization, differential systems and so on. For the sake of conciseness, only those topics that are of interest in this paper are briefly introduced hereinafter and we remind the reader to related works (Buzzi-Ferraris and Manenti, 2010b, 2010c) to get more details about them.

Starting from OPTNOV's variant (Buzzi-Ferraris, 1967) up to the most recent improvements (Buzzi-Ferraris and Manenti, 2010a), numerically robust and efficient optimizers are implemented for unconstrained optimization and linear/nonlinear programming by exploiting the multi-processor structure by means of *OpenMP* directives. For example, very robust methods perform additional (parallel) searches even far from the point which the algorithm is converging to. The selection of the new starting point for parallel searches is not carried out randomly, but it is performed by means of criteria similar to the ones adopted to tackle the optimal experimental design paradigm (Buzzi-Ferraris and Manenti, 2009, 2010c; Manenti and Buzzi-Ferraris, 2009a, 2009b) to reasonably examine the function at the macro-scale without leaving any sub-region unexplored. Robust algorithms find immediate application even in the solution of nonlinear systems. For example we quote the very large-scale system (sparse blocks matrix with a number of equations in the order of some tens of millions) that is behind the kinetic post-processor already discussed elsewhere (Cuoci *et al.*, 2007).

To solve the nonlinear model predictive control (NMPC) problem, differential solvers are required to integrate the convolution model representing the process and they must be opportunely coupled with optimizers (Binder *et al.*, 2001). BzzMath library includes very performing algorithms for solving ordinary differential equation (ODE) systems and differential and algebraic equation (DAE) systems. Moreover, by performing opportune checks on the system sparsity and, if sparse, even on its structure, C++ classes are able to automatically select the best solver to reduce computational efforts. Recently, a solver to efficiently integrate partially structured DAE systems has been introduced in BzzMath library (Manenti *et al.*, 2009). Such a solver is particularly performing to face process control problems where the (usually sparse and structured) matrix structure is spoiled by the integral terms of proportional-integral loops if conventional controls are adopted to manage process units and plant-wide operations.

3. Generalized Class Architecture for Nonlinear Model Predictive Control

NMPC is an appealing and well-established control methodology as: (i) it is intrinsically able to manage nonlinearities in process dynamics; (ii) it allows easily implementing constraints, not only those ones concerning the model, but even bounds of manipulated variables; and (iii) its receding horizon structure could be extended to the highest levels of the supply chain management.

The basic architecture of NMPC application is reported in Fig. 1. Assuming an on-line implementation of this technique, the plant provides data to the model predictive control at each sampling time. Specifically, plant data are firstly processed and reconciled (Signor *et al.*, 2010) and then they are sent to the optimizer, which includes an objective function, a dynamic model and usually, according to the mathematical model type, a numerical integrator to solve specific differential systems, such as ordinary differential, differential-algebraic, partial differential, and partial differential-algebraic equations systems. If anyone needs to go beyond the classical quadratic formulation of NMPC, economic data, scenarios, and demand market could be added to the objective function.
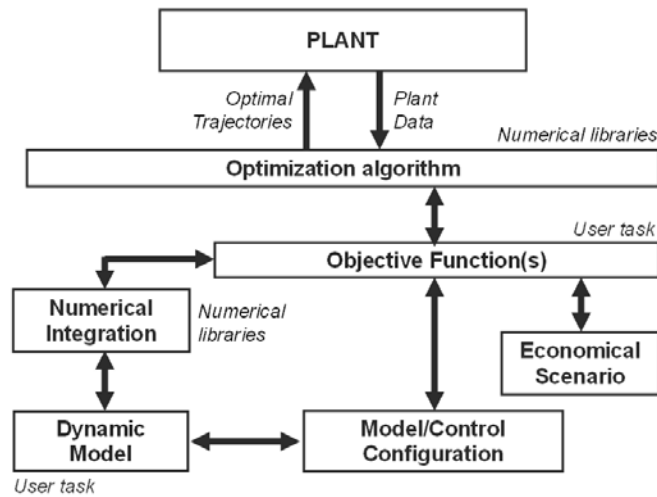
Figure 1: Architecture of nonlinear model predictive control

3.1. Objective Function

Many optimal control problems can be formulated as a minimization of a least squares objective function subject to equality/inequality constraints. NMPC enters this family and its generalized formulation is often the following one:

$$\min_{u(k)....u(k+h_c-1)} \left\{ \sum_{j=k+1}^{k+h_p} \omega_j \left(y_j - y_j^{set} \right)^2 + \sum_{l=k}^{k+h_p-1} \omega_{1,l} \left(u_l - u_l^{tar} \right)^2 + \sum_{l=k}^{k+h_p-1} \omega_{2,l} \left(u_l^{(i)} - u_l^{(i-1)} \right)^2 \right\}$$

subject to: (1)

$$\begin{cases} y_j^{min} \leq y_j \leq y_j^{max} \\ u_l^{min} \leq u_l \leq u_l^{max} \\ \Delta u_l^{(i),min} \leq \Delta u_l^{(i)} = u_l^{(i)} - u_l^{(i-1)} \leq \Delta u_l^{(i),max} \\ \textit{Convolution Model} \end{cases}$$

where $y_j - y_j^{set}$ is the deviation between the $j-th$ controlled variable and its setpoint; $u_l - u_l^{tar}$ is the deviation between the $l-th$ manipulated variable and its steady-state target; $u_l^{(i)} - u_l^{(i-1)}$ is the incremental variation between $i-th$ and $(i-1)-th$ time intervals of the $l-th$ manipulated variable; coefficients of the diagonal semipositive definite matrices ω are the weighting factors; *HP* is the prediction horizon; *HC* is the control horizon; min and max superscripts indicate lower and upper bounds, respectively, for both manipulated and controlled variables; and *convolution model* represents constraints dictated by mathematical model of the plant/process unit to be controlled.

3.2. The NMPC Algorithm and Preliminary Operations

The aforementioned formulation can be converted into an algorithm based on differential solvers and optimizers. Nevertheless, before initializing the NMPC, it is necessary to process raw measures coming from the plant so to detect possible outliers and to reconcile data (Bagajewicz, 2003; Manenti, 2009). To accomplish these preliminary operations, two sets of classes were developed: the first is to reconcile raw data set by basing on **QR** factorization and overdimensioned linear systems; the second is based on a novel technique to handle outliers (Buzzi-Ferraris and Manenti, 2010c). Reconciled data is then used to initialize NMPC structure: the optimizer is called the first time to evaluate the best manipulated variables **u** by minimizing the objective function (1). To do so, all constraints (including the differential system) are evaluated and an opportune differential solver should be invoked. The differential system is then integrated on a specific prediction horizon h_p to foresee the future behavior of the plant according to different values of **u**. After an iterative procedure, the optimal vector **u** is implemented in the plant and a new set of measures are acquired to restart the cycle.

3.3. Additional Constructors for the NMPC Class (Exploiting the Polymorphism)

It is well-known that a relevant advantage of object-oriented programming is the polymorphism that is the possibility to define more constructors for the same class so to have a single class that works as many functions with different arguments. Some additional constructors were developed to solve issues of the same family such as input–blocking, Δ–input blocking, and Δ–offset blocking (reduced nonlinear programming problems for the NMPC).

4. Class Validation

The class was validated on different case studies (ODE/DAE systems) already proposed in literature. Specifically, the polyethylene terephthalate model characterized by a band diagonal block matrix is reproposed here to validate the class (for more details, refer to Manenti and Rovaglio, 2008). Numerical results by the generalized class corroborate previous trends obtained by procedural structure. Fig. 2 shows trends of the polymer intrinsic viscosity (IV) of the polymer exiting the final crystallizer (PHCR) and the solid state polymerizer (SSP) during a grade change production. Trends obtained by the generalized class for NMPC are compared to the plantwide control consisting of conventional proportional-integral loops. Facing the well-known superior performances of NMPC, the proposed comparison validates the proposed generalized class.
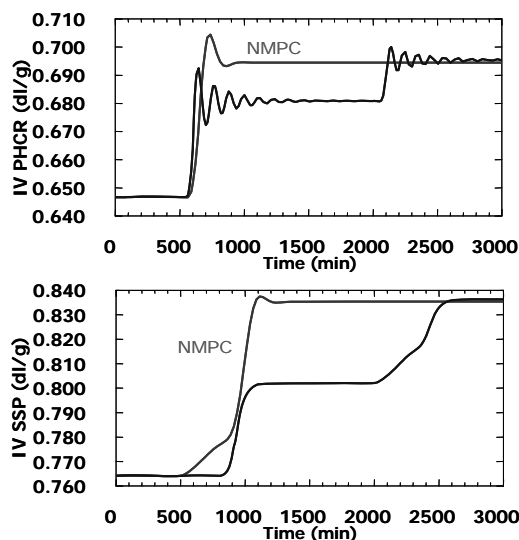
Generalized Classes for Lower Levels of Supply Chain Management: Object-Oriented Approach

Figure 2: PET plant: generalized class against plantwide conventional control

5. Conclusions

A generalized class for solving different problems belonging to the family of nonlinear model predictive control and dynamic optimization problems, based on BzzMath library, and exploiting multi-processor machines by means of OpenMP directives was proposed and validated on different case studies. Such a class allows the implementation of nonlinear NMPC once adequate objective function and differential system are both well-defined by the user. It can be easily used in FORTRAN and C++ code without the need to worry about differential solvers and optimization algorithms, since the object-oriented nature of the class as well as the philosophy adopted in BzzMath library synergistically drive the self-selection of algorithms to solve these issues.

Advanced users may also think to use commercial dynamic simulators such as DYNSIM, UNISIM, ASPENHYSYS… or any kind of home-made software to get a detail model of the process/plant and to implement the classes proposed here to solve model predictive control and dynamic optimization problems. In this specific case, the differential solver is the one included in the commercial package and not those implemented into the generalized class by limiting the class flexibility.

6. Acknowledgements

References

Bagajewicz, M.J. (2003). Data Reconciliation and Instrumentation Upgrade. Overview and Challenges. FOCAPO 2003. 4th International Conference of Computer-Aided Process Operations, Coral Springs, Florida, 103-116.
Bauer, M., I.K. Craig. (2008). Economic Assessment of Advanced Process Control - A Survey and Framework. Journal of Process Control, 18, 2-18.

Biegler, L.T. (2007). Large-scale Nonlinear Programming: an Integrating Framework for Enterprise-wide Dynamic Optimization. Proceedings of ESCAPE-17, Bucharest, Romania, 575-582.

Binder, T., L. Blank, H.G. Bock, R. Bulirsch, W. Dahmen, M. Diehl, T. Kronseder, W. Marquardt, J.P. Schloder, O. Van Stryk. (2001). Introduction to Model Based Optimization of Chemical Processes on Moving Horizons. M. Groetschel, S. O. Krumke and J. Rambau Eds., Online Optimization of Large Scale Systems: State of the Art.

Buzzi-Ferraris, G. (1967). Ottimizzazione di funzioni a più variabili. Nota I. Variabili non vincolate. Ing. Chim. It., 3, 101.

Buzzi-Ferraris, G. (2009). BzzMath: Numerical library in C++. Politecnico di Milano, http://chem.polimi.it/homes/gbuzzi.

Buzzi-Ferraris, G., F. Manenti. (2009). Kinetic models analysis. Chemical Engineering Science, 64(5), 1061-1074.

Buzzi-Ferraris, G., F. Manenti. (2010a). A Combination of Parallel Computing and Object-Oriented Programming to Improve Optimizer Robustness and Efficiency. Computer Aided Chemical Engineering, to appear.

Buzzi-Ferraris, G., F. Manenti. (2010b). Fundamentals and Linear Algebra for the Chemical Engineer Solving Numerical Problems. ISBN: 978-3-527-32552-8, WILEY-VCH, Germany.

Buzzi-Ferraris, G., F. Manenti. (2010c). Interpolation and Regression Models for the Chemical Engineer Solving Numerical Problems. ISBN: 978-3-527-32652-5, WILEY-VCH, Germany.

Cuoci, A., A. Frassoldati, G. Buzzi-Ferraris, T. Faravelli, E. Ranzi. (2007). The ignition, combustion and flame structure of carbon monoxide/hydrogen mixtures. Note 2: Fluid dynamics and kinetic aspects of syngas combustion. International Journal of Hydrogen Energy, 32(15), 3486-3500.

Lima, N.M.N., F. Manenti, R. Maciel Filho, M. Embiruçu, M.R. Wolf Maciel. (2009). Fuzzy Model-Based Predictive Hybrid Control of Polymerization Processes. Industrial & Engineering Chemistry Research, 48(18), 8542–8550.

Manenti, F. (2009). From reacting to predicting technologies: A novel performance monitoring technique based on detailed dynamic models. Chemical Product and Process Modeling, 4(2).

Manenti, F., G. Buzzi-Ferraris. (2009a). Criteria for Outliers Detection in Nonlinear Regression Problems. In J. Jezowski & J. Thullie (Eds.), Computer Aided Chemical Engineering (Vol. 26, pp. 913-917).

Manenti, F., G. Buzzi-Ferraris. (2009b). A New Strategy to Improve Parameters Estimation. Chemical Engineering Transactions, 17(3), 1335-1340.

Manenti, F., I. Dones, G. Buzzi-Ferraris, H.A. Preisig. (2009). Efficient Numerical Solver of Partially Structured Differential and Algebraic Equation Systems. Industrial & Engineering Chemistry Research, 48(22), 9979-9984.

Manenti, F., D. Manca. (2009). Transients modeling for enterprise-wide optimization: Generalized framework and industrial case study. Chemical Engineering Research and Design, 87(8), 1028-1036.

Manenti, F., M. Rovaglio. (2008). Integrated multilevel optimization in large-scale polyethylene terephthalate plants. Industrial and Engineering Chemistry Research, 47(1), 92-104.

Morari, M., J.H. Lee. (1999). Model predictive control: past, present and future. Computers & Chemical Engineering, 23(4-5), 667-682.

Qin, S.J., T.A. Badgwell. (2003). A survey of industrial model predictive control technology. Control Engineering Practice, 11(7), 733-764.

Signor, S., F. Manenti, M.G. Grottoli, P. Fabbri, S. Pierucci. (2010). Sulfur Recovery Units: Adaptive Simulation and Model Validation on Industrial Plant. Industrial & Engineering Chemistry Research, to appear.

Simulation-based Optimization Approach to Clinical Trial Supply Chain Management

Ye Chen*, Linas Mockus, Seza Orcun, Gintaras V. Reklaitis

Purdue University, West Lafayette, IN 47907, USA, chen231@purdue.edu

Abstract

The development activities required to bring a new drug to market involve considerable expense ($1+ Billion) and can take in excess of ten years. Clinical trials constitute a critically important and very expensive part of this development process as it encompasses producing, distributing and administering the candidate therapy to volunteer patients located in different geographic zones. A number of different approaches are being pursued to reduce clinical trial costs, including innovations in trial organization and patient pool selection. In this work, we focus our attention on improved management of the supply chain which provides the dosage required by the clinical sites. A simulation-based optimization approach is presented, which includes patient demand forecasting, mathematical programming based planning, and discrete event simulation. The objective is to enhance the robustness of the supply chain under different sources of uncertainties. A case study is reported which demonstrates the application of the proposed approach.

Keywords: Clinical Trial, Supply Chain, Optimization, MILP, Simulation

1. Introduction

New drug development follows an extended sequence of steps (discovery, animal trials, FDA application, product and process development, three phases of clinical trials, FDA filing and approval, and launch). As a result it takes many years and considerable expense ($1+ Billion) to bring a new drug to market. The clinical trials themselves constitute a very expensive part of this process. Normally, clinical trials with different test objectives (e.g. safety, efficacy, side effects) are conducted at the same time to expedite the new drug development process, which further complicates the clinical trial supply chain. While clinical trials are in progress, the development team also continues work towards improving the manufacturing processes.

The clinical trial material supply chain management problem is composed of the planning and scheduling of all transactions, operations and organizations during a trial, beginning with active ingredient manufacturing, followed by drug manufacturing and distribution to the clinical sites, and ending with dispensing the drugs to patients at each clinical site. A substantial amount of work has been reported on process industry supply chain optimization, but only a limited literature has addressed the issues faced in the pharmaceutical industry. Shah (2004) presented a review paper, categorizing previous work and analyzing the key issues for pharmaceutical supply chain optimization. There have been research activities on management of the product development pipeline, capacity planning, risk management, process development and plant design, as well as production planning and scheduling, but the issue of materials management for clinical trials has not been studied. Monkhouse et al (2006) discussed the design and development of clinical trials in some detail, but they provided little information about the actual management of the clinical trials material supply chain.

Traditionally, the pharmaceutical industry uses batch processes in the manufacture of pharmaceutical products both at the pilot and the commercial scale. Since these batch facilities are usually shared across various products, especially for the quantities needed for clinical trials, it is necessary to decide on the order and timing of the products to be produced. These decisions can have a large economic impact on the company at the clinical trials stage, because missing the delivery of trial dosage to patients can significantly delay completion of the trial and hence delay the time to market which in turn can mean significant loss of revenue. Deterministic mixed integer linear programs (MILP) and mixed integer nonlinear programming (MINLP) optimization methods have been proposed and used to solve resource constrained project planning and scheduling problem. Floudas and Lin (2004) presented a comprehensive review of these approaches. Most of the work reported is confined to a deterministic context. While some approaches have addressed uncertainties to generate robust schedules and plans, none of them are equipped to deal with the uncertainties faced in clinical trial supply chains.

The key technical challenges in managing a clinical trial materials supply chain are to meet the needs from clinical sites, so that patients are fully supplied once they are enrolled while minimizing oversupply since unused materials cannot be re-routed to other sites due to regulatory restrictions. Not only is patient enrollment highly variable, but uncertainties also arise in manufacturing and shipment lead times, in process failures and in production yields. Furthermore, the life of a clinical trial materials supply chain, which is around 1-2 years, is significantly shorter than that of a commercial supply chain, which usually exceeds 10 years. Therefore, the strategies utilized to buffer the uncertainties in commercial supply chains become ineffective as expected values cannot be effectively used as targets. Subramanian, Pekny & Reklaitis (2001) propose a computational architecture called "Sim-Opt", which combines mathematical programming and discrete event system simulation to assess the uncertainty and control the risk present in the new product development pipeline problem. Simulation-based optimization methods were found to be efficient and effective alternatives to solving a large stochastic decision problem. In this work, we propose a simulation-based optimization approach combining mathematical programming-based planning, and discrete event simulation to deal with our clinical trial materials supply chain management problem where uncertainties cannot be modeled analytically in a computationally tractable way.

2. Problem definition and assumptions

2.1 Multi-echelon production-distribution supply chain

The production of drug begins with active ingredient manufacturing (API), which normally involves either a series of chemical synthesis and separation processes, or fermentation and purification processes. The API is next converted to a new drug product (NDP) by adding "excipients" and conducting a series of additional processing steps, followed by packaging and labelling (PL) to obtain the final drug product form. In addition to the new drug product, a placebo (the product without the API) and a comparator (a form containing a commercial drug targeting the same disease) are also produced and used. To avoid psychological biases, the placebo and comparator undergo the same manufacturing, packaging and labelling stages as the target drug to make sure the appearance of these three types are the same to insure effectiveness in double blinded clinical trials. The finished drug product forms are shipped to various clinical sites worldwide. Therefore, a clinical trial materials supply chain can be treated as a

multi-echelon production/distribution supply chain including the API-NDP-PL manufacturing stages and the product distribution network.

For purposes of this study we assume there are no feed material constraints. The API, NDP and PL stages are conducted in the same facility and share the same inventory location in the US. Furthermore, shipment times between these three production stages are neglected. Since, compared to commercial drug manufacturing, the volume of drugs used in a clinical trial is small, we assume there is no inventory capacity limit. All finished drugs (target drug, placebo, and comparator) will be kept in the same distribution center with a certain shelf life, and must be disposed of after their expiration date. The distribution network starts at the US distribution center and covers various clinical sites used in the clinical trial located around the world.

2.2 Batch operation of manufacture process

Traditionally, the pharmaceutical industry uses the batch-campaign mode. In our models, the batch manufacturing process is described by campaign start time, number of batches in each campaign, batch size, batch processing time, drug type (target drug, placebo and comparator), and yield. Uncertainties exist in processing time and yield. Within each stage, there are multiple production lines of processing units working in parallel, and each production line could be utilized for different products. API stage only produces the active ingredient for the target drug of the trial, but there will be multiple product types at the NDP stage: target drugs at different dosage levels, placebo and comparator. Since the clinical trials will be conducted all over the world, drugs sent to a certain country should satisfy its country specific packaging and labelling requirements. Therefore, the number of stock keeping units (SKU) can grow significantly, depending on the design and topology of the clinical trial.

3. Simulation-based optimization approach

3.1 Computational framework

The framework proposed for this study consists of a simulation of demands (by forecasting methods), planning method, and a discrete event simulation for assessing the robustness of the supply chain under different sources of uncertainties as depicted in Fig. 1. The forecasting function uses a simulation model to determine the demand profile for each drug product. Given demand forecasts, a planning model is used to determine the manufacturing campaign details and shipping plans. The model is implemented as a Mixed-Integer-Linear-Programs (MILP) and solved using CPLEX. A simulation model of the entire supply chain, which is developed using the discrete event simulation software, ExtendSim, captures all activities, operations and processes involved in the clinical trial. The operational plans developed via the MILP planning models serve as drivers for the execution of supply chain simulation. The quality and robustness of the plans are assessed by replicated simulation runs. Upon convergence to appropriate statistical criteria, the supply chain performance is improved by adjusting the key system parameters and repeating the Simulation-Optimization cycle.
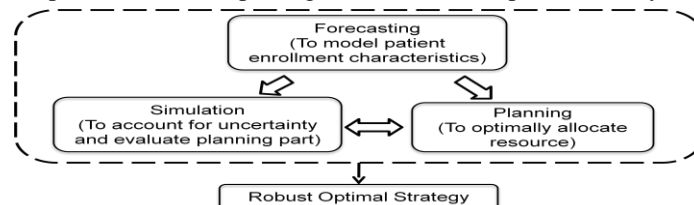

Fig. 1 Clinical trial supply chain management computational framework

3.2 Demand forecasting

The demand of each product, which is non-stationary, is obtained from detailed clinical site simulations. The arrival of patients can be treated as a Poisson process, and every patient is randomly assigned to different clinical trial dosages: target drug, placebo or comparator. During the treatment period, patients are required to follow pre-set visit profiles, which also determine the drug dispensation schedules. However, some patients may drop out during the course of the treatment for various reasons, such as loss of interest, dissatisfaction due to no observed improvement, or changes in personal life. The mean and variance of demand for each drug SKU are obtained from these simulations and are in turn used in the other supply chain decision models.

3.3 Planning

As noted above, the entire clinical trial materials supply chain is divided into the API, NDP, PL, and Distribution network components. Under typical industry practice, a global coordinator works within a decentralized control supply chain, which coordinates each stage towards to a common objective. The global objective of a clinical trial materials supply chain is to satisfy the patient demand with minimum cost. The downstream demands along with campaigning/shipping plans create the demands for the upstream stages in terms of material requirement.

Eqn. 1 and Eqn. 2 represent objective functions of the production and distribution sub-models, respectively. Each sub-problem seeks to minimize an objective function representing the total expected cost, consisting of several sub-problem specific cost factors. Demand data obtained from detailed patient enrollment forecasts and their simulations are aggregated into three discrete demand profile scenarios, each with certain probability. With distribution objective and constraints, an optimal shipment plan is obtained by formulating the distribution process as an MILP model solved by CPLEX. The shipment plans generate the demands for the manufacturing stages. Due to the space limitation the complete model equations are not presented herein.

Min *Cost = expected (Waste cost + Production cost + Holding cost)* (Eqn. 1)
Min *Cost = expected (Waste cost + Penalty cost + Fixed cost + Variable Cost)*
= (cost of destruction of material + cost of product + cost of packaging component)+
(Inventory opportunity cost) + (cost of direct labor for entering shipment +
cost of direct labor for processing shipment + cold chain container cost) +
(cost of direct labor of selecting and picking + shipment cost + container cost) (Eqn. 2)

3.4 Discrete event simulation

To investigate the quality of the plans generated, we represent each batch as a single transaction with specific properties such as start time, duration, batch type and size. Five simulation sub-models: API, NDP, PL, Distribution and clinical sites have been implemented. These models can be assembled to define any clinical trials supply chain. Within each sub-model, the batch is the flowing entity, moving through the network model. A batch waits for a specified period (could be sampled from a distribution or predefined as a property) of simulation time before proceeding to the next block. Also, this model dynamically communicates with decisions models through Excel files storing the manufacturing and distribution plans.

To capture the effects of uncertainties in this supply chain, the complete supply chain simulation is repeated many times for different sampled values of the uncertain parameters to generate the distribution data with which to verify and assess the efficiency and quality of the plans generated by the decision models. The simulation model records the number of missed patients, the number of patients who successfully

finished the treatment, the number of patients who drop out, and the average inventory at each clinical site and distribution center. The simulation results are used to restart the planning model to produce revised production and distribution plans. The planning and simulation loop is continued until the performance of the entire supply chain improves and converges to a satisfactory level.

4. Case study

The proposed approach is demonstrated by a case study outlined in this section. The topology of the case study is shown in Fig. 2. There is only one active ingredient produced in the API stage, but four SKU's need to be produced in the NDP stage: placebo, comparator, high dosage and low dosage target drug. Since this clinical trial will be conducted in two continents (US and European), two different types of packaging and labeling are used: one for the Americas (countries A and B) and the other for the European (countries C and D). Thus, there will be eight SKUs in the final distribution center to be shipped to various clinical sites. The shelf life of these drugs is 8 months, treatment lasts for 6 weeks, and the enrollment period of this clinical trial is 24 months. There are 75 clinical sites in total: 36% of them are in country A, 24% in country B, 21% in country C, and 19% in country D. Patients arriving at each clinical site will be assigned to take either placebo or high-dose target drug or low-dose target drug or comparator randomly following 1:2:2:2 enrolment ratio.
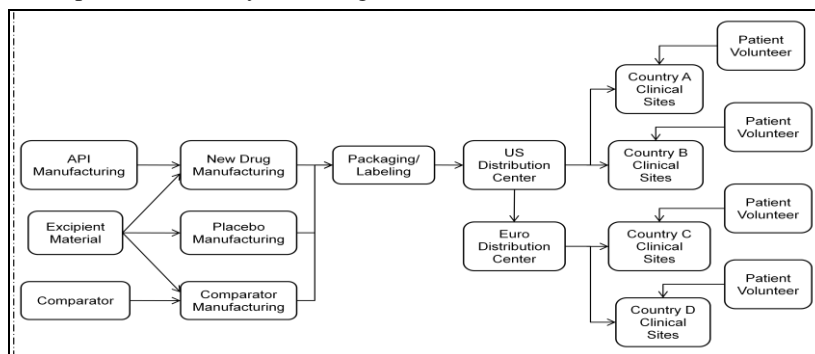

Fig. 2 Network of clinical trial supply chain case study

Fig. 3 is the demand profile obtained from the demand simulation (see section 2.3). The increasing nature of the demand is due to the fact that enrollment is low at the beginning since it takes time to generate patient awareness of this clinical trial. With advertisement more and more patients enroll to the clinical trial. However, the enrollment rate drops as the trial nears the end. The valley in the figure is as a result of a combination of factors such as promotional incentives offered and variability of the enrollment start in clinical sites. The dropout rate of patients is 45% in this scenario. A patient is missed if there are not enough drugs available in that clinical site at the time of the visit. The results of the approach described in section 3 are shown in Fig. 4 and Table 1. As Fig. 4 demonstrates the inventory profiles vary significantly over time.

5. Conclusion and future work

The clinical trial materials supply chain management problem is discussed and a simulation-based optimization approach, which combines stochastic mathematical planning with discrete event simulation, has been proposed. The quality and robustness of the plans generated by the planning model are assessed by replicated simulation runs.

We demonstrated our approach with a case study: a worldwide operated clinical trial materials supply chain management problem. The proposed approach yielded a production and distribution plan with 90% service level (Table 1). Also from the simulation, we can generate the inventory information at each clinical site. This information will be used in continuing research utilizing risk pooling strategies (e.g. Vidyarthi et al (2007)) to further mitigate the risks in clinical trials materials supply chain operation.
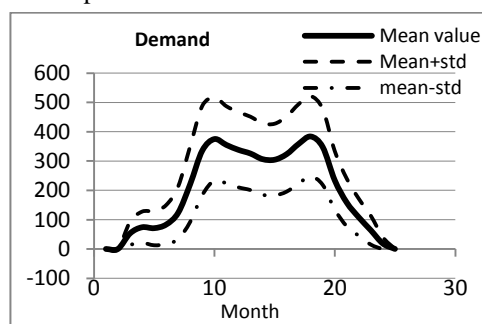

Fig. 3 Drug demand profile from simulation
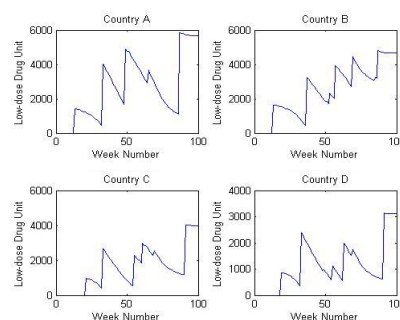
Fig. 4 Drug inventories of various countries

Table 1 Simulation Result

Patient Number		country A	country B	country C	country D
Placebo	missed	40	32	28	26
	dropped	410	314	275	245
	successful	495	377	331	294
Dose1	missed	129	93	121	136
	dropped	816	621	528	464
	successful	967	741	630	541
Dose2	missed	126	93	96	147
	dropped	813	619	541	457
	successful	969	739	643	530
Comparator	missed	84	55	58	54
	dropped	828	633	554	491
	successful	998	766	667	590

6. Acknowledgements

References

1. N. Shah, 2004, Pharmaceutical supply chains: key issues and strategies for optimization, Computers and Chemical Engineering, 28, 929–941.
2. D. C. Monkhouse, C. F. Carney, J. L. Clark, 2006, Drug Products for Clinical Trials, Informa Health Care.
3. C. A. Floudas, X. Lin, 2004, Continuous-time versus discrete-time approaches for scheduling of chemical processes: a review, Computers and Chemical Engineering, 28, 2109-2129.
4. N. Vindyarthi, E. Celebi, S. Elhedhili, E. Jewkes, 2007, Integrated Production-Inventory-Distribution System Design with Risk Pooling: Model Formulation and Heuristic Solution, Transportation Science, 41, 3, 392-408.
5. D. Subramanian, J. F. Pekny, G. V. Reklaitis, 2001, A Simulation-optimization Framework for Research and Development Pipeline Management, AIChE Journal, 47(10), 2226-2242.

Risk Management in Production Planning under Uncertainty by Multi-Objective Hybrid Evolutionary Algorithms

Thomas Tometzki[a] and Sebastian Engell[a]

[a]*Process Dynamics and Operations Group, Technische Universität Dortmund, 44221 Dortmund, Germany,{thomas.tometzki | sebastian.engell}@bci.tu-dortmund.de*

Abstract

We consider production planning problems with uncertainties in the problem data. The optimization problems are formulated as two-stage stochastic mixed-integer models in which some of the decisions (first-stage) have to be made under uncertainty and the remaining decisions (second-stage) can be made after the realization of the uncertain parameters. The uncertain model parameters are represented by a finite set of scenarios. The production planning problem under uncertainty is solved by a stage decomposition approach using a multi-objective evolutionary algorithm which takes the expected scenario costs and a risk criterion into account to compute the first-stage variables. The second-stage scenario decisions are handled by mathematical programming. Results from numerical experiments for a multi-product batch plant are presented.
Keywords: Risk conscious planning, two-stage mixed-integer programming, stage decomposition, hybrid algorithm, multi-objective evolutionary algorithm

1. Introduction

In production planning problems, a large number of decisions have to be made in short time and under significant uncertainties. Predictions about the evolution of the demands, the availability of the processing units and the performance of the processes are necessarily based on incomplete data. Resource assignment decisions must be made at a given point of time despite the fact that their future effects can not be foreseen exactly. The existing approaches to address planning under uncertainty can be classified into reactive approaches and stochastic approaches. The former use deterministic models and modify a nominal decision if an unexpected event occurs, whereas the latter approaches include descriptions of the the uncertainty in the models. A recent overview of relevant solution techniques for the class of stochastic approaches was provided in [1,2]. Handling of uncertainties in the dynamic scheduling context is discussed in [3].

The focus of this work is on the solution of two-stage stochastic mixed-integer problems. They are solved by a stage decomposition based hybrid evolutionary approach which was published first in [4]. For two-stage stochastic mixed-integer programs with a large number of scenarios, or when good solutions are needed quickly, the hybrid approach can provide better results than the formulation and solution of monolithic large MILPs [5]. The solution of two-stage stochastic mixed-integer programs in [4, 5, 6] aims at maximizing the expected profit. But plant managers frequently also try to avoid the occurrence of very unfavorable situations, e.g. heavy losses. Naturally, they aim at a compromise between expected profit and accepted risk. Using the scenario based two-stage stochastic approach the risk can be controlled. This contribution introduces a hybrid multi-objective evolutionary approach to two-stage stochastic mixed-integer problems with additional risk objectives.

2. Two-stage stochastic mixed-integer programs

A two-stage stochastic mixed-integer program is used to model uncertainties in problem data. It is assumed that the uncertainties have a finite number of realizations that can be modeled by a discrete set of scenarios $\omega = 1, \ldots, \Omega$. The decisions are divided into the first-stage decisions $\mathbf{x}$ which have to be taken before the uncertainty is disclosed and second-stage decisions $\mathbf{y}_\omega$, which have to be taken after the uncertainty is realized. In this paper, we consider two-stage stochastic mixed-integer programs that involve linear constraints on the first- and second-stage decisions and linear disjunctions on the first-stage decisions which explicitly model operational constraints:

$$\min_{\mathbf{x}, \mathbf{y}_1, \ldots, \mathbf{y}_\Omega} \mathbf{c}^\mathrm{T}\mathbf{x} + \sum_{\omega=1}^{\Omega} \pi_\omega \mathbf{q}_\omega^\mathrm{T}\mathbf{y}_\omega \tag{1}$$

$$\text{s.t.} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b} \tag{2}$$

$$\mathbf{W}_\omega \mathbf{y}_\omega \leq \mathbf{h}_\omega - \mathbf{T}_\omega \mathbf{x} \tag{3}$$

$$\mathbf{x} \in X, \mathbf{y}_\omega \in Y, \forall \omega = 1, \ldots, \Omega.$$

The objective of the problem (1) consists of the first-stage costs and the expected value of the second stage costs. The costs are calculated as linear functions of the first-stage variables $\mathbf{x}$ and the second-stage variables $\mathbf{y}_\omega$ with vectors of cost parameters $\mathbf{c}$ and $\mathbf{q}_\omega$. The two-stage model consists of inequality constraints in both stages (2, 3). The finite sets X and Y may contain integrality requirements.

3. Stage decomposition based hybrid evolutionary approach

The main idea of stage decomposition is to remove the ties between the second-stage scenario subproblems by fixing the first-stage decisions. The scenario subproblems are of significantly smaller size than the full two-stage problem. The master problem is a function of the vector of first-stage variables $\mathbf{x}$ only:

$$\min_{\mathbf{x}} f(\mathbf{x}) = \mathbf{c}^\mathrm{T}\mathbf{x} + \Phi(\mathbf{x}) \tag{4}$$

$$\text{s.t.} \quad \mathbf{A}\mathbf{x} \leq \mathbf{b}, \mathbf{x} \in X \tag{5}$$

The second-stage value function $\Phi(\mathbf{x})$ for a first-stage decision $\mathbf{x}$ is given by the expected value of the Ω independent second-stage functions $Q_\omega(\mathbf{x})$:

$$\Phi(\mathbf{x}) = \sum_{\omega=1}^{\Omega} \pi_\omega Q_\omega(\mathbf{x}) \tag{7}$$

The evaluation of $\Phi(\mathbf{x})$ requires the solution of Ω subproblems over the second-stage variables $\mathbf{y}_\omega$:

$$Q_\omega(\mathbf{x}) = \min_{\mathbf{y}_\omega} \mathbf{q}_\omega^\mathrm{T}\mathbf{y}_\omega \tag{8}$$

$$\text{s.t.} \quad \mathbf{W}_\omega \mathbf{y}_\omega \leq \mathbf{h}_\omega - \mathbf{T}_\omega \mathbf{x}, \mathbf{y}_\omega \in Y, \forall \omega = 1, \ldots, \Omega. \tag{9}$$

The constraints of the master problem (4 - 5) are scenario independent, while the parameters of the second-stage problems in (8 - 9) may vary from scenario to scenario. The vector of the first-stage variables $\mathbf{x}$ appears as a vector of fixed parameters in the constraints of the second-stage scenario problems. The challenge of the master problem is that $\Phi(\mathbf{x})$ in general is discontinuous and non-convex due to integrality requirements and the minimization in the second stage. Additionally, first-stage feasible solutions do not necessarily have feasible solutions in the second-stage due to the implicit constraints. The main algorithmic idea of the hybrid evolutionary approach is to address

the master problem given by (4 - 5) by an evolutionary algorithm. To evaluate f(**x**), the Ω subproblems given by (8 - 9) are solved independently by a MILP solver.

4. Risk Management under Uncertainty

The two-stage stochastic optimization approach described above accounts for uncertainty by optimizing the expected profit without reflecting and controlling the variability of the performance associated with each specific scenario. Thus, there is no guarantee that the process will perform at a certain level for all uncertain scenarios. However, for the solution with the best expected cost there may exist scenarios with poor outcomes, i.e. high costs. From an economic point of view a high economic loss or other disadvantageous outcomes should be avoided. The measure of the occurrence of such disadvantageous events or their degree of damage is termed *risk*. For given first-stage variables **x**, the scenario costs are random variables, thus the consequences of a decision are given by the distribution of the scenario costs and can be graded according to various risk measures.

   Incorporation of the trade-off between risk and profit leads to a multi-objective optimization problem in which the expected performance and the risk measure are the two objectives. Different criteria for assessing risk have been proposed in the literature [7]. The standard deviation for a given set of scenarios is one of the metrics commonly used for quantifying variability. Alternative approaches for integrating risk have been proposed, e.g. the value at risk (VaR), the conditional value at risk (CVaR) and the worst case performance. The risk conscious criteria are expressed in the two-stage stochastic program by a second master problem $\min_x$ r(**x**). The function r(**x**) is determined by the formal definition of the risk function which is based on the scenario cost values $\pi_\omega(\mathbf{c}^T\mathbf{x} + Q_\omega(\mathbf{x}))$, $\omega = 1,\ldots,\Omega$. Risk measures used in this work are:

- **Standard deviation:** The standard deviation is a measure of how broad the distribution of the scenario costs is. It reflects the chance that the actual costs may differ largely from the expected costs.
- **Value at risk (VaR$_\alpha$):** For a given scenario cost distribution and a confidence level α, the value of VaR$_\alpha$ is the cost of the most favorable scenario of the $(1-\alpha)\cdot100\%$ most unfavorable scenarios.
- **Conditional value at risk (CVaR$_\alpha$):** For a given scenario cost distribution and a confidence level α, the value of CVaR$_\alpha$ is the mean cost of the $(1-\alpha)\cdot100\%$ most unfavorable scenarios.
- **Worst-case cost:** Cost of the scenario with the worst performance. A major difference with respect to other risk measures is that the probability information is not used.

5. Multi-Objective Evolutionary Approach

In optimization problems with multiple objectives in general no solution exists for which all objectives are optimal. Therefore the goal of multi-objective optimization is to compute the set of the Pareto-optimal solutions. This is a set of solutions, where no improvement in one objective can be achieved without downgrading another objective. The attractive feature of multi-objective evolutionary algorithms (MO-EA) is their ability to find a set of non-dominated solutions close to the Pareto-optimal solutions. Instead of using classical multi-objective optimization approaches (i.e. weighted sum approach, ε-constraint method) which convert a multi-objective optimization problem into a single-objective optimization problem, the evolutionary approach finds a number of trade-off solutions in one single optimization run (for an overview see [8]).

The two objectives f(**x)** and r(**x)** are addressed by a MO-EA. In this contribution, an integer evolutionary algorithm is used. The selection is adapted to the multi-objective environment by using the elitist non-dominated sorting concept (NSGA-II) of [9].

5.1. Representation and Initialization

Each individual of the population represents a search point $\mathbf{x}_k = (x_1, \ldots, x_n)$ by its object parameters, and mutation strength parameters $\mathbf{s}_k = (s_1, \ldots, s_n)$ which affect the mutation operator. A population of μ individuals is initialized randomly within the bounds of the box-constrained first-stage decision space $\mathbf{x}_{min} \leq \mathbf{x} \leq \mathbf{x}_{max}$. The mutation strength parameters are initialized randomly in the range of the corresponding object parameter bounds $\mathbf{1} \leq \mathbf{s} \leq \mathbf{x}_{max} - \mathbf{x}_{min}$.

5.2. Evaluation

After each generation, an evaluation of the individuals is performed. For first-stage feasible solutions **x** the Ω scenario subproblems are solved independetly by a MILP solver. After this both fitness values f(**x)** and r(**x)** are calculated. If the first-stage constraints $\mathbf{Ax} \leq \mathbf{b}$ are not satisfied, the fitness functions f(**x)** and r(**x)** are replaced by the penalty function $g(\mathbf{x}) + f_{max}$ which is the sum of first-stage constraint violations $g(\mathbf{x}) = \sum_j (\mathbf{A}^j \mathbf{x} - \mathbf{b}^j)$ and an upper bound f_{max} of f(**x)** for feasible solutions **x**. Due to this choice feasible solutions are always preferred over infeasible solutions.

5.3. Mutation

After the evaluation, for each subproblem λ offspring are generated by λ-fold application of the mutation operator. It perturbs each variable x_i by a random number drawn from a normal distribution with an expected value of zero. For integer variables, the random numbers are rounded to the nearest integer value. The distribution variance depends on parameter s_i which is modified log-normally. To maintain the bounds for x_i, values outside the bounds are mapped onto the next bound.

5.4. Selection for population replacement

The non-domination selection chooses the μ best ($1 \leq \mu \leq \lambda$) individuals out of the union of μ parents and λ offspring which do not exceed the maximum age of κ for the next iteration loop. First, the populations of the μ parents and the λ offspring are combined. If the age of an individual equals κ, this individual is not further considered in the selection. Then the entire population is sorted into different front sets based on non-domination. All non-dominated solutions in the population are assigned to the first front F_1. The non-dominated individuals of the remaining population are then assigned to the second front F_2. This procedure continues until all population members are classified into fronts. The new population for the next generation is successively filled up with individuals from the fronts starting with the first front F_1. This procedure continues until the individuals of a front F_j can no longer be accommodated in the new population. To choose exactly μ individuals, the solutions of the front F_j are sorted using a crowded-distance comparison operator [9] in descending order and the best solutions are used to fill up the new population. After a new population of μ individuals is generated, the age of the individuals is increased by one and a new iteration loop starts if the termination criterion is not fulfilled.

6. Numerical Study

The performance of the hybrid multi-objective evolutionary approach is evaluated by the quality of the non-dominated solutions. Convergence comparisons to other approaches or statistical analysis of the random behavior of the algorithm are outside the scope of this contribution.

Risk Management in Production Planning under Uncertainty by Multi-Objective Hybrid Evolutionary Algorithms

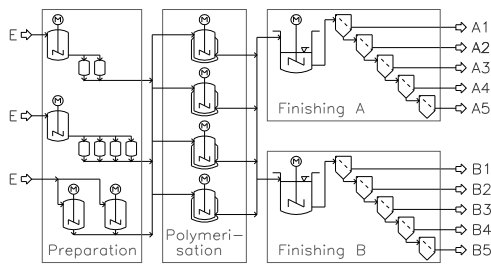### 6.1. Production Planning Example

Fig. 2: The flow sheet of the multi-product batch plant.

Fig. 2 shows the layout of a multi-product batch plant for the production of expandable polystyrene (EPS) [5, 6]. Two types A and B of the polymer in five grain size fractions are produced from raw materials E. The preparation stage is not considered here. The polymerization stage is operated in batch mode and is controlled by ten recipes. Each recipe defines the product (A or B) and its grain size distribution. Each batch yields a main product and four coupled products. The capacity of the polymerization stage constrains the number of batches to 12 in each two-day period. The batches are transferred into two semi-continuously operated finishing lines which fractionate the grain sizes. The capacity of each finishing line is between 5 and 12 batches per period in case it is operated, and 0 otherwise. The operation mode can be changed every second period. The planning decisions which have to be made are *batching decisions* on the numbers of polymerizations of each EPS-recipe in each period. The decisions in periods 1 to 3 are considered as first-stage, those in periods 4 and 5 as second-stage decisions. The uncertainty in the demands is represented by 64 scenarios of equal probability such that the expected average total product demand in each period is between 75% and 100% of the capacity of the polymerization stage. The profit is calculated from sales revenues, production costs, storage costs, and penalties for lateness and for finishing line start-ups and shut-downs. The full mathematical description of the mixed-integer model can be found in [6].

6.2. Computational results

The MO-EA was implemented in MATLAB 7.3. All MILPs were solved using CPLEX 10.2. The algebraic models to be solved by CPLEX were formulated using GAMS distribution 22.5. The computational equipment for all the experiments performed was a dual Xeon machine with 3 GHz speed, 1.5 GB of main memory with Linux operating system. For all experiments the calculation time was limited to 4 CPU-hours per setting. All parameters are chosen as in [5].

The plots in Fig. 3 show the results obtained by the MO-EA for different risk measures r(x) in the order: standard deviation of scenario costs, worst scenario costs, VaR_α (for $\alpha = 0.6$, 0.8, 0.9, and 0.95), and $CVaR_\alpha$ (for $\alpha = 0.6$, 0.8, 0.9, and 0.95). The results for the minimization of the standard deviation show that the MO-EA was able to generate a relatively large number of non-dominated solutions which are disconnected to two sets. The results for the optimization of the worst-case scenario costs, VaR_α and $CVaR_\alpha$ show smaller non-dominated solutions sets. In the lower plots, the results shown for VaR_α and $CVaR_\alpha$ are quite similar. The number of non-dominated solutions in the sets increases with the α-value. The best solutions obtained by CPLEX for the single-objective problems are only slightly better than the best expected costs obtained by the MO-EA for the multi-objective problems. This indicates that the MO-EA approach is able to find solutions for the multi-objective problem which are of similar quality as the CPLEX solutions for the single-objective problem. On the other side, however, since the values obtained by CPLEX are not reached by the MO-EA, the exact Pareto-optimal solutions are not provided by the MO-EA.
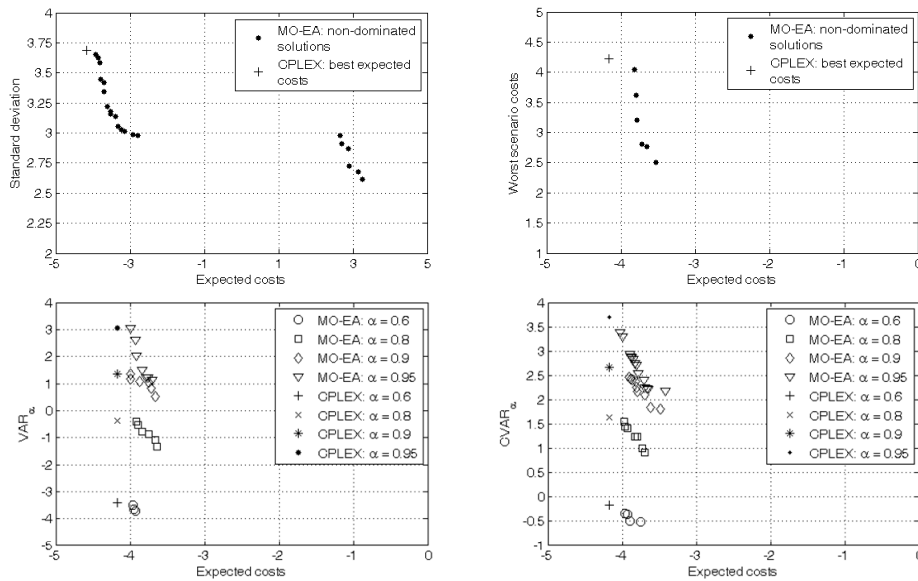
Fig. 3: Non-dominated solutions from the multi-objective optimization of the expected costs and
different risk measures and best solutions obtained by CPLEX for the expected costs problem.

7. Conclusion

The paper describes the solution of risk conscious planning problems under uncertainty
by a multi-objective evolutionary algorithm combined with MILP solutions of the
scenario subproblems. The proposed approach reflects risk conscious decisions under
uncertainty. The results from the case study show that its application may have a great
practical benefit on the risk conscious planning under uncertainty. The decision maker
gets a set of solutions among which he can choose according to his risk aversion.

References

[1]   A. Ruszczynski and A. Shapiro, editors. Stochastic Programming. Handbooks in
Operations Research and Management Science. Elsevier, Amsterdam, The Netherlands, 2003.
[2]   Z. Li and M. Ierapetritou. Process scheduling under uncertainty: Review and challenges.
Computers and Chemical Engineering (CACE), 32:715–727, 2008.
[3]   Engell, Sebastian. Uncertainty, decomposition and feedback in batch production
scheduling. In Proc. ESCAPE-19, pages 43-62, 2009.
[4]   J. Till, G. Sand, S. Engell, M. Emmerich, and L. Schönemann. A hybrid algorithm for
solving two-stage stochastic integer problems by combining evolutionary algorithms and
mathematical programming methods. In Proc. ESCAPE-15, pages 187–192, 2005.
[5]   T. Tometzki and S. Engell. Hybrid evolutionary optimization of two-stage stochastic
integer programming problems: An empirical investigation, Evolutionary Computation, 17 (4):
511-526, 2009.
[6]   J. Till, G. Sand, M. Urselmann, and S. Engell. Hybrid evolutionary algorithms for solving
two-stage stochastic integer programs in chemical batch scheduling. CACE, 31: 630–647, 2007.
[7]   A. Bonfill, M. Bagajewicz, A. Espuna, and L. Puigjaner. Risk management in the
scheduling of batch plants under uncertain market demand. Industrial and Engineering
Chemistry Research, 43: 741–750, 2004.
[8]   K. Deb. Multi-Objective Optimization using Evolutionary Algorithms. Wiley-Interscience
Series in Systems and Optimization. John Wiley & Sons, Chichester, 2001.
[9]   K. Deb, A. Pratap, S. Agarwal, and T. Meyarivan. A fast and elitist multiobjective genetic
algorithm: NSGA-II. IEEE Transactions on Evolutionary Computation, 6 (2):182–197, 2002.

Risk Management Framework for the Petroleum Supply Chain

Leão J. Fernandes,[a,b], Ana Paula Barbosa-Póvoa,[b*] Susana Relvas, [b]

[a]CLC, EN 366, Km 18, 2050 Aveiras de Cima, Portugal, leao.fernandes@clc.pt
[b]CEG-IST, UTL, Av.Rovisco Pais, 1049-001 Lisboa, Portugal, susanaicr@ist.utl.pt
[*]Corresponding author: apovoa@ist.utl.pt

Abstract

Complex plants and investments are commonplace in the Petroleum Supply Chain (PSC), known for its highly automated infrastructures and processes. Expensive equipment items like drilling rigs, offshore platforms, oil tankers, refineries, pipelines, petroleum depots and transport equipment are critical to this industry. The petroleum supply chain appears as a significant risk and high impact industry at the micro and macro economic level. Although, risk management bears prime importance for this industry, there is notorious absence of quantitative modeling. This paper introduces the relevance of a systematic approach for the identification, quantification and mitigation of risk and presents a practical framework for risk management. A PSC example is used to demonstrate its utilization and the resulting information identifies modeling data for a PSC risk management tool.

Keywords: petroleum, risk management, uncertainty, framework, roadmap

1. Introduction

The Petroleum Supply Chain (PSC) is a complex assortment of infrastructures and processes whose mainstream begins with the exploration of crude oil and finalizes with the delivery of petroleum products to consumers. This industry moves huge quantities of products and value and is backbone to almost all economic activity. This strategic sector is highly automated and optimized, so disruptions can rapidly escalate to an industry-wide or nation-wide crisis. Oil companies, aware of these risks, have put significant effort in Risk Management, however most of the work is qualitative and is still at the initial stage. Nevertheless, some advances have been done in quantitative risk management for pipeline integrity, Muhlbauer (2004) and Alvino (2003). Besides, considerable research can be observed in Supply Chain Risk Management (SCRM) which has led to the publication of important reviews. Categorization of these developments can be found in Tang (2006) and Peidro et al. (2008). However there is no direct method to identify possible uncertainties, risks and mitigation strategies for a particular situation.

This investigation builds on earlier research and constructs a framework that is then tested specifically for the petroleum supply chain. This framework appears as a practical method that assists in structuring the activities and the information of the risk management process. The following sections describe the petroleum supply chain, present the developed framework for risk management, demonstrate its utilization using a PSC real case, identify modeling directions and finally present the conclusions and proposals for future research.

2. Problem Statement and Background

Fig. 1 resumes the petroleum supply chain that divides into two major areas: upstream and downstream. The upstream comprises of crude oil exploration, production and transportation. The downstream industry involves product refining, transport, storage, distribution and retail. These are major activities, which aggregate several hundreds of processes and thousands of equipment items where availability is of paramount importance.

Figure 1: The Petroleum Supply Chain

The PSC activities are sequential in nature and as such any failure is critical to the next stage and more so as this implies huge working capital that is blocked in petroleum inventories. The prevalent risks in business, operations, finance, environment, safety and security provide a huge potential for risk optimization. The problem scope will focus on risk management for the PSC. Hence the problem frontiers include processes, equipment items, activities and costs while the main drivers are risk sources, impacts and mitigation strategies.

3. A Hierarchical Framework

Literature on PSC risk management is mainly confined to qualitative approaches concerning the process of risk analysis and assessment, thereby exposing an absence of quantitative modeling. This lack of risk structuring and breakdown methodology provides the motivation to develop a framework that could provide a structured method for the risk identification, quantification and mitigation process. Research on SCRM and PSC literature and investigation on PSC risk management (Fernandes et al., 2009) has lead to the development of a simple framework that assists in capturing and building quantitative data through a well defined risk management process. Fig. 2 presents a new risk management hierarchical framework. The framework builds an information model using two processes: Risk identification process and the Risk mitigation process.

The risk identification process utilizes the framework to identify and hierarchically relate first the risk agents, second the risk sources, the risk objects and finally the risk events. Risk elements are identified by asking the following questions: Who initiates the risk? Risk agent; What are the causes of the risk? Risk source; Which resources are affected? Risk object; and How does the risk manifest? Risk event. More specifically, risk agents are the drivers of organizational risks, for example the finance area or the transport activity. Risk sources are the causes of risks, which provide a negative impetus to the risk objects thereby generating a risk. An example of a risk source is the increase in value added tax. Risk objects are resources of an organization, whose malfunctioning would originate a consequence. An example of risk objects is the country tax structure. Risk events are the factual occurrence of the risk thereby resulting in the effective loss, for instance reduced profits. In a nutshell, the framework indicates

that the *risk agent* (financial area) includes a *risk source* (value added tax) which can affect the *risk object* (tax structure) thereby generating a *risk event* (reduced profit).

Figure 2: Risk management hierarchical framework

The above risk identification process culminates into the identification or computation of the *consequence* estimated for each risk agent/source/object/event combination. The framework follows the risk mitigation process to complete the risk information. Each risk agent initiates a hierarchical identification of a planning level and an appropriate mitigation strategy that could reduce the potential risks of the risk sources. *Planning level* provides a timeframe for the mitigation activity, which could be Strategic or long-term, Tactical in the case of mid-term, Operational or short-term planning and Contingencial or post-occurrence planning. *Mitigation strategies* are counter measures that could reduce the likelihood and the consequences of the risk events triggered by the risk sources. Expected *payoff* estimates should be computed for the combination risk source/mitigation strategy to provide the quantitative data.

Fig. 3 presents an influence diagram of the proposed risk management framework. The circular chance nodes represent the uncertain events and the square deterministic nodes represent the decision events. Finally, the diamond-shaped element depicts the result or the expected payoff for the enterprise. Hence, the risk agent stochastically influences the risk source and the planning level, which consequently direct the mitigation strategy. The risk source stochastically determines the risk objects affected which consequently undermine the occurrence of a specific risk event. The mitigation strategy influences the risk object and the risk event thereby affecting the risk outcome.
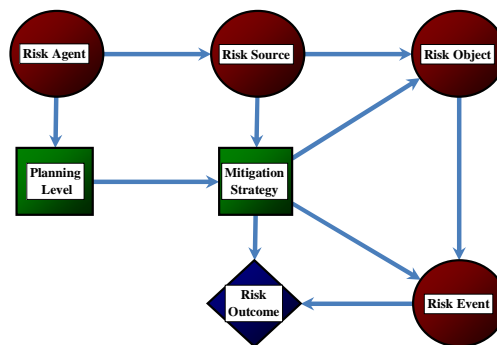

Figure 3: Risk management framework influence diagram

The risk elements are presented in their aggregated form, however these elements can be further decomposed, based on their importance and the depth of quantitative modeling required. Hence a *risk agent α* (Operation) could be sub-divided into *risk agents α1, α2 and α3*, more precisely Transport, Storage and Production operations. The transport operation *α1* can be further divided into pure products transfer *α11* and

interface products transfer *α12*. The pure product transfer operation *α11* can be further sub-divided into *α111..α115,* namely, Butane, Propane, Diesel, Gasoline and Jet transfer operations and so on. The same theory can be applied to *risk sources*. Analogously, the *risk object* storage can be decomposed into individual product subsystem. Each product subsystem can be decomposed into the dynamic, static and instrumentation equipment. Static equipment could divide further into tubes, tanks and spheres. Finally *risk events* can also be hierarchised, for instance BLEVE(boiling liquid expanding vapor explosion) can be subdivided in accordance to its intensity, duration and/or time of occurrence.

The ongoing demonstrates that the developed framework supports a flexible and interactive process for risk identification and risk mitigation information gathering that could be used to construct a holistic risk management decision tree or a decision matrix. The following section provides a demonstrative example of the construction of a detailed risk management decision tree for the PSC using the RM framework.

3.1. A PSC-RM example

As mentioned earlier, the hierarchical framework could be used to construct a decision tree to guide building of an information database that could drive a quantitative model such as a mathematical model to optimize the risk management process. A real case example is used to demonstrate the building of a decision tree using the framework. Companhia Logística de Combustíveis (CLC) is a strategic lean member of the Portuguese petroleum supply chain which owns and operates a petroleum products pipeline and a storage and expedition infrastructure.

Some risks of this organization are visited, though not in a detailed form due to lack of space, using the simplified decision tree presented in Fig. 4 that was built using the RM framework. Although the strategic, tactical, operational and contingencial *planning levels are* considered, the focus is on the operational risk management. Operational risks are seen to stem from various risk agents in the PSC, which could generally arise form the Business, Condition, Operations, Hazard and Finance agents. Each of these risk agents can be subdivided to observe a detailed view. For instance, the risk agent of condition could be subdivided into pipeline, storage and bottle filling conditions. Further exemplification concentrates on the pipeline operation sub-tree.

The pipeline operations are decomposed as pure products and interface operations. Pure products include the diesel transfer operation where risk sources such as third-parties, construction, corrosion, ground movement and operator errors are identified. Focusing on the corrosion risk source, two mitigation strategies are identified, namely product buffering and risk based inspection. Corrosion affects various risk objects including inventory, sales, pump station and transport duct resulting in risk events like ruptures, holes in pipe, cracks and coating damages. These risk events are categorized as ignition, no ignition, reported and unreported and again subdivided as detonation, high thermal damage, torch fire, product spill, corrosion leaks and no leaks. As statistical correlation is observed between the risk elements, a detailed classification although industry dependent, is crucial to risk quantification. Bayesian theory can be applied to corrosion determining factors like product, soil, construction material, protection and age to estimate the probability of the possible risk events and outcomes referred earlier.

The decision tree in Fig. 4 provides a risk profile of a petroleum tank farm and pipeline company using the new risk management framework. The risk probabilities and consequence costs are calculated, based on the literature and empirical estimates in the petroleum sector. Costs include lost sales, product, equipment, environmental fines, repairs and casualties. For generalization purpose, the monetary unit (m.u.) used equates

Risk Management Framework for the Petroleum Supply Chain

to one day's EBITDA of the enterprise. This quantification is yet at its early stage and requires further research. Parameter estimation will be focused in future investigation to obtain robust generalized measures for risk probabilities, mitigation and consequences.
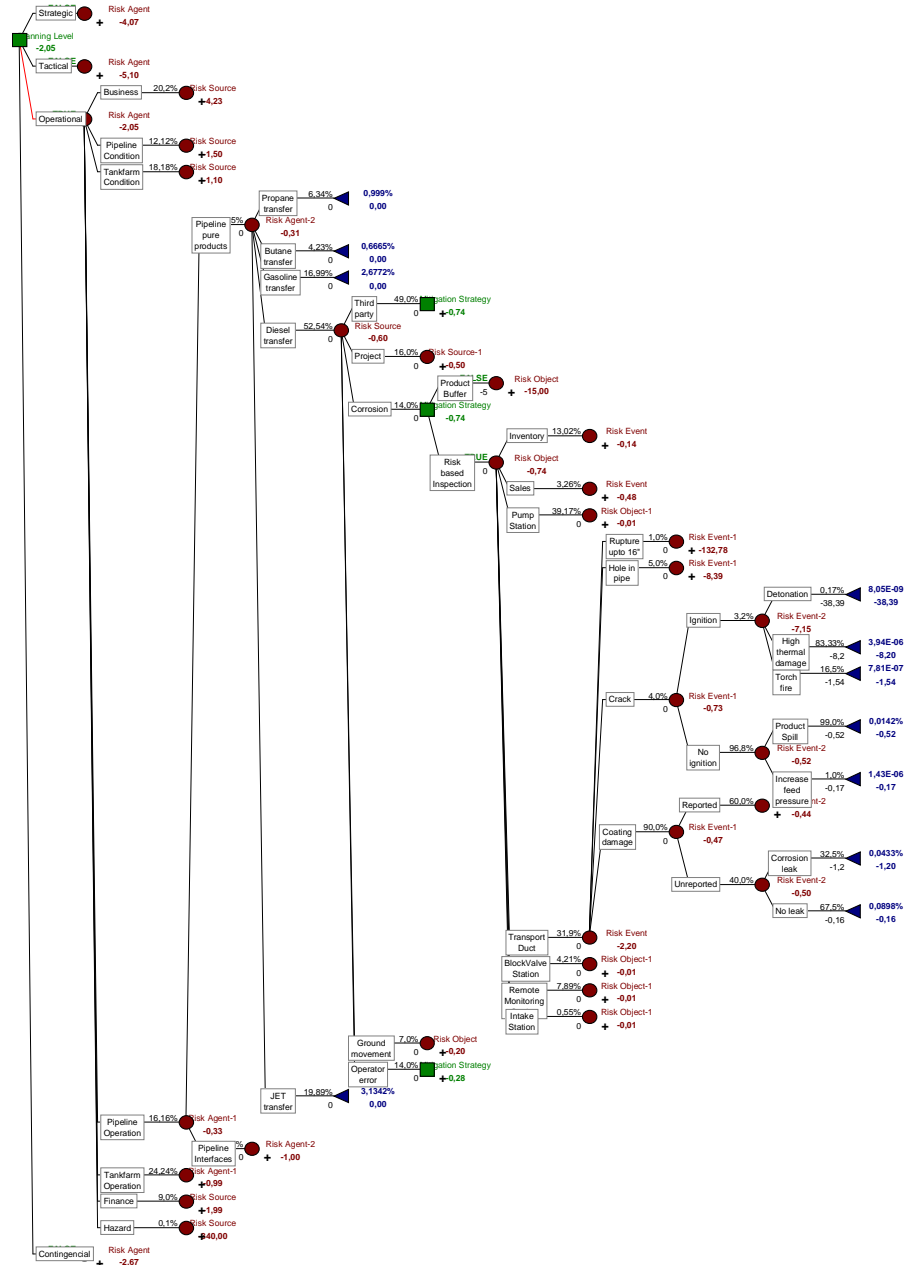

Figure 4: Framework decision tree for the Petroleum Supply Chain

4. Modeling approaches

The framework presented in earlier sections assists us in structuring the risk sources and mitigation information. Nevertheless the development of a risk management model is of

prime importance in order to generate maximum potential from this information. The framework may require enhancement to feed appropriate modeling techniques which in turn should be selected taking the framework into consideration.

Various modeling approaches can be found in the literature that model supply chain and risk management problems. Stochastic dynamic programming, two-stage multi-objective modeling, decomposition models, mixed integer linear and non-linear programming, scenario simulation, genetic algorithms, decision tree analysis, and agent-based and artificial intelligence are seen to have potential for risk modeling. More specifically the agent-based modeling used in Adhitya et al. (2007), the multiobjective model used in You et al. (2009) and the model predictive control from Puigjaner (2008) have special interest for the PSC risk management due to their proved applicability in the process industry.

The presented work will evolve from the translation of the presented framework into a quantitative model approach. Further investigation and testing is required to develop this framework methodology into a functional model.

5. Conclusions and future work

The current investigation adds to earlier research on methodologies and models on supply chain risk management which now results in the development of a new SCRM framework. The framework is used to define a decision tree for the deployment of a risk methodology. The framework is demonstrated using typical risk and mitigation scenarios from the petroleum supply chain. Potential modeling directions are considered for the quantitative risk management model.

Future research will focus on building an elaborate risk profile for the PSC. Major impact agents such as business, operations and finance risk areas will be targeted. This effort is directed to developing and case testing an integrated model for risk management in the petroleum supply chain.

6. Acknowledgements

References

A. Adhitya, R. Srinivasan, I.A. Karimi, 2007, Heuristic rescheduling of crude oil operations to manage abnormal supply chain events, AIChE Journal, Vol. 53, No. 2, 397-422.

A.E.I. Alvino, 2003, Aplicação da Lógica Nebulosa ao Modelo Muhlbauer para análise de risco em dutos, Doctoral Thesis, Pontfícia Universidade Católica do Rio de Janeiro.

L.J. Fernandes, A.P. Barbosa-Póvoa, S. Relvas, 2009, Risk Management in Petroleum Supply chain, in Barbosa-Póvoa A.P., Salema M.I., (Eds.), Proceedings of the 14th Congress of APDIO, Vencer Novos Desafios nos Transportes e Mobilidade, 59-66.

W.K. Muhlbauer, 2004, Pipeline risk management manual ideas techniques and resources (3rd. Edition), Elsevier.

D. Peidro, J. Mula, R. Poler, F. Lario, 2008, Quantitative models for supply chain planning under uncertainty: a review, Int. J. Adv. Manuf. Tech.,Vol. 43, No. 3-4, 400-420.

L. Puigjaner, 2008, Capturing dynamics in integrated supply chain management. Comput. Chem. Eng., 32, 11, 2582-2605.

C.S. Tang, 2006, Perspectives in supply chain risk management, Int. J. Prod. Econ., Vol. 103, Issue 2, 451-488.

F. You, J.M. Wassick, I.E. Grossmann, 2009, Risk Management for a global supply chain planning under uncertainty: Models and Algorithms, AIChE Journal, 55, 4, 931–946.

A New Scheme for Management-of-Change Support Based on HAZOP Log

Teiji Kitajima[a], Tetsuo Fuchino[b], Yukiyasu Shimada[c], Yuji Naka[d],

Ling Yuanjin[e], Kensuke Iuchi[e], Koji Kawamura[e]

[a]*Institute of Symbiotic Science and Technology, Tokyo University of Agriculture and Technology, 2-24-16, Naka-cho, Koganei, Tokyo, 184-8588, Japan, teiji@cc.tuat.ac.jp*
[b]*Chemical Engineering Department, Tokyo Institute of Technology, 2-12-1, O-okayama, Meguro-ku, Tokyo, 152-8552, Japan, fuchino@chemeng.titech.ac.jp*
[c]*Chemical Safety Research Group, National Institute of Occupational Safety and Health, 1-4-6, Umezono, Kiyose, Tokyo, 204-0024, Japan, shimada@s.jniosh.go.jp*
[d]*Chemical Resources Laboratory, Tokyo Institute of Technology, 4259, Nagatsuda, Midori-ku, Yokohama, 226-8503, Japan, ynaka@pse.res.titech.ac.jp*
[e]*Techno Management Solutions Ltd., 4259-3, Nagatsuda, Midori-ku, Yokohama, 226-8510, Japan, ling@techmas.co.jp, iuchikj@utopia.ocn.ne.jp, kawamura@techmas.co.jp*

Abstract

Consistent and effective Management of Change (MOC) is the most important for process safety. Therefore, MOC should be logically and explicitly realized based on To-be reference model. In this study, a new scheme for MOC support based on engineering activity model through the plant-lifecycle is proposed. Practical and effective utilization of HAZOP logs obtained at plant/process design phase in reusable form by using an intelligent support tool (HazopNavi) is archived.

Keywords: Management of Change (MOC), Plant Lifecycle Engineering (Plant-LCE), HAZOP Log, PDCA Cycle, Engineering Activity Model

1. Introduction

Management of change (MOC) is a process for evaluating and controlling modifications to facility design, operation, organization, or activities. Consistent and effective MOC is one of the most important elements of process safety management (PSM) system, because uncontrolled changes can directly cause or lead to catastrophic events as well as degrading the quality of manufacturing operations. The importance of MOC is well recognized and enormous efforts have been making by many companies. However, present MOC scheme is not sufficient to achieve process safety excellence, especially for a whole lifecycle of chemical plants.

On the other hand, in Japan, as the necessity of PSM based on plant lifecycle engineering (Plant-LCE), which is from development to design, construction, production, has been realized, various intelligent engineering methodologies and tools with information/knowledge management are being developed. Therefore, it is reasonable that we assume the following engineering environment for MOC.

- HAZOP logs with propagating behavior models of failure modes obtained at process/plant design phase are available (Kawamura, 2008).
- Plant-LCE is based on logical activity model, which implements PDCA (Plan, Do, Check, Act) cycle in all hierarchical layers of activities (Fuchino, 2008).
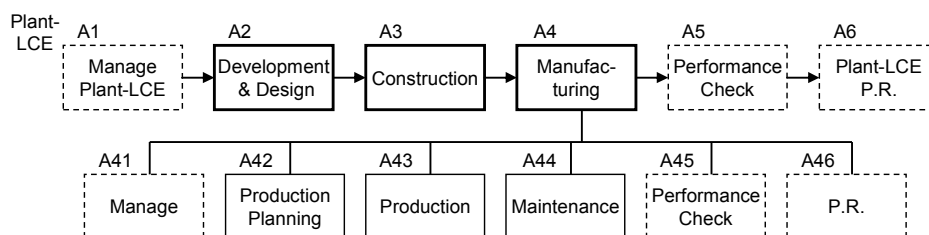
Figure 1: Hierarchical structure of Plant-LCE

In this study, a new support scheme for MOC through the Plant-LCE supposing the above engineering environment is developed. Especially, we focused on the practical and effective utilization of HAZOP logs obtained at plant/process design phase. There is an advantage from the existence of To-be models, it enables to extract the hidden "design rationale" which is considered carefully at process/plant design phase. Because the activity model explains logical structure and information-flow among the engineering activities, the design rationale is found out directly from activity model. Additionally, it is not necessary to care for logical omission, which is unavoidable when we use only as-is model based on restrict experiences.

2. Reference Activity Model for Plant-LCE

2.1. Activity Model for Plant-LCE
MOC with consideration for the whole plant lifecycle is essential for realizing the practical process safety. Although standard model for enterprise manufacturing control function has been discussed (CEI/IEC 62264-1:2003), appropriate reference model for engineering activities achieved by the function does not exist. Therefore we are trying to develop an activity model for Plant-LCE to prevent industrial accidents in chemical processes as shown in Figure 1.
Development and application of PSM system based on reference model
● enable to correspond immediately to change issues occurred from various causes;
● make it easy to understand for operator and engineer, because the system is explicitly represented;
● make it easy to modify/correct the management scheme even if any problem is found; and etc.
Reference activity model for Plant-LCE plays an important role oriented to representing the design rationales, supporting the logical MOC, realizing the consistent engineering from plant design to operation and maintenance, and maintaining the integrity of process information. The reference model may not be the same with present As-is model.

2.2. PDCA Template for Activity Modeling
We developed a template as meta-model for activity modeling in the form of IDEF0 format as shown in Figure 2, because there is no widely accepted principle or manner for modeling such engineering activities. A similar model for plant operation was developed by PIEBASE (Process Industry Executive for achieving Business Advantage using Standard data Exchange) project (PIEABASE, 2005), but the relation between the Plant-LCE and the model is not considered.

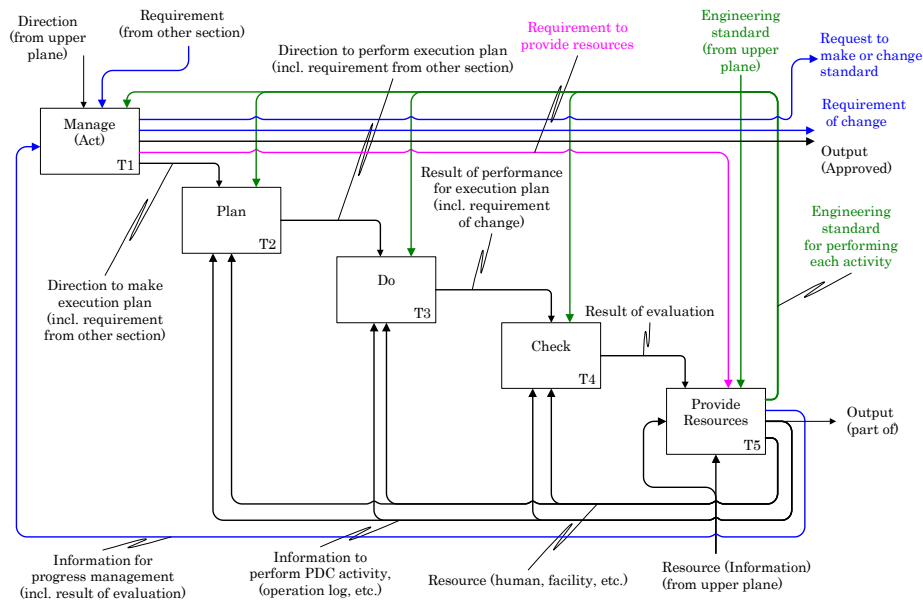*A New Scheme for Management-of-Change Support Based on HAZOP Log*

Figure 2: Template for representing PDCA Cycle and 'Provide Resources'

Based on the proposed PDCA template, practical engineering activities for production management were analyzed and extracted. Then, relation among the activities and information flow as To-be model for Plant-LCE was described (Shimada, 2009).

This template consists of 'Performance in the form of PDCA cycle' and 'Resource Provision'. The most important benefit of using IDEF0 standard for developing the activity model is to enable function-based discussions. In other words, discussions based on IDEF0 standard focus on 'what is needed as function or activity' at all times and it can clarify what to do for Plant-LCE.

- **Performance in the form of PDCA cycle**:
  Basically, each activity should be carried out according to engineering standards given by 'Provide Resources' activity. 'Manage (Act)' activity manages the progress of overall activities within the same plane, including the requirement of resource provision, improvement of engineering standards, and the decision making of next action for change requirement.

  'Plan' activity makes execution plan. 'Do' activity executes plan and gives requirements for administrative defect factors, if any. 'Check' activity evaluates the results and the performance.

- **Resource Provision**:
  This activity provides resources to support 'Plan', 'Do', 'Check', 'Act' activities and maintains the consistency of engineering standards through the Plant-LCE. These resources include a) educated and trained people and organizations, b) facilities and equipment, tools and methods for supporting activities, c) information to perform PDC activities, d) information for progress management, and e) engineering standards for controlling each activity which are distributed from upper level.
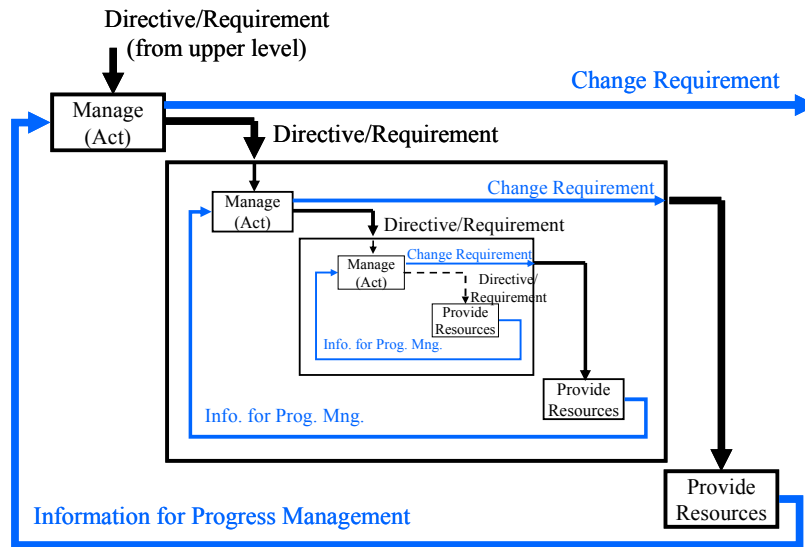
Figure 3: Directive and change requirement flow

2.3. MOC Based on Activity Model

According to the proposed PDCA template, MOC is executed as following procedure (Figure 3).

- 'Management' activity derives the directive and requirement ordered from upper level plane, and outputs directives to 'Plan', 'Do' and 'Check' activities within the same plane.
- Each activity directs to 'Management' activity in the sub-level, and outputs the execution result to 'Provide Resource' activity. If any trouble occurs during the executions, requirement of change may be included in the outputs.
- 'Provide Resource' activity arranges the results included requirements of change.
- 'Management' activity pulls up information for progress management from 'Provide Resource' activity, and compares it with new directive/requirement form upper plane. If the requirements of change can be solved within the same plane, 'Management' activity outputs directives to act for requirement. Otherwise, 'Management' activity outputs the requirement to upper, if necessary.

3. MOC Support Based on Hazop Log

3.1. Intelligent HAZOP Support System: HazopNavi

At the present work, typical MOC has a flow, which defines MOC procedures for every cases in advance, and executes tasks according to the procedures. Almost problems occurred at the present situation are that either necessary work flow (for example, derived from engineering activity model) is not exist or explicit, or support environment for the work is not available.

To support such work, an intelligent HAZOP system (HazopNavi) has been developed (Kawamura, 2008). HazopNavi shows plant areas where failures may propagate and indicates equipment of plant structure where may be changed to other failure. In addition, it lists candidates of sensors to detect failures. Sensor information is used to

recognize that each deviation can be detected by DCS or local instrument, and to decide whether or not protection control logic and countermeasure against the failure works. If it is not sufficient, further study should be required so as to improve facilities or additional instrument, interlock or alarm system. All these propagation behaviors are confirmed graphically.

3.2. MOC Based on Engineering Activity Model and HAZOP logs

Principally, all related PDCA cycles from bottom to top planes on the Plant-LCE should be executed. Summary of proposed MOC procedure is as follows:

1) Identify the related engineering activities and standards from the engineering activity model described To-be practical work, and confirm the design rationales represented as various design documents involving HAZOP logs.

2) Assess the process safety by taking into account of the design rationales involved with changes through the use of HazopNavi.
   - Refer the obtained HAZOP logs maintained in the 'Provide Resources' section, and reexamine a HAZOP study for the target facilities and/or process.
   - Correct the HAZOP logs related to the target.

3) Revise the related engineering standards for plant operation and maintenance to satisfy the supposed function or performance.

If appropriate engineering standards are improved, it makes easy to check the satisfaction of the function and performance. Consideration of the deviation from standard values which is estimated from HAZOP logs is just needed.

4. Case Study

4.1. Scenario

Effectiveness of the proposed scheme for MOC is evaluated through the case study. "Process equipment change required from maintenance activity in hydro-desulfurization (HDS) plant" is chosen as practical scenario. We assumed that HAZOP logs shown in Table 1 are available as the results of PHA executed at the plant design. Detailed HAZOP logs held in HazopNavi are reusable and understandable because it clearly keeps relationships between plant structure (P&ID) and failure propagation paths.

Table 1: HAZOP logs for HDS Plant

No.	Trigger Equip.	Failure Mode	Detection	Deviation	Event	Severity Class
1	P-201A	Down	O	F0	H-201 Tube overheats …	C:(2)
2			O	F0	Feed oil to R-201 is lost, …	D:(1)
3	H-201 Tube	Heat Hi	O	T ↑	R-201 Temp. up.	B:(3)
4	H-201 Tube	Heat Low	O	T ↑	Desulfurization fails in …	D:(1)
5	FCV-201	Close	×	F0	H-201 Tube Temp. Hi.	C:(2)
…	…	…	…	…	…	…

No.	Consequences	Action/ Safeguards	Recommendations	Path
1	Tube is broken, furnace …	Sensor setting around …	N/A	--, Va1009, FCV-201, …
2	Desulfurization is stoped, …	FSL-201A/B Flow L-alarm …	N/A	--, Va1009, FCV-201, …
3	Reaction is runaway and …	Sensor setting in R-201 …	Confirm …	--, --, --, P-1078, --, --, --
4	No Problem for Safety.	Operate R-201 Temp. …	N/A	--, --, --, P-1078, --, --, --
5	H201 Tube is broken, …	Sensor setting around …	N/A	--, Va1009,
…	…	…	…	...

4.2. PHA by reusing HAZOP logs

From 'A4: Perform Production' including 'A44: Perform Maintenance' activity, the following requirement is reported as 'Difference between Forecasting and Performance'. Local heat-up occurred at tube in the reactor-charge-furnace: H-201, thus replacement of the tube is necessary to avoid damage of the furnace. 'A1: Manage Plant-LCE' pulls up the requirement via 'A6: Provide Resources for Plant-LCE', and directs the review of the replacement to 'A2: Develop and Design' to execute PHA.

According to the information from maintenance activity, we found that this issue is out of tolerance without revamping. This was decided by reference of HAZOP logs, when maintenance activity requires the replacement as 'Difference between Forecasting and Performance'. By searching the tube to be replaced on HazopNavi, the regarding HAZOP logs can be retrieved (No.3 log in Table 1). Taking into account of the deviation of combustion, the regarding HAZOP logs is picked up to review the change of heat-up capacity occurred by the tube replacement.

On HazopNavi, the propagation of deviations is described in the form of tracing the connection of the process equipments on P&ID. Therefore all the deviations which involve the replacement target can be identified immediately one after another. In this case, since the necessary log which can be reused without any modifications is left, it is easy to execute PHA for this replacement. Through this case study, we found that HAZOP logs are helpful to identify the target deviations and the scope of its influences.

5. Conclusion

Logical reference activity model involving PDCA cycle must be clarified to realize the practical MOC lied on whole engineering phases through plant-lifecycle. In this paper, a new scheme for MOC support based on utilization of HAZOP logs is developed. Referring the explicit logical model, HAZOP logs can be reused to execute PHA easily when change requirements are occurred during ordinal work.

6. Acknowledgment

References

CEI/IEC 62264-1:2003: Enterprise-control System Integration – Part 1: Models and Terminology, ISO/IEC, 2003.

Tetsuo Fuchino, Yukiyasu Shimada, Masazumi Miyazawa, Yuji Naka: Business Process Model for Knowledge Management in Plant Maintenance, ESCAPE18, Computer-Aided Chemical Engineering, Vol.25, pp.955-960, 2008.

Koji Kawamura, Yuji Naka, Tetsuo Fuchino, Atsushi Aoyama and Nobuo Takagi: Hazop Support System and Its Use for Operation, ESCAPE18, Computer-Aided Chemical Engineering, Vol.25, pp.1003-1008, 2008.

PIEBASE - Process Industries Executive for Achieving Business Advantage Using Standards for Data Exchange, http://www.posc.org/piebase/, 2005.

Yukiyasu Shimada, Teiji Kitajima, Kazuhiro Takeda: PRACTICAL FRAMEWORK FOR PROCESS SAFETY MANAGEMENT BASED ON PLANT LIFE CYCLE ENGINEERING, 3rd International Conference on Integrity, Reliability and Failure, Paper Ref: S1707_A0531, Porto/Portugal, 20-24 July 2009.

Fault Diagnosis in a Heat Exchanger using Process History Based-Methods

Juan C. Tudón-Martínez,[a] Ruben Morales-Menendez,[b] Luis E. Garza-Castañón,[c]

[a]*PhD Student,* [b]*Associate Director of Research,* [c]*Graduate Program Coordinator,*
Tecnológico de Monterrey, Av. E. Garza Sada 2501, 64849, Monterrey N.L., México,
{A00287756, rmm,legarza}@itesm.mx

Abstract

A comparison of fault diagnosis systems based on Dynamic Principal Component Analysis (*DPCA*) method and Artificial Neural Networks (*ANN*) under the same experimental data is presented. Both approaches are process history based methods which do not assume any form of model structure, and rely only on process historical data. The comparative analysis shows the online performance of both approaches when sensors and/or actuators fail. Robustness, quick detection, isolability capacity, false alarm rates and multiple faults identifiability are considered for this experimental comparison. An industrial heat exchanger was the experimental system. *ANN* showed instantaneous detection for actuator faults; however, with greater (22%) false alarm rate. *ANN* can isolate multiple faults; whereas, *DPCA* did not show this property, but required a minor training effort.

Keywords: *DPCA*, *ANN*, Fault Detection, Fault Diagnosis

1. Introduction

Early detection and diagnosis of abnormal events in industrial processes can represent economic, social and environmental profits. Generally, the measuring and actuating elements of a control system fail causing abnormal events. When the process has a great quantity of sensors or actuators such as chemical processes, the Fault Detection and Diagnosis (*FDD*) task is very difficult. Advanced methods of *FDD* can be classified into two major groups (Venkatasubramanian, et al., 2003), process history-based methods and model-based methods.

Most of the existing Fault Detection and Isolation (*FDI*) approaches tested on Heat Exchangers (*HE*) are based on model-based methods. Ballé et al. (1997) presented an adaptive control using normal/faulty fuzzy-models. Krishnan and Pappa (2005) used quantitative models for fault signatures. Morales-Menéndez, et al. (2003) used particle filtering for predicting the fault probability distribution in a *HE*.

A comparative analysis between two *FDI* systems is proposed in this paper. One of them is based on the Dynamic Principal Component Analysis (*DPCA*) and another one on Artificial Neural Networks (*ANN*). Both methods are designed to online detect and isolate faults related to sensor or actuators in an industrial *HE*.

Recently, *DPCA* and Correspondence Analysis (*CA*) have been compared (Detroja, et al., 2005). *CA* shows shorter detection delay and lower false alarm rates; however, *CA* needs greater computational effort. Mina and Verde (2007) proposed an adaptive standardization for *DPCA* which allows detecting faults and avoiding normal variations. Many approaches based on *ANN* have been proposed to diagnosis faults in nonlinear systems (Korbicz et al., 2004). Tan et al. (2009) used an *ANN* to model a *HE* including detection and fault classification.

The aforementioned works are tested under different faults and non-uniform process conditions, i.e. it is impossible a comparative analysis. This paper shows a comparison between *DPCA* and *ANN* under the same experimental data provided from a *HE*.

This paper is organized as follows: in the next section, *DPCA* formulation is presented. Section 3 describes the *ANN* design procedure. Section 4 shows the experimentation. Section 5 presents the results. Final conclusions of this work are presented in section 6.

2. *DPCA* Formulation

Process data in the normal operating point must be acquired. Process variables can have different ranges of values, thus the data matrix $\boldsymbol{X}$ must be normalized. In chemical processes, serial and cross-correlations among the variables are very common. To overcome the limitations of normality and statistical independence of the samples, the column space of the data matrix $\boldsymbol{X}$ must be augmented with a few past observations for generating a static context of dynamic relations.

$$\boldsymbol{X_D}(t) = [X_1(t), X_1(t-1), \dots, X_1(t-w), \dots, X_n(t), X_n(t-1), \dots, X_n(t-w)] \qquad (1)$$

where w represents the quantity of time delays to include in n process variables. The main objective of *DPCA* is to get a set of a smaller number ($r < n$) of variables by solving an optimization problem which involves the maximization of the explained variance in the data matrix. r must preserve most of the information given in these variances. *DPCA* formulation can be reviewed in detail in (Tudón, et al., 2009). Once the scaled data matrix $\overline{\boldsymbol{X}}$ is projected by a set of orthogonal vectors, called *loading vectors* ($\boldsymbol{P}$), a new and smaller data matrix $\boldsymbol{T}$ is obtained: $\boldsymbol{T} = \overline{\boldsymbol{X}}\boldsymbol{P}$. $\boldsymbol{P}$ contains the eigenvectors with the largest eigenvalues. Matrix $\boldsymbol{T}$ can be back-transformed into the original data coordination system as, $\boldsymbol{X}^* = \boldsymbol{T}\boldsymbol{P}^T$. Fig. 1 shows the block diagram for getting the characterization of the normal operating point using *DPCA*.
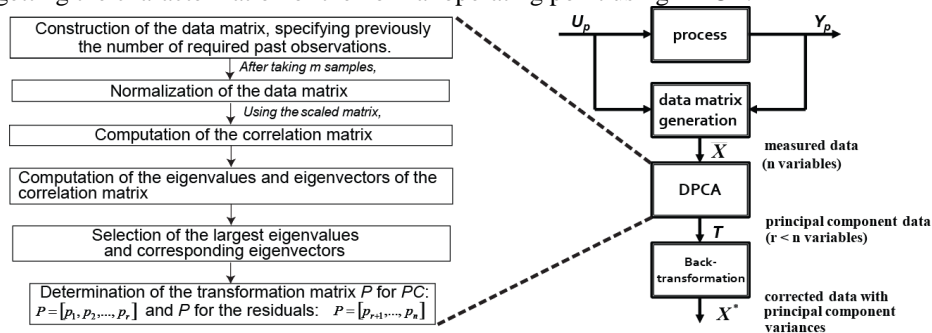

Figure 1. *DPCA* Algorithm.

2.1. FDD using DPCA

Normal operating conditions can be characterized by T^2 statistic (Hotelling, 1993) based on the first r Principal Components *(PC)*, or by Q statistic (Jackson, et al., 1979) based on the residual space. Left plot in Fig. 2 shows the block diagram for applying *DPCA* in the online fault detection task. New measurements are projected in both spaces. A fault is correctly detected when both statistics overshoot their respective thresholds, see details in (Tudón, et al., 2009). Once a fault is detected, contribution plots are used to isolate the most relevant cause of fault (Miller, et al., 1998). Fig. 2 (right) shows a block diagram for achieving the fault diagnosis using contribution plots. Since the residual space is more sensible to the faults than the *PC* space (Isermann, 2006), residual space is used to generate the residue which is used to compute the error contributions.
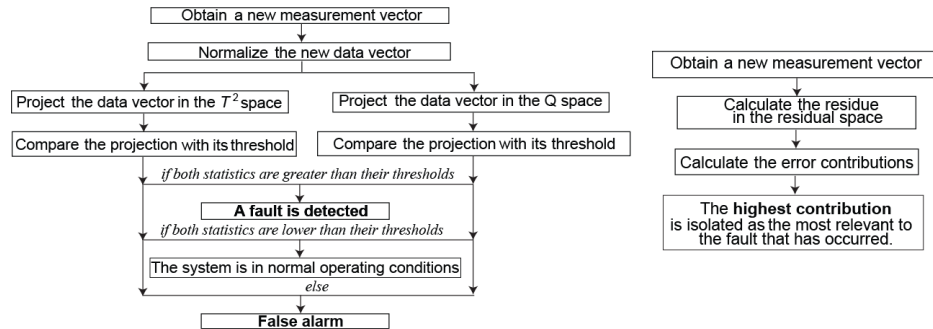
Figure 2: *FDD* algorithm using *DPCA* (left figure) and contribution plots (right figure).

3. Design of the *ANN*

An *ANN* is a computational model inspired on biological neural networks, it is composed by several basic elements named *neurons* which are clustered in *layers* and highly interconnected. These connections are used to store knowledge. The *ANN* architecture determines the neurons distribution into three types of layers: input layer, hidden layers and output layer. Based on the flow of signals, the *ANN* architecture can be classified into two major groups: *feedforward* and recurrent networks. Feedforward networks project the flow of information only in one way, i.e. the output of a neuron feeds to all neurons of the following layer (Hagan, et al., 1996).

The *ANN* training is defined as the adaptation process of the synaptic connections under external stimulations. The *backpropagation* algorithm is the most used training method since it allows to solve problems with complex net connections; its formulation can be reviewed in detail in (Freeman and Skapura, 1991).

Korbicz et al. (2004) established that the *ANN* can be used to model nonlinear, complex and unknown dynamic systems, as well as to take decisions (e.g. in faulty scenarios). A Multilayer Perceptron (*MLP*) has been proposed to diagnosis sensor and actuator faults. The main objective of the *MLP* is to classify the inputs into a specific class of faults (binary values) using a nonlinear transformation function. For the *HE* data, the *MLP* is designed to classify the set of the inputs in normal or faulty operating conditions. *ANN* inputs correspond directly to the process measurements. *ANN* outputs generate a fault signature which must be codified into pre-defined operation states. When the *ANN* output is zero the sensor or actuator is free of faults; otherwise it is faulty. All neurons of the input and hidden layers have an activation function of hyperbolic tangent; while the neurons of the output layer have sigmoid function. The trained network can be subsequently validated by using unseen process data. *Crossed validation* is used to assess the ability of the *ANN* to generalize the process it has been trained to represent.

4. Experimental System

4.1. Industrial Heat Exchanger
An industrial shell-tube *HE* was the test bed. Fig. 3 (left side) shows a photo of the *HE*; while right picture shows a diagram of the main instrumentation: 2 temperature sensor /transmitters (TT_1, TT_2), 2 flow sensor/transmitters (FT_1, FT_2) and their control valves (FV_1, FV_2). Instrumentation is connected to a data acquisition system (NI USB-6215).

4.2. Design of Experiments
Abrupt faults in sensors and actuators (*soft faults*) have been implemented in additive form. Also, the process always was free of disturbances. Multiple faults identifiability is
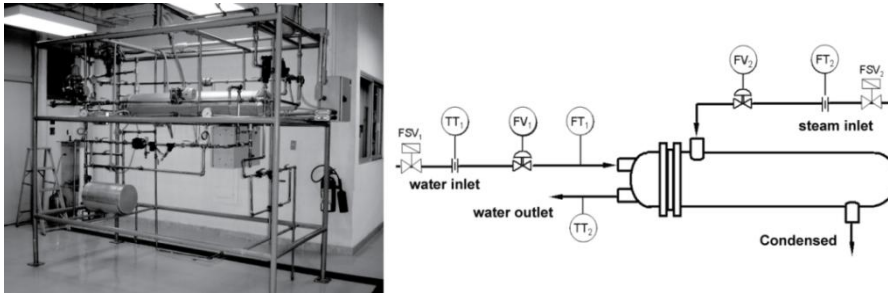
Figure 3. Experimental System.

analyzed under sensor faults, which simulate transmitter biases. The fault magnitude for each sensor is: $FT1 \rightarrow 6\%$ (5σ), $FT2 \rightarrow 8\%$ (5σ), $TT1 \rightarrow 2°C$ (8σ), $TT2 \rightarrow 2°C$ (8σ). Four types of faults have been implemented in the steam (cases 1 and 2) and water (cases 3 and 4) control valves. The faults are considered like low or high pressure in the valves ($\pm10\%$). The case 0 corresponds to the normal operating point: steam valve in 70% and water valve 38%.

4.3. Implementations

DPCA uses 1 second as time delay in the training stage, and 1900 experimental data for each sensor. Thus, the measurement vector is: $x(t) = [FT2(t)\ FT1(t)\ TT1(t)\ TT2(t)]$. In the *DPCA* training, all data vectors are considered in the normal operating point. While, a *MLP* (4-10-4) is used to detect sensor faults, i.e. an *ANN* with 1 hidden layer of 10 neurons. 3000 experimental data for each sensor are used to train the *ANN*. For actuator faults, a *MLP* (4-10-10-3) is used, i.e. the *ANN* has 1 more hidden layer in order to increase the network expressivity. 1000 measurements of each sensor were taken from the normal and faulty process conditions for training the *ANN*.

5. Results

For comparison, same metrics have been monitored in both approaches: quick detection, isolability capacity, false alarm rates and multiple faults identifiability.

5.1. DPCA approach

Taking one time delay of each measurement, it is possible to explain a high quantity of variance including the possible auto and cross correlations. The normal operating conditions can be explained with 5 principal components (99.95%). When a fault is implemented in FT_1 at time 10, both statistics clearly overshoot their thresholds (left plot in Fig. 4). Contribution plot shows correct fault isolation, 59% of total error corresponds to water flow signal. When the remainder faults are introduced, the statistics belong still upon their control limits; however, they move more away from them. After the 1st fault,
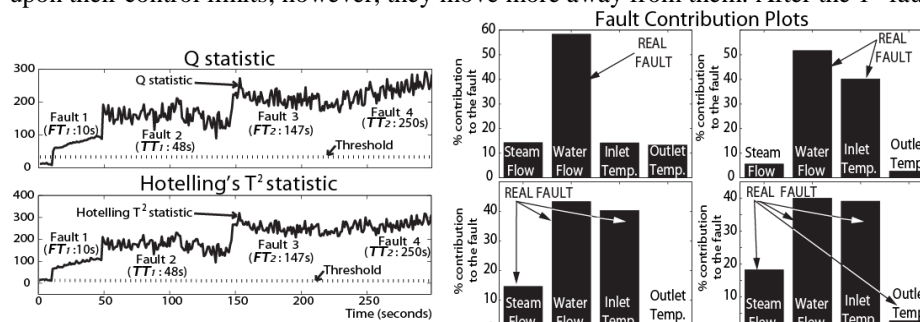
Figure 4. *FDD* analysis for multiple faults using *DPCA*

Fault Diagnosis in a Heat Exchanger using Process History Based-Methods

contribution plots can not associate the error to a specific variable (right plot in Fig. 4). A similar result to Fig. 4 is obtained for actuator faults. No matter if the bias is positive or negative, both statistics overshoot their control limits when a fault is detected. For the cases 1 and 2, the steam flow signal has the greatest error contribution because these faults are associated to changes in the pressure of steam valve. Similarly, the water flow has the greatest error contribution when the water valve fails (cases 3 and 4).

5.2. ANN approach

Fresh data are used in the *ANN* testing. Fig. 5 shows the *ANN* performance for sensor faults. After $t = 1750$, multiple faults are considered. The output layer of the *MLP* has 4 neurons, one for each sensor signal. When the signal is equal to 1, a fault has occurred.
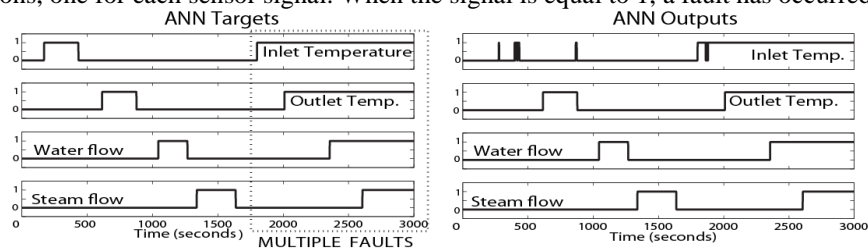

Figure 5. *FDD* analysis for sensor faults using *ANN*

Fig. 6 shows the *ANN* targets and outputs under faulty actuators. 3 neurons are considered in the output layer. A codifier is used to translate the 5 different operating points to a set of binary fault signature, e.g. the case 3 in the binary signature is: $(3)_{10} = (011)_2$. The binary numbers 110 and 111 are considered as false alarms.
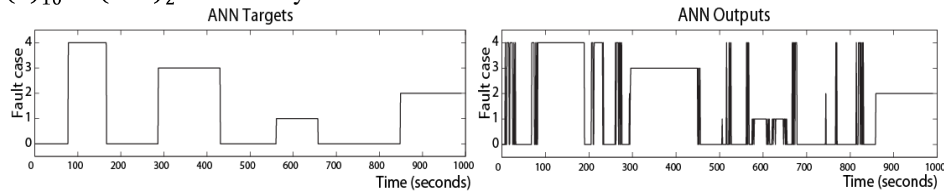

Figure 6. *FDD* analysis for actuator faults using *ANN*

5.3. Comparison of the methods

The false alarm rate is the index of false events which occurs when: (1) the *FDI* system does not detect an occurred fault or (2) the *FDI* system detects a fault which did not happen. For actuator faults, the false alarm rate is greater 22% in *ANN* because the isolation decision is ON/OFF; while in *DPCA*, the false alarm is caused by the time delay of both statistics. Furthermore, the detection time is greater in *DPCA* even when these faults are abrupt; whereas, *ANN* shows an instantaneous detection (Table 1).

Both methods can detect multiple faults; however, *DPCA* can not isolate correctly two or more sequential faults.

Table 1. Comparison of *DPCA* and *ANN* approaches.

Approach	Location	Detection (s)	Isolability	False alarm (%)
DPCA	Actuators	9 – 18	✓	14.13
	Multiple sensors	Instantaneous	–	0
ANN	Actuators	Instantaneous	✓	18.30
	Multiple sensors	Instantaneous	✓	2.01

According to computational requirements, the *ANN* training needs greater resources (memory) since historical data of the normal and faulty conditions must be known. While, the *DPCA* training is quickly executed; it only requires historical data of the normal operating point. Although both methods are based on historical data, they have

different frameworks: *DPCA* is a multivariate statistical method; whereas, *ANN* is based on pattern recognition. When unknown soft faults are presented, the performance of *ANN* can be deteriorated; whereas, the *DPCA* method does not suffer this limitation. Although, the *ANN* has the extrapolation property in nonlinear systems as well as the capacity to interpolate unknown results, a new *ANN* training can be necessary under these new malfunctions. On the other hand, *DPCA* does not require more training effort because only deviations from the normal operating point are considered.

Both methods are deteriorated when the process is time-variant; however, the training stage of both methods (retraining) can be easily applied every specific time window.

6. Conclusions

A comparison between *DPCA* and *ANN* under the same experimental data provided from an industrial heat exchanger is presented. *DPCA*, which do very well on fast detection of abnormal situations, is easier to implement in industrial applications. A process model is not required; however, a broad acquisition of the historical process measurements under only normal conditions is necessary; whereas, the *ANN* training requires a priori knowledge of normal and faulty conditions.

Under new and unknown faults, the training stage of the *ANN* must be computed again including all faulty behaviours; while, *DPCA* does not need more training effort because only deviations from the normal operating point are considered. Therefore, unknown soft faults can be detected and isolated using *DPCA*; while, *ANN* can diverge.

References

P. Ballé, M. Fischer, D. Füssel, and R. Isermann, 1997, Integrated Control, Diagnosis and Reconfiguration of a Heat Exchanger, ACC, Albuquerque New Mexico, pp. 922–926.

K. Detroja, R. Gudi and S. Patwardhan, 2005, Plantwide Detection and Diagnosis using Correspondece Analysis, Control Engineering Practice.

J. A. Freeman and D. M. Skapura, 1991, Neural Networks: Algorithms, Applications and Programming Techniques, Adisson-Wesley.

M. Hagan, H. Demuth and M. Beale, 1996, Neural Network Design, PWS Publishing Company.

H. Hotelling, 1993, Analysis of a Complex of Statistical Variables into Principal Components, J. Educ. Psychol., 24.

R. Isermann, 2006, Fault-Diagnosis Systems, Springer, Germany, 1st edition.

J. Jackson and G. Mudholkar, 1979, Control Procedures for Residuals Associated with Principal Component Analysis, Technometrics, 21, pp. 341–349.

J. Korbicz, J. M. Koscielny, Z. Kowalczuk and W. Cholewa, 2004, Fault Diagnosis Models, Artificial Intelligence, Applications, Springer.

R. Krishnan and N. Pappa, 2005, Real Time Fault Diagnosis for a Heat Exchanger A Model Based Approach, IEEE Indicon Conference, Chennai India, pp. 78–82.

P. Miller, R. Swanson and C. Heckler, 1998, Contribution Plots: A Missing Link in Multivariate Quality Control, Appl. Math. and Comp. Sci., 4(8), pp. 775–792.

J. Mina and C. Verde, 2007, Fault Detection for MIMO Systems Integrating Multivariate Statistical Analysis and Identification Methods, ACC, New York U.S.A., pp. 3234–3239.

R. Morales-Menendez, N. D. Freitas and D. Poole, 2003, State Estimation and Control of Industrial Processes using Particles Filters. ACC, Denver Colorado U.S.A., pp. 579–584.

C. K. Tan, J. Ward, S. J. Wilcox and R. Payne, 2009, Artificial Neural Network Modelling of the Thermal Performance of a Compact Heat Exchanger, Applied Thermal Eng., 29, pp. 3609-3617.

J. C. Tudón-Martínez, R. Morales-Menendez and L. E. Garza-Castañón, 2009, Fault Detection and Diagnosis in a Heat Exchanger, 6th ICINCO'09, Milan Italy, pp. 265-270.

V. Venkatasubramanian, R. Rengaswamy, S. Kavuri and K. Yin, 2003, A Review of Process Fault Detection and Diagnosis Part I Quantitative Model-Based Methods, Computers and Chemical Eng., 27, pp. 293–311.

Improving Process Safety and Product Quality using Large Databases

Ankur Pariyani,[a] Warren Seider,[a] Ulku Oktem,[b] Masoud Soroush[c]

[a]*Department of Chemical and Biomolecular Engineering, University of Pennsylvania, Philadelphia, PA 19104, USA; pariyani@seas.upenn.edu , seider@seas.upenn.edu*
[b]*Risk Management and Decision Processes Center, Wharton School, University of Pennsylvania, Philadelphia, PA 19104, USA; oktem@wharton.upenn.edu*
[c]*Department of Chemical and Biological Engineering, Drexel University, Philadelphia, PA 19104, USA; masoud.soroush@drexel.edu*

Abstract

This paper introduces a novel modeling and statistical framework (based on Bayesian theory) that utilizes extensive distributed control system and emergency shutdown databases, to perform thorough risk and vulnerability assessment of chemical/petrochemical plants. Quality variables are utilized, in addition to safety (or process) variables, to enhance both process *safety* and product *quality*. To effectively achieve these objectives, new concepts of *abnormal events* and *upset states* are defined, which permit the identification of near-miss events from the databases.   The databases for a fluid catalytic cracking unit at a major petroleum refinery are used to demonstrate the application and performance of the techniques introduced herein. The results show that with the novel utilization of near-miss data, one can perform robust risk calculations using both product-quality and safety data.

Keywords: Process safety, product quality, risk assessment, Bayesian theory, chemical process industries.

1.   Introduction

In the chemical process industries (CPIs), the extent of human and financial losses due to incidents is staggering – the U.S. Chemical Safety and Hazard Investigation Board website [1] lists about 65 serious accidents that occurred over the past decade, with their consequences and key technical findings. Based on the severity levels, incidents can be broadly classified as *accidents* or *near-miss consequences*. Though accidents have low probabilities of occurrence, they have high severities, often accompanied by on-site and/or off-site major impacts. Near-miss consequences, on the other hand, have much higher probabilities of occurrence, but have limited, or sometimes no, impact. Recent studies have demonstrated the importance of identifying near-miss consequences to predict the probability of accidents [2, 3], and reporting, monitoring, and investigating them to reduce the potential of accidents [4-6].

Because the CPIs have been adapting plants with minimal design changes to produce higher-quality products at higher production rates, the rate of reporting of near-miss consequences has increased in recent years, with more companies seeking to improve their reporting and investigation of incidents [7, 8]. Several industry-standard software packages, which perform quantitative risk assessment using incident (or consequence) databases, are widely used. In addition, several papers and books ([2, 3, 9-11]) have

proposed the analysis of incident databases using fault-trees, hazard and operability (HAZOP) studies, failure mode and effects analysis (FMEA) and Bayesian theory, among other approaches, to gain predictive insights concerning incidents. However, because most chemical processes have hundreds of variables to monitor their dynamics (especially large-scale continuous processes), in our views, a significant amount of *precursor* information, on unsafe conditions, is overlooked and un-utilized as it is immersed in their large dynamic databases. The dynamic data associated with their distributed control systems (DCSs) and emergency shutdown (ESD) systems, which help human operators assess and control plant performance, especially in the face of potential *safety* and *quality* problems, contain real-time information on the progression of disturbances (or special-causes) and the performance of their *safety and/or quality systems* (barriers to protect the process from abnormal behavior). To our knowledge, existing techniques do not adequately utilize this information. In this paper, new methods are presented that utilize this information to assess the risk levels and predict the probability of the occurrence of incidents.

Furthermore, in this paper, quality variables are utilized, in addition to safety (or process) variables, to enhance both plant *safety* and product *quality*. This approach improves the methods for integrating safety and quality management systems introduced over the past two decades [12, 13].

2. Model Development

Modern chemical processing units are equipped with distributed control systems (DCSs) and emergency shutdown (ESD) systems to assure safe operation and high product-quality performance. The DCSs involve controller elements distributed throughout the units, with central servers that issue controlling actions. This allows the operators (human + machine) to control the variables well within their defined operating envelopes to optimize the profitability, safety, quality, and flexibility of the units. Depending upon the type of measurements, variables are defined as *process* or *quality variables*. The former are online measurements that track the dynamics of the process (for example, temperatures, pressures, flow rates and their rates of change, etc.) whereas, the latter are often estimated measurements (using mechanistic and/or statistical models) related to the *quality* of the products (for example, viscosity, density, average molecular weight, etc.).

Based on their sensitivity and importance, variables are classified into two categories: primary and secondary variables. Primary, or key, variables are most crucial for the *safety* of the process and are associated with the ESD system. Whenever these variables move beyond their ESD limits, emergency shutdowns or 'trips" are triggered shortly after a small time-delay. Secondary variables, on the other hand, are not associated with the ESD. For large-scale processes, typically 150-300 operating variables are monitored; however, only a small percentage (<10%) are associated with the ESD.

The control chart of any primary variable is divided into four zones, beginning with its green-belt zone (normal operation), during which the process variable lies within acceptable limits. When the variable moves beyond these limits, into its yellow-belt zones, high/low alarms are triggered. When it moves beyond the limits of its yellow-belt zones, into its orange-belt zones, high-high/low-low alarms are triggered. The border between its orange- and red-belt zones is the threshold limit for the triggering of

the ESD system. *Abnormal events* begin when process (or quality) variables move from their green-belt zones to their yellow-, orange-, or red-belt zones, triggering alarms. Clearly, these departures can be interpreted as precursors to undesirable consequences or accidents, when safety and/or quality systems fail to maintain normal operation. Consequently, in this paper, abnormal events, for variables that return to their green-belt zones, are recognized as *near-miss events*, which could have propagated to incidents. As a result, vast amounts of near-miss information become available for dynamic risk assessment.

Depending upon their criticality, abnormal events are classified in three categories: *least-critical abnormal events* that cross the high/low alarm thresholds, but do not cross the high-high/low-low alarm thresholds; *moderately-critical abnormal events* that cross the high-high/low-low alarm thresholds, but do not cross the ESD thresholds; and *most-critical abnormal events* that cross the ESD thresholds. Secondary variables don't have red-belt zones, and consequently, most-critical abnormal events cannot occur.

2.1. Dynamic Databases: DCS and ESD Logs

The use of incident databases only by the industry-standard software packages restricts their abilities to achieve a high degree of predictive accuracy of the frequencies and consequences of incidents. The current lack of dynamic analyses to identify and target near-miss events has contributed to many serious accidents over the last decade [1]. Had there been systematic procedures for analyzing dynamic data identifying near-miss events and the performance of their safety and/or quality systems, it seems clear that a large fraction of these accidents (and shutdowns) would have been avoided by alerting the plant-management well in advance.

This work relies primarily on the efficient extraction of knowledge from *dynamic databases*, namely DCS and ESD databases. These databases are inherently heterogeneous and describe the state of a plant instantaneously or over a period of time. To our knowledge, the utilization of dynamic databases has not previously been achieved as prior analyses [2, 3, 9-11]) have focused on the usage of incident (consequence) databases. Typically, the DCS database contains abnormal event data; that is, alarm identity tags for the variables, alarm types (low, high, high-high, etc.), times at which the variables cross the alarm thresholds (in both directions), variable priorities, etc. Its associated ESD database, of greater consequence, contains trip event data, timer-alert data, etc. As discussed earlier, the framework utilizes the DCS and ESD databases for the FCCU at a major petroleum refinery over an extended period. The DCS databases are vast, with 5,000-10,000 alarm entries recorded every day, equaling 500-1,000 abnormal events.

2.2. Safety and/or Quality Systems

A safety and quality management structure of any process handles abnormal events with various *safety and/or quality systems (SQSs)*, which are components of the DCS, the human operators, and the ESD system. Five SQSs are defined, as follows: a) the basic process control system (BPCS) – SQS^1, b) the operator (machine + human) corrective actions, Level I – SQS^2, c) the operator (machine + human) corrective actions, Level II – SQS^3, d) the override controller – SQS^4, and e) the emergency shutdown (ESD) system – SQS^5. They are usually activated sequentially and the corrective actions become more rigorous down the sequence. The actions of the automatic safety and/or

quality systems influence the actions of human operators and this causative relationship is modeled using *copulas* (discussed next).

2.3. Industrial FCCU as Case Study

The FCCU studied herein converts low-value, heavy oil into lighter and more valuable products and processes more than 250,000 barrels of oil per day. In summary, for the study period, a total of 2,036 abnormal events occurred for the primary variables (1,527 associated with process variables, 509 with quality variables). Variables, equipped with all 5 safety and/or quality systems, experienced 1,857 abnormal events (1,720 least-critical, 21 moderately-critical and 116 most-critical abnormal events, with 2 leading to 2 ESDs). Those variables not equipped with high-high/low-low alarms and an override controller, experienced 179 abnormal events (176 least-critical and 3 most-critical leading to 3 ESDs).

Using the event-tree approach presented in [2], Table 1 is obtained, which is used as likelihood data for a Bayesian simulation of the entire study period.

Table 1. Failure (F) and Success (S) Counts for Safety and/or Quality Systems

I	II		III		IV		V	
Abnormal events	S	F	S	F	S	F	S	F
2,036	1,896	140	21	116	114	2	5	0

3. Bayesian Simulation

In our previous work, Bayesian analysis was used for dynamic risk analysis [2, 14, 15]. Assumed prior distributions were updated dynamically with data (incident consequences), to yield mean failure and incident probabilities. Instead, herein, the complete posterior distributions are computed. Furthermore, the failure counts of safety and quality system 1 are modeled using a Poisson distribution (likelihood) and the failure rate (θ_1) using a Gamma distribution. The failure probabilities of the other four safety systems ($\theta_j, j = 2,\ldots,5$) are modeled using Beta distributions and their failure and success counts are modeled using Binomial distributions (likelihood).

The joint failure probability distribution of the failure rate and probabilities of safety and/or quality systems is given as:

$$p(\theta_1, \theta_2, \ldots, \theta_{Ns}, \mathbf{W} \mid Data) \propto \underbrace{(\theta_1)^{n_i} e^{-(\theta_1)} \prod_{j=2}^{Ns} (\theta_j)^{K_j} (1-\theta_j)^{L_j}}_{\text{Likelihood}} \times \underbrace{p(\theta_1, \theta_2, \ldots, \theta_{Ns} \mid \mathbf{W}) \times p(\mathbf{W})}_{\text{Prior}}$$

where, $\mathbf{W}$ is the Spearman rank correlation matrix between the failure rate and probabilities of safety and/or quality systems, and Ns (=5) is the number of safety and/or quality systems. The copulas (*c*), which are multivariate functions, are used to model the joint probability distribution of the random variables, as a function of one-dimensional, marginal, cumulative distribution functions and their correlations [16, 17]. Primarily, they are tools for modeling the dependences of the random variables. Therefore, the joint prior distribution of the failure rate and probabilities, conditional on the correlation matrix is given by:

$$p(\theta_1, \theta_2, ..., \theta_{Ns} \mid \mathbf{W}) \propto c(F_1(\theta_1), F_2(\theta_2), ..., F_{Ns}(\theta_{Ns})) \times \prod_{k=1}^{Ns} p(\theta_k)$$

The simulation results (posterior distributions of failure rate/probabilities and their correlation matrix) are obtained using the random-walk, multiple-block, Metropolis-Hastings Algorithm [18] – with Gamma and Beta distributions as *proposal* distributions for the failure probabilities of the safety and/or quality systems, and the Inverse-Wishart distribution for the correlation matrix, and their means at the previous iteration values (random-walk distributions). The Cuadras and Augé copula [19] is used to model the dependences among the failure probabilities.

3.1. Results and Key Findings

The histograms of the marginal posterior failure rate and probabilities of safety and/or quality systems (θ) associated with the primary variables for the studied unit, based on the likelihood data in Table 1 and calculated using Cuadras and Augé copula, are presented in Figure 1.
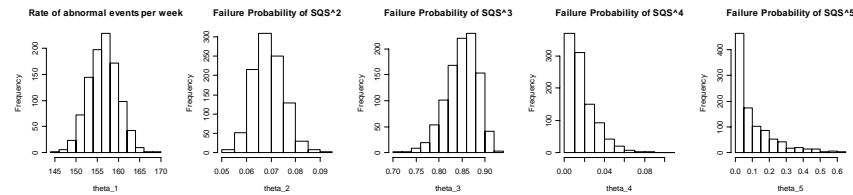
Figure 1. Marginal posterior distributions for θ using Cuadras and Augé copula

Next, the mean failure rate/probabilities are presented in Table 2 when the quality data in the DCS and ESD databases are not included in the Bayesian analysis. These are compared with the results when both the safety and quality data are included.

Table 2. Comparison of the mean failure probabilities and probabilities of occurrence of incidents – with and without quality data

Mean	θ_1 *(Rate/week)*	θ_2	θ_3	θ_4	θ_5	*Prob. of ESD**	*Prob. of Accident**
Safety data only	119.17	0.085	0.869	0.025	0.14	1.58E-3	2.58E-4
Safety and quality data	156.36	0.068	0.838	0.024	0.145	1.1E-3	1.9E-4

* Given an abnormal event

Clearly, Table 2 shows that the exclusion of quality data results in the overestimation of the failure probabilities of the safety/quality systems, θ_2, θ_3, θ_4, and θ_5, the probabilities of an emergency shutdown, and the occurrence of an accident. The quality data complement the safety data to yield more near-miss information, as well as more realistic and reliable results.

Figure 1 presents the performances of the SQSs and the probabilities of the occurrence of shutdowns and accidents for the study period and the associated uncertainties. As expected, when data points are limited, the variance increases. The main advantage of

using a copula model is that it allows the sharing of information (through correlations), embedded in the dataset. These and other results show that the means of the posterior distributions for safety systems I-III, calculated using different copulas, do not differ significantly due to the availability of many data points. However, for safety systems IV and V, with limited data points, the choice of copula, and consequently, the amount of information transmitted, has a significant impact. This is being examined closely in our current research. In addition, this framework calculates finite posterior distributions of the correlations between the failure rate/probabilities and demonstrates that the safety and quality systems are well correlated in practice.

4.   Concluding Remarks

This paper presents an overview of new techniques to: (i) utilize the vast dynamic databases recorded in the CPIs for dynamic risk analysis, and (ii) assess and enhance process safety and product quality in a unified way. To effectively achieve these objectives, new concepts of *abnormal events* and *upset states* are introduced, which permit the identification of near-miss events from the databases. A combined modeling and statistical framework (based on Bayesian theory) is developed to obtain thorough and robust risk estimates (with associated uncertainties) using the DCS and ESD system databases. As the failure probabilities of SQSs increase, the recognition of near-misses increases – alerting operators and management to consider corrective actions; e.g., improved (1) DCS configurations and tuning, (2) operator training, (3) operating regimes, (4) process designs, and (5) alarm system configurations. The technique presented herein improves upon existing techniques by accounting for the near-misses experienced by individual variables, thereby leading to improved risk estimates. More specifics are provided in recent papers submitted for publication.

5.   References

[1]   U.S. Chemical Safety and Hazard Investigation Board, http://www.csb.gov/.
[2]   A. Meel and W.D. Seider, Chem. Eng. Sci., 61 (2006) 7036.
[3]   W. Yi and V.M. Bier, Management Science, 44 (1998) S257.
[4]   J.R. Phimister, U. Oktem, P.R. Kleindorfer, and H. Kunreuther, Risk Anal., 23 (2003) 445.
[5]   I. Rosenthal, P.R. Kleindorfer and M.P. Elliott, Proc. Safety Prog., **25** (2006) 135.
[6]   S. Jones, C. Kirchsteiger, and W. Bjerke, J. of Loss Prev. in the Process Ind., 12 (1999) 59.
[7]   UNEP/ILO/WHO International Programme on Chemical Safety, http://www.who.int/ipcs/emergencies/identifying/en/index.html.
[8]   K. Rasmussen, The experience with Major Accident Reporting System from 1984 to 1993. European Commission, Joint Research Center, EUR 16341 EN, 1996.
[9]   J.M. Santamará Ramiro and P.A. Braña Aísa, Risk Analysis and Reduction in the Chemical Process Industry, Blackie Academic and Professional, New York, 1998.
[10] J. Steinbach, Safety Assessment for Chemical Processes, Wiley-VCH, Weinheim, 1999.
[11] L.M. Morrison, J. Hazard. Mater., 111 (2004), 161.
[12] S.G. Herrero, M.A.M. Saldana, M.A.M. Del Campo, D.O. Ritzel, J. Saf. Res., 33 (2002) 1.
[13] M.M. Williamsen, Prof. Safety, 50(2005), 41.
[14] A. Meel, L.M. O'Neill, W.D. Seider, U. Oktem, and N. Keren, J. of Loss Preven. in Proc. Ind., 20 (2007) 113.
[15] A. Meel and W.D. Seider, Comput. Chem. Eng., 32 (2008), 827.
[16] R.B. Nelsen, An Introduction to Copulas, Lecture Notes in Statistics, Springer, 1999.
[17] P.Embrechts, F. Lindskog, and A. McNeil, Handbook of Heavy Tailed Distributions in Finance, Elsevier, 2003, 331-385.
[18] A. Gelman, J.B. Carlin, H.S. Stern and D.B. Rubin (Second Edition), Bayesian Data Analysis, Chapman & Hall/CRC, 2004.
[19] C.M. Cuadras, and J. Augé, Commun. in Statistics-Theory and Methods, A10(1981), 339.

A Knowledge-based Framework for Incident Management of Pharmaceutical Processes

Jyunpei Sugiura,[a] Hirofumi Kawai,[b] Yukiyasu Shimada,[c] Tetsuo Fuchino,[d] and Rafael Batres[a*]

[a]*Toyohashi University of Technology, Toyohashi 441-8580, Japan*
[b]*Process Systems Engineering Division, Tokyo Institute of Technology*
[c]*National Institute of Occupational Safety and Health, Tokyo 204-0024, Japan*
[d]*Tokyo Institute of Technology, Tokyo 152-8552, Japan*
[*]*Corresponding author: rbp@tut.jp*

Abstract

Current GMP regulations require pharmaceutical, medical device, and food manufacturers to keep incident records that document incipient faults, near-misses, and incidents which have a potential impact on the quality and safety of their products. This paper presents a framework for storing, maintaining, and retrieving information about such past incidents and their solutions. Extracted information can be used to identify what went wrong and what solutions were effective in order to avoid similar incidents. This paper focuses on one of the components of the framework, namely the identification of association rules.

Keywords: FCA, ontologies, CBR, GMP, incident logs, product safety

1. Introduction

Despite the emphasis on R&D in the pharmaceutical industries, quality-related deficiencies contribute to more than 25% of the total revenues (Winkle, 2007). Good Manufacturing Practice (GMP) regulations apply to pharmaceutical, medical device, and food manufacturers to ensure that their products are processed reliably, repeatedly, consistently, safely and to a high quality. Current GMP regulations (CGMPs) require manufacturers to investigate incipient faults, near-misses, and incidents that have an impact on the product quality and safety. These data are used to produce incident reports that are maintained with the intent of extracting information that can be used to avoid similar incidents while improving product quality and productivity. Incident reports contain data such as materials and equipment involved in the incident (e.g. process lines, products, batches and raw materials), possible or actual consequences, possible causes, products made prior to and during the event and corrective or preventive actions. Incident reports are written the form of textual natural language descriptions which limit the ability to use past data in an efficient way. Some research has been done on mining maintenance logs of discrete manufacturing plants, but no research has been reported on the mining of incident logs of batch or continuous plants. Specifically, Devaney et al. (2005) discuss a project for mining maintenance logs using text processing, text clustering and case-based reasoning but no specific results are reported. Anand et al (2006) use association rules for mining a subset of incidents stored in the National Response Center incident database. The association rule extraction is performed by exploring all the combinations resulting from different kinds of equipment and 12 chemicals, an approach that works well for small datasets but

requires a computational effort that increases exponentially. This computational effort can be alleviated by Formal Concept Analysis which includes a number of algorithms for a more efficient mining (Estacio-Moreno, 2008) (Lakhal and Stumme, 2005).

This paper presents a framework for storing, maintaining and retrieving information about past decisions on incident resolution based on case-based reasoning, ontologies, and formal concept analysis. This paper focuses on one of the components of the framework, namely the identification of association rules.

2. Methodology

The core of the methodology is based on case-based reasoning (Fig. 1). In case-based reasoning problems are solved "by using or adapting solutions to old problems" (Riesbeck and Schank, 1989). A case is a representation of the problem and a solution to that problem. In this paper, a case is made up with information contained in an incident report. Incident reports are stored in a case base where incident information follows a predefined structure which includes definitions from domain ontologies. Ontologies define a set of classes and a set of relations between these classes for things such as equipment, processing activities, and causality. When a new problem arises (for example when new incident occurs) the information that is available about this target incident is entered to retrieve similar cases from which the best matching case is selected. Subsequently, domain knowledge is used to complete missing information or adapted by using domain-specific knowledge. The resulting case is also adapted by means of association rules generated by analyzing the case base using a technique called Formal Concept Analysis. The retrieved case and the adapted case are shown to the user who uses this information to take preventive or corrective actions. Subsequently, the user confirms the case with on-site investigation and stored as a new case.
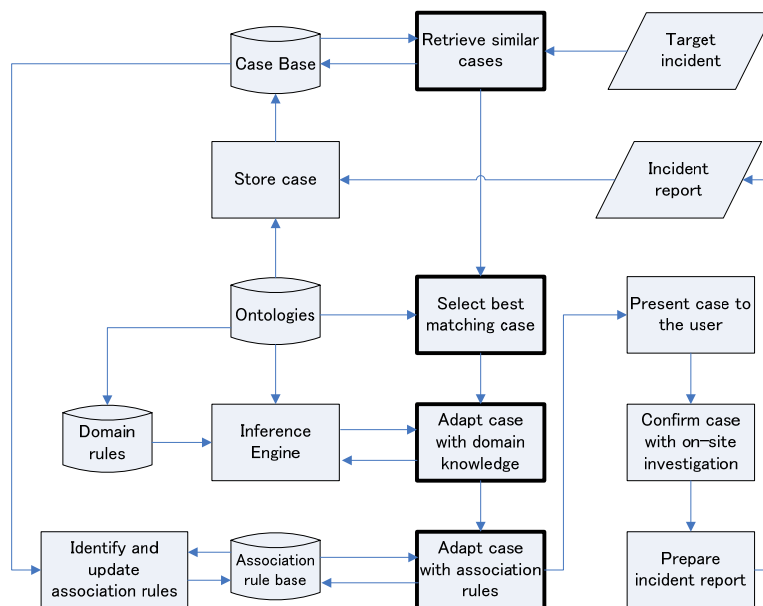

Figure 1. Methodology of the proposed framework

3. Ontologies

Ontologies are used in the "Select best matching case" step to find similarities between parts of the previous cases and the target problem. An ontology describes a shared and common understanding of a domain that can be communicated between people and heterogeneous software tools. An ontology is constructed by defining classes of things, their taxonomy, the possible relations between things and axioms for those relations. A class represents a category of things that share a set of properties. The classes in the ontology can be represented as a tuple $\langle C, \leq c \rangle$ where C is a set in which each element is a class with partial order $\leq c$ on C. A relation is a function that maps its arguments to a Boolean value of true or false. Examples of relations are *connected_to*, and *part_of*. Class taxonomies are defined with the use of the *subclass* relation. A class is a subclass of another class if every member of the subclass is also a member of the superclass. ISO 15926 describes classes and relations that can be used to represent things such as processing activities, personnel, plant equipment, chemical processes, batch recipes and engineering diagrams (Batres *et al.*, 2007).

4. Best case selection

The best case is selected using a similarity function that evaluates the differences between the target problem and a case stored in the case base. A number of different similarity functions for text, quantities, and ontology objects are investigated. A number of well-known similarity functions exist for text and quantities. For similarity between two ontology objects $a \in C_1$ and $b \in C_2$ is measured by the distance between classes C_1 and C_2 in regards to their lowest common ancestor.

Let C denote the set of all classes in the ontology. Given two classes C_1, $C_2 \in C$, let lca(C_1, C_2) be the common ancestor of C_1 and C_2. We define similarity between C_1 and C_1 as:

$$\text{sim}(C_1, C_2) = 1 - \frac{\text{dist}(C_1, \text{lca}(C_1, C_2))}{\text{dist}(C_2, \text{lca}(C_1, C_2))}$$

where dist(C_1, lca(C_1, C_2)) and dist(C_2, lca(C_1, C_2)) are the number of classes + 1 in the shortest path from the class C_1 to lca(C_1, C_2) and from C_2 to lca(C_1, C_2).

5. Adaptation with domain knowledge

Domain knowledge can be used to identify and modify parts of the best case that are not applicable to the target problem. If the framework attempts to re-apply a corrective action and discovers that one part of the corrective action is not applicable for the current target problem, the corrective action will modified by finding a different alternative for that part. For example, suppose that the framework finds that by "numbering the trays when more than one sample is analyzed" help to prevent an incident that is caused by "mistakenly switching samples" when carrying out a near-infrared analysis. The solution presented by the framework suggests to "number the cuvettes when more than one sample is analyzed," when carrying out a similar analysis using UV. However, the current target problem uses sampler trays but not cuvettes. In this case, the solution would be an adapted transformation by replacing cuvettes by trays.

6. Adaptation with of association rules

The identification of association rules refers to the extraction of hidden relations in the logs that permit the discovery of knowledge that includes the identification of patterns in the records, the prediction of the probability that events will occur, and the identification of strong relations between causes and effects. An association rule is a relation between two sets of items A, B, that indicates that cases involving A tend also to involve B.

Identification of association rules is done by processing the incident data stored in the case base using Formal Concept Analysis (FCA). FCA is an analysis technique for knowledge processing based on applied lattice and order theory (Wille, 1982). FCA assumes that data is represented in as a tuple $K := \langle O, A, Y \rangle$ where O is a set of objects, A is a set of attributes, Y a set of binary relations $Y \subseteq O \times A$ containing all pairs $\langle o, a \rangle \in Y$ such that the object o has the attribute a such as in as in (incident1, caused by step started too late). The initial step in FCA is to find all pairs $\langle O_i, A_i \rangle$ that satisfy $O_i \subseteq O$, $A_i \subseteq A$, $O' = A_i$ and $A' = O_i$ where A' is the set of attributes common to all objects in O_i, and O' represents the set that has all attributes in A_i. Each pair $\langle O_i, A_i \rangle$ is called a formal concept. O_i and A_i are respectively the *extent* and the *intent* of the formal concept. The hidden relations become apparent by analyzing the so-called concept lattice. A concept lattice is a partially ordered set in which a $\langle O_i, A_i \rangle \subseteq \langle O_j, A_j \rangle$ iff $O_i \leq O_j$. Several algorithms for lattice-construction are available.

Typically, the set $K := \langle O, A, Y \rangle$ is represented by a cross table. In this paper, O represents the set of incidents and A denotes the set of attributes that include specific causes and consequences, equipment categories, product names, raw materials used, products made prior to the event, products made during the event, and impacts to product.

Preliminary observations indicate that it is possible to identify mutually exclusive classes of events, direct causality (when the intent includes only one cause and only one consequence), and potential multiple causality (when the intent includes two or more causes and one consequence).

In FCA, association rules are expressed as $A_1 \Rightarrow A_2$ where A_1 and A_2 are sets of attributes that are disjoint ($A_1 \cap A_2 = \varnothing$). Unlike the domain rules (explained in section 5) association rules are probabilistic in nature. The level of support of $A_1 \Rightarrow A_2$ is defined as the proportion of cases that include all the attributes in A_1 and A_2. The level of confidence is the proportion of cases that include all the attributes in A_2 within the subset of those cases that include all the attributes in A_1 . Here we define

$$f(O_i) := \{a \in A \mid \langle o, a \rangle \in Y \; \forall o \in O_i\} \text{ and } g(A_i) := \{o \in O \mid \langle o, a \rangle \in Y \; \forall a \in A_i\}.$$

A Knowledge-based Framework for Incident Management of Pharmaceutical Processes

Formally, the level of support of rule $A_1 \Rightarrow A_2$ is defined as $supp(A_1 \Rightarrow A_2) = \dfrac{|g(A_1 \cup A_2)|}{|O|}$. Based on this definition, the confidence of rule $A_1 \Rightarrow A_2$ is defined as $conf(A_1 \Rightarrow A_2) = \dfrac{supp(A_1 \cup A_2)}{supp(A_1)}$.

6.1. Example

Fig. 2 shows an illustrative example of past cases D1,…,D11 organized as a context table. Incidents are reported to occur when carrying out recipes R1 and R2 with cases C1,…,C5 and consequences E1,…,E3. It is immediately apparent that cases labeled D1, D2, D3 constitute a formal concept because they share exactly the same attributes {R1, C1, C2, E1} not shared by any other object. Similarly, cases labeled D11, D12 constitute another formal concept with attributes {R1, C2, E1}. From the lattice it can be seen that {R1, C1, C2, E1} $\subseteq$ {R1, C2, E1}.

	R1	R2	C1	C2	C3	C4	C5	E1	E2	E3
Obj1	×		×	×				×		
Obj2	×		×	×				×		
Obj3	×		×	×				×		
Obj4		×		×				×		
Obj5		×				×	×		×	
Obj6		×				×	×		×	
Obj7		×				×	×		×	
Obj8		×			×			×	×	
Obj9	×	×						×		×
Obj10	×	×						×		×
Obj11	×			×				×		
Obj12	×			×				×		
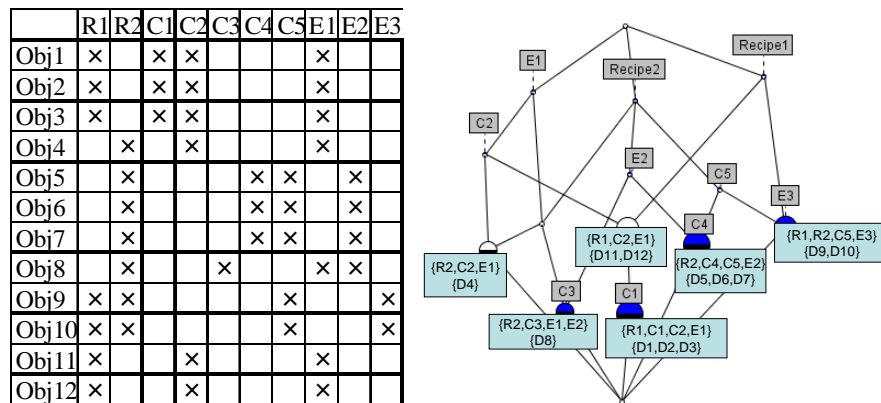
Figure 2. A context table incident cases and its corresponding lattice

There are three cases related to recipe R1 in which C1, C2 are the cause of E1. From a simple look at these individual cases, three causality alternatives are possible: {R1, C1, E1}, {R1, C2, E1} or {R1, C1, C2, E1}. However, by taking into account other cases, FCA concludes that C1 alone cannot be considered a cause of E1. Note that {R1, C1, E1} is not included in the lattice, while keeping {R1, C2, E1} and {R1, C1, C2, E1}. This observation is also supported by the association rules in Fig. 3.

Notice that rules 3, 5 support the causality of {R1, C2, E1} and rule 8 supports {R1, C1, C2, E1}. Specifically, rule 8 suggests that C1 and C2 are probably interrelated.

6.1.1. Direct causality

Rule 6 indicates that there is a strong support to conclude that independently of which recipe is carried out C2 causes E1.

6.1.2. Relations between recipes and events

From rule 6 and from the lattice it can be seen observed that E2 occurs only when the production is carried out with recipe R2.

1 < 6 > C2 =[100%]=> < 6 > E1;	8 < 3 > C1 =[100%]=> < 3 > R1 C2 E1;
2 < 7 > E1 =[86%]=> < 6 > C2;	9 < 3 > C4 =[100%]=> < 3 > R2 C5 E2;
3 < 5 > R1 E1 =[100%]=> < 5 > C2;	10 < 2 > R1 R2 =[100%]=> < 2 > C5 E3;
4 < 5 > C5 =[100%]=> < 5 > R2;	11 < 2 > E3 =[100%]=> < 2 > R1 R2 C5;
5 < 6 > C2 E1 =[83%]=> < 5 > R1;	12 < 1 > R2 E1 E2 =[100%]=> < 1 > C3;
6 < 4 > E2 =[100%]=> < 4 > R2;	13 < 1 > C3 =[100%]=> < 1 > R2 E1 E2;
7 < 3 > R2 C5 E2 =[100%]=> < 3 > C4;	

Figure 3. Association rules

6.1.3. Mutually exclusive events

From the lattice it can be noticed that consequences E1 and E2 can occur simultaneously. Rule 12 indicates that there is a probability although small that E1 and E2 are related. Contrasting with this, neither the lattice nor the association rules indicates that E3 can take place along with E1 and E2. The latter situation is an example of exclusive events which can be used to adapt a case.

7. Conclusions

This paper presents a framework for storing, maintaining and retrieving information about past corrective actions of incidents by combining case-based reasoning, ontologies, and formal concept analysis. Preliminary observations indicate that it is possible to identify mutually exclusive classes of events, direct causality (when the intent includes only one cause and only one consequence), potential multiple causality (when the intent includes two or more causes and one consequence) and the relations between events and other entities. However, much work is needed in several components of the framework.

8. Acknowledgements

References

R. Batres, M. West, D. Leal, D. Price, K. Masaki, Y. Shimada, T. Fuchino, Y. Naka. An Upper Ontology based on ISO 15926, Computers and Chemical Engineering, 31, 519–534 (2007)

M. Devaney, A. Ram, H. Qiu, J. Lee. Preventing Failures by Mining Maintenance Logs with Case-based reasoning. 59th Meeting of the Society for Machinery Failure Prevention Technology (2005)

A. Estacio-Moreno, Y. Toussaint, C. Bousquet. Mining for Adverse Drug Events with Formal Analysis. 21st Int. Congress of the European Federation for Medical Informatics (2008)

L. Lakhal and G. Stumme. Efficient Mining of Association Rules Based on Formal Concept Analysis. In B. Ganter, G. Stumme, R. Wille (Eds.) Formal Concept Analysis, Foundations and Applications. Lecture Notes in Computer Science 3626 Springer 2005

J.F. van Leeuwen, M. J. Nauta, D. de Kaste, Y.M.C.F. Odekerken-Rombouts, M.T. ldenhof, M.J. Vredenbregt and D.M. Barends. Risk analysisnext term by previous termFMEA as an element of analytical validationnext term. Journal of Pharmaceutical and Biomedical Analysis, 50(5), (2009) 1085-1087

C. Riesbeck, and R. Schank. Inside Case-based Reasoning. Northvale, NJ, Erlbaum (1989)

R. Wille. Restructuring lattice theory: an Approach based on Hierarchies of Concepts, In I. Rival (Ed.), Ordered sets. Reidel, Dordrecht-Boston, (1982) 445–470

H. N. Winkle. Implementing Quality by Design. Evolution of the Global Regulatory Environment: A Practical Approach to Change, http://www.fda.gov/downloads/AboutFDA/ CentersOffices/CDER/ucm103453.pdf (2007)

A Quantitative Assessment of the Viareggio Railway Accident

Davide Manca, Sara Brambilla, Roberto Totaro

CMIC Department, Politecnico di Milano, 20133 MILANO, ITALY,
davide.manca@polimi.it

Abstract

The manuscript focuses on the LPG railway accident occurred in Viareggio at the end of June 2009. The accident investigation highlights the uncertainties and complexities in modeling the accident in the urban environment close to the railway station. The analysis allows assessing the sequence of events, the way they influenced each other, and the alternative possible evolutions. In particular, the paper describes and discusses suitable models to quantify the consequences of most probable accident dynamics and their results.

Keywords: Railway accident; Train derailment; LPG emission, explosion and fire; Accident investigation; Complexity and uncertainty

1. Introduction

On June 29th, 2009 a freight train with 14 tank cars carrying LPG went off the rails while it was travelling through the Viareggio station. The first tank crashed into a signaling stake and the LPG was released from a crack in the vessel. Consequently, the LPG flashed partially and the remaining liquid fraction spread onto the ballast. The dense-gas cloud originated after the emission moved towards the surrounding quarter that overlooks the station, penetrated in some basements and ground floors, and eventually it was ignited. The following explosion(s) and fires produced a high number of victims (thirty-one fatalities and more than thirty casualties) and major infrastructural damages (32 M€).

2. Accident modeling

This section of the paper focuses on a quantitative assessment of the accident from the overturning of the tank cars to the explosion and the burn down of some houses, automobiles, and people. In a number of points, we had to investigate different accidental scenarios because the exact timeline of events is unknown. This analysis, assisted by the witnesses and reports on the injuries, and damages, is the basis for a reliable and detailed accident investigation.

A quantitative assessment of the consequences of the Viareggio accident has to account for several interlinked and consequential phenomena, where the events to model are:

- the release of liquid propane from the crack in the tank car;
- the flash of the liquid jet in the atmosphere;
- the spreading and boiling of the LPG pool on the ballast;
- the dispersion of vapors emitted from the tank car and of those evaporated by the pool;
- the dilution of the gas cloud in presence of obstacles such as the permeable fence at

the railroad borders, and the houses on the cloud path;
- the ignition of gas pockets inside the houses and the magnitude of the consequent explosion;
- the ignition of the liquid pool, and the consequent pool fire.

3. The LPG release

The derailed and damaged tank car was a horizontal cylinder 15.95 m long and 3.04 m large and carried about 45 t of LPG. Since the hole was in the bottom part of the tank car, the release was liquid. During the discharge, no air could enter the car, consequently a fraction of the liquid evaporated to preserve the internal pressure, *i.e.* the vapor pressure at the liquid temperature. The evaporation subtracted energy to the liquid fraction that cooled down. The evaluation of the dynamics of the LPG inside the car, in terms of liquid and vapor masses, and temperature, calls for the solution of the following ordinary differential equation (ODE) system:

$$
\begin{cases}
\dfrac{dm_L}{dt} = -\dfrac{\dot{m}_{disch}}{\left(1 - \dfrac{\rho_V}{\rho_L}\right)} \\[4mm]
\dfrac{dm_v}{dt} = \dot{m}_{evap} \\[4mm]
\dfrac{dT}{dt} = -\dfrac{\dot{m}_{evap}\Delta H_{ev}}{m_L cp_L + m_V cp_V}
\end{cases}
\tag{1}
$$

Where $\dot{m}_{disch} = \rho_L A_h c_D \sqrt{\dfrac{2}{\rho_L}\left[P - P_a\right] + 2gh_L}$ and $\dot{m}_{evap} = -\dfrac{dm_L}{dt}\dfrac{\rho_V}{\rho_L}$ (the assumption that

the densities do not vary with the time is reasonable due to the small variation of the liquid temperature as shown in the following). The level of the liquid inside the carrier (h_L) can be evaluated from the volume of liquid (V_L) and the tank car geometrical dimensions, by considering it as a horizontal cylinder with flat bottoms:

$$
V_L = L\left[R^2 \operatorname{acos}\left(1 - h_L/R\right) - \left(R - h_L\right)\sqrt{2Rh_L - h_L^2} \right]
\tag{2}
$$

When all the liquid was released, the tank car was still under pressure and the left vapors were emitted through the hole in the choked flow regime. The discharge rate can be evaluated as (van der Bosch and Duijm, 2005):

$$
\dot{m}_{disch} = A_h\, c_D \sqrt{\rho_V\, P\, \gamma \left(\dfrac{2}{\gamma - 1}\right)^{\frac{\gamma+1}{\gamma-1}}}
\tag{3}
$$

Figure 1 shows the decrease of the liquid mass in time and the corresponding discharge rate. It follows that the 45 t of LPG were released in 262 s, *i.e.* 4 min and 22 s.

The evaluated temperature decrease due to the partial evaporation of the liquid fraction inside the tank car was ~11 K. This analysis is valid if the ignition occurred after the release of all the LPG from the tank car. Otherwise, the dynamics would change because of the heat exchanged between the damaged tank car and the fire. Actually, there was not a complete agreement among the witnesses of the Viareggio accident

about the ignition time. Probably the ignition occurred between 2 and 5 min after the release start.
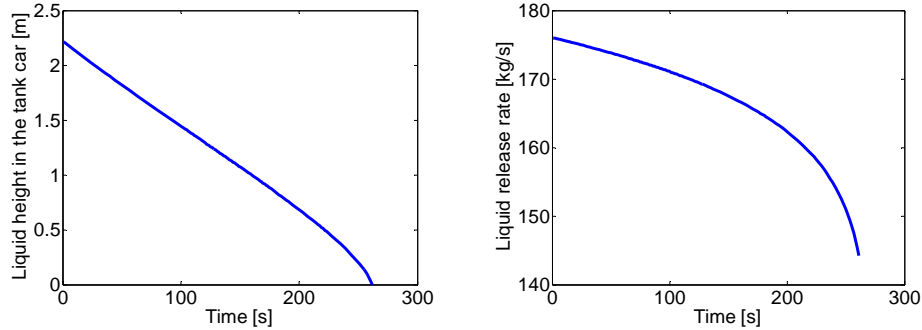

Figure 1. Dynamics of the liquid holdup in the tank (left) and the liquid release flowrate (right)

By assuming that the ignition occurred after 2 min, the pressure inside the tank car would have increased (during the release) up to ~18 bar, from the initial storage conditions of about 9 bar. In this case, the release time would have fallen to 236 s, *i.e.* 3 min and 56 s (Figure 2).
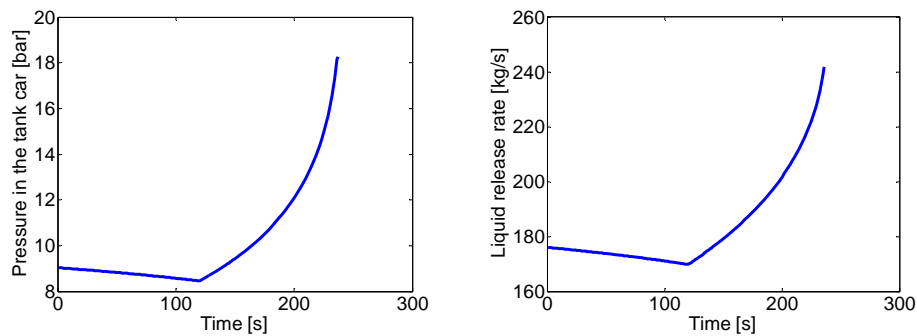

Figure 2. Internal pressure and released liquid flowrate in case of pool ignition after 2 min

4. Flash of the liquid jet

Once at ambient conditions, the liquid jet flashed and produced a two-phase jet. According to Hanna and Drivas (1987), the fraction of vapor can be evaluated as:

$$x_V = \frac{cp_L \left(T - T_{eb}\right)}{\Delta H_{ev}} \tag{4}$$

where T is the temperature of the jet and $T_{eb} = 231.1\,\text{K}$ is the pure propane boiling temperature since the real composition of LPG was unknown.

Eq. (4) allows evaluating the liquid and vapor mass fractions and the corresponding flow rates in case of ignition after both 2 min and 5 min.
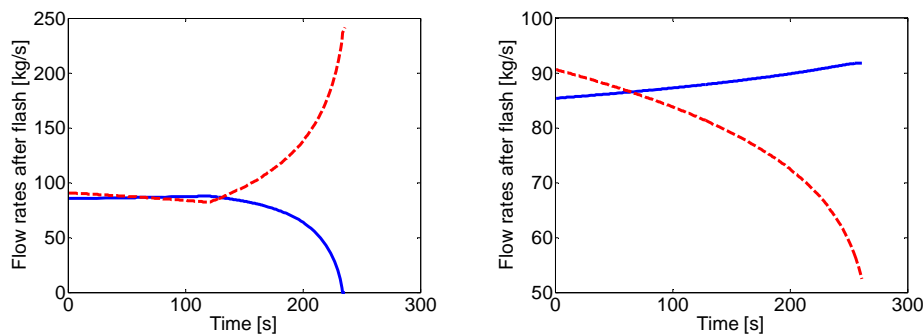
Figure 3. Liquid and vapor flow rates after the flash of the liquid jet emitted by the crack in case of ignition after 2 min (left) and 5 min (right)

5. Pool spreading and pool-fire

Since the crack in the tank car was close to the ground, the liquid jet impacted directly on the ballast and did not have the time and space to break up into drops. Therefore, the vapor fraction did not entrain any liquid fraction. Consequently, the liquid spread on the ground and formed a pool. The spreading of the LPG on the ballast, its evaporation, and delayed ignition were simulated with AXIM™, a software tool for the simulation of chemical accidents (Brambilla and Manca, 2009). By assuming that the pool was free to spread and expand, AXIM™ determined that, in case of ignition after 2 min, the pool reached a diameter of ~20 m, whilst the flame reached a maximum drag diameter of ~23 m and a height of ~41 m. In case of ignition after 5 min, the pool reached a larger diameter (~23 m), whilst the flame reached a maximum drag diameter of ~25 m and a height of ~44 m.

Probably, the aforementioned values overestimate the pool and flame dimensions because the liquid permeation into the ground and the presence of objects acting as a confinement were not taken into account. A reduced flame height of about 25 m, as reported by some eyewitnesses, was possible only if the pool spreading was inhibited by the morphology and features of the ground. Actually, the ballast was not a flat ground of pebbles. Instead, there were piles of rocks, depressions, and some obstacles. Therefore, the aforementioned obstacles, acting as an irregular confinement of about 10 m diameter, together with the permeable ballast, could have led to flames as high as 25 m.

Another issue to be considered is related to the hypothesis that all the liquid discharged from the crack in the tank car contributed to the pool.

Some photos taken immediately after the release show that separate pool-fires formed on the ballast and these cannot be ascribed to the combustion of the sleepers. On the contrary, they could be due to the spray of the liquid released to different places and directions because of the unevenness of the ballast, and the high release velocity (30 m/s). This explanation is supported by the train drivers, who told that they ran away trampling on the LPG pools on the ballast (Dellacasa, 2009). The locomotive head was approximately at 35 m from the accident and the pool (under the hypothesis of a single pool) did not reach so far. Probably, the liquid discharged from the crack did not form probably only one pool. Instead, it formed a main pool, and a set of different, isolated smaller pools.

6. The dispersion of the dense gas-cloud

The simulation of the dispersion of the dense gas cloud flashed from the liquid jet emitted by the punctured tank, has to account for the following phenomena:

- the gravity slumping, due to the high density of the gas cloud;
- the entrainment of fresh air that dilutes and heats up the cloud;
- the change of the cloud temperature due to heat exchange with the ground;
- the motion of the cloud with the local wind.

Due to the rather calm weather conditions of that night, the motion of the cloud by the wind was negligible respect to the gravity slumping. AXIM™ simulated the gas dispersion on the railroad, the ballast, and finally through the rows of surrounding buildings by means of the shallow water equations model (Hankin and Britter, 1999). We simulated the congested environment in terms of buildings and infrastructures (*e.g.*, fences, cabins, pylons). The positions of the obstacles in the surrounding of the accident were determined with a georeferenced software (Google Earth™). For the sake of simplicity, we assumed that the buildings are 10 m high (Figure 4).
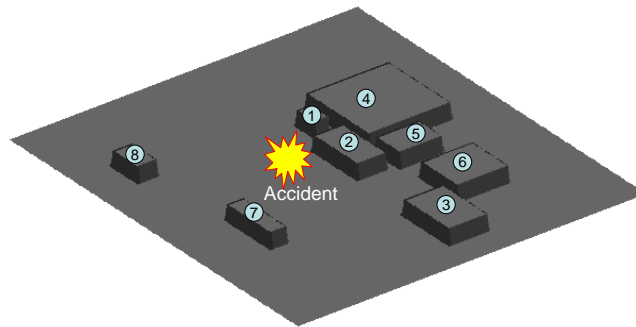
Figure 4. Simplified 3D reconstruction of the buildings in the surroundings of the accident
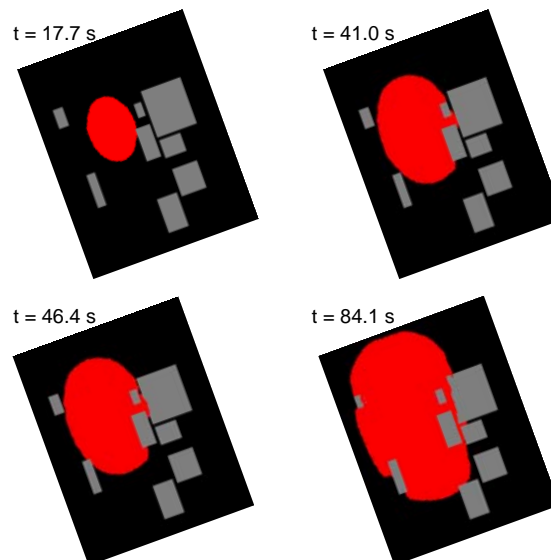
Figure 5. Dispersion of the dense gas cloud at different times after the release start. It is possible to observe the channel effect of the dense gas cloud among the buildings

Figure 5 shows the dispersion of the gas cloud in the area close to the accident and its spreading over the buildings and in the street canyons. It is possible to notice that the cloud arrived at the closest building after only 17.7 s, then it reached the second rows of buildings at about 41 s. Afterwards, the cloud reached the buildings on the west side of the accident after ~46 s. The picture on the right bottom corner shows that after about 84 s the cloud arrived also at the farthest building on the right bottom. The modeling of the dense gas dispersion motion within the rows of houses shows that after approximately 90 s the propane cloud arrived at all the locations where the explosions occurred. This is in good agreement with the witnesses who reported that the explosions occurred after only a couple of minutes from the train derailment.

7. The explosions

The evidence tells us that the explosions occurred inside the houses due to the penetration of the dense gas through the open windows in the basement and ground floors. Buildings are generally not very resistant to overpressures due to explosions that are triggered from inside. In fact, an overpressure of 7 kPa is often enough to destroy a typical brick building (Lees, 2004). On the other hand, the presence of weaker elements in the walls (such as windows) which fail first provides vents to the explosion, resulting in lower overpressures. Conventional windows fail at overpressures of 3–4.6 kPa, strained windows at 0.2–1 kPa, brick walls (114 mm thick) at 35 kPa, and brick walls (228 mm thick) at 105 kPa (Lees, 2004).

The dense gas model discussed in the previous section cannot simulate the penetration into buildings. Consequently, we did not determine the amount of LPG that permeated into the houses and eventually exploded. In addition, it is supposed that there were not explosions external to the houses, consequently, conventional methods as the TNT equivalent method or the multienergy method (Lees, 2004; Mercx and Duijm, 2005) do not apply to the Viareggio accidental event.

8. Conclusions

The dynamic analysis of the accident showed how fast were the emission and dispersion of the LPG gas cloud towards the surrounding houses. It took less than 90 s for the dense-gas cloud to reach the furthest house that eventually exploded. Such a short time inhibits any emergency-response activities aimed at reducing the accidental outcomes.

References

Brambilla S., D. Manca, "Accidents Involving Liquids: a Step Ahead in Modeling Pool Spreading, Evaporation and Burning", Journal of Hazardous Materials, 161, 1265–1280, 2009
Dellacasa E., Corriere della Sera (Italian newspaper), 1-Jul-2009, 2009
Hankin R.K.S., R.E. Britter, "TWODEE: the Health and Safety Laboratory's Shallow Layer Model for Heavy Gas Dispersion. Part 1. Mathematical Basis and Physical Assumptions", Journal of Hazardous Materials, **A66**, 211–226, 1999
Hanna S.R., Drivas P.J., "Guidelines for use of Vapor Cloud Dispersion Models", American Institute of Chemical Engineers, New York (NY – USA), 1987
Lees F.P., "Loss Prevention in the Process Industries", Third Edition, Elsevier, Oxford, 2004
Mercx W.P.M., Duijm N.J., "Vapor Cloud Explosion", in Methods for the Calculation of Physical Effects due to Releases of Hazardous Materials (Liquid and Gases), Ed. van der Bosch C.J.H., Weterings R.A.M.P, Netherlands, 2005
van der Bosch C.J.H., N.J. Duijm, "Outflow and Spray Release", in Methods for the Calculation of Physical Effects due to Releases of Hazardous Materials (Liquid and Gases), Ed. van der Bosch C.J.H., Weterings R.A.M.P, Netherlands, 2005

Input Reconstruction for Statistically Enhanced Fault Detection and Isolation

Udo Schubert[a,b], Uwe Kruger[b], Harvey Arellano-Garcia[a], Günter Wozny[a]

[a]*Chair of Process Dynamics and Operation, Berlin Institute of Technology, Str. d. 17. Juni 135, 10623 Berlin, Germany,* *udo.schubert@zmms.tu-berlin.de*
[b]*Department of Electrical Engineering, The Petroleum Institute, P.O. Box 2533, Abu Dhabi, United Arab Emirates*

Abstract

This paper proposes a unified framework for actuator and process fault detection and isolation. This framework combines methods from model-based FDI and multivariate statistical process control. The complementary nature of both disciplines is utilized to facilitate the interpretation of non normal operation conditions (NNOC) in complex large-scale processes by plant operators. The main part of the unified approach is the reconstruction of system inputs for enhanced fault diagnosis. The utility of the proposed unified approach is demonstrated using a simulation example and the analysis of industrial process data, exemplifying that improved fault isolation can be achieved.
Keywords: Fault Detection and Isolation, Fault Diagnosis, Input Reconstruction, Operator Performance

1. Introduction

Several factors contribute to the difficulties in monitoring complex processes in the chemical and manufacturing industries. From the system's perspective, the continuously increasing complexity (e.g. through higher levels of automation, process intensification, optimal operation at the constraints) is the major contributor. From the personal perspective of plat operators, however, effects arising from out-of-the-loop phenomena also have to be taken into account when designing process control systems. Together with the ever more stringent legislation on emissions and process safety, this yields the requirement to implement automated process monitoring methods.

The development of suitable and efficient methods must, on the other hand, support both, the monitoring of dynamic operation by plant personnel and, at the same time, offer the ability of transparent fault detection and isolation (FDI) to mediate a deeper process understanding. Significant progress has been achieved in this area by developing approaches that utilize causal models in model-based fault detection and identification (MBFDI), as well as statistical models in multivariate statistical process control (MSPC). Whereas MSPC can handle large scale systems including many variables, the treatment of pronounced dynamic behavior is not satisfactory [1]. In contrast, causal models utilized in MBFDI typically aim for an accurate description of complex dynamic behavior, but the expenses for model development substantially increases with process complexity [2].

To overcome the limitations of individual approaches, several contributions suggest a combination of MBFDI and MSPC, e.g. reference [3]. However, a framework that utilizes methods from both domains to detect and diagnose different faults has not yet been proposed in the literature. This article proposes a unified approach for actuator and process FDI and shows the utility of this approach using two application studies.

2. Unified Approach

This article proposes the detection and subsequent diagnosis of abnormal operation conditions in complex chemical and manufacturing processes using MBFDI and MSPC by introducing a unified approach. While MBFDI is well suited for fault diagnosis because of the causal nature of utilized models, it is usually not feasible to create first-principles models for large scale processes [1]. Furthermore, a priori knowledge about the dynamic properties of predetermined disturbances is required and therefore MBFDI schemes are often restricted to the detection and diagnosis of specific fault scenarios [2]. In contrast to MBFDI, MSPC based process monitoring is robust in detecting departures from steady state operation. However, the conventional fault diagnosis using contribution plots [4] or variable reconstruction based contributions [5] can show ambiguous or misleading results, if the abnormal condition is caused by a complex fault. This stems from the associated complex interactions between recorded process variables, which can yield that a faulty variable affects several other output variables at the same time [1]. A contribution plot consequently associates a large number of contributing variables in such a case. This, however, complicates isolation of a root cause, because the pattern of contribution variables may not be distinctive enough for different fault conditions.
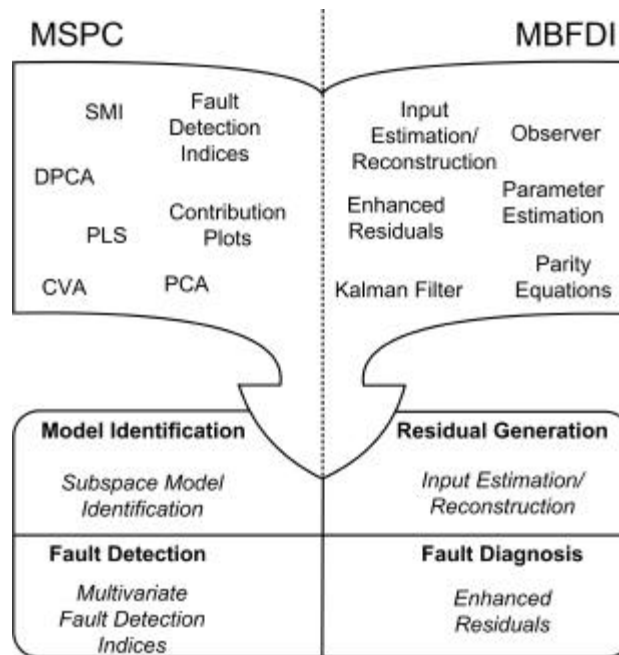
Fig. 1: Unified approach with methods from MBFDI and MSPC to fulfill the individual tasks in automated FDI.

While the assistance of operators through MBFDI or MSPC provides a promising perspective for human-factors engineering, it is lacking a generic fault diagnosis capability. The output that is generated does not support the prevention of fading mental models that operators are faced with, due to automation induced skill degradation, e.g. out-of-the-loop phenomena [6].

Therefore the core of the unified approach consists of a generic fault diagnosis capability that enables the operator to acquire knowledge about causal relations among

process variables and integrates expertise into process supervision. This is realized through the reconstruction of process inputs using measured outputs and a comparison with the measured inputs.

The unified approach is schematically depicted in Fig. 1. Existing methods in both domains at the top are assembled in a unified MBFDI/MSPC scheme below to achieve efficient supervision of complex dynamic processes with respect to robust fault detection, as well as a generic fault diagnosis capability. Using subspace model identification (MSPC based) to create suitable process models, the problem of existing first-principle models is circumvented. State observers (MBFDI based) together with recordings of input/output variables use these models to track the process' behavior with respect to common cause variations, model uncertainty and also during faulty conditions. The generated model residuals are further analyzed by multivariate fault detection indices (MSPC based) to detect abnormal behavior with a single measure.

3. Input Reconstruction for Enhanced FDI

This work is based on the discrete time-invariant causal models in state-space representation following Eqn (1).

$$\mathbf{x}(k+1)=\mathbf{A}\mathbf{x}(k)+\mathbf{B}\mathbf{u}(k)+\mathbf{E}\mathbf{w}(k), \ \mathbf{y}(k)=\mathbf{C}\mathbf{x}(k)+\mathbf{D}\mathbf{u}(k)+\mathbf{F}\mathbf{v}(k) \tag{1}$$

Where $\mathbf{x} \in \mathrm{R}^{n}$ is the state vector, $\mathbf{y} \in \mathrm{R}^{l}$ is the output vector, $\mathbf{u} \in \mathrm{R}^{m}$ is the input vector and $\mathbf{v} \in \mathrm{R}^{l}$ is measurement noise, whereas $\mathbf{w} \in \mathrm{R}^{n}$ represents process noise for the *kth* sample. Practical experience has shown, that industrial processes can be approximated with sufficient accuracy for specific operation regions by the model structure in Eqn (1) [7]. Using recorded input/output data, subspace model identification (SMI) methods simultaneously estimate the appropriate system order, calculate system matrices $\mathbf{A}, \mathbf{B}, \mathbf{C}$ and $\mathbf{D}$, and also extract statistical parameters describing the disturbance distribution during normal operation [7].

Using state observers, the output measurements can be used to track the dynamic behaviour of a process in the state space. Regarding common cause variations like process or sensor noise the state-to-output relations in (1) will remain consistent. With upcoming abnormal behaviour (e.g. caused by model error, unmeasured disturbances or faults), however, the state-to-output relation will not yield the fault condition by examining output residuals because of the observer feedback. Instead, the input-to-state relationship in (1) must describe the mismatch. Therefore, we assume that the nominal input $\mathbf{u}_{0}(k)$ is superimposed with an input residual vector $\Delta\mathbf{u}(k)$ that is required to compensate the presence of a fault condition (2).

$$\mathbf{u}(k) = \mathbf{u}_{0}(k) + \Delta\mathbf{u}(k) \tag{2}$$

$$\Delta\hat{\mathbf{u}}(k) = \mathbf{B}^{\dagger}\left(\hat{\mathbf{x}}(k+1) - \mathbf{A}\hat{\mathbf{x}}(k) - \mathbf{B}\mathbf{u}(k)\right) = \mathbf{B}^{\dagger}\left(\hat{\mathbf{x}}(k+1) - \hat{\mathbf{x}}(k+1\,|\,k)\right) \tag{3}$$

Inserting (2) into Eqn. (1) and inversion w. r. t. the residual input vector and neglecting $\mathbf{w}(k)$ yields the reconstruction of the input residual (2) from estimated states (3), given a system for which $n \geq m$. The input residuals (3) can be assumed to follow a multinormal Gaussian distribution in the fault free case, because $E\{\Delta\hat{\mathbf{u}}\} = \mathbf{0}$, following

from $E\{\mathbf{v}\} = E\{\mathbf{w}\} = \mathbf{0}$. We can therefore calculate a T^2 statistic that is sensitive to process and actuator faults that affect the state-to-input relationship (5). The error covariance matrix can be determined from normal operation condition (NOC) reference data. The corresponding control limit is determined from a F-distribution using m degrees of freedom, the number of reference samples K, as well as a given sensitivity a (e.g. 99%). The statistic then yields $T^2(k) \le T_a^2$ during NOC, whereas $T^2(k) > T_a^2$ implies the presence of a fault (NNOC).

$$T^2(k) = \Delta\hat{\mathbf{u}}^T(k)\hat{\mathbf{S}}_{\Delta u \Delta u}\Delta\hat{\mathbf{u}}(k) \tag{5}$$

A subsequent diagnosis consists of a thorough analysis of generated input residuals. Because the model describes the relationship between the input and output variables, the input residuals provide a quasi-causal interpretation that allows for a diagnosis of the root cause behind detected faults. The following case studies demonstrate the usefulness of analyzing the patterns described by the $\Delta\hat{\mathbf{u}}(k)$ sequence during fault conditions.

3.1. Simulation Study

A simulation example is discussed first, to illustrate the calculation of T^2 statistics using input reconstruction. The SMI procedure is omitted, because a suitable state space model is available. The simulation model is the discrete-time version of the fourth-order process found in [8]. The discrete model has been obtained using a zero order hold with $dt = 0.01\ sec$. This process has a single input and the corresponding actuator is considered to be faulty here. From the output variables, only the third state variable cannot be measured.
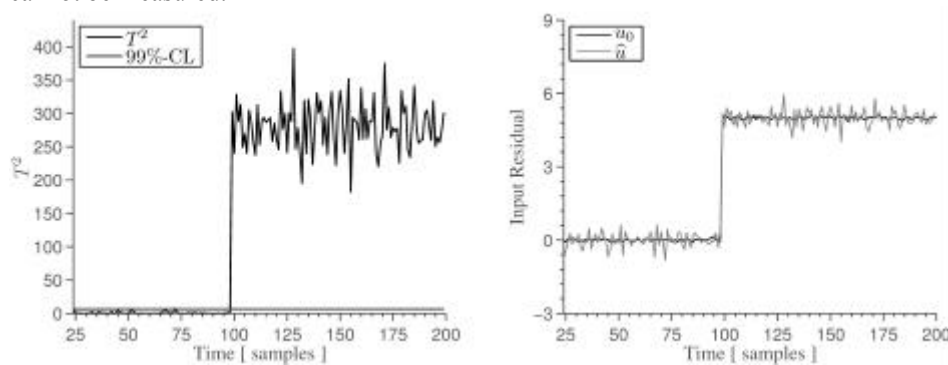

Fig. 2: T^2 statistic (left) for detection of an actuator fault and subsequent fault diagnosis (right).

A period of *1000* intervals representing NOC has been simulated with sinusodial input superimposed with gaussian noise and output measurements were also corrupted with noise. Here, an unknown input observer has been designed for state estimation to be robust against deviations of the true input from the measured, nominal input $\mathbf{u}_0$. The left plot in Fig. 2 clearly indicates a violation of the T^2 control limit for a test dataset, in which an actuator fault has been introduced after *100* samples of NOC. Using the input residual (3), the fault signature could be reconstructed successfully. Whereas this example does not include a difficult fault isolation step, this problem is covered in the next section together with the model identification step.

Input Reconstruction for Enhanced Multivariate Statistical Process Control and Fault Detection
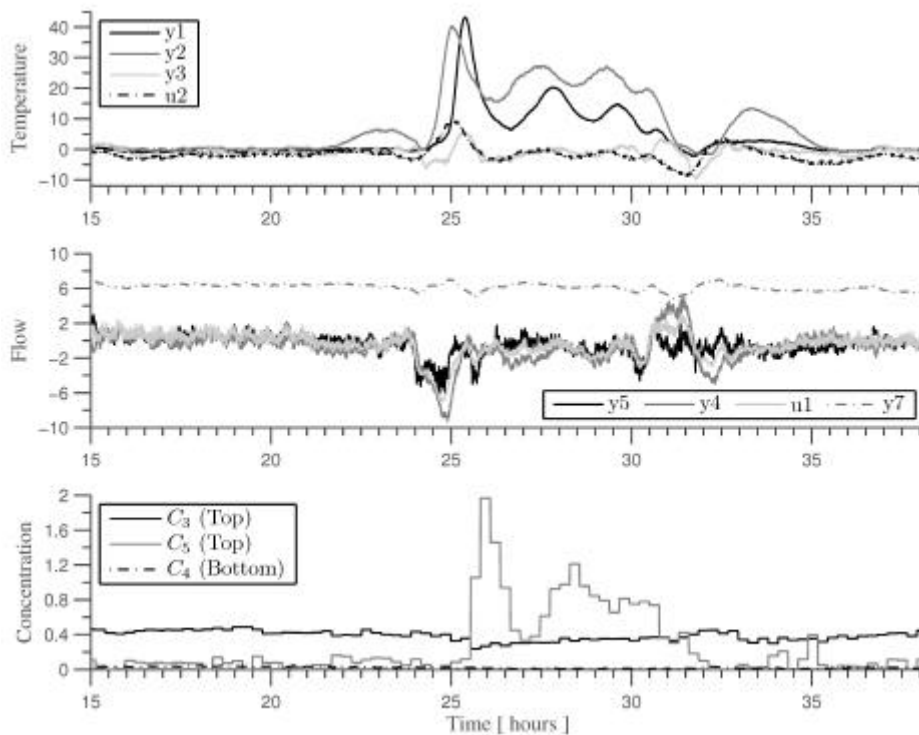

Fig. 2: Disturbance propagation resulting from a feed flow disturbance (scaled units for temperatures and flows, +6 mean shift for u3).

3.2. Industrial Distillation Column

The distillation column separates butane C_4 from heavier hydrocarbons with quality constraints for top and bottom product compositions. The plant is operated in closed loop configuration, corresponding to the *LV-configuration* with an additional temperature control loop, which is known to be sensitive towards feed flow disturbances at high frequencies [9]. In this specific case, the constant feed flow (u1), frequently drops in level by up to 30% and the dynamic response, depicted in Fig. 2, shows a delayed increase of heavy components in the distillate which violates operational constraints. The reason is inadequate controller action, in terms of excessive steam flow (y7) in the reboiler. However, a simple monitoring of concentration or feed flow measurements is not sufficient, because it either introduces a large detection delay, or restricts the column operation to a single steady state. Therefore, a model-based monitoring approach is required, to assess whether a significant departure from NOC arises, with dynamic properties, that are likely to affect the product composition in the near future, or if a rather gradual transition is observed, that will not have a similar effect.

Following [10], the feed flow is the main source for a disturbed energy balance if it does not provide a constant feed temperature or enthalpy. Both a reduced feed flow, and an increased feed temperature require less steam flow and more reflux to preserve the established temperature profile and compositions. From a period of NOC that consists of *1850* samples with *dt = 30* sec, *l = 7* and *m = 3*, SMI estimates a state space model with order *n = 8*. The T^2 statistic for input residuals calculated from (5) in the left plot

of Fig. 3. shows a clear violation of the control limit after *10-15 min*. From the right plot in Fig. 3 it is obvious, that the estimated trajectory in the state space can be explained by a feed temperature that is considerably higher than the measured value (u2). In contrast, there is only little uncertainty about the feed flow itself (u1) and the reflux flow (u3). This clearly indicates a disturbed energy balance due to inappropriate control action. The timely alarm through the single statistic in the left plot gives an operator sufficient time for situation assessment and taking compensating measures to stay in specification.
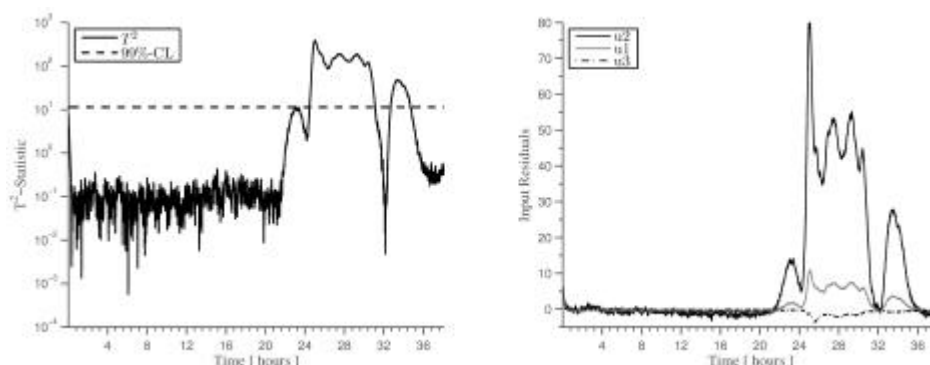
Fig. 3: T² statistic and input residuals for detection and diagnosis of a disturbed energy balance.

4. Conclusions

In this work, a unified MBFDI/MSPC scheme for actuator and process fault detection has been proposed. The combination of methods from both domains clearly improves the performance of the individual approaches and helps to overcome their limitations. The application of MBFDI for large scale processes is greatly simplified through the use of identified models estimated with SMI. The integration of properly defined statistical indices and control limits for multivariate residual evaluation enhances robustness. The potential for ambiguous output in MSPC has been circumvented through the generation of input residuals. At the same time, this form of visual aid supports the operator during situation assessment after a control limit violation but also helps to recapture well-known causal relations and update established mental models.

5. Acknowledgements

References

[1] S. Yoon, J.F. MacGregor, J Process Contr, 11, 4, (2001) 387-400
[2] S.X. Ding, Model-Based Fault Diagnosis Techniques (2008)
[3] V. Venkatasubramanian, et al., Comp Chem Eng, 27, 3, (2003) 293-311
[4] T. Kourti, J.F. MacGregor, J Qual Tech, 28, 4, (1996) 409-428
[5] D. Lieftucht, et al., Cont Eng Prac, 17, 4, (2009) 478-493
[6] C.D. Wickens, et al., Introduction to Human Factors Engineering, Prentice Hall (2003)
[7] S.J. Qin, Comp Chem Eng, 30, 10-12 (2006) 1502-1513
[8] C. Edwards, Decis Contr, (2004)
[9] S. Skogestad, Chem Eng Res Des, 85, 1 (2007) 13-23
[10] P.S. Buckley, et al., Design of Distillation Control Systems, Edward Arnold (1985)

The Estimation of Hazardous Gas Release Rate Using Optical Sensor and Neural Network

Won So[a], Jamin Koo[a], Dongil Shin[b], En Sup Yoon[a]

[a]*Seoul National University, 311-504, Shillim-dong, Kwanak-gu, Seoul 151-744, Korea, sowon3036@pslab.snu.ac.kr*
[b]*Myongji University, Yongin, Kyunggido 449-728, Korea*

Abstract

A real-time monitoring and estimation technique is proposed for managing facilities that store hazardous materials. It relies on Gaussian dispersion model, optical sensor and neural networks for the detection and analysis of hazardous (gas) releases. We used an optical sensor network that has already been set surrounding the area where hazardous gas releases can occur. From the real-time sensor data, we detected and analyzed releases of harmful materials and their concentrations. Based on the results, the release rate is estimated using the neural network. This method consists of 14 input variables (sensor data, material properties, process information, and meteorological conditions) and one output (release rate). The dispersion model then performs the simulation of the expected dispersion consequence. The proposed technique provides calculated values that are within acceptable differences from those of commercial software like PHAST and TRACE. In comparing with commercial software, a series of scenarios are developed that involve commonly used hazardous materials such as ammonia and chlorine. The results indicate that the proposed technique can estimate release rates effectively within seconds, with 14 input variables that are far less in number than those required by previous methodologies.

Keywords: hazardous gas, release rate, neural network, optical sensor

1. Introduction

Economic development and ever-increasing adoption of technology for everyday life have unavoidably introduced many dangerous facilities inside and nearby city. Accidental discharges of hazardous gases, either flammable or toxic, more likely to occur during manufacture, storage or transport can be considered as one of those cases.

A feeling of insecurity for residents living nearby dangerous facilities increased, but response technologies are not accepted to be adequate [1, 2]. Many studies on real-time monitoring are progressing. The research of monitoring system based on sensor network is practical to be applied in the field. However, existing methods are not easy to predict the release of toxic material using sensor-network information. Thus, this research proposes a tacking methodology for estimating release rates of toxic gases.

Initially, the research focus was on designing and developing sensors that can meet industrial, practical needs; these included desirable characteristics such as automatic and immediate fault indication, fast response time, and elimination of the poisoning problem [3]. In these days, sensors were assumed to be placed in the proximity of facilities known to contain and/or process hazardous materials. With the advent of global terrorism, however, the assumption could not hold any more and consequently, a far greater area of territories became in need of monitoring. This promoted additional research on the placement of sensors in the network [4].

In this paper, an efficient, real-time monitoring technique is proposed that can estimate the hazardous gas release rate by using optical sensors and neural networks. Case studies with chlorine and ammonia will be given to illustrate the implementation process. The proposed technique can provide estimates of release rates that can improve the accuracy and availability of information key to the success of emergency information delivery system.

2. Theory

The availability of various dispersion models, artificial neural networks, and sensory equipments allowed the development of the monitoring technique proposed here. Thus, conceptual and practical description of each will be given throughout the rest of this section.

2.1. Dispersion models

There are three types of dispersion models that vary in complexity and application domains: three-dimensional, slab, and Gaussian. Three-dimensional models solve conservation of mass, momentum and energy equations and are especially effective in modelling the dispersion of clouds in complex geometry with obstacles. Slab models are particularly useful with heavy gas dispersion. Gaussian models assume the distribution of dispersed molecules with standard deviations dependent on the atmospheric conditions and distance from the source [1].

In the context where the proposed technique is to be applied, geometries of areas of interests may be complex but can be of a wide variety. Furthermore, hazardous gases whose release the technique attempts to monitor may be light or heavy. As a result, Gaussian model is selected as the dispersion model and the commercially available software PHAST is used in training the neural networks.

2.2. Neural networks

Neural networks are composed of simple elements operating in parallel. These elements are inspired by biological nervous systems. As in nature, the network function is determined largely by the connections between elements. One can train a neural network to perform a particular function by adjusting the values of the connections (weights) between elements. In general neural networks are adjusted, or trained, so that a particular input leads to a specific target output. The network is adjusted based on a comparison of the output and the target iteratively until the network output matches the target. Typically a large number of input/target pairs are needed to train a network. Neural networks have been trained to perform complex functions in various fields, including pattern recognition, identification visions and control systems [5].

2.3. Sensory Equipments

Different types of sensors are now available that can detect and measure concentrations of gaseous molecules in the perimeter. Among them, optical sensors appear attractive for their low cost, applicability and high sensitivity [6]. These sensors use the changes of optical absorption spectrum in reaction with (gaseous) molecules for detecting the target concentration. The ones used in the proposed technique rely on 4 subunits for measuring the concentration: light-emitter, optical receiver, spectrum analyzer and concentration computing machine (Fig. 1).
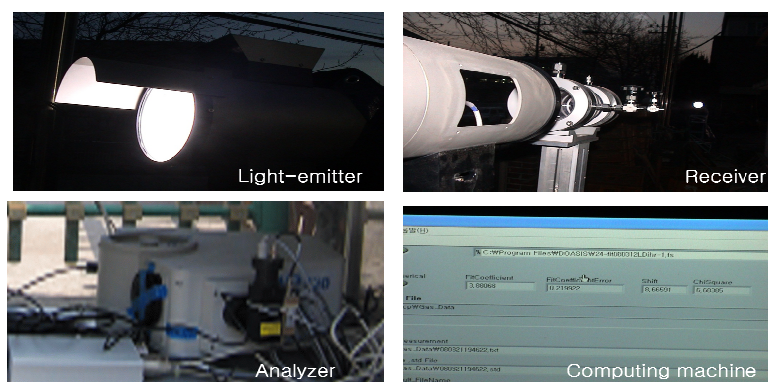
Fig 1. Composing parts of the singular sensor network

3. Methodology

The combination of optical sensor network, back-propagation neural networks, and the Gaussian dispersion model can build an effective, real-time monitoring system for storage vessels containing hazardous materials. For this to be true, the combined monitoring system needs to meet the following conditions. First, it has to be able to detect and measure concentrations of the releases of hazardous materials in an isotropic manner, which means regardless of the location of source or direction of release; it needs to do so in a cost-effective way. Second, the combined monitoring system should be able to predict accurately the dispersion of the released hazardous material(s) in a timely manner—that is as soon as possible, so that immediate response can be taken. Lastly, it needs to meet the first two conditions by using real-time input data that are available with acceptable costs. The proposed technique satisfies these conditions in the manner described in the following sub-sections.

3.1. The selection of input data
In order to predict accurately the dispersion of the released hazardous material(s) in a timely, cost-effective manner, an accurate dispersion model and its necessary input data are required for training neural networks. For the reasons discussed in Section 2.1, the Gaussian model has been selected as the dispersion model. Consequently, the necessary input data stem from the variables required by the model in predicting the dispersion of the released molecules. In terms of PHAST, 14 input data of three types—source information, GIS (Geographic Information Systems) data, and meteorological conditions—are needed.

If all the necessary data are available at zero or negligible costs, the dispersion model and neural networks can be run in their maximum accuracy. This is rarely the case in practice, however, as it sometimes costs a significant amount of capital and labor for collecting all of the necessary data. For example, instead of installing its own anemometer, the user of the technique may rely on the wind speed of the city where it belongs to that is given by the weather casting center; this is likely to make the prediction of dispersion less accurate since the specific value of wind speed may show a significant variance within the city.

As a result, one may be able to use only some of the data if the costs and other conditions appear unaffordable. If this is the case, it is necessary to choose the variables with the highest parametric sensitivity so that the greatest accuracy of the results is

obtained. In doing so, the data in Table 1 will be helpful. It is organized in terms of input variables' parametric sensitivities s_1 and s_2 with respect to release rate, and dispersed concentration to be calculated by the following equation:

$$s_{1i} = \Delta Q_m / \Delta d_i \tag{1a}$$

$$s_{2i} = \Delta C_{100} / \Delta d_i \tag{1b}$$

where s_1 parametric sensitivity of an input variable with respect to release rate, s_2 parametric sensitivity of an input variable with respect to dispersed concentration, Q_m represents the release rate in kg/s, C_{100}[1] the concentration (ppm) of dispersed molecules at 100 meters away from the source, d_i an input data of type i, and operator Δ meaning the change (difference) over the calculated interval.

In preparing the sample data for calculating the sensitivities it is important to distinguish the data of light gases from those of heavy ones. The two show significant difference in their dispersion behavior, as well as the relative influences of parameters on the dispersion phenomena. For example, light gas molecules forms a cloud having similar vertical and horizontal dimensions when released while the heavy cloud slumps toward the ground under the influence of gravity and thus has reduced height. Taking this into account, the parametric sensitivities in Table 1 have been divided into two groups, one for light gases and the other for the heavy. As one may observe, the magnitude and relative ranking of parametric sensitivities differ for the two groups, with the meaningful exception that wind velocity has the greatest sensitivity for both. The detailed order may not be the same if different materials are used for calculating the parametric sensitivities; the important point here is that it is necessary to check and make use of the parametric sensitivities of input data when applying the proposed technique.

Table 1. Parametric sensitivities of input variables with respect to release rate and effect distance

Data	Release rate		Dispersion	
	Light gas	Heavy gas	Light gas	Heavy gas
Distance from source	0.0000	0.0000	0	0
Concentration	0.0000	0.0000	1	1
Wind velocity	0.0000	0.0000	649	375
Atmospheric stability	0.0000	0.0000	220	88
Temperature	0.0000	0.0000	67	21
Relative humidity	0.0000	0.0000	14	3
Storage pressure	0.0444	0.0623	118	46
Storage temperature	0.0024	0.0029	13	10
Storage quantity	0.0000	0.0000	3	5
Molecular weight	0.0443	0.0443	105	105
Density	0.0298	0.0674	157	251
Boiling point	0.0000	0.0000	0	0
Melting point	0.0000	0.0000	0	0
Vapor pressure	0.0000	0.0000	0	0

3.2. Training the neural network

With With the installation of optical sensors in the manner described in Section 3.1 and the use of a variety of input data determined according to the criteria in Section 3.2, the remaining issue is how to train the neural network. Since the proposed technique is designed to estimate the hazardous (gas) release rate, this becomes the target output layer in the neural network. The input layer consists of the selected input data from Table 1. Then, the back-propagation algorithm is applied to calculate the weights between the input layer and the hidden layer of 20 units, and between the hidden layer and the output layer, under the MATLAB software package.

There are not many data available on the dispersion phenomena for two reasons. First, there have been a limited number of release accidents in the past. In Korea, for example, only 47 release cases have been reported in a major industrial complex during the past 35 years. Moreover, only a small portion of the reported cases have all of the necessary data. Consequently, it is recommended to use the data generated from the Gaussian dispersion model as the initial data necessary for training the neural network.

In fact, a total of 1,600 data generated from PHAST have been used for calculating the aforementioned parametric sensitivities. This is a size large enough to give R greater than 0.900 when a single kind of hazardous molecule is to be monitored; the specific value shows slight variance for different types of hazardous materials (Table 2). When more than one hazardous material is monitored, R remains acceptable while MSE increases like in Table 2. In this study, release rates of four different chemicals were estimated separately first, and then they were added to see how MSE changes (Fig. 2). As shown in the graphs, deviations from actual release rates do not increase substantially when the number of materials does not exceed 4. Indeed, MSE manages to be less than 0.1, which is still acceptable under the general view.

Table 2. Changes in MSE of the trained neural network with respect to number of target

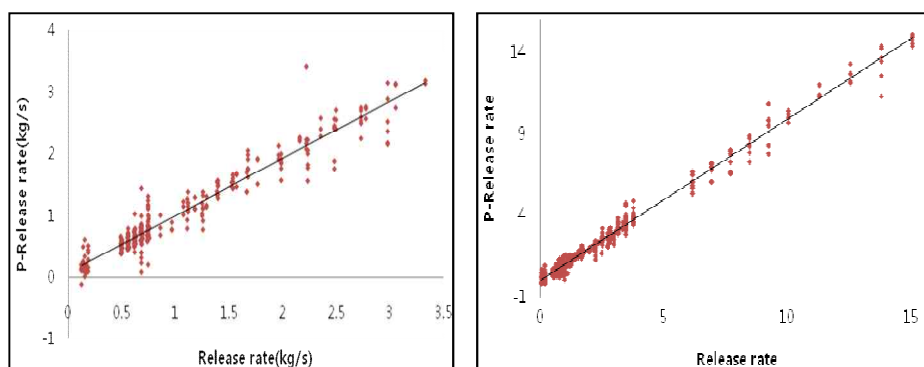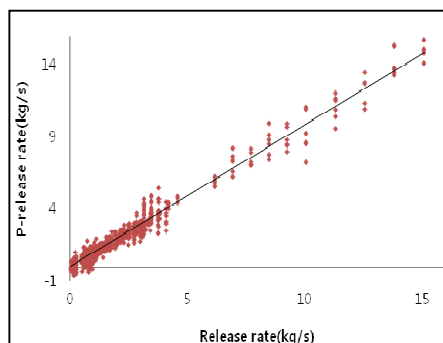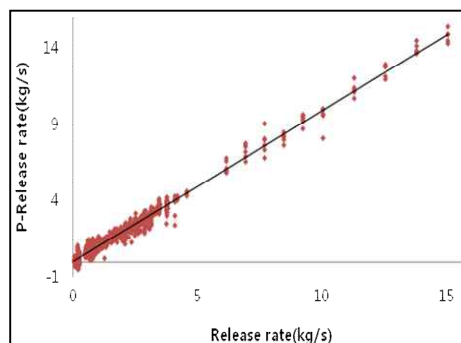|	1 (Cl_2)	1 (NH_3)	1 (SO_2)	1 (C_6H_6)	2	3	4
R	0.999	0.971	0.981	0.941	0.996	0.985	0.992
MSE	0.0225	0.0316	0.0399	0.0722	0.0734	0.0906	0.0988

Fig. 2. Fitness analysis between predicted and actual release rates for (a) 1 material, (b) 2 materials,

(c) 3 materials,   (d) 4 materials

4. Conclusion

An efficient, real-time monitoring technique is proposed that can estimate the hazardous gas release rate by using optical sensor and neural networks. The comparison against widely used commercial software like TRACE and PHAST illustrate that the proposed technique can estimate release rates within acceptable differences; the proposed technique does so with a far less number of input variables in a shorter period of time—within seconds. Consequently, it offers advantages against the traditional monitoring systems in which release rates are, in the first place, assumed rather than estimated. The accurate, spontaneous measurement of release rates is crucial for taking immediate responses in the outbreak of release accidents. The proposed technique can be used to estimate this value, thereby contributing to the building of a more effective information delivery system. Additional validation of the technique with more materials will ensure its applicability in broader situations.

Acknowledgment

References

[1] A. Dandrieux, J. P. Dimbour, G. Dusserre, 2006, Are dispersion models suitable for simulating small gaseous chlorine releases, J. Loss Prevention, 19, 683-689.

[2] M. Abbaspour, N. Mansouri, 2005, City hazardous gas monitoring network, J. Loss Prevention, 18, 481-487.

[3] R. K. Chohan, 1989, Safety and fault detection in process control systems and sensors, Fire Safety J., 14, 167-177.

[4] R. W. Lee, J. J. Kulesz, 2008, A risk-based sensor placement methodology, J. Hazard. Mater. , 158, 417-429.

[5] S. Alhajraf, L. Al-Awadhi, S. Al-Fadala, A. Al-Khubaizi, A. R. Khan, S. Baby, 2005, Realtime response system for the prediction of the atmospheric transport of hazardous materials, J.Loss Prevention, 18, 520-525.

[6] C. Massie, G. Stewart, G. McGregor, J. R. Gilchrist, 2006, Design of a portable optical sensor for methane gas detection, Sens. Actuators, B, 113, 830-836.

[7] A. D. Belegundu, T. R. Chandrupatla, 1999, Optimization concepts and applications in engineering, Prentice-Hall, New Jersey.

Validation of CFD simulation results in case of portal vein blood flow

C.C. Botar[a]*, T. Vasile[b], S. Sfrangeu[b], S. Clichici[c], P.S. Agachi[a], R. Badea[b], P. Mircea[d], M.V. Cristea[a]

[a]*Department of Chemical Engineering and Oxide Materials Science, "Babes-Bolyai" University, Cluj-Napoca, Romania, cbotar@chem.ubbcluj.ro*
[b]*Department of Medical Imaging, University of Medicine and Pharmacy "Iuliu Hațieganu", Cluj-Napoca, Romania*
[c]*Department of Physiology, University of Medicine and Pharmacy "Iuliu Hațieganu", Cluj-Napoca, Romania*
[d]*Department of Gastroenterology, University of Medicine and Pharmacy "Iuliu Hațieganu", Cluj-Napoca, Romania*

Abstract

Promising applications such as disease research and medical diagnostic, where fluid mechanical conditions are correlated to regions prone to different pathologies represented during the time the driving force to study the blood flow and its relation with the vessels walls in the human circulatory system. The hemodymanics simulation studies have been frequently used to gain a better understanding of functional, diagnostic and therapeutic aspects of the blood flow. Therefore a precise quantification of the blood flow in vessels could constitute a strong basis for diagnosis, prediction or evolution estimation of blood vessels or associated organ diseases. Due to the fact that many fundamental issues of the blood flow, like phenomena associated with pressure and viscous forces fields, are still not fully understood or entirely described through mathematical formulations the characterization of blood flow is still a challenging task. The computational modeling of the blood flow and mechanical interactions that strongly affect the blood flow patterns, based on medical data and imaging represent the most accurate analysis of the blood flow complex behavior. But, in order to represent a valuable non-invasive tool capable to provide comprehensive insights of the overall phenomena taking place at the most intimate level inside the sanguine vessels and which could serve to medical purposes the computational modeling of the blood flow needs accurate validation.

In the present paper the mathematical modeling of the blood flow in the portal vein has been addressed, the computational fluid dynamic (CFD) technique has been used and the model simulation results have been validated using in vivo Echo-Doppler measurements.

Keywords: Computational Fluid Dynamics, hemodynamics, blood flow, flow/vessel wall interaction, results validation

1. Introduction

The blood flow complex characteristics have been investigated along the time through simulations based on mathematical models that include constitutive equations describing the hemodynamics and its relations with the deformable vessels wall. The computational techniques applied to model the blood flow in the circulatory system

have been used to investigate either the velocity field or the pressure field. The vessel walls have been treated as rigid [1] ones or considering significantly simplified or reduced geometries for the deformable wall models [2]. The approximation of rigid-walls was made mostly due to the difficulty of solving the coupled blood flow/vessel deformation problem and was justified by the observation that, under normal conditions, wall deformability does not significantly alter the velocity field [2]. Modeling the three-dimensional blood flow in compliant vessels is extremely challenging for a number of additional reasons such as: geometry acquisition, accurate constitutive description of the behavior and induced movement of the tissue, inflow and outflow boundary conditions, etc. The vessel wall, due to its elasticity, changes shape in response to flow field dynamics and surrounding organs movement. Thus, the homodynamics is inherently a fluid–structure interaction phenomenon, which supplementary requires knowledge about the flow rate time evolution. When simulating venous regions the flow rate time dependency is reduced and the flow regime could be considered in steady state. Therefore, the most realistic methods, for hemodynamics description in veins, as in this case, are the real fluid – structure interaction ones [3, 4]. In these methods the geometry is modeled by conventional Finite Element Method (FEM) and the fluid flow by CFD technique.

Tue to its capability the CFD technique have been mostly used for studying the complex behavior of the blood flow. The first CFD studies [5, 6] used idealized geometries to calculate the blood flow characteristics and properties like wall shear stress and residence time. Later, due to the development of medical imaging techniques, more accurate and realistic geometries [7] have been used in blood flow simulation studies.

2. Problem statement

The computational fluid dynamics (CFD) technique has been applied to describe the blood flow in vessels and the elastic wall conditions have been introduced. Medical imaging techniques such as MRI and Echo-Doppler have been used for a more accurate acquisition of the portal vein geometry and in vivo blood velocity measurement respectively. The 3D CAD techniques have been used for the reconstructed model of the vessels. The reconstructed 3D portal vein model provides geometric boundaries for the CFD blood flow model. In this respect a finite difference grid has been generated over the finite element model geometry. Homodynamic parameters such as velocity magnitude, pressure and wall shear stress have been computed and the homodynamic validation has been done by comparing the mathematical model simulation results with the values of the blood velocity obtained by Echo-Doppler measurements.

3. Paper approach

The paper approach consists on three main issues: portal vein geometry acquisition, blood flow modeling, simulation and results validation.

3.1. The portal vein geometry acquisition

The portal vein geometry and the flow related data have been acquired using a 1 T MRI system (Sigma LX, GE Medical Systems) with a 9.1.0723d software, 1270 MHz IP30 processor, main memory size 512 Mbytes, and a phase array TORSOPA receiver. The geometrical information have been extracted from a 3D TOF SPGR vascular sequence acquisition, using the SmartPrep option based on bolus detection (gadolinium). The apparatus settings consisted in: Angio TOF SPGR – 3D acquisition; maximum monitor period 12 s; image acquisition delay 4 sec; imaging options: Fat sat; SmartPrep; TE Min; Prep Time 21; FA 35; Bandwidth 41,67; FOV 40; Sl Th 2.2; Locs per slab 34;

Validation of CFD simulation results in case of portal vein blood flow

Freq/Phase 256/192; NEX 0.50; PhaseFOV 0.90; scanning time 0:23, breath hold. The geometry of the portal vein has been reconstructed using the 3D computer-aided capabilities of Solid Edge V20 software (Figure 1).
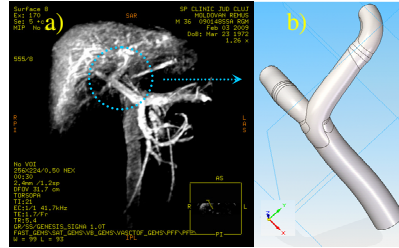

Figure 1. The geometry of the portal vein: a) RMI image; b) Solid Edge reconstruction

Table 1. Portal vein model dimensions

Inlet main branch diameter	10.52 mm
Middle main branch diameter	9.91 mm
Distal main branch diameter	9.41 mm
Main branch length	45.02 mm
Left branch inlet diameter	10.14 mm
Left branch middle diameter	8.53 mm
Left branch distal diameter	7.69 mm
Left branch length	29.66 mm
Right branch inlet diameter	7.45 mm
Right branch middle diameter	8.62 mm
Right branch distal diameter	7.48 mm
Right branch length	27.15 mm

The volume geometry has been imported in GAMBIT, a software package designed to build and mesh models for computational fluid dynamics (CFD) software packages. The surfaces mesh has been generated, using the face surface Quand/Pave algorithm and smoothed using the length-weighted Laplacian algorithm. The volume mesh was generated using a Tet/Hybrid/Tgrid meshing algorithm and boundary zone and continuous zone have been defined.

The geometry parameters are presented in Table 1. The volume mesh has been exported to CFD software and the mathematical model was applied to the 3D geometry.

3.2. Blood Flow Modeling

The flow model considers the thixotropic characteristics of the blood. In a non-Newtonian fluid, as in the blood case, the relation between the shear stress and the strain rate is nonlinear, and can be even time-dependent; therefore a ratio between shear stress and strain rate (or shear-dependent viscosity) has been defined. Moreover, this concept is more useful for fluids without time-dependent behavior, as in the case of portal vein circulation.

The computational modeling of the blood flow has been done considering three-dimensional incompressible Navier-Stokes equations written in a strong conservation form for mass and momentum. Finding the solution of the governing equations is difficult using traditional analytical techniques, especially for such a complicated

system, which involves 3D irregular geometry, complex fluid-structure interaction and non-Newtonian viscosity. Numerical techniques have been required, hence the need for CFD. The CFD software used for simulations was the ANSYS's FLUENT.

The model implemented to describe the hemodynamics was the Reynolds stress model (RSM), which is the most elaborate turbulence model that FLUENT software provides.

The RSM model is abandoning the isotropic eddy-viscosity hypothesis, and closes the Reynolds-averaged Navier-Stokes equations by solving transport equations for Reynolds stresses together with an equation for the dissipation rate. Since the RSM accounts for effects of streamline curvature, swirl, rotation, and rapid changes in strain rate, in a more rigorous manner than the one-equation and the two equation flow models, it has been used in simulation due to its greater potential to give accurate predictions for complex flows [FLUENT 6.3.26 user guide].

A steady state model has been used; the reduced time dependency of the blood flow in portal vein system has been neglected. The differential equations have been discretized in a manner of finite element method. The operation and the boundary conditions have been specified. The vessel wall has been treated as elastic. A dynamic mesh model has been used in order to address the movement of the mesh in the steady state solver. The no-slip condition has been imposed, meaning that the speed of the fluid relative to the boundary is 0, but at some distance from the boundary the flow speed must equal that of the fluid bulk. The blood viscosity has been defined according to the non-Newtonian power law.

3.3. Simulation Results

The following results of the complete process of preprocessing, solving and postprocessing following the procedure discussed above enable the visualization and quantification of the biologic phenomena taking place in the portal vein.

The portal vein (pv) hemodynamics simulation has been initialized with a static pressure of 2737.1 Pa and a blood velocity of 25 cm/s. These parameters have been considered constants at the entrance of the portal vein. The resulted data, provided by the computer simulation, supply the values of the blood velocity along the entire portal vein geometry.
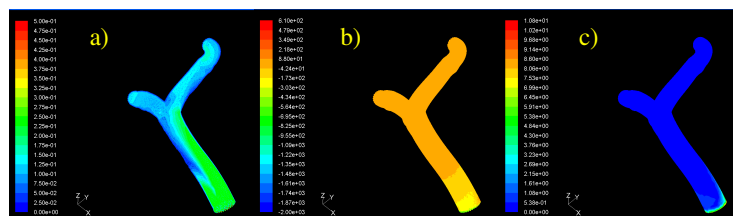

Figure 2. Contours of velocity magnitude (m/s) (a), static pressure (in Pascal) (b), wall shear stress (in pascal) along the portal vein (c).

In Figure 2 the contours of velocity magnitude (m/s), static pressure (in Pascal) and wall shear stress (in pascal) along the portal vein are represented. The results suggested laminar flows along the entire structure, uniformity in static pressure and wall shear stress distribution along the main braces of the portal vein.

3.4. Model validation

The blood flow modeling approach has been validated through comparison between simulation results and clinical investigations data. The velocity of the blood flow was measured in vivo using the Eco-Doppler technique. The medical investigations have

Validation of CFD simulation results in case of portal vein blood flow

been conducted using an Ultrasound Logiq 7 BT 06 machine, a Convex probe 4C, with the following B mode settings (Freq 4.0 MHz, AO=100%, Gn 78, DR 111), Doppler color settings (Freq 3,3 MHz, PRF 1.2 KHz, Gn 29, WF 175 KHz) and Pulse Doppler settings (SVL 4 mm, GN 23, PRF 3,5 KHz, DR 40, WF 69 Hz).

The investigated areas were the common portal vein, the left and the right branches. The data obtained by computer simulation are well fitted on those obtained by in vivo Eco-Doppler measurements, for all investigated portal vein domains, as can be seen in Figures 3 (I-III) (a) Eco-Doppler measurements; b) simulation results). Figure 3-I shows the Eco-Doppler blood velocity measured values in the portal vein main branch and the blood velocity values resulted from mathematical model simulation. The velocity value obtained by medical investigations ranges from 22.9 cm/s to 25 cm/s and the velocity value obtained by simulation ranges in the domain 20-27.5 cm/s along the main flow path lines.

Figure 3-II and Figure 3-III show the same type of data as Figure 3-I but for the right branch and respectively for the left branch of the portal vein. The velocity value obtained by medical investigations in the right branch ranges from 10.8 cm/s to 13.2 cm/s and the velocity value obtained by simulation ranges in the domain 9-13.5 cm/s along the main low path lines. In case of left branch the velocity value obtained by medical investigations ranges from 12.4 cm/s to 16 cm/s and the velocity value obtained by simulation ranges in the domain 12.5-17.5 cm/s along the main flow path lines.
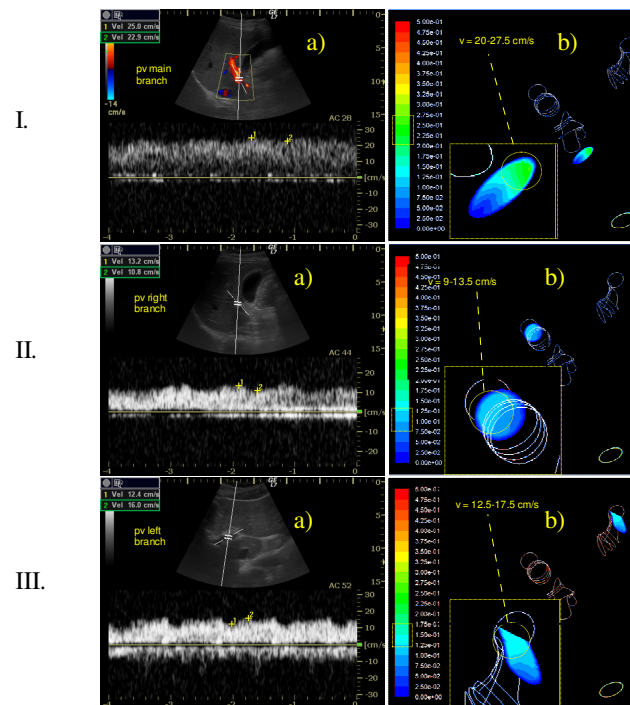

Figure 3 I-III. Contours of velocity magnitude (m/s) (a), static pressure (in Pascal) (b), wall shear stress (in pascal) along the portal vein (c).

4. Conclusions

The data for the 3D reconstruction of the portal vein have been provided by MRI. The values of the macroscopic parameters needed in simulations have been provided by Doppler ecography. The mathematical modeling method used is able to supply valid information about portal vein hemodynamics. The modeling approach was validated by comparison with the Eco-Doppler data. A very good correlation between simulation results and clini-cal data has been obtained.

The results of this study, in association with other medi-cal data, configure the support for the development of soft-ware assistance solutions for patients' therapy by providing useful information regarding the hemodynamic characteris-tics and dynamics associated with portal vein hypertension evolution.

References

1. Perktold K, Rappitsch G (1995) Computer simulation of local blood flow and vessel mechanics in a compliant carotid artery bifurcation model, J. Biomech., 28 (7): 845–856
2. Cairncross RA, Schunk PR, Baer TA, Rao RR, Sackinger PA (2000) A finite element method for free surface flows of incompressible flu-ids in three dimensions, Part I: boundary fitted mesh motion., Int. J. Numer. Meth. Fluids, 33:375–403
3. Vierendeels JA, Riemslagh K, Dick E (2000) Computer simulation of intraventricular flow and pressure gradients during diastole. J. Bio-mech. Eng., 122:667–674
4. Watanabe H, Hisada T, Sugiura S, Okada J, Fukunari H (2002) Computer simulation of blood flow, left ventricular wall motion and their interrelationship by fluid-structure interaction finite element method. JSME Int. J. Ser. C – Mech. Syst. Mach. Elem. Manufact. 45(4):1003–1012.
5. Finnigan P, Hathaway A, Lorensen W (1990) Merging CAT and FEM, Mech. Eng.,112: 28–32
6. Taylor CA, Hughes TJR, Zarins CK (1998) Finite element modeling of blood flow through arteries. Comput Methods Appl Numer Eng, 158:155–96
7. Perktold K, Resch M, Peter RO (1991) Three dimensional numerical analysis of pulsatile flow and wall shear stress in carotid artery bifur-cation. J. BioMech., 24:409–20

Particle Transfer and Deposition Using an Integrated CFD Model of the Respiratory System

Aleck H. Alexopoulos[a], Paraskevi Karakosta[a], Costas Kiparissides*[a]

[a]*Department of Chemical Engineering and Chemical Process Engineering Research, Aristotle University of Thessaloniki, P.O. Box 472, Thessaloniki, Greece, cypress@cperi.certh.gr*

Abstract

The present work describes an integrated CFD model of the respiratory system from the nasal cavity down to the bronchioli. The model is comprised of nine sequential computational blocks corresponding to the nasal cavity, the pharyngo-trachea, and a series of branches in the pulmonary system. Steady-state turbulent flow is employed to describe the inspiration flow and deposition of particles of different sizes. Local deposition efficiency is found to increase with particle size and flow rate. The deposition profiles are in accordance to experimental and computational results available in the literature. The proposed integrated respiratory model describes the flow, penetration, and deposition of particles in the respiratory system accounting for the influence of the nasal cavity and the pulmonary branches.

Keywords: respiratory, pulmonary, CFD, particles, deposition

1. Introduction

The transfer and deposition of particles in the respiratory system is of major interest for the development of targeted drug delivery formulations but also due to the increasing concerns over the potential toxicity of natural and engineered particles. Experimental and theoretical work have focused on different regions of the respiratory system, e.g., oropharyngeal, nasal, pulmonary, and alveolar, where many aspects of particle penetration are fairly well understood. However, several issues remain to be elucidated including the deposition of non-spherical particles and fibers, particle dispersion and aggregation, changes in particle size and shape and finally the fate of deposited particles. To further improve our understanding on particle penetration and deposition an integrated CFD model of the entire respiratory system has been developed and the detailed penetration and deposition of particles has been investigated.

In this work the respiratory system is assumed to consist of ten sequential computational blocks. The first block corresponds to the inflow cavity, e.g., the nasal or the oral cavity. The second block connects the inflow cavity to the lower trachea just above the first pulmonary bifurcation. The following seven blocks correspond to branching structures of the pulmonary system. The final computational block corresponds to the alveoli sacs and individual alveoli where gas exchange occurs. Flow and deposition of particles in the respiratory system is determined by performing CFD simulations in each computational block. The outflow conditions of a specific computational block are used as inflow conditions of the subsequent computational block. This paper deals with the simulation of steady-state nasal inspiration and particle deposition in the first nine blocks of the respiratory model.

2. CFD Simulations of the Respiratory System

The nasal cavity consists of two nasal air paths converging posteriorly to a single pathway, the nasopharynx, which is then directed downwards to the pharynx and the trachea. The two nasal air paths are highly curved and convoluted in shape providing a total surface area of about 150 cm^2. The nasal walls are covered by a mucous layer which moves to the posterior clearing deposited particles. The flow and deposition of particles in the nasal cavity has recently been investigated by several groups (Liu et al., 2007; Shi et al., 2008; Wen et al., 2008)**.**
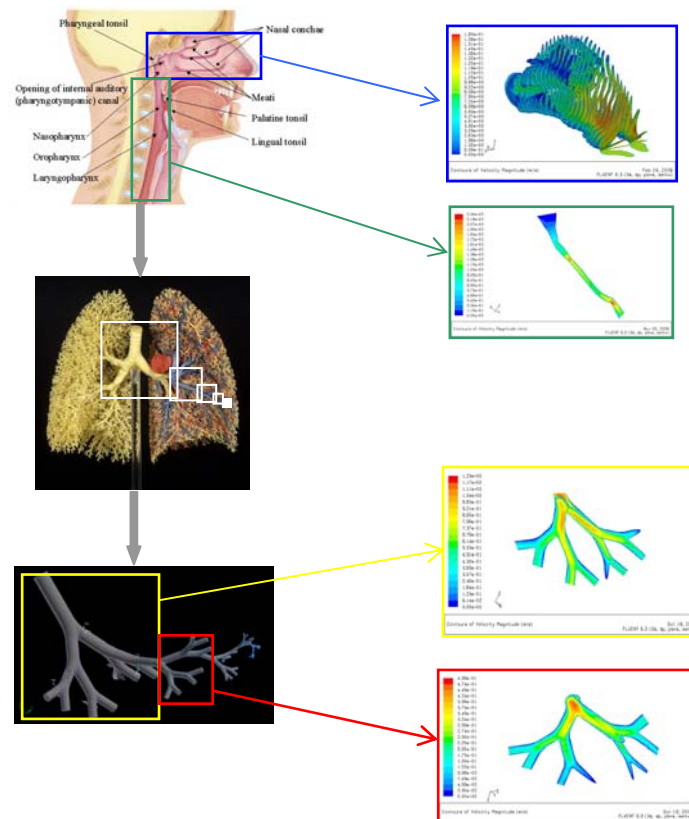
Figure 1. Block Computational Structure of the Respiratory System

The pharyngotracheal route can be studied not only to determine particle depositions but also to determine the function of the larynx. In terms of airflow it can be simplified to a curved conduit of changing cross-sectional area and shape. There are two main constrictions: one at the top of the oral cavity and the second near the larynx (Fig. 1). Most of the deposition occurs in these two regions.

The pulmonary system consists of a multitude of nonsymmetrical branches of progressively smaller diameter. There are a total of 23-24 branch generations leading to about 10^8 simple branches in the entire pulmonary system (Finlay, 2004). This limits the number of branch generations that can be completely simulated to around 5-6 (Longest and Vinchurkar, 2007; van Ertbruggen et al., 2005; Zhang et al., 2002). However, if a single pathline down to the alveolar sacs is considered, a successive simulation approach can be followed (Nowak et al., 2003).

Particle Transfer and Deposition Using an Integrated CFD Model of the Respiratory System.

In the present work, a model of the pulmonary tract is developed based on a sequence of seven consecutive "blocks" from the bronchi down to the alveolar sacs. Each block consists of four generations of symmetrical branches with one side rotated 90 degrees relative to the other. The inlet flow and particle motion conditions of each block are obtained from the outlet conditions of the previous block and the inlet conditions of the first block of the pulmonary system are obtained from the outlet conditions of the pharygotracheal block. Particle depositions in the pulmonary tract favor the larger particles with only the smallest of particles reaching the lower respiratory tract and alveolar sacs at significant concentrations.

3. Results and Discussion

Results are presented for the velocity magnitude and the local particle deposition efficiency in the nasal cavity, the pharyngotrachea, and the initial blocks of the pulmonary system. Particle deposition is determined by a Eulerian/Lagrangian tracking scheme based on the steady-state solutions for flow. Both uniform and Rosin-Rammler size distributions are employed.

3.1. Results for the Nasal Cavity

In this work, the nasal cavity geometry is reconstructed based on a series of medical images and is similar to the geometry described in Shi et al. (2008). A number of different computational grids consisting of tetrahedral cells were generated varying from about $3 \ 10^5$ to $2 \ 10^6$ in number with a distortion of less than 0.8. Because of the transitional nature of the flow in the nasal cavity the transitional k-ω model was employed as it is a good compromise between laminar and fully-developed turbulent flow. Inflow conditions to the nasal cavity were taken to be either constant or parabolic.
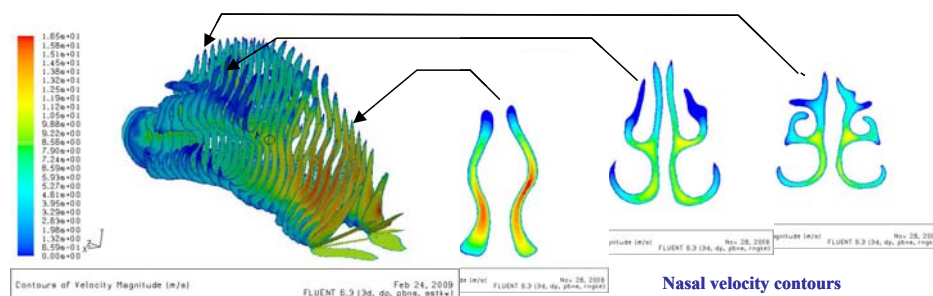

Figure 2. Velocity magnitude in the nasal cavity. Coronal sections. Inlet velocity V_{in}=2m/s.

The velocity magnitude at different coronal slices is obtained for different inlet velocites. As can be seen in Fig. 2 the incoming air-flow accelerates up to the nasal valve after which it decelerates and is directed towards the intersection of the nasal meatuses where the cross-sectional area is largest and the resistance to flow the lowest. This flow pattern permits adequate mixing up to the sensitive olfactory region, situated in the upper region of the nasal cavity, while at the same time leads to a large capture rate of large (i.e., >5μm) particles.

For inertia-dominated cases, particle deposition efficiency is typically described in terms of the impaction parameter, QD^2, where Q is the volumetric flow rate and D the particle diameter. Comprehensive information on the axial and size distribution of deposited particles is obtained (Fig. 3). It is found that larger particles are mostly

deposited in the anterior region of the nasal cavity while smaller particles were deposited less yet more evenly throughout the entire nasal cavity.
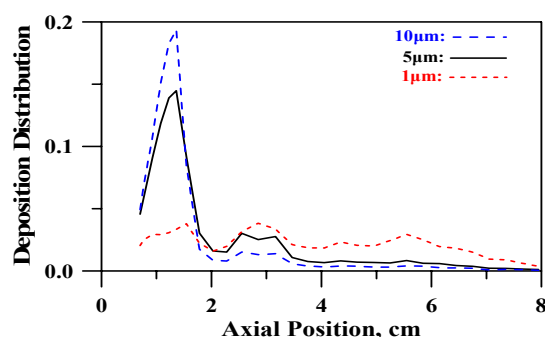

Figure 3. Particle Deposition Profiles in the Nasal Cavity. Inlet velocity $V_{in}=2m/s$.

3.2. Results for the PharyngoTrachea

The computational geometry was obtains by extending the nasopharyngeal region of the posterior nasal cavity downwards towards the pharynx and trachea. The geometric description of Heenan et al. (2003) was employed without the oral cavity and with the laryngeal region smoothed to obtain a computational grid consisting of 88,000 tetrahedral cells.
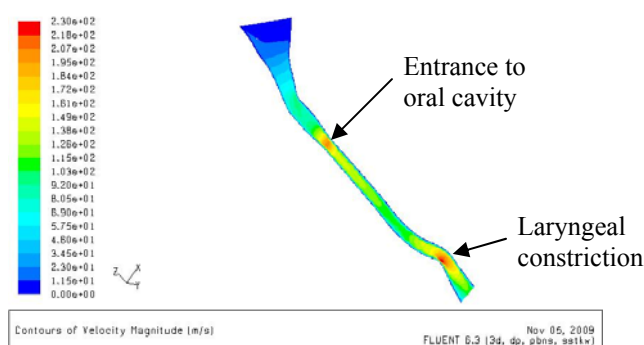

Figure 4. Velocity Magnitude, $V_{in}=5m/s$.

As can be seen in Fig. 4 the velocity of airflow accelerates in the converging nasopharyngeal region. Peak velocities are observed at pharyngotracheal positions corresponding to narrow cross-sectional positions observed just above the oral cavity and in the laryngeal region. Particle depositions increased with particle size and flow rate and occurred mostly in the regions of constricted and fast flow.

3.3. Results for the Pulmonary System

In the present work, a model of the pulmonary tract is developed based on 7 consecutive "blocks" of the pulmonary system. Each block consists of 4 generations of symmetrical branches with one side rotated 90 degrees relative to the other and is discretized into 260,000 tetrahedral cells. To examine the sensitivity of block selection on the final solution different computational geometries and grids were constructed by altering the sequence of blocks.

Particle Transfer and Deposition Using an Integrated CFD Model of the Respiratory System.

In Fig. 5 the velocity magnitudes are shown for blocks one and four of two different geometries (i.e., block sequence). It is clear that the velocity magnitudes are significantly different between block one and block four for both geometries. Moreover, the differences in the calculated velocity magnitudes between two different geometries are significant for the first block. On the other hand, the flows are nearly identical in the fourth block for both geometries. Therefore, block selection during the assembly of the pulmonary geometry is important only in the first two blocks of the pulmonary system.
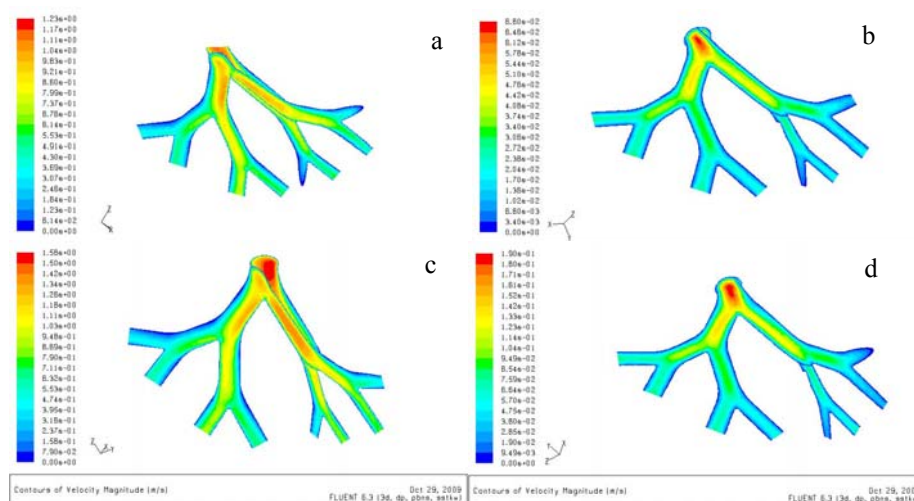

Figure 5. Velocity magnitudes for different blocks and geometries, V_{in}=2.7m/s. First block (a) and (c). Fourth block (b) and (d). Geometry one (a) and (b). Geometry two (c) and (d).
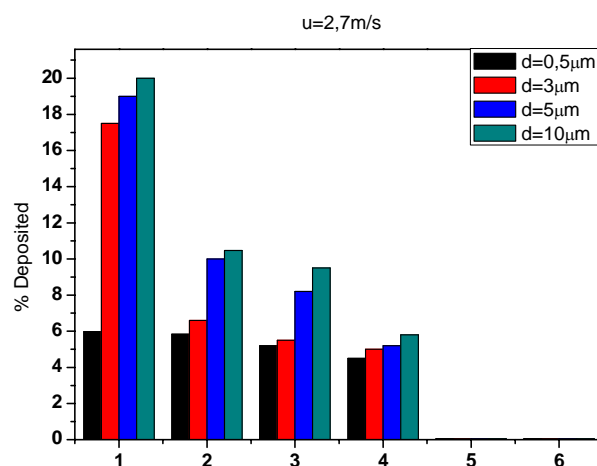

Figure 6. Local Deposition Efficiency in the Pulmonary System V_{in}=2.7m/s.

The calculated local deposition efficiencies for each computational block are displayed in Fig. 6. As can be seen, the local deposition efficiency decreases with each successive block due to a decrease in the value of the flow rate. On the other hand, for each block the local deposition efficiency increases with particle size from 500nm to 10μm. The results for blocks five and six are not shown due to the inadequate resolution of particle

deposition given that only a small number of particles reached these blocks. These results for local particle deposition are in qualitative agreement to the results reported in the literature, e.g., Nowak et al. (2003), and are consistent to an inertia dominated deposition mechanism which is expected for micron-sized particles.

4. Conclusions

In conclusion the sequential block approach, despite its many simplifications, can describe flow and deposition profiles in the respiratory system. The choice of pulmonary blocks is shown to be important only in the first two blocks of the pulmonary system. Therefore, realistic physiological representations should only be performed for the upper respiratory tract. Deposition of large (i.e., $>1\mu m$) particles are inertially dominated and scale with QD^2. The deposition of smaller particles (i.e., $<1\mu m$) is more complicated as diffusional forces becomes important. The computed particle deposition profiles in terms of local deposition efficiency appear to be in qualitative agreement with experimental and computational data.

Future work will involve the description of ellipsoidal particles and investigate the effect of surface charge. The alveolar sacs and individual alveoli will also be included to the integrated model of the respiratory system and the simulations will be extended to a full breathing cycle (i.e., inhalation and exhalation).

References

W.H. Finlay, 2004, Lung Deposition Simulation, in A.J. Hickey (ed.) Pharmaceutical Inhalation Aerosol Technology, Marcel Dekker, NY 2004.

A.F. Heenan, E. Matida, A. Pollard, and W.H. Finlay, 2003, Experimental measurements and computational modeling of the flow in an idealized human oropharynx. Experiments in Fluids, 35, 70-84.

Y. Liu, E.A. Matida, J. Gu and M.R. Johnson, 2007, Numerical simulation of aerosol deposition in a 3-D human nasal cavity using RANS, RANS/EIM, and LES, Aerosol Sci., 38, 683-700.

P. Longest and S.Vinchurkar, 2007, Validating CFD predictions of respiratory aerosol deposition: Effects of upstream transition and turbulence, J. of Biomechanics, 40(20), 305-316.

T.B. Martonen, Z. Zhang, G. Yue, and C.J. Musante 2002. 3-D particle transport within the human upper respiratory tract. Aerosol Science, 33, 1095-1110.

N. Nowak, P.P. Kakade and A.V. Annapragada, 2003, Computational Fluid Dynamics Simulation of Airflow and Aerosol Deposition in Human Lungs", Annals of Biomed. Engineering, 31, 374-390.

H.W. Shi, C. Kleinstreuer and Z. Zhang, 2008, Dilute suspension flow with nanoparticle deposition in a representative nasal airway model, Physics of Fluids 20, 013301.

C. van Ertbruggen, C. Hirsch and M. Paiva, 2005, Anatomically based three-dimensional model of airways to simulate flow and particle transport using computational fluid dynamics, J. Appl. Physiol., 98, 970-980.

J. Wen, K. Inthavong, J. Tu and S. Wang, 2008, Numerical simulations for detailed airflow dynamics in a human nasal cavity, Respiratory Physiology & Neurobiology.

G. Yu, Z. Zhang and R. Lessmann, 1998, Fluid Flow and Particle Diffusion in the Human Upper Respiratory System, Aerosol Science and Technology, 28(2),146-158.

Z. Zhang, C. Kleinstreuer and C.S. Kim, 2009, Comparison of analytical and CFD models with regard to micron particle deposition in a human 16-generation tracheobronchial airway model, Aerosol Science, 40, 16-28.

Stochastic Optimal Control for the Treatment of a Pathogenic Disease

Vicente Rico-Ramirez[a], Fabricio Napoles-Rivera[a], Guillermo González-Alatorre[a] and Urmila M. Diwekar[b]

[a]*Instituto Tecnologico de Celaya, Departamento de Ingenieria Quimica, Av. Tecnologico y Garcia Cubas S/N, Celaya, Guanajuato, Mexico 38010*
[b]*Vishwamitra Research Institute, 368 56-th Street, Clarendon Hills, IL, 60514, USA*

Abstract

Mathematical modeling as a tool for the treatment of a pathogenic disease has been widely proposed in the literature. Most of the modeling approaches represent the immune system dynamics as deterministic optimal control problems. Deterministic approaches, however, do not consider uncertainties in model parameters and variability among different individuals. To include uncertainties in the formulation, the aim of this paper has been using stochastic optimal control theory to develop protocols for the treatment of human diseases. We model time dependent uncertainties as Ito processes. That results in an optimal control problem where the constraints are stochastic differential equations and the objective function is an integral equation. The optimality conditions of the problem are obtained through the stochastic maximum principle, which results in a boundary value problem. The boundary value problem is solved iteratively by using a combination of the gradient method and a stochastic version of the Runge-Kutta method derived in this work. As an illustration of the proposed approach, we solve a mathematical model to determine the evolution of a generic disease and obtain regimens for applying therapeutic agents in a manner that maximizes efficacy while minimizing side effects. We show that stochastic optimal control theory can indeed help develop clinical insight in treating illness under uncertainties in model parameters.

Keywords: Stochastic optimal control, disease dynamics, real options theory

1. Introduction

Numerous models of immune response to infection have been proposed in the literature (Stengel et al., 2002). In such problems, model constraints describe the evolution of the disease, which is characterized by a non-linear set of ordinary differential equation. The dynamic equations are then controlled by therapeutic agents that affect the rate of change of the system variables. Objective functions are generally integral equations which model the tradeoff between pathogen concentration, organ health, and use of therapeutics.

1.1. Model of Disease Dynamics(Stengel et al., 2002)
Equations (1) through (4) described the generalized model provided by Asachenkov et al. (1994) and further analyzed by Stengel et al. (2002).

$$\frac{dx_1}{dt} = (a_{11} - a_{12} * x_3)x_1 + b_1 u_1 \tag{1}$$

$$\frac{dx_2}{dt} = a_{21}(x_4)a_{22}x_1(t-\tau)x_3(t-\tau) - a_{23}(x_2 - x_2^*) + b_2 u_2 \tag{2}$$

$$\frac{dx_3}{dt} = a_{31}x_2 - (a_{32} - a_{33}x_1)x_3 + b_3u_3 \tag{3}$$

$$\frac{dx_4}{dt} = a_{41}x_1 - a_{24}x_4 + b_4u_4 \tag{4}$$

where x_1 represents the concentration of pathogen cells, x_2 is the concentration of plasma cells, x_3 is the concentration of antibodies which kill the pathogen and x_4 is the relative characteristic of a damaged organ (0= healthy, 1=dead). $a_{21}(x_4)$ is a non-linear function that describes the immune deficiency triggered by damage to the organ. Each u_i represents a suitable therapeutic agent that controls the rate of change of the state variable x_i. A natural response of the system with no application of therapy implies $u_i=0$. Table 1 shows the values of the rest of the parameters used in the model. Four cases are considered depending upon the initial conditions of x_1: *i)* In the sub-clinical ($x_1(0)=1.5$) case the immune system integrity is never threatened, and pathogens are easily destroyed, *ii)* The clinical case ($x_1(0)=2.0$) reflects momentary threat to the patient's health and thus justifies medical treatment, *iii)* In the lethal case ($x_1(0)=3$), the antibodies are not sufficient to overcome the infection which leads to fatal damage of the organ, *iv)* In the chronic case ($x_1(0)=2.7$) the pathogen reaches a steady-state value without the patient being completely cured. Figure 1 illustrates the dynamics of each of the state variables for the chronic case (a) and the lethal case (b).

1.2. An Optimal Control Problem
Given the disease dynamics, optimal values of the control variables u_i (optimal treatment) can be determined by representing the model as an optimal control problem. Model constraints are described by Equations (1) through (4). To complete the mathematical model, an objective function that penalizes large values of pathogen concentration, poor organ health, and excessive application of therapeutic agents is defined by Equation (6).

$$J = \frac{1}{2}\left(p_{11}x_{1f}^2 + p_{44}x_{4f}^2\right) + $$
$$\frac{1}{2}\int_{t_0}^{t_f}(q_{11}x_1^2 + q_{44}x_4^2 + r_{11}u_1^2 + r_{22}u_2^2 + r_{33}u_3^2 + r_{44}u_4^2)\ dt \tag{5}$$

1.3. Solving the Optimal Control Problem
The optimal control problem defined by Equations (1)-(5) can be solved by applying the Maximum principle (Stengel, 1994). First, the Hamiltonian function is defined as given by Equation (6).

$$H = q_{11}x_1^2 + q_{44}x_4^2 + r_{11}u_1^2 + r_{22}u_2^2 + r_{33}u_3^2 + r_{44}u_4^2 + \lambda_1\big((a_{11} - a_{12}x_3)x_1 + b_1u_1\big) + \lambda_2(a_{21}(x_4)a_{22}x_1(t-\tau)\ x_3(t-\tau) - a_{23}(x_2 - x_2^*) + b_2u_2) + \lambda_3(a_{31}x_2 - (a_{32} - a_{33}x_1)x_3 + b_3u_3) + \lambda_4(a_{41}x_1 - a_{24}x_4 + b_4u_4) \tag{6}$$

Derivation of the Hamiltonian function with respect to the state variables provides the dynamics of the adjoint variables λ_i (a new set of ordinary differential equations) represented by Equations (7)-(10).

Table 1. Model Parameters

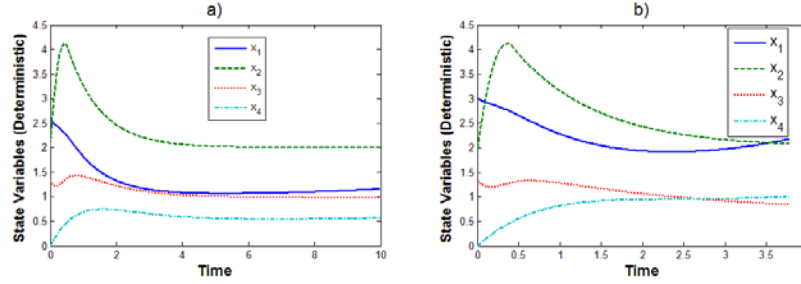|				
a_{11}=1	a_{12}=1		b_1=-1	$x_2(0)$=Each case
a_{21} (Equation 5)	a_{22}=3	a_{23}=1	b_2=1	$x_2(0)$=2
a_{31}=1	a_{32}=1.5	a_{33}=0.5	b_3=1	$x_3(0)$=$2x_2^*/3$
a_{41}=0.5	a_{42}=1		b_4=-1	$x_4(0)$=0

Figure 1. Two cases in disease dynamics: a) chronic and b)lethal

$$\frac{d\lambda_1}{dt} = -[q_{11}x_1 + \lambda_1(a_{11} - a_{12}x_3) + \lambda_2 a_{21}(x_4)a_{22}x_3 - \lambda_3 a_{33}x_3 + \lambda_4 a_{41}] \tag{7}$$

$$\frac{d\lambda_2}{dt} = \lambda_2 a_{23} - \lambda_3 a_{31} \tag{8}$$

$$\frac{d\lambda_3}{dt} = \lambda_1 a_{12}x_1 - \lambda_2 a_{21}(x_4)a_{22}x_1 + \lambda_2 a_{33}x_1 \tag{9}$$

$$\frac{d\lambda_4}{dt} = -\left[q_{44}x_4 + \lambda_2 \frac{\partial a_{21}}{\partial x_4}a_{22}x_1 x_3 - \lambda_4 a_{42}\right] \tag{10}$$

The final conditions of the adjoint variables are found by deriving the salvage function of objective function (non-integral term) with respect to the state variables:

$$\lambda_1(t_f) = p_{11}x_1(t_f), \lambda_2(t_f) = \lambda_3(t_f) = 0, \lambda_4(t_f) = p_{44}x_4(t_f) \tag{11}$$

The resulting problem is a two-boundary value problem which includes the state dynamics (Equations 1-4) and the adjoint dynamics (Equations 7-10). The optimal control profiles are obtaining through the optimality conditions defined by Equation (12), which results in Equation (13) for each control variable u_i and adjoint variable λ_i.

$$\frac{\partial \mathcal{H}}{\partial u} = 0 \tag{12}$$

$$\frac{dH}{du_i} = r_{ii}u_i + b_i\lambda_i = 0 \tag{13}$$

2. Methodology: A Stochastic Approach for the Treatment of a Pathogenic Disease

The approach described in Section 1 assumes deterministic behavior of the model parameters. Deterministic approaches, however, do not consider uncertainties and variability among different individuals. In practice, significant variability of relevant parameters among patients and for a given patient during the course of the disease has been reported. The success of an optimal control method depends on the accuracy of the model. If uncertainties are omitted, this can lead to significant performance degradation. Therefore, the inherent uncertainties in the patient need to be addressed. To include uncertainties in the formulation, the aim of this paper has been using stochastic optimal control theory to develop protocols for the treatment of human diseases.

2.1. Disease Dynamics as Ito Processes

Based on the so called Real Options Theory, we model time dependent uncertainties as Ito processes (Dixit and Pindyck, 1994). Equations (14)-(17) are Ito processes including a Wiener process to model uncertainties in the behavior of the state variables.

$$dx_1 = [(a_{11} - a_{12} * x_3)x_1 + b_1 u_1]dt + \sigma_1 \epsilon_1 \sqrt{dt} \tag{14}$$

$$dx_2 = [a_{21}(x_4)a_{22}x_1(t - \tau)x_3(t - \tau) - a_{23}(x_2 - x_2^*) + b_2 u_2]dt + \sigma_2 \epsilon_2 \sqrt{dt} \tag{15}$$

$$dx_3 = [a_{31}x_2 - (a_{32} - a_{33}x_1)x_3 + b_3 u_3]dt + \sigma_3 \epsilon_3 \sqrt{dt} \tag{16}$$

$$dx_4 = [a_{41}x_1 - a_{24}x_4 + b_4 u_4]dt + \sigma_4 \epsilon_4 \sqrt{dt} \tag{17}$$

Figure 2 illustrates how uncertainties impact the behavior of the system dynamics. The initial concentration of pathogen corresponds to the deterministic chronic behavior described in Figure 1a. However, by assuming the same initial conditions, but uncertain behavior in one of the state variables (x_1), different realizations of the uncertain parameters result in any of the four possible cases. Figure 2 shows the lethal case (a) and the clinical case (b). Note that both cases start form the same initial conditions for x_1, but x_1 is now assumed to behave as an Ito process (σ_1=0.25, σ_2=σ_3=σ_4=0).

2.2. Solving the Stochastic Optimal Control Problem

Equations (14)-(17) introduce uncertainties to the optimal control problem of Section 1. Solving the stochastic optimal control problem requires the combined use of the stochastic maximum principle (Rico-Ramirez and Diwekar, 2004) and Ito´s Lemma (the fundamental theorem of stochastic calculus), which also results in a two-boundary value problem. For simplicity, the derivations are not presented here. However, by assuming that the diffusion coefficients (σ_i) of Equations (14)-(17) depend neither on the state nor on the control variables, the same equations (adjoint equations and optimality conditions) as those derived for the deterministic case can be applied (Poznyak, 2002). A difference, however, is that the integration of a stochastic variable requires a stochastic integration method. Based on the derivation of the Euler-Murayama method used for the integration of stochastic differential equations, a stochastic version of the Runge-Kutta method has been derived in this work. Such a derivation, as well as the derivations from the stochastic maximum principle, are omitted seeking simplicity, but they will be presented in an extended version of this manuscript.

2.3. Solution Algorithm

A numerical implementation of the iterative gradient method is used for solving the two-boundary value problem; the algorithmic steps are as follows:

a) Sample the random parameter of Equations (14)-(17), $\in$, for each time step. Set *k=0*.

b) Provide a guess for the control profile u_{ik}.

c) Perform forward integration of the state variables using the stochastic Runge-Kutta method.

d) Similarly, perform backward integration of the adjoint equations.

e) Update the control profiles by using Equation (18), where ε is a small positive number (0.1 was used in most of our calculations).

$$u_{ik+1} = u_{ik} - \varepsilon \frac{\partial H}{\partial u_i} \tag{18}$$

f) if $u_{ik+1} - u_{ik}$ is greater than a specified tolerance, set *k=k+1* and go back to step c) using u_{ik+1} as new guess for the control profile. Otherwise, convergence has been achieved and u_{ik+1} is the optimal control profile.

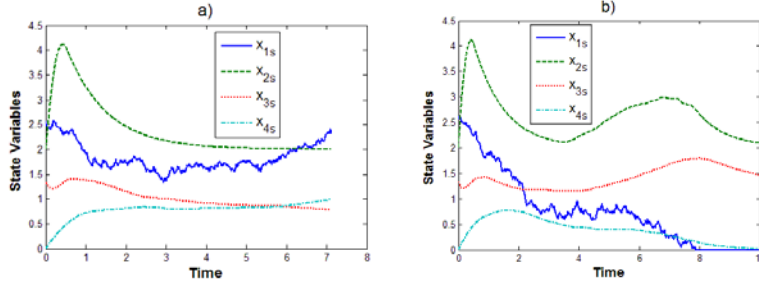*Stochastic Optimal Control for the Treatment of a Pathogenic Disease*

Figure 2. Effect of time dependent uncertainties in system dynamics

3. Numerical Results

24 different cases have been analyzed for various parameter values, initial conditions, and control variable combinations in of the stochastic pathogenic disease model. As an illustration of a stochastic simulation, a typical solution is depicted in Figure 3. If no control is used, this example corresponds to the deterministic lethal case of Figure 1b. However, the pathogen concentration is assumed to present stochastic behaviour. Also, the use of the 4 control variables is assumed. Figure 3 shows the behaviour of sate variables x_1 (Pathogen concentration) and x_2 (Plasma cells). Further, Figure 4 shows the optimal profiles for the control variables u_2 and u_4 as well as the profiles for the adjoint variables. Figure 4a also compares the stochastic and the deterministic profiles. Figure 3 and 4 are examples about how uncertainty may impact the treatment of diseases through therapeutic agents. Even by assuming only one stochastic variable (x_1), the difference in the values of the stochastic and deterministic control agents can be as much as 50%.

4. Conclusions and Future Work

This work uses real option theory to develop protocols for the treatment of human diseases. Time dependent uncertainties are modeled as Ito processes. The optimality conditions of the problem are obtained through the stochastic maximum principle, which results in a boundary value problem. The boundary value problem is solved iteratively by using a combination of the gradient method and a stochastic version of the Runge-Kutta method derived in this work. We have solved a mathematical model for the evolution of a generic disease. We show that stochastic optimal control theory can indeed help develop clinical insight in monitoring and treating illness under uncertainties in model parameters.

4.1. Ongoing Research

Current research continues our effort for analyzing practical real-world applications of stochastic optimal control in therapeutic optimization. An application to HIV treatment has been studied by using our solution approach. The model constraints given by Equations (19)-(22) and the model parameters are the same as those recommended by Kirschner et al. (1997).

$$\frac{dT}{dt} = \frac{s}{1+V} - \mu_T T + rT\left(1 - \frac{T+T^*+T^{**}}{T_{max}}\right)x_1 - k_1 VT \qquad (19)$$

$$\frac{dT^*}{dt} = k_1 VT - \mu_T T^* - k_2 T^* \qquad (20)$$

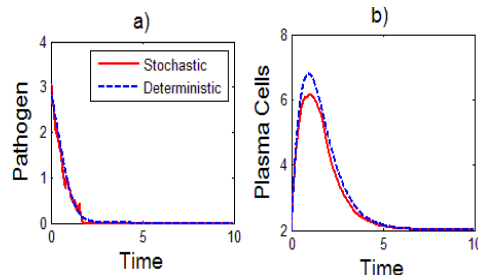$$\frac{dT^{**}}{dt} = k_2 T^* - \mu_{bT} T^{**} \qquad (21)$$

Figure 3. Dynamics in the pathogenic disease model. x_1 (Pathogen concentration) is an Ito Process
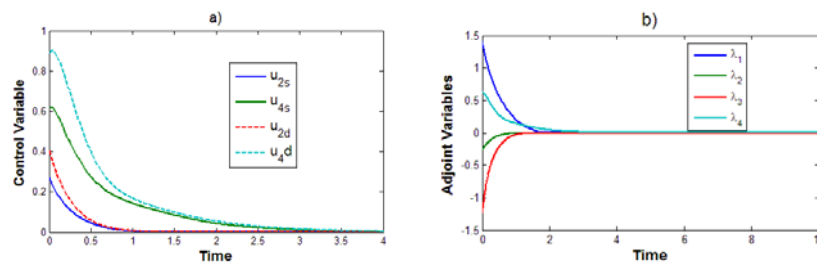

Figure 4. Optimal profiles in the pathogenic disease model: a) Control variables u_2 and u_4 for the stochastic and deterministic cases, b) Adjoint variables

$$\frac{dV}{dt} = N\mu_{bT}T^{**} - k_1VT - \mu_V V \qquad (22)$$

Results obtained so far with in the HIV model suggest that our approach can be used to handle not only qualitative models but also practical applications.

5. Acknowledgements

References

A. Asachenkov, G. Marchuk, R. Mohler, S. Zuev, 1994, Disease Dynamics, Birkhauser, Boston, USA.

A. K. Dixit, R. S. Pindyck, 1994, Investment under Uncertainty, Princeton University Press, Princeton, New Jersey, USA.

D. Kirschner, S. Lehart, S. Serbin, 1997, Optimal Control of the Chemotherapy of HIV, Journal of Mathematical Biology, 35, 775-792.

V. Rico-Ramirez, U. M. Diwekar, 2004,Stochastic Maximum Principle for Optimal Control Under Uncertainty, Computers and Chemical Engineering, 28, 2845-2849.

R. F. Stengel, 1994, Optimal Control and Estimation, Dover, New York, USA.

R. F. Stengel, R. Ghigliazza, N. Kulkarni and O. Laplace, 2002, Optimal Control of Innate Immune Response. Optimal Control Applications and Methods, 23, 91-104.

A. Poznyak, 2002, Robust Stochastic Maximum Principle: Complete Proof and Discussions. Mathematical Problems in Engineering 8(7), 389-411.

Optimal Catheter Placement for Chemotherapy

Dongning Li, Oleksandr Ivanchenko, Nikhil Sindhwani,
Eric Lueshen, and Andreas A. Linninger

*Laboratory for Product and Process Design, Departments of Chemical &
Bioengineering, University of Illinois at Chicago M/C 063, 851 S. Morgan St. - 218
SEO Chicago, Illinois, 60607-7000 USA, linninge@uic.edu*

Abstract

This paper addresses the problem of optimal administration of chemotherapeutic agents for the treatment of brain tumors by convection-enhanced drug delivery. The optimal catheter position is located by a novel optimization technique, which simultaneously maximizes drug concentration in the desired brain region, while ensuring that the final drug concentration does not fall below a therapeutically effective level or rise above the toxic threshold in non-treatment areas. A modified finite volume discretization method is used inside a nonlinear hybrid optimization algorithm. The distributed optimization problem with an embedded transport problem is solved on a coarse computational mesh, while searching for the optimal catheter position in a separate continuous coordinate system. In order to obtain continuous positional dependency of the objective function, two reference systems are used for solving the transport equations. The first analytical method projects the outflow from a specific catheter position inside the coarse finite volume cell onto its vertices. Once the cell face flux resulting from a specific continuous catheter positions are thus determined, the remaining two-dimensional transport problem is solved rigorously with a classical finite element method. A score function φ evaluates the match between the drug distributions achieved by a particular catheter placement with the therapy goals. Genetic inheritance adjusts the catheter locations to identify the globally optimal solution. Using the novel multi-scale algorithm, it is possible to optimize catheter placement and design, as well as to control drug distribution volume without the need for mesh refinement for different catheter positions.

Keywords: Computational analysis, CFD, Genetic algorithm, Optimization, Drug delivery

1. Introduction

Delivery of chemotherapeutics for treating malignant gliomas requires developing new computational techniques that maximize the effectiveness of therapy, while minimizing the risks associated with such therapies [1]. A glioma is a type of tumor that arises from glial cells in the brain or spine. In recent years, convection-enhanced drug delivery (CED) has received attention as means to deliver therapeutic agents directly to the target site inside the brain. This approach circumvents the problems posed by the impermeability of the blood brain barrier to macromolecules delivered systemically [2-3]. Moreover, CED achieves greater volume of distribution compared to diffusion-driven drug administration [3-4]. Many different protocols and catheter designs have been developed. However, successful chemotherapy requires effective drug dosage to destroy all cancerous cell at the tumor site as well as malignant cells spreading diffusely along white matter fiber tracts, as found in the clinical studies [1]. White matter is one

of the two components of the central nervous system, composed of bundles of myelinated nerve cell processes (or axons). White matter has anisotropic transport properties due to fiber like axonal bundle structure. The effective chemotherapeutic range is a strong function of the catheter placement within the brain parenchyma, which represents the bulk of brain tissue. The parenchyma is also the functional parts of a brain. We propose a novel mathematical method that suggest catheter placement for optimal distribution of agent to treat brain cancers. Rigorous determination of the optimal catheter location is a challenging distributed optimization problem. Optimal chemotherapy is a multi-objective problem. On one hand, drug concentration should be maximum in the vicinity of the tumor and the surrounding areas where recurrence may occur. The design has to ensure that in these regions the final drug concentration does not fall below a therapeutically effective level. On the other hand, critical brain areas such as the sensory cortex areas should not be harmed by hazardous agent levels to hold off side-effects. The second challenge stems from the distributed nature of the problem. The computational domain for drug distribution is typically solved over a coarse mesh with methods of finite differences, volumes or elements. For a specific catheter location, the domain would have to be re-meshed to adjust for proper boundary conditions at the interface between catheter outlet and porous brain tissue. Accordingly, the entire transport problem would have to be solved for each candidate solution; but more inconveniently, a different computational mesh would have to be set up with new boundary conditions for each and every experiment. To avoid this impractical approach, we propose in this paper a multi-scale technique using two computational reference frames.

A modified finite volume discretization method was used to avoid grid re-meshing. An analytical solution of the transport equations was used to calculate the flux inside a cell from the catheter tip to the cell walls. With the face fluxes thus determined, the convection diffusion problem was solved with the classical finite volume method. Combining these two computational levels, only a single mesh of the brain could be used, while varying the catheter position to find the optimal location. This technique avoided the need to re-mesh the computational domain for every new catheter position. A hybrid algorithm with a stochastic optimization and rigorous transport computation was used to solve the resulting nonlinear transport problem. The following sections show solutions to the distributed optimization problem coupled with the embedded transport problem.

2. Computational Model

Two dimensional models were reconstructed from MRI images of a human brain. The images were imported into Mimics, reconstruction software (Materialize, Inc) to accurately delineate boundaries and surfaces of the specific individual brain [5]. This information was imported into Gambit, and a computational mesh was generated [6]. Gambit is a meshing tool provided with ANSYS Fluent computational fluid dynamics package. Figure 1 shows a coronal slice from an MRI image of a human brain and the corresponding mesh reconstructed with the image reconstruction software as described above. More details about this technique can be found elsewhere [7-10].

Figure 1 also describes in simplified form the therapy target. Locations I_1 is close to the tumor in which high therapeutic levels should be ensured. In addition, the white matter tracts I_2 should be treated. Areas, V_1 and V_2, are critical cortical areas in which dosage must not reach harmful levels. The next section describes the process of optimizing the catheter outlet position in order to achieve the desired drug distribution for the chemotherapy.

Optimal Catheter Placement for Chemotherapy

2.1. Fluid motion and drug transport

Computational analysis was used to formulate the first principles conservation balances for mass, momentum and species transport over a mesh accurately representing the patient's specific brain geometry. This study focused on the methodology of optimizing the catheter position based on the patient specific geometry. For simplicity, the simulation was performed on a 2D coronal section of the brain. Darcy's law was employed to model the bulk fluid transport processes in the brain [7]. The governing equations are given in (1-2),

$$\nabla(K\nabla P) = 0 \tag{1}$$

$$\nabla(D\nabla C) - \overset{u}{U}\nabla C = 0 \quad where \quad \overset{u}{U} = K\nabla P \tag{2}$$

where P represents pressure, U is the velocity, D is the drug diffusion coefficient, and K is the hydraulic conductivity. The equations were discretized using the finite volume method using the computational domain shown in Figure 1.
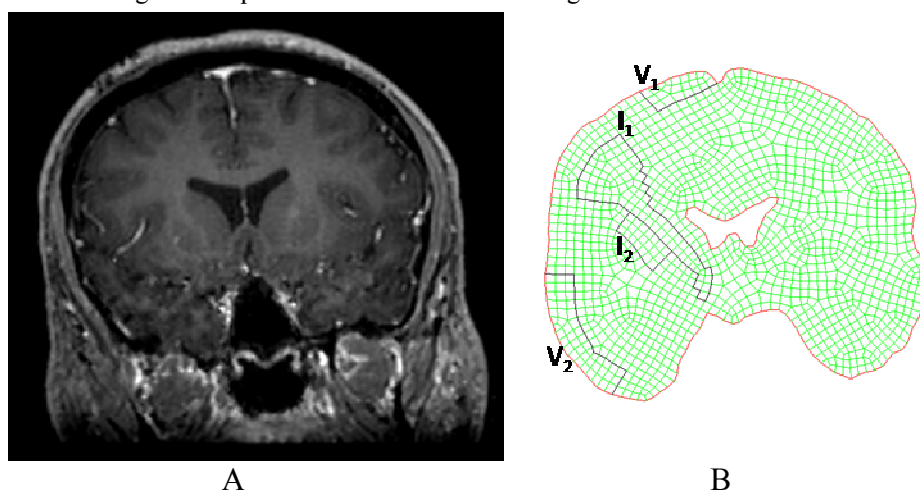

| A | B |

Figure 1: A. Coronal MRI slice of human brain used to reconstruct 2-D mesh; B. Computational domain and mesh reconstructed from MRI image. The mesh also shows the target area of the therapy. Locations I_1 is close to the tumor in which high therapeutic levels should be ensured. In addition, the white matter tracts I_2 should be treated. Areas, V_1 and V_2, are critical cortical areas in which dosage must not reach harmful levels.

Catheter flux. For a specific catheter position, the boundary conditions at the interface would have to be adjusted to match eqs. (1) and (2). To avoid the need to adjust boundary conditions and remesh the computational domain for every possible catheter position, we used an approximate method to superimpose the local flux from a catheter at position (x, y) onto the underlying coarse mesh. The local computation uses an analytical method to calculate flux from catheter to the cell faces. The continuous field solution is projected to the small region of the mesh cells close to the catheter ports. The combination of the two computational levels produces results similar to a mesh
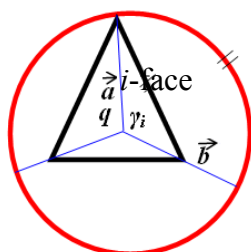

Figure 2: Every cell face is projected radially to the red circle. Flow from the catheter tip is projected onto each face and used in the solution of the transport equation for the rest of the domain.

refinement without actually having to reconstruct different meshes. The influence of the catheter port was modeled as an incoming flow with constant flow rate, q, with drug concentration of 3.69 mol/L. According to Green's theorem, the flux due to the source q, is equal to the outward flux through the closed red circle, with arc length 2π, as shown in Figure 2. Every cell face, or edge of the triangle, can be projected radially at the circle, producing a corresponding arc. Flux via every cell side is then exactly the same as via the correspondent arc. This is illustrated by using two vectors, $\vec{a}$ and $\vec{b}$ that connect the source q with the cell corners, directed towards the corner from source. If the flow is assumed to be radially homogeneous, the flux through face i can be calculated by using γ_i, the angle subtended by the face arc at the center of the circle. The flux calculated in this way is shown in eq. 3:

$$flux_{cell\ side,\ i} = q\,\frac{\gamma_i}{2\pi} \qquad \gamma_i = arc\cos(\frac{\vec{b}\cdot\vec{a}}{|\vec{b}||\vec{a}|}) \tag{3}$$

2.2. Therapy Optimization

The optimal catheter location was found by a hybrid algorithm composed of a genetic search with embedded transport calculations. The genetic algorithm generated an ensemble of candidate catheter positions. For each position, the resulting drug distribution was computed using the multi-scale method described above. Each resulting drug distribution was evaluated using a score function φ, for which a lower value indicates superior alignment with the chemotherapy targets. The genetic algorithm then adjusted the catheter outlet position based on the previous score result in order to minimize the score function. These steps were repeated for thirty generations until a minimum value for the score function was found. A more detailed description of the score function follows.

In order to achieve the best treatment effect, the drug concentration needed to be higher than the effective therapy concentration C_t=1.5 in the tumor region (I$_1$) and the white matter region (I$_2$), and lower than the toxic concentration C_p=0.7 in the cortex region (V$_1$ and V$_2$) as showing in Figure 1B. A score function φ was constructed as a penalty function to achieve the distribution described above. A function φ is calculated for each cell and is defined in eq. 4:

$$\varphi = \sum_{i\ in\ V_1\ or\ V_2} \left(1000\left(C_i - C_p\right)\right) - \sum_{i\ in\ I_1\ or\ I_2} \left(1000\left(C_i - C_t\right)\right) \tag{4}$$

where C_i is the drug concentration in ith-cell, i is the therapeutic concentration range defined for a particular region, and C_p is the pathogenic concentration that has to be avoided in the defined regions. For every cell i in V$_1$ or V$_2$, the concentration is compared to the desired maximum allowed concentration. A positive value is added to the score function if the criterion is not satisfied, as shown by the first term. Since a required therapeutic concentration C_t needs to be reached for every cell in I$_1$ or I$_2$, a negative score is added to the score function value when the concentration C_i crosses this level. Thus, the score function will assign a positive score every time the concentration levels drop below the therapeutic range in I$_1$ and I$_2$, or when they reach the toxic levels in V$_1$ and V$_2$. The process is repeated for every cell until the optimum

placement is obtained which has the least value of the score function. A schematic of this process is shown in Figure 3.

3. Results and Discussion

The distribution of drug is a function of infusion flow rate, infusion drug concentration, material property of the brain and the catheter outlet position. For this study, the catheter outlet position was chosen as the optimization parameter. The flow rate and inlet drug concentration were considered given parameters.

Single catheter. The results of the optimization for both a catheter placement with one port are presented in Figure 4 - left. For a single catheter, the optimal position is a compromise between reaching the tumor region (I_1) and also covering the white matter (I_2). Yet, even optimal placement is unsatisfactory, because only 70% of the target region can be reached with a therapeutically effective dosage. The damage to V_1 and V_2 is minimal with 99% of the area below toxic levels. The overall score of 82% indicated that a one catheter solution cannot meet the chemotherapy specifications, and that the treatment will most likely not be effective in treating the brain tumor.

Two catheters solution. A much more comprehensive drug volume distribution can be achieved by using two catheters as shown in Figure 4. Two catheters manage to completely cover the tumor region (I_1) and the white matter region (I_2). Clinical studies suggest that tumors usually propagate

Figure 3: Flowchart of the algorithm used to find catheter placement optimization.

along white matter tracts and the tumor and propagation path should be covered by treatment. The single catheter produces limited drug distribution which is not able to cover the tumor area and propagation path simultaneously. For the optimal solution to be safe, it should deliver the drug only in effective therapy concentration of $C_t=1$. At the same time, it keeps drug level below toxic levels in the cortex region, V_1 and V_2, below $C_p=0.7$. The score function corresponds to 100% satisfaction of the treatment objectives.

Our rigorous mathematical optimization approach could be extended by taking into account the anisotropy of brain properties [10]. This new computational approach would combine advanced imaging technologies to predict a more realistic distribution of drug macromolecules injected into the brain.
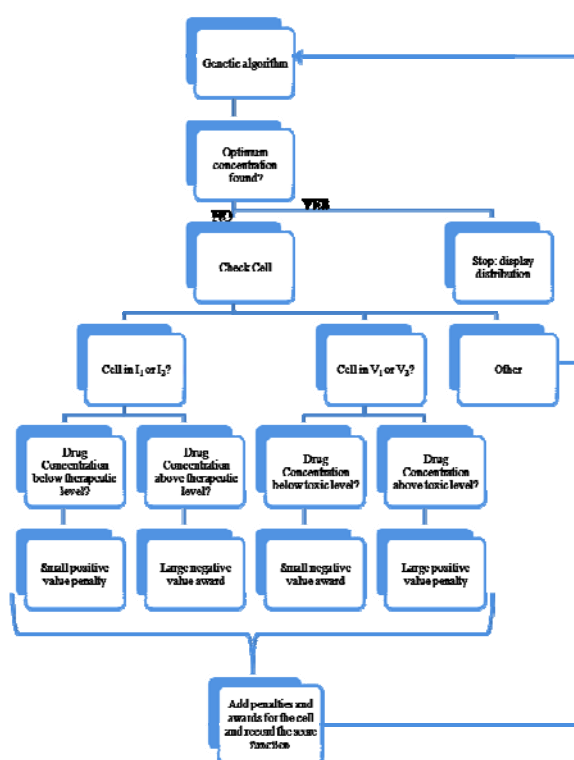
4. Conclusions

Detailed computational models for brain structures were constructed and then drug distribution was predicted with infusion catheters placed in specific brain regions. The results suggest that given desired drug distribution boundary values, it is possible to calculate the optimal placement of catheters in the parenchyma. Also, the possibility of using two different infusion catheters was explored and its advantages were discussed.
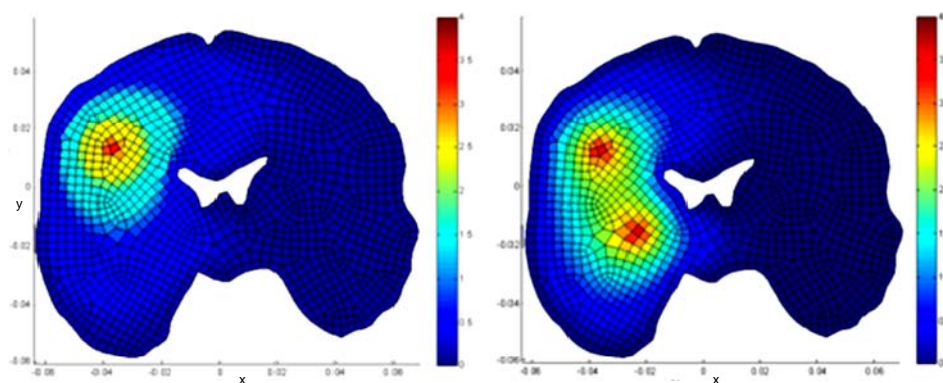

Figure 4. Optimization results for one and two catheter placement.

5. Acknowledgements

References

1. Hall, W. A., Rustamzadeh, E. & Asher, A. L. (2003) Convection-enhanced delivery in clinical trials, Neurosurg Focus. 14.
2. Bobo, R. H., Laske, D. W., Akbasak, A., Morrison, P. F., Dedrick, R. L. & Oldfield, E. H. (1994) Convection-enhanced delivery of macromolecules in the brain, Proceedings of the National Academy of Sciences of the United States of America. 91, 2076-2080.
3. Raghavan, R., Brady, M. L., Rodreguez-Ponce, M. a. I., Hartlep, A., Pedain, C. & Sampson, J. H. (2006) Convection-enhanced delivery of therapeutics for brain disease, and its optimization, Neurosurgical FOCUS. 20.
4. R. Penn, and A. Linninger; Intraparenchymal Drug Delivery For Parkinsons's Disease. In Stereotactic and Functional Neurosurgery; A. M. Lozano, P. L. Gildenberg, R. R. Tasker (Eds.) ISBN: 978-3-540-69959-0, Springer Verlag, NY, 2009.
5. Mimics Medical Imaging Software, Materializa Inc. http://www.materialise.com/mimics/.
6. Gambit meshing software, Fluent Inc, http://www.fluent.com/.
7. Linninger, A. A., Somayaji, M. R., Zhang, L., Hariharan, M. S. & Penn, R. D. (2008) Rigorous Mathematical Modeling Techniques for Optimal Delivery of Macromolecules to the Brain, Biomedical Engineering, IEEE Transactions on. 55, 2303-2313.
8. Morrison, P. F., Chen, M. Y., Chadwick, R. S., Lonser, R. R. & Oldfield, E. H. (1999) Focal delivery during direct infusion to brain: role of flow rate, catheter diameter, and tissue mechanics, Am J Physiol Regul Integr Comp Physiol. 277, R1218-1229.
9. Netti, P. A., Baxter, L. T., Boucher, Y., Skalak, R. & Jain, R. K. (1997) Macro- and microscopic fluid transport in living tissues: Application to solid tumors, AIChE Journal. 43, 818-834.
10. A. Linninger, M. Somayaji, X. Guo, T. Erickson, X. Guo and R. Penn. Computational Methods for Predicting Drug Transport in Anisotropic and Heterogeneous Brain Tissue. J of Biomechanics, 41: 2176-2187, 2008.

Estimation of seasonal transmission parameters in childhood infectious disease using a stochastic continuous time model

Daniel P. Word[a], James K. Young[a], Derek Cummings[b], Carl D. Laird[a]

[a]*Artie McFerrin Department of Chemical Engineering, Texas A&M University, MS 3122 TAMU, College Station, TX 77843, USA,*
carl.laird@tamu.edu (corresponding author)
[b]*Johns Hopkins Bloomberg School of Public Health, 615 N. Wolfe Street, Rm E6541, Baltimore, MD 21205, USA, dcumming@jhsph.edu*

Abstract

The development of accurate disease models is desirable for the purposes of gaining a better understanding of the underlying dynamics of infectious disease spread and for designing and implementing appropriate control measures to curb infectious disease spread. In this work we develop an estimation framework for long-term continuous time infectious disease models that considers both model and estimation noise. We present a nonlinear programming approach for efficient estimation of model parameters, including seasonal transmission profiles. We then demonstrate the effectiveness of this framework using measles data from New York City and Bangkok, and show that a strong correlation exists between estimated seasonal parameters and school term holidays.

Keywords: Measles, continuous, disease models, seasonal transmission parameters

1. Introduction

One goal of public health programs is to control the spread of infectious diseases and minimize the impact of disease on the population through various control measures such as vaccination programs. However, there are several social, environmental, and biological factors affecting the spread of infectious disease, and the observed temporal dynamics are not always well understood. The development of reliable mechanistic models for the spread of infectious diseases is needed both for aiding public health decision-making and for improving our understanding of factors affecting infectious disease spread. Childhood infectious diseases, such as measles and chickenpox, remain a serious public health concern, especially in developing countries, and are commonly used as a test bed for developing disease models.

Compartment-based disease models are commonly used to describe the dynamics of the disease within the population. In these models the population is assumed to be well mixed with individuals placed into various compartments based on their status with respect to the disease. For example, individuals can be classified as being susceptible to the disease (S), infected but not infectious (E), infected and infectious (I), or recovered and immune (R). This type of model is typically classified according the progression of the population through the compartments [1].

Much work has been done to model measles incidence using discrete time models [2-4]. The time-series SIR (TSIR) model, introduced by Finkenstädt and Grenfell [4], is a discrete-time model that incorporates a seasonal transmission parameter and an exponential mixing in the infection term. This model is capable of

capturing the biennial dynamics seen in cities with low birth rates, and it can quantitatively explain the annual cycle seen in measles incidence in cities with high birth rates [4-5]. The TSIR model and estimation procedure described for measles assumes a two-week reporting interval, which is similar to the serial interval for measles. If the reporting interval is different than the serial interval of the disease a different estimation procedure is needed. A continuous time model and estimation framework, on the other hand, can accommodate the common situation where the reporting interval and the serial interval are not similar. In previous work, we addressed the estimation of continuous time deterministic models for infectious disease spread with seasonal transmission parameters [7]. While deterministic models can reasonably capture incidence dynamics in large cities where the disease is endemic, infectious diseases are inherently stochastic in nature, and in communities below around 300,000 people stochastic fadeout of measles cases is commonly observed [6]. Deterministic models are incapable of capturing the disease dynamics in these cases. For these reasons, a model with both measurement and dynamic noise is desired. In this paper, we present a framework for estimation of continuous time infectious disease models from long-term time series data that considers both model and measurement noise. We demonstrate the effectiveness this approach using measles data from New York City and Bangkok.

2. Problem Formulation

The classic SIR framework model with a seasonal transmission parameter is used in our study. This model is sufficient to capture the key features of measles dynamics since life-long immunity is typically retained following infection. It has long been observed that measles incidence exhibits a seasonal pattern that appears to be correlated with school terms [3,9]. In the continuous time SIR model, we include a seasonal transmission parameter $\beta(y(t))$, also called the contact rate. Here, the function $y(t)$ maps the overall horizon time to the elapsed time within the current year. This forces β to have a yearly periodicity.

2.1. Stochastic Continuous Time Formulation

The differential equations describing the seasonal stochastic continuous time SIR model can be written as,

$$\frac{dS}{dt} = \frac{-\beta(y(t))S(t)I(t)}{N} + \mu(t)N + \varepsilon_S \tag{1}$$

$$\frac{dI}{dt} = \frac{\beta(y(t))S(t)I(t)}{N} - \gamma N + \varepsilon_I \tag{2}$$

where S is the number of susceptibles and I is the number of infectives. System parameters include the birth rate, $\mu(t)$, which is known and time varying, and N and γ, the reported population and the recovery rate respectively, which are known scalar inputs. Dynamic noise terms ε_S and ε_I are included for the susceptible balance and the infective balance equations respectively.

Prevalence refers to the number of individuals in the population who are infected at a given point in time, whereas incidence is the number of new infectious occurring over a given time interval. The available reported case data is measles incidence, but the state variable $I(t)$ represents the measles prevalence. Over a

particular reporting interval, *i*, the incidence can be calculated by integrating the rate of infection,

$$\int_{t_{i-1}}^{t_i} \frac{\beta(y(\lambda))I(\lambda)S(\lambda)}{N} d\lambda. \tag{3}$$

To include this in the estimation formulation, a new state variable $\phi(t)$ is introduced to represent the cumulative incidence at time *t*.

$$\frac{d\phi}{dt} = \frac{\beta(y(t))S(t)I(t)}{N} \tag{4}$$

Since not every individual who becomes infected will seek medical assistance, not every infection is reported. This underreporting can be significant and must be considered in the estimation. The output equation for the reported cases is given by,

$$\Phi_i = \eta_i \left(\phi_i - \phi_{i-1} \right) + \varepsilon_\phi \tag{5}$$

where ε_ϕ is the measurement noise term, ϕ is the estimated cumulative incidence at a point in time, Φ_i is the reported incidence over a given time interval, and η_i is the time varying reporting factor that accounts for the degree of underreporting. We use a standard susceptible reconstruction procedure to estimate this reporting factor.

The estimation problem can be written as a nonlinear programming problem with the differential and algebraic constraints described above. There are a number of techniques for solving this class of optimization problem. Here, we discretize all the state and algebraic variables and include the complete set of discretized equations as constraints in the nonlinear programming problem. The Explicit Euler technique was used to discretize the system, however, we have used Radau collocation techniques with similar results. Without further restriction of $\beta(y(t))$, this estimation problem has all the challenges of classic inverse problems including ill-conditioning and non-uniqueness. The seasonal parameter $\beta(y(t))$ was discretized less finely than the differential equations and the profile was regularized. Total variation regularization is used since it allows for discontinuous jumps as expected from a seasonal transmission parameter correlated with school term holidays. Combining the regularization term with a least-squares objective for the noise terms and initial state conditions gives the following objective function,

$$\min \ a \sum_{i \in \Im} \varepsilon_{Si}^2 + b \sum_{i \in \Im} \varepsilon_{Ii}^2 + c \sum_{i \in \Re} \varepsilon_{\phi i}^2 + d \left(S_0 - S_{init} \right)^2 + e \left(I_0 - I_{init} \right)^2 + \frac{1}{\rho} \| \Delta \beta \|_1 \tag{6}$$

where $\Im$ is the set of finite elements, $\Re$ is the set of reporting intervals, *a*, *b*, and *c* are weighting terms based on assumed standard deviations of the residuals, and *d* and *e* are weights placed on the residuals of the initial conditions. In the regularization term, $\Delta\beta$ is a first order approximation of $d\beta/dt$, and ρ is the regularization parameter that is calculated using the standard L-curve method [10]. As written, the regularization term is non-differentiable, however, it is easily reformulated by writing,

$$\frac{1}{\rho} \| \Delta\beta \|_1 = \frac{1}{\rho} \sum_{j \in B} \Delta\beta_j^+ + \Delta\beta_j^- \tag{7}$$

and including the following constraints,

$$\Delta\beta_j - \Delta\beta_j^+ + \Delta\beta_j^- = 0 \tag{8}$$

$$\Delta\beta_j^+ , \Delta\beta_j^- \geq 0 \quad \forall\, j \in B \tag{9}$$

where B is the set of discretizations for β within the year. This reformulated objective function with constraints (8) and (9), along with the constraints arising from the discretization of (1), (2), and (4), and the reporting factor adjustment (5) give rise to a large-scale nonlinear programming problem with purely algebraic constraints.

2.2. Data

The data sets used in this work contain yearly population and birth rate data, monthly measles case count data from New York City from the years 1947-1965 [11], and monthly measles case count data from Bangkok for the years 1975-1984. In the Thai data, there is regular passive surveillance for measles coupled with active surveillance to assess the performance of the passive surveillance system. All data is anonymized, and laboratory confirmation is reported when available. These two locations have very different school term holidays, allowing us to show the correlation between school terms and seasonal transmission. New York city has a long summer school holiday lasting from the end of June to mid September, while Bangkok has two long school holidays: one from the beginning of March to the end of April and one the entire month of October.

An additional challenge in the Bangkok data is missing information for the year 1979. To account for this, our model is integrated through this period, however the estimation is weighted to exclude these points from the objective function. Both data sets suffer from significant under-reporting. Susceptible reconstruction techniques have shown that about 1 in 9 cases are reported in New York City across the entire time horizon studied. Bangkok, however, has a varying reporting fraction that for this work is assumed to be linear over the time horizon and ranging from about 1 in 100 cases reported at the start of the time horizon to about 1 in 20 cases at the end.

3. Estimation Results

Estimations were performed for New York City and Bangkok using 240 finite elements per year and 60 discretizations of the seasonal transmission parameter per year. The problems were formulated in AMPL [12] and solved using the nonlinear interior-point method IPOPT [13].

3.1. New York City

The estimation for New York City produced essentially zero mean profiles for all model and measurement noise terms. The characteristics of these estimated noise terms are shown in Table 1. The model noise in the susceptible balance equation (ϵ_S) however, showed an apparent correlation in time. This would indicate that the susceptible dynamics are not being appropriately captured with this simple model, and future work will address improvements in this area. Nevertheless, estimated mean percentage of susceptibles in the population was 4.3%, which is similar to other literature values for measles [8].

Table 1. Noise terms for New York City

Residuals	ϵ_S (1/day)	ϵ_I (1/day)	ϵ_ϕ (1/day)
Mean	-14.973	-0.896	-0.477
Stand. Dev.	172.331	41.076	10.478

Estimation of seasonal transmission parameters in childhood infectious disease using a stochastic continuous time model

Fig. 1 shows the estimated seasonal transmission profiles, *β*. The profile shows seasonality that coincides almost perfectly with the school term summer holiday that occurs from the end of June to mid September.
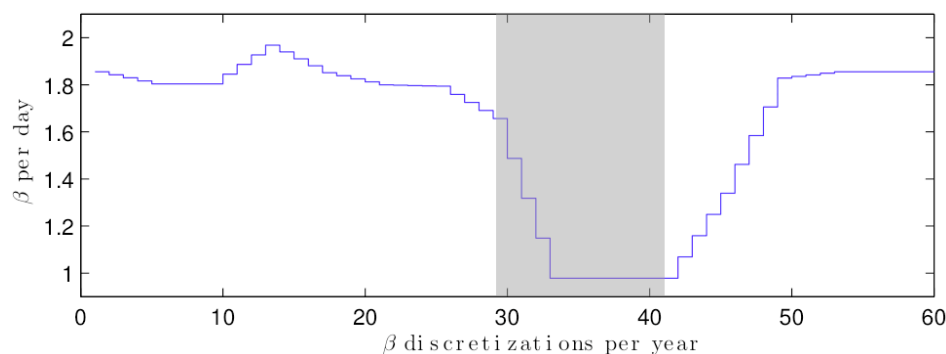

Figure 1: Seasonal transmission parameter for NYC with the school holiday indicated in grey

3.2. Bangkok

The estimation for Bangkok gave results similar to the New York City estimation. The reporting fraction is very low for Bangkok, and the data contains significant noise, but the model and measurement noise terms all have near zero mean as shown in Table 2. Again, the model noise corresponding to the susceptible balance equation do not appear to be independent in time, however, the estimated mean percentage of susceptibles in the population was still 4.4%.

Table 2. Noise terms for Bangkok

Residuals	ε_S (1/day)	ε_I (1/day)	ε_ϕ (1/day)
Mean	-14.262	-0.480	-0.104
Stand. Dev.	225.630	24.386	1.902

The estimated profile for the seasonal transmission parameter is shown in Fig. 2. The profile shows strong agreement with the school holidays that occur from the beginning of March through the end of April and the whole of October, although a delay of approximately one month is apparent. This may be due to delayed reporting or an artifact of the long reporting interval.

4. Conclusions

The usefulness of reliable disease models for further understanding the dynamics of infectious diseases and planning public health policy is apparent. The model described here appears to capture the dynamics of measles effectively for two diverse cities, and the continuous time formulation allows us to make immediate use of the data with larger reporting intervals. The estimated transmission parameter profiles for both cities demonstrate strong seasonality correlated with school term holidays despite the very different school holiday schedules for these locations. These estimations not only help improve our understanding of infectious disease spread, but also help quantify the effect of closing schools, a commonly proposed control measure for emerging infectious diseases.
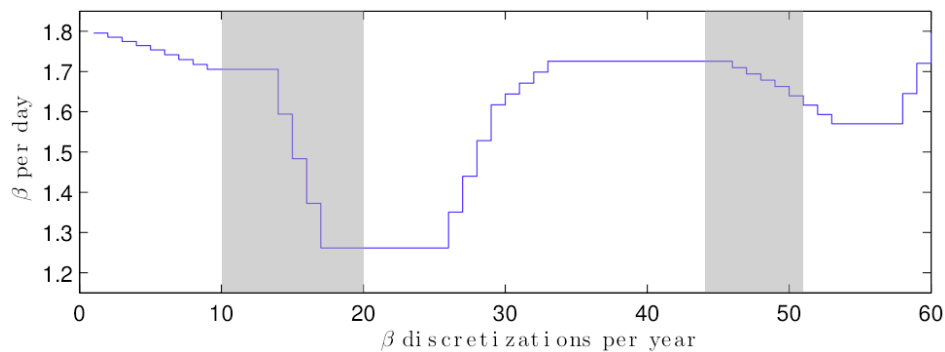
Figure 2: Seasonal transmission parameter for Bangkok with school holidays indicated in grey

For large cities where diseases are endemic, deterministic models can effectively reproduce the observed dynamics, however, in smaller communities were stochastic fadeout is evident, stochastic models are necessary. Future work will include a thorough analysis of the estimation framework on smaller community sizes. Furthermore, the approach is suitably efficient and flexible, and future work will investigate more complex model structures. Continued study will improve our understanding of the system, improve prediction, and improve our ability to control endemic and emerging infectious diseases.

5. Acknowledgements

References

[1] H.W. Hethcote, 2000, The mathematics of infectious diseases, SIAM Review, 42, 4, 599-653

[2] W.H. Hamer, 1906, Epidemic Disease in England, Bedford Press

[3] P.E. Fine, J.A. Clarkson, 1982, Measles in England and Wales I: An Analysis of Factors Underlying Seasonal Patterns, Int. J. Epidemiology, 11, 1, 5-14

[4] B.F. Finkenstädt, B.T. Grenfell, 2000, Time series modelling of childhood diseases: a dynamical systems approach, J. Royal Stat. Soc., Ser. C, 49, 187-205

[5] A.J.K. Conlan, B.T. Grenfell, 2007, Seasonality and the persistence and invasion of measles, Proc. R. Soc. B, 274, 1133-1141

[6] M.J. Keeling, 1997, Modelling the Persistence of Measles, Trends Microbiol., 5, 12, 513-518

[7] G.H. Abbott III, D.P. Word, D. Cummings, C.D. Laird, 2009, Estimating Seasonal Drivers in Childhood Infectious Diseases with Continuous Time and Discrete-Time Models, submitted to the 2010 American Control Conference (invited)

[8] RM. Anderson, R.M. May, 1991, Infectious Diseases of Humans: Dynamics and Control, Oxford University Press

[9] H.E. Soper, 1929, The interpretation of periodicity in disease prevalence, J. Royal Stat. Soc., 92, 34-73

[10] R.C. Aster, B. Borchers, C.H. Thurber, 2005, Parameter Estimation and Inverse Problems, Elseiver Academic Press

[11] J.A. Yorke, W.P. London, 1973, Recurrent outbreaks of measles, chickenpox, and mumps, Am. J. Epidemiology, 98, 6, 469-482

[12] R. Fourer, D.M. Gay, B.W. Kernighan, 1993, AMPL: A Modeling Language for Mathematical Programming, The Scientific Press

[13] A. Wächter, L.T. Biegler, 2006, On the Implementation of an Interior-Point Filter Line-Search Algorithm for Large-Scale Nonlinear Programming, Math. Programming, 106, 25-57

Optimization of Cultivation Conditions in Spin Tubes for CHO Cells Producing Erythropoietin

Jure Strnad[a,c], Matjaž Brinc[a], Vatroslav Spudić[a], Nadja Jelnikar[b], Lidija Mirnik[b], Barbara Čarman[b], Zdravko Kravanja[c]

[a]*Department of Bioprocess Development, Lek Pharmaceuticals d.d., a Sandoz Company, Biopharmaceuticals, Kolodvorska 27, SI-1234 Mengeš, Slovenia*

[b]*Analytical Department, Lek Pharmaceuticals d.d., a Sandoz Company, Biopharmaceuticals, Kolodvorska 27, SI-1234 Mengeš, Slovenia*

[c]*University of Maribor, Faculty of Chemistry and Chemical Engineering, Smetanova 17, SI-2000 Maribor, Slovenia*

Abstract

This paper describes the optimization of cultivation factor settings, i.e. the shaking rate and working volume in 50 mL spin tubes for a Chinese hamster ovary (CHO) cell line expressing recombinant human erythropoietin, using a response D-optimal surface method. The main objectives of the research were, firstly, to determine a setting in which the product titer and product quality attributes in spin tubes are equivalent to those in 250 mL shake flasks in a seven day batch and, secondly, to find a setting in which the product titer is maximal. The model for product titer prediction as a function of shaking rate and working volume in the defined design space was successfully applied to the optimization of cultivation conditions in spin tubes for the tested cell line. Subsequently, validation experiments were carried out simultaneously in spin tubes, shake flasks and bench scale bioreactors in order to compare cell culture performance parameters such as growth, productivity and product quality attributes in the form of isoform profiles and glycan antennarity structures. The results of the experiments showed that similar cell culture performance and product quality could be achieved in spin tubes when compared to shake flasks. Additionally, bioreactor titers could be reproduced in spin tubes at high shaking rates and low working volumes, but with differing product quality. Cultivation at lower shaking rates in spin tubes and shake flasks produced a glycoprotein with a product quality slightly comparable to that from bioreactors, but with titers being only two thirds.

Keywords: D-optimal design, Design Expert, Spin Tube, Glycosylation, Erythropoietin

1. Introduction

The presented paper gives an overview on the application of DOE (design of experiments) methods in the optimization of process parameter values for small scale bioreaction units. Experimental build up, data analysis and model validation using Design Expert software are described. Chinese hamster ovary cells are the workhorse in the production of biopharmaceutical products as they are widely used in research. Their most important characteristic is the ability to produce highly complex glycoproteins that are similar to the native biomolecules in the human body. One such glycoprotein is recombinant human erythropoietin, a hormone that acts on the bone marrow stem cells to stimulate erythropoiesis [1]. The glycoprotein consists of 165 amino acids that form a polypeptide chain with two disulfide chains at Cys7-Cys161 and Cys29-Cys33, the former being essential for the biological activity of the glycoprotein [2]. Laboratory research and development work on suspended CHO cell growth optimization is mostly carried out in shake flask vessels [3]. However, their geometry is not favourable with respect to capacity considerations as only approximately 50 shake flasks with a 250 mL nominal volume can simultaneously be placed in a standard incubator unit. By using spin tubes with the nominal volume of 50 mL and a tubular geometry, the incubator capacity can be tripled. Previous research work on spin tube characteristics was carried out by De Jesus et al. [4]. The main objective of our research work was to define cultivation conditions where the performance of spin tubes is equal to that of shake flasks. Additionally, the performance in smaller systems (spin tubes and shake flasks) and bench scale bioreactors was compared. Another important aspect of our research work was the use of DOE methods, since they are in the pharmaceutical industry highly favored by the Food and Drug Administration (FDA) in its Process Analytical Technology (PAT) initiative. In the pharmaceutical industry and especially in bioprocess development, factorial and response surface methods for process analysis and optimization have seldom been used [5]. They were usually applied in the optimization of medium formulations to increase cell growth or production [6-9]. Some researchers used DOE methods to investigate the combined effects of medium composition and process parameters like pH, temperature, etc. [10-12]. DOE methods can be divided into three functional classes, namely screening, full factorial and response surface designs [13]. One of the more complex response surface methods, a computer generated D-optimal design was applied to optimize the cultivation conditions in spin tubes for a CHO cell line expressing recombinant human erythropoietin.

2. Experimental design

Design Expert version 7.1.3 (Stat-Ease, Inc., USA) was used to build the response surface design and to perform the statistical analysis of the experimental results. A D-optimal design type was used because it enables the imposing of different restrictions on the design space. It is a computer generated design and is mostly used when the experimental region is irregular, e.g. if the region is not a cube or a sphere [14]. In such cases standard designs like e.g. central composite, face-centered composite, Box-Behnken etc. may not be the best choice. Irregular regions of interest, as is the case in this paper, occur fairly often. In generating a D-optimal design the user has to predict the form of the final statistical models e.g. linear, two factor interaction (2FI), quadratic

Optimization of Cultivation Conditions in Spin Tubes for CHO Cells Producing Erythropoietin

or cubic. If the form were too simple, this would result in the statistical models making a poor description of the system. On the other hand, if it were too complex, some experimental runs would become redundant. Shaking rate and working volume were predicted to be the most influential factors due to the differences in the geometry of spin tubes and shake flasks. The shaking rate was restricted to an interval from 180 rpm to 300 rpm and the working volume from 14 mL to 34 mL. In addition, a linear constraint of shaking rate and working volume, was imposed for settings where the spin tube screw cap with an in-built gas-permeable membrane could be moistened due to culture splashing. It should be noted that the restriction regarding screw cap membrane moistening was experimentally found as being non-linear (dotted line on Fig. 1). Since the software allows only linear constraints to be inserted, the slope of the linear constraint was adjusted so that high working volumes were retained in the design space, but a part of the space around the smaller working volumes and the higher shaking rates was therefore lost. The design space is presented in Fig. 1.
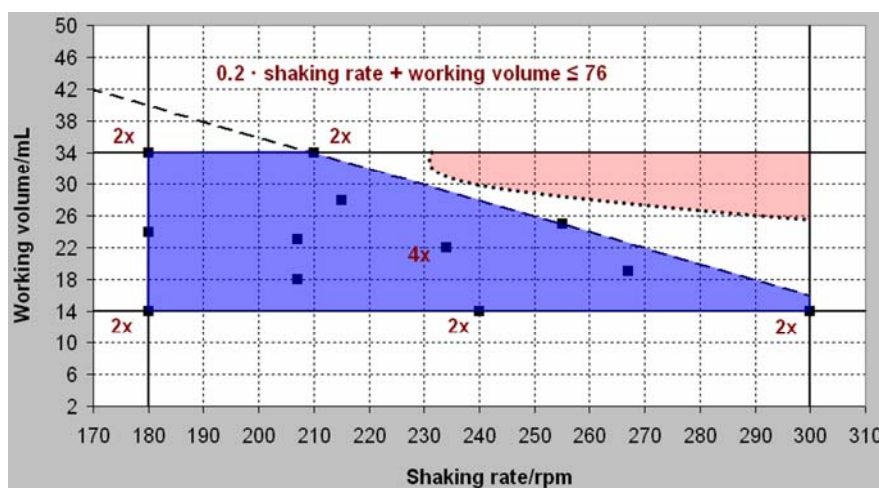

Figure 1. A graphical representation of the design space (dark shaded area): Non-linear restriction (dotted line), linearized non-linear constraint (dashed line), unfeasible region (light shaded area), working volume (*V*) and shaking rate (*n*) bounds (bold solid lines). Boundary and central experimental settings were replicated.

3. Results and discussion

After executing all 20 experimental runs, data was analysed using the Design Expert software. Altogether, the data for 20 responses (titer, metabolite concentrations, etc.) were inserted. The analysis of adjusted and predicted coefficients of determination (R^2-values) which should be higher than 0.6 indicated that the majority of responses cannot be described with the chosen statistical models. Therefore, only the model for the product titer (*P*) was used for further analysis. The model is presented in equation 1.

$$P = \frac{1}{0.056806 - 0.000223 \cdot n - 0.002421 \cdot V + 0.000015 \cdot n \cdot V} \qquad \text{Eq. (1)}$$

The overall statistical analysis showed that the model that describes the product titer value on the seventh day of cultivation is significant and can be used for modelling and optimization purposes in the defined design space.

4. Model validation

The main objective was to define the operational settings in which spin tubes have approximately the same titer as the shake flasks at standard conditions (90 rpm, 10 % CO_2 and 37 $^{\circ}$C). The average titer from all reference shake flask runs was 40 mg/L. Alongside the equality setting, the present research also wished to determine those settings in which maximal product titer could be obtained. The two computerized optimal configurations were i) shaking rate of 180 rpm and working volume of 30 mL for the point of equality and ii) shaking rate of 300 rpm and working volume of 14 mL for the point of the maximal titer. The predicted and experimental results for the product titer from the validation step are compared in Table 1.

Table 1. Validation for calculated 95 % prediction intervals (PI) for product titer.

Setting		n (rpm)	V (mL)	Average titer/(mg/L)			
				Low PI	Predicted	High PI	Observed
Spin Tube	A	180	30	30	40	60	54
Shake Flask	B	90	50	/	/	/	50
Spin Tube	C	300	14	36	53	97	81
Bioreactor	D	/	5000	/	/	/	78

Table 1 shows that the spin tube system at setting A and the shake flask system at setting B have almost equal product titers. Also, the spin tube setting at high shaking rates (C) produces a higher product titer than settings A and B, as predicted by the model. All the results fall into the software calculated 95 % prediction intervals. Based on these results it can be concluded that the statistical model for product titer prediction in spin tubes is valid for the specified design space.

The validation run was additionally used to compare product quality attributes between systems of different scales. Isoform distribution (Fig. 2A) representing the charge distribution of the intact glycoprotein and glycan group distribution according to the degree of glycan group antennarity (Fig. 2B) were determined as two possible quality attributes, where higher ranked isoforms and tetra- and triantennary groups are favoured. Figure 2 shows evidence confirming the possibility that the product quality in the spin tubes and that in the shake flasks (settings A and B) are equal. Slightly larger differences were observed only at the values of isoform 8 (Fig. 2A) and the biantennary glycan structures (Fig. 2B).
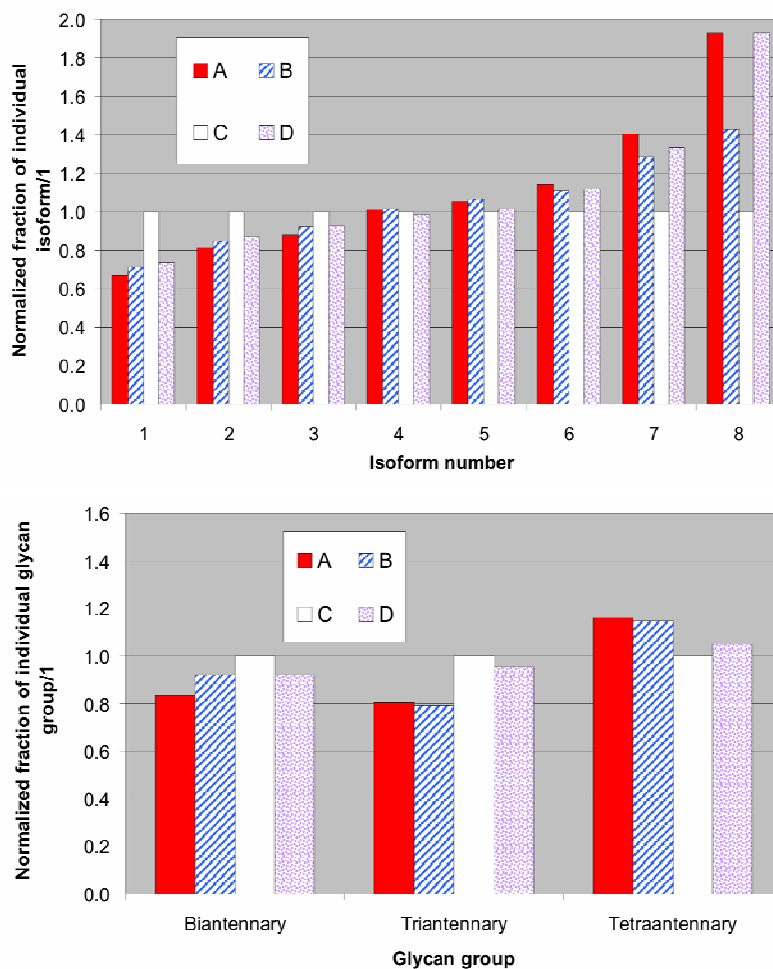
Optimization of Cultivation Conditions in Spin Tubes for CHO Cells Producing Erythropoietin

Figure 2. A) Relative proportions of product isoforms and B) Relative proportions of product glycan groups for the settings discussed in Table 4.

Higher ranking isoforms are desired as the lack of terminal sialic acids from carbohydrate chains decreases the *in vivo* activity of the glycoprotein [15]. The results of the isoform distribution (Fig. 2A) show that at setting C more lower ranking isoforms are produced than in other settings, while the distribution in the bioreactor setting (D) is essentially equal to those in settings A and B. It is interesting to note that a significant difference in triantennary and tetraantennary structures can be observed between the bioreactor on the one side and the spin tube and shake flask systems one the other side (Fig. 2B). The proportion of higher antennary structures in bioreactors is lower than in spin tubes and shake flasks at settings A and B, but their isoform distribution is quite

similar. It is possible that, although the glycan structures produced in bioreactors are less branched, the sialylation is better than in spin tubes and shake flasks at the used settings. In the spin tube and shake flask systems the glycoprotein could have more branched glycans chains, but they are not all sialylated. Based on these observations, it can be concluded that the distribution of glycan structures for the bioreactor is not comparable to those of settings A and B, while it is comparable to the one of the spin tube at the point of maximal titer (C). The distribution of glycan structures for setting A and B is also comparable, a slight discrepancy occurring at the fractions of the biantennary structures.

5. Conclusion

A response surface D-optimal design was used to optimize the cultivation conditions in spin tubes for a CHO cell line producing erythropoietin. Two distinct settings were determined, namely one where the product titer was equal in the spin tubes and the shake flasks, and one where a maximal titer in the spin tubes could be produced. Product glycosylation patterns were found to be equal for validated settings in spin tubes and shake flasks, but are not fully comparable to the glycosylation of the product from the bioreactors at the chosen similar settings. Although these conclusions are valid only for the cell clone used in this research and further work has to be done to asses the influence of the chosen factors on different clones, the methodology used and presented in this paper is generally applicable.

References

[1]   G. Walsh, Proteins: Biochemistry and Biotechnology, John Wiley & Sons, USA, 2005.
[2]   M.C. Flickinger and S.W. Drew (eds.), Encyclopedia of bioprocess technology: Fermentation, Biocatalysis and Bioseparation, John Wiley & Sons, USA, 1999.
[3]   J. Büchs, Biochemical Engineering Journal, 7 (2001) 91.
[4]   M.J. De Jesus, P. Girard, M. Bourgeois, G. Baumgartner, B. Jacko, H. Amstutz, F.M. Wurm, Biochemical Engineering Journal, 17 (2004) 217.
[5]   G. Hanrahan and K. Lu, Critical Reviews in Analytical Chemistry, 36 (2006) 141.
[6]   R. Kammoun, B. Naili, S. Bejar, Bioresource Technology, 99 (2008) 5602.
[7]   C. Chun, K. Heineken, D. Szeto, T. Ryll, S. Chamow, J.D. Chung, Biotechnology Progress, 19 (2003) 52.
[8]   S. Sandadi, S. Ensari, B. Kearns, Biotechnology Progress, 22 (2006) 595.
[9]   G.M. Lee, E.J. Kim, N.S. Kim, S.K. Yoon, Y.H. Ahn, J.Y. Song, Journal of Biotechnology, 69 (1999) 85.
[10] M.D. Gupte, P.R. Kulkarni, Letters in Applied Microbiology, 35 (2002) 22.
[11] B. Chauhan, R. Gupta, Process Biochemistry, 39 (2004) 2115.
[12] J. De Coninck, B. Leclercq, J.M. Exbrayat, F. Duyme, Journal of Industrial Microbiology and Biotechnology, 31 (2004) 204.
[13] D.C. Montgomery, Design and Analysis of Experiments, John Wiley & Sons, USA, 2005.
[14] R.H. Myers, D.C. Montgomery, Response Surface Methodology: Process and Product Optimization Using Designed Experiments, John Wiley & Sons, USA, 2002.
[15] E. Tsuda, G. Kawanishi, M. Ueda, S. Masuda, R. Sasaki, European Journal of Biochemistry, 188 (1990) 405.

A systematic synthesis and design methodology to achieve process intensification in (bio) chemical processes

Philip Lutze, Alicia Román-Martinez, John M. Woodley, Rafiqul Gani

Department of Chemical & Biochemical Engineering, Technical University of Denmark, Soltofts Plads, Building 227, DK-2800 Lyngby, Denmark, rag@kt.dtu.dk

Abstract

Process intensification (PI) has the potential to improve existing processes or create new process options which are needed in order to produce products using more sustainable methods. PI creates an enormous number of process options. In order to manage the complexity of options in which a feasible and optimal process solution may exist, the application of process synthesis tools results in the development of a systematic methodology to implement PI. Starting from an analysis of existing processes, this methodology generates a set of feasible process options and reduces their number through a number of screening steps until from the remaining feasible options, the optimal is found. The application of this systematic methodology through a computer-aided framework is presented through a case study, the chemo-enzymatic synthesis of N-acetyl-D-neuraminic acid (Neu5Ac).

Keywords: Methodology, Process synthesis, Process intensification, Biocatalysis, N-acetylneuraminic acid (Neu5Ac)

1. Introduction

Process intensification (PI) provides opportunities as well as challenges to satisfy the needs for significant improvement or development of new process options in the (bio) chemical industry (Stankiewicz and Moulijn, 2000) in order to achieve a more sustainable production. Process intensification can be defined as the improvement of a process by adding/enhancing phenomena in a process through the integration of operations, integration of functions, integration of phenomena or alternatively through the sole enhancement of phenomena in a given operation. Using this definition, PI potentially creates a large set of alternative process options. Hence, obtaining a feasible intensified process option that optimally improves the process is difficult.

Process synthesis is the systematic strategy to identify the optimal path to reach a given product in the desired quality and quantity with respect to defined constraints on the process. Therefore, application of process synthesis tools incorporating PI leads to the development of a PI synthesis/design methodology in which redundant intensified process options are systematically removed by checking against predefined constraints through a decomposition approach of the optimization problem. Also, it consists of an inner method for a detailed synthesis/design of specific sub-problems of PI options, for example, investigation of PI options considering only specific reaction-separation techniques.

In this paper, the developed systematic synthesis methodology to achieve PI together with some of the tools and algorithms needed is applied to the chemo-enzymatic synthesis of N-acetyl-D-neuraminic acid.

2. General synthesis/ design framework to achieve PI

The focus of the methodology is two-fold. First, all intensified options are generated and second, through screening with respect to feasibility and performance, the search space is reduced stepwise in order to locate candidate process options from which the optimal improved process is found. In the proposed methodology this is achieved by employing a hierarchical sequence of steps where the lower level steps employ simple and easy calculations, while the higher level steps employ more rigorous and detailed calculations, as shown in Figure 1. In this way, as in the techniques for computer aided molecular design (Harper and Gani, 2000), large numbers of infeasible options are screened out early and the remaining are gradually removed through further (more detailed) analysis. Such a methodology is reliant on structured knowledge, which is provided in this work through retrieval from a computer-aided knowledge base where relevant information obtained through a literature survey is classified in terms of known intensified processes and the principles on which the intensification are based, different methods for achieving PI and the tools employed to achieve them. Also, if a certain principle for PI is identified to have the potential for the targeted improvement, it is possible to consider existing PI methods for the corresponding operational tasks. The work and data flow for the integration of specific algorithms, here for the integration of reaction with reaction/separation for bio-based processes, is presented (Figure 1).
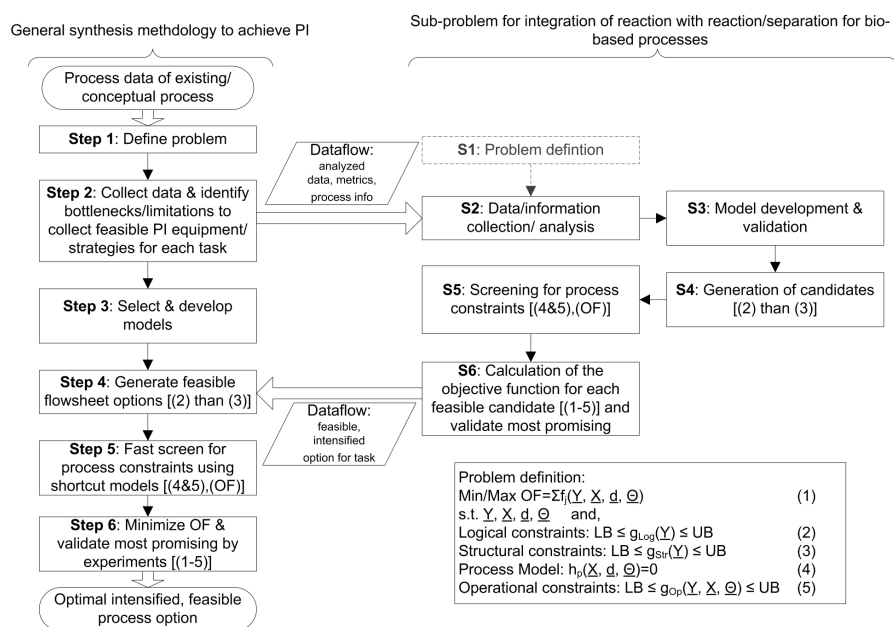

Figure 1. The integration of specific methods into general PI synthesis methodology - Work & Data flow, Binary decision variables $\underline{Y}$, Design variables $\underline{X}$, Equipment parameters $\underline{d}$ & product parameters $\underline{\theta}$

The PI-methodology follows a rationale which may be common sense but has not been established with respect to PI yet. The input to the methodology (and the specific synthesis methods,) is information about an existing or a conceptual process. Subsequently, the PI-synthesis/design problem is defined by an objective function (OF)

A systematic synthesis and design methodology to achieve process intensification in (bio)chemical processes

subject to a set of optimization variables (design $\underline{X}$ & binary decision variables $\underline{Y}$), equipment parameters $\underline{d}$ and product parameters $\underline{\theta}$) and constraints (such as logical, structural, operational). Objective function and metrics for evaluation are selected based on sustainability requirements such as operational costs, capital costs, safety, energy consumption, waste generation, efficiency, development time as well as intensified metrics, such as simplification, residence time, volume, flexibility. Next, available data about the process is collected in order to analyze the process behavior. In the case of the PI synthesis/design methodology this information is used to identify limitations/ bottlenecks to target in which steps the process needs improvement. The knowledge base tool is consulted to determine intensified equipment/principles which overcome identified limitations/ bottlenecks. If a PI option is identified for which a specific synthesis option exists, all information obtained together with the problem definition are delivered to PI sub-problems. Next, all equipments used in the PI-options are screened for feasibility through information obtained in the process analysis. Because data is usually incomplete, process models are developed and used to supplement the missing data. In particular, based on analysis of reported bio-processes, a superstructure representing all intensified bioprocess options (see Figure 2) involving reaction-separation has been developed. A generic process model has been derived for this superstructure consisting of mass and energy balance equations and connection equations. From the generic model, different specific process\operation sub-models are generated for the subsequent screening steps. Process options based on unreliable models are removed.
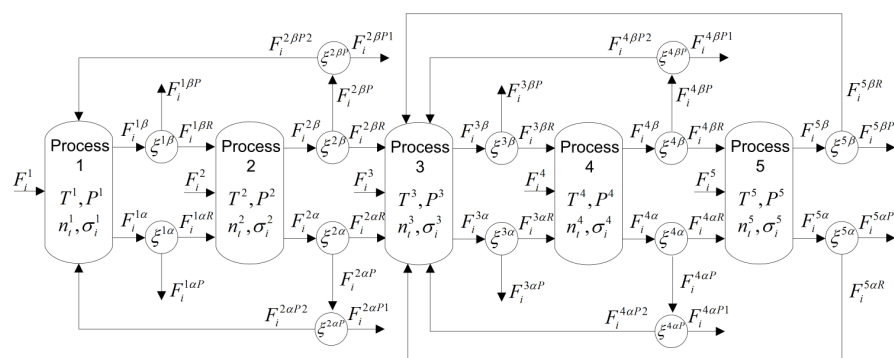

Figure 2. Superstructure for the integration of reaction with reaction/separation for bio-based processes.(Symbols: streams F, temperature T, pressure P, molar hold up's n, separation factors σ,, binary existence variable ξ [0,1]; Subscripts: Bottom flow α, top flow β, product P)

In step 4, all process options, including those identified by the PI sub-problems, are synthesized through use of logical constraints and screened with respect to structural constraints (for sets of fixed binary decision variables). Redundant options are removed. In step 5, for the remaining PI options (for each subset of Y's), the process model and operational constraints are solved simultaneously determine a feasible set of variables $\underline{X}$ (if existent). The surviving PI-options are now benchmarked in terms of performance metrics & OF. In step 6, if some design variables $\underline{X}$ are still undetermined, the most promising process options, are optimized with respect to the OF (if not, only OF is calculated). Finally, a small number of PI-options are selected for validation by rigorous simulation and experiments, thereby, reducing the number of experiments, the cost and time for development.

3. Case study

The methodology is highlighted through the production of Neu5Ac which is an important pharmaceutical intermediate due to its anti-viral, anti-cancer and anti-inflammatory effect (Zimmermann et al., 2007). One of the current synthesis routes for Neu5Ac synthesis, run in batch mode, is presented in Figure 3 (Blayer et al., 1999).
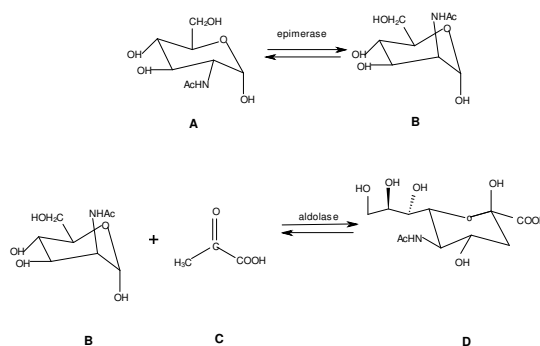

Figure 3. Synthesis of Neu5Ac from GlcNAc in two reaction steps; A: N-acetyl-D-glucosamine (GlcNAc); B: N-acetyl-D-manosamine (ManNAc); C: pyruvic acid (Pyr); D: N-acetylneuraminic acid (Neu5Ac); Enzymes: N-acylglucosamine-2-epimerase, N-acetylneuraminic acid aldolase

Starting with the PI-methodology, in the first step the problem is defined .The objective function is defined to maximize the yield of the process (Blayer et al., 1999)

$$Max \, OF = yield = f(\underline{Y}, \underline{X}, \underline{\theta}) \tag{6}$$

subject to binary decision, design variables and specifications and, process model h_p, logical constraints such as connection rules for synthesis, structural constraints related to metrics for evaluation of PI such as "Not use two different solvents for base catalysed epimerization due to enormous increase in waste generation" and operational constraints. The yield is defined as:

$$yield = \frac{n_{product,end} - n_{product,0}}{n_{substrate,0}} . \tag{7}$$

In step 2, all data with respect to the process scenario, such as, enzyme properties, solubilities, charges, properties with respect to inhibition of the enzyme, are collected. Through analysis of the current synthesis route, limitations of the process are identified to be an unfavourable equilibrium, high amounts of waste per kilogram of product and difficult downstream processing. Consulting the knowledge base tool, two possible PI strategies are retrieved for which a specific PI sub-problem with respect to bio-processes has been developed: the integration of both reaction steps in a single pot reactor and the integration of reaction/separation (in situ product removal, ISPR).

Hence, the defined problem as well as all obtained data is delivered to the second step S2 of the sub-problem (see Figure 1). Following the workflow of the PI sub-problem, the reported PI process options from the knowledge-base are collected together with conventional equipment and information related to the process tasks.

In S3 (specific method), the generic model is developed, represented by the superstructure containing all intensified/conventional equipments (Figure 2), and is validated based on experimental data for one pot synthesis obtaining an acceptable relative deviation in the product yield of less than 5%.

A systematic synthesis and design methodology to achieve process intensification
in (bio)chemical processes

The goal of S4 is to generate the set of feasible processes from the generic superstructure by fixing the binary decision variables *Y*. First, all feasible options to produce Neu5Ac are generated with respect to logical constraints. In total 622 feasible options are generated. The number of process steps for each option varies between 2 process units (e.g. one-pot synthesis followed by different possible downstream processes such as crystallization) and 5 process units (e.g. epimerization reaction, ManNAc enrichment, aldolase reaction, pyruvate treatment, downstream processing). Furthermore, the search space is reduced through structural constraints related to the PI metric. For example, the use of evaporation for ManNac enrichment is not recommended prior to the aldolase reaction step, since the biotransformation occurs only at dilute concentrations and the amount of energy needed for evaporation would be very high. The results related to PI metrics are presented in Table 1

Table 1. Results of the stepwise screening through set of structural constraints

PI Metric	Number of redundant options	Example of structural constraint:
Waste	104	Use of two different solvents for base catalysed epimerization
Energy	103	High energy use for evaporation for ManNAc enrichment
Efficiency	311	Enzyme in solution
Maturity	30	Difficult reported operation for whole-cell biotransformation
Operational cost	52	Chromatography discarded
Flexibility	12	Enzyme in solution

At the end of S4, the number of feasible process options has been reduced to 10.
The remaining options are further screened in S5 solving the process models with respect to defined operational constraints (fixing T, pH, substrate concentration) in order to find processes with a feasible set of *X*. These are benchmarked in this case with respect to OF and process time. The two most promising options, compared to a currently used configuration, are presented in Table 2.

Table 2. Simulation results of step 5 of the specific methodology. Conditions: Initial concentration of substrates: GlcNAc: 1.3 M, Pyr: 1.0 M. Enzyme concentrations: 12000 U/l for epimerase and aldolase. T=25°C, pH=7.5

Process	Scenario	Process time [h]	OF [%]	Number of processing units
Currently used	Epimerization, Aldolase reaction, Crystallization	10.45	37.8	3
Option 1	Epimerization, ManNAc enrichment, Aldolase reaction, Crystallization	12.982	81.1	4
Option 2	One-pot synthesis & ISPR by crystallization	8.27	75.7	2

In S6, the most promising options are optimized with respect to the objective function because two optimization variables (the two enzyme concentrations) have not yet been determined. At this point, all other binary and process variables have already been determined. The result of the optimization of the most promising option, which have been obtained in S5 (see Table 2), is graphically presented in Figure 4, indicating that, as expected, the most efficient process is found at the upper bound of loading for both enzymes. The value of the objective function is 81.8%.
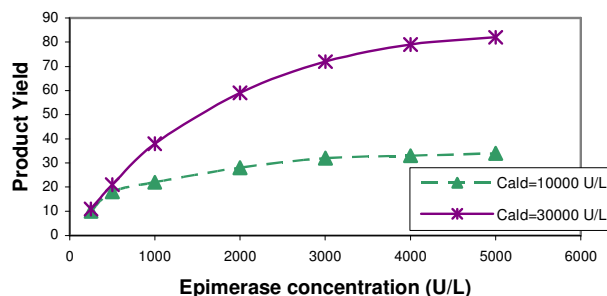

Figure 4. Product yield with respect to enzyme concentration; Initial concentration of substrates:
GlcNAc: 1.3 M, Pyr: 1.0 M., T=25°C, pH=7.5

The optimized PI-process obtained in this PI sub-problem is returned to the general PI-methodology where this option is benchmarked against other PI options/strategies obtained from other sub-problems. For this case study, no other strategies were under investigation (see step 2, PI methodology). Hence, the optimal solution with respect to yield found for the production of Neu5Ac is the option 1 (Table 2), which increases the product yield from 37.8 to 81.8, compared to an existing process.

4. Conclusion

It has been shown that process intensification has the potential to improve chemical and bio-based processes. The PI-methodology (together with the PI sub-problems) provides the means to generate, evaluate and identify PI-options, from which the optimal is found. That is, it is able to manage the complexity associated with finding the optimal intensified process. The developed methods and tools have been successfully applied to a relevant case study involving an important product from the pharma-sector. Current and future work is extending the PI-methodology in terms of extending the knowledge base, the models library and therefore the ability to handle a wide range of PI synthesis/design problems as well as further testing and validation of the combined methodology with case studies from different industrial sectors.

References

S. Blayer, J.M. Woodley, M.J. Dawson, M.D. Lilly. 1999, Biotechnology and Bioengineering, 66, 131-136.
P. M. Harper, R. Gani, 2000, Computers and Chemical Engineering, 24, 677-683.
A. Stankiewicz, J. Moulijn, 2000, Chemical Engineering Progress, 96, 22-34.
V. Zimmermann, H. Hennemann, T. Dauβmann, U. Kragl, 2007, Applied Microbial Biotechnology, 76, 597-605.

Multivariate Statistical Monitoring of Wine Ageing Processes

Ana C. Pereira[a,b], Marco S. Reis[a], Pedro M. Saraiva[a], José C. Marques[b]

[a]*Chemical Engineering Department, University of Coimbra, Pólo II - Rua Sílvio Lima, 3030-790 Coimbra, Portugal., apereir@eq.uc.pt*
[b]*Madeira Research Chemistry Center, Department of Chemistry, University of Madeira, Complexo da Penteada, 9000-390 Funchal, apereira@uma.pt*

Abstract

The flavor pattern is a key quality feature in the wine industry. Being the result of a complex interplay of different classes of volatile compounds, it presents an important evolution during the final and longer phase of the wine production process: the ageing period. In this paper, we present a data analysis framework for supporting the proper monitoring of the quality of wine products, during the ageing process, focusing on their flavor characteristics. We focus our analysis on a high quality Portuguese wine, the fortified Madeira wine, for which samples were collected over an extended time period, and analyzed in terms of their flavor composition using GC-MS. Then, several classification methodologies for age prediction were developed and evaluated, after a preliminary feature extraction stage using partial least squares for discriminant analysis (PLS-DA). Our analysis shows that it is indeed possible to identify the relevant ageing patterns and trends in a lower dimensional subspace, and to achieve very interesting classification performances, despite the natural variability present in wine products. The proposed framework also offers the potential to be applied in identity assurance and fraud detection tasks.

Keywords: Wine ageing process, multivariate analysis, partial least squares for discriminant analysis, classification.

1. Introduction

The whole wine industry has gone through massive changes in recent years and the challenges facing producers have increased significantly as the market becomes global and more competitive. In this context, we have been witnessing an increased investment of resources in research activities aimed at supporting wine production, in order to deepen our understanding about relevant phenomena going on during the winemaking process and enhance the quality assessment of the final product. Furthermore, the high economic value associated with some wines makes them rather vulnerable to adulteration practices and fraudulent copies. These potential dangers call for new and reliable tools for the proper assessment of the winemaking process and to provide the necessary identity assurance of high value wine products.

The final phase of the wine production process, called "ageing", deserves special attention and calls for strict process control. Most wines are consumed after a pre-specified period of ageing, during which the wine's identity is supposed to become

established. This period implies a significant financial overhead that is only recovered after several years, when the wine is sold. Given these premises, a lot is in stake during the ageing process, as significant quantities of wine are maturing and developing the final features that will grant them their final value in the market. In this context, the development of methods to assess and classify the quality of aged wines becomes of central importance. Therefore, in the present work, we have used a flavor chromatography technique for collecting the chemical data that characterizes wines, in a fundamental way, at their different stages of ageing (Figure 1).
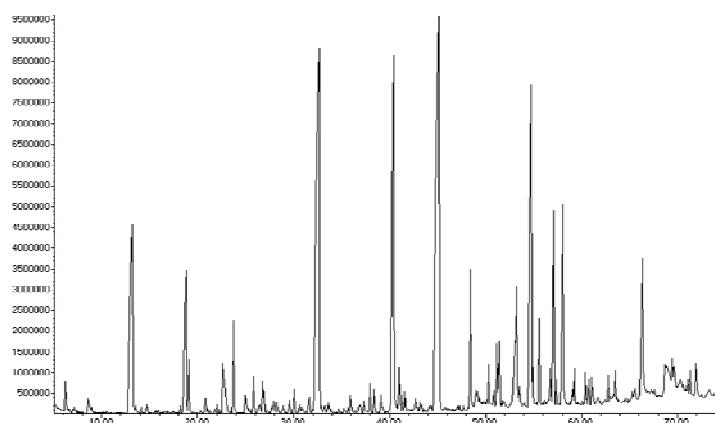

Figure 1. Total ion chromatogram from aroma profile of 1988 harvest Madeira wine obtained by SPE/GC-MS.

Then, we have applied multivariate statistical analysis techniques as well classification approaches (to be described later on in the paper), in order to compress all the information to a lower dimensional space, where monitoring and classification tasks can be efficiently and effectively conducted.

This paper is organized as follows. We first present our case study and address the main methodological steps considered in our analysis. Then, we present the results obtained through the application of feature extraction methods coupled with classification schemes. Finally, we discuss the predictive reliability of the classification models achieved using cross validation methodologies (k-fold Monte Carlo).

2. Data set

A total of 26 wine samples with known ageing periods (ranging from 1988 to 2006 in intervals of 2 years), produced from the same type of grape variety and kept under similar conditions, were analyzed in duplicate through gas chromatography coupled with mass spectrometry (GC-MS). This is a critical step, as it enables the identification and quantification of all relevant compounds, even those present in quite low concentrations but that might still be quite important, as a consequence of, for instance, their strong impact in sensorial detection. In this regard, an extensive and artifact-free enrichment of the flavor compounds followed by an effective separation by means of

gas chromatography was optimized and established in order to properly identify those compounds qualified for characterizing wine quality. In summary, our dataset is composed by 26 samples of Madeira wine, representing different ageing periods (ten harvest years grouped in five classes), with information regarding 83 compounds in each, obtained by GC-MS.

3. Methods

The basic methodological steps consider in our analysis are the following ones: the analysis of volatile compounds was carried out using the solid phase extraction method and the extract thus obtained, was analyzed by an adequate GC-MS methodology (Agilent 6890N coupled to an Agilent 5975 quadrupole inert mass selective detector); after the identification and quantification of the compounds for all the samples in duplicate, a (52×83) data matrix was obtained, which was then analyzed through several statistical and classification approaches, as described in the next paragraphs.

3.1. Feature extraction method

In order to handle the high dimensionality and complex nature of collected data, a preliminary stage of feature extraction is advised. In the present situation, we tested several methods, after which we opt by the partial least squares method for discriminant analysis (PLS-DA) (Geladi and Kowalski 1986), given its improved ability to simultaneously compress the features space and achieve good discrimination among the several groups. PLS-DA essentially consists in estimating a linear model through the PLS algorithm (Barker and Rayens 2003, Martens and Naes 1989) relating the predictor variables, X (which in our case are the chemical composition data), with a response variable in the form of an indicator matrix, that codifies the class a given sample belongs to, by inserting a "1" in the corresponding column, all the others being "0" (the number of columns in the indicator matrix is usually equal to the number of classes). In this regard, this method estimates a low dimensional subspace that still enables a good class separation while still explaining a good fraction of X-variability.

3.2. Classification methods

The development of a suitable monitoring approach for wine ageing encompasses building a methodology for predicting the wine ageing period from the chemical measurements made available by flavor chromatography. Such methodology essentially consists of a classification problem, where the classes are the ageing periods (as presented in Table 1) and the predictors are the chemical composition data or a set of features extracted from them.

In our study, we have tested several classifiers, and will refer here the results obtained with a parametric methodology (the linear discriminant classifier, LDA) as well as a non-parametric representative (the k-nearest neighbor method, kNN).

Table 1: Wine samples analyzed and respective class in the classification problem

	Classes				
	1	2	3	4	5
Harvest year	1988 1990	1992 1994	1996 1998	2000 2002	2004 2006

Being a parametric technique, Linear Discriminant Analysis relies on the estimation of statistical parameters for the class distributions of objects (a normal distribution is assumed), leading to the derivation of a linear decision function of the type $a^T X$, that maximizes the ratio of between-class variance to within-class variance, when the covariance matrices of the class dependent distributions are assumed to be all equal (Johnson and Wichern 2007). The non-parametric k Nearest Neighbours classification algorithm kNN does not assume any underlying probability distribution for the observations, but rather employs a distance metric for selecting the k nearest objects to the observation to be classified. Once the k nearest objects have been found, the observation is classified according to the category that prevails in such set of neighbors.(Hastie et al. 2001)

All data analysis tasks were performed in the computational platform MatLab (version 7.6, The Mathworks, Inc.).

4. Results

4.1. Feature extraction

Eighty-three variables were initially considered from the twenty-six samples analyzed. A PLS-DA model was built, whose X-scores for the first two latent variables (LV) are presented in Figure 2.a), which account for approximately 67% of the total system X-variability. From this plot, a natural separation of different aged wines is quite apparent, as well a consistent evolution trend. Occasionally, in such a 2-dimensional subspace, the evolution trend identified does not always have a good resolution, and there are some overlapping classes indicating that with only two LV and the available amount of samples, the ageing period distinction can be difficult to establish. However, this can be easily improved by increasing the numbers of LV used to develop the classification models, as will be seen later on. In Figure 2b), we present the individual evolution of the first two LV, by plotting their mean values and standard deviation, for each class. An evolution trend regarding the ageing processes is identified across the 1st LV – decreases during ageing, while for the 2nd LV there is a trend reversal in the second part, after class 3. These deterministic patterns extracted in the latent variables spaces, are likely to have a deeper meaning in terms of the chemical reactions underlying the ageing process, and what is cooperatively being produced and consumed in the cascades of reactions going on. This analysis was carried out with the use of contribution plots, but is outside the scope of the current paper.

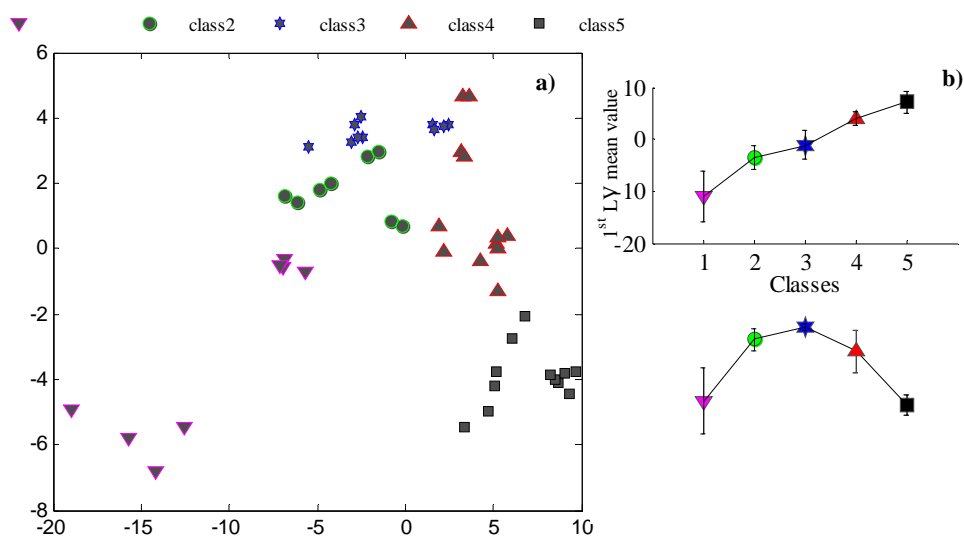*Multivariate Statistical Monitoring of Wine Ageing Processes*

Figure 2. a) LV1 versus LV2 scores obtained through PLS-DA. b) Evolution of the mean values of the different classes under study, regarding LV1 and LV2.

4.2 Model assessment and evaluation

Now we will focus on the assessment and validation of the classification models implemented over the scores obtained from the selected feature extraction technique (PLS-DA). The prediction performance of the classifiers LDA and *k*NN is assessed through a k-fold Monte Carlo Cross-Validation (MC-CV) methodology. This strategy consists of leaving aside a prescribed number of randomly selected samples (in our analysis, two observation from each class) to be used to test the classifiers, while the remaining ones are used to estimate its parameters; this process is repeated a large number of times, so that all samples were set aside approximately the same number of times, and finally the overall class-prediction error rate is computed. This approach was adopted not only to test the method but also to select the number of latent variables to consider in the model and their composition, i.e., the best combination of latent variables for a given size, after considering all possible combinations. The global error rates computed (for all such latent variables combinations) correspond to the mean value of 1000 Monte Carlo trials, for which the standard deviations were also computed. These results are presented in Table 2, where it is pointed out the results of this resampling scheme for the best three latent variable combinations regarding the global error rate and its standard deviation for classification.

Analyzing Table 2 it is possible to verify that both classifiers can achieve quite good performances, and that the *k*NN was the classifier leading to better classification results. We should point out that, in this study, there are replicates present in the dataset, which could positively bias the *k*NN classifier towards more optimistic results. Therefore, in order to avoid this effect, all the replicates of a given sample were left out from the analysis together with the sample under consideration, when the *k*NN classifier is

implemented. In this scenario the best three latent variables combination still lead to low classification error rate, about 1%.

Table 2. Mean and standard deviation (in bracket) of the MC-CV global error rate for the best combination of PLS-DA latent variables, using the Linear and *k*NN classifier. Also shown are the 3 best combinations of latent variables obtained for each size considered.

		1LV	2LV	3LV	4LV	5LV	6LV	7LV
Linear		0.461 (0.083)	0.309(0.108)	0.245(0.107)	0.144(0.096)	0.099(0.085)	0.059(0.074)	0.033(0.0551)
	1		1-3	1-4-7	1-4-6-7	1-4-6-7-9	1-4-5-6-7-9	1-3-4-5-6-7-9
	2		1-7	1-3-7	1-3-7-9	1-4-5-6-7	1-3-4-6-7-9	1-3-4-6-7-8-9
	3		1-4	1-4-6	1-4-7-9	1-4-5-7-9	1-3-5-6-7-9	1-3-4-5-6-7-9
kNN		0.401(0.113)	0.104(0.082)	0.028(0.04)	0.019(0.037)	0.015(0.037)	0.013(0.033)	0.018(0.030)
	1		1-4	1-3-4	1-2-3-4	1-2-3-4-7	1-2-3-4-5-6	1-2-3-4-5-6-7
	2		1-3	1-2-4	1-3-4-5	1-3-4-5-6	1-2-3-4-5-7	1-2-3-4-5-6-10
	4		2-3	1-4-7	1-3-4-7	1-2-3-4-6	1-2-3-4-9-10	1-2-4-5-6-7-10

5. Conclusions

In this paper, we have presented a classification approach for predicting the ageing period of wine, which is instrumental in the future development of wine monitoring techniques. The proposed methodology consists in combining a preliminary stage of feature extraction using PLS-DA, in combination with a classifier, such as LDA and *k*NN. Results achieved show that it is possible to achieve quite good classification accuracy in a reduced dimensional space, especially using *k*NN on latent features obtained through PLS-DA. Therefore, we can conclude from the results presented here, that we are indeed one step closer to be able to effectively follow the ageing wine process using chemical information properly processed by the proposed methods. Furthermore, the same framework offer very good characteristics to be applied in identity assurance and fraud detection tasks, namely by exploring the information contained in the residuals, as well as other informative statistics easily accessible.

6. Acknowledgements

References

M., Barker, and W. Rayens, 2003, Partial least squares for discrimination, Journal of Chemometrics, 17, 166-177.

P., Geladi, B. R. Kowalski, 1986, Partial Least-squares regression: a tutorial. Anal. Chim. Acta, 185, 1-17.

T. Hastie, R.Tibshirani, J. Friedman, 2001, The elements of statistical learning - Data Mining, Inference, and Prediction, Springer, New York.

R. A. Johnson, D. W. Wichern, 2007, Applied Multivariate Statistical Analysis, Pearson Education.

H.Martens, T. Naes, 1989, Multivariate Calibration, Wiley, Chichester.

Designing Eco-Efficient Biodiesel Production Processes from Waste Vegetable Oils

Sérgio Morais,[a,b] Sandra Couto,[a] António A. Martins[a] , Teresa M. Mata,[a*]

[a] *Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias, 4200-465, Porto, Portugal,* [*] *tmata@fe.up.pt*

[b] *REQUIMTE/Instituto Superior de Engenharia do Porto (ISEP), Rua Dr. António Bernardino de Almeida, 431 4200-072, Porto, Portugal*

Abstract

In this work the conventional alkali-catalyzed transesterification process for biodiesel production from waste vegetable oils is studied considering the two process alternatives normally used industrially: with and without free fatty acids (FFA) pre-treatment. Simulation models of these process alternatives are developed using the chemical process simulator ASPEN Plus® and their potential environmental impacts (PEIs) and economic potentials are determined and compared. Results show that the contribution to total PEIs of the process alternative with the FFA pre-treatment is 25% higher than the alternative without pre-treatment. Concerning the economic potential the process alternative with the FFA pre-treatment is greater showing a net present value of about 1.8 times higher than the alternative without the FFA pre-treatment. The comparison using plant data will be performed as future work.

Keywords: Biodiesel, Waste vegetable oil, Waste Reduction (WAR) algorithm, Alkali-catalyzed process, Free Fatty Acids Pre-treatment.

1. Introduction

Biodiesel is currently the most important alternative diesel fuel in EU, contributing to reduce the external dependence on fossil fuels and simultaneously the environmental impacts of the transportation sector, since it emits substantially lower quantities of most of the regulated pollutants compared to mineral diesel. However, biodiesel industry has some significant difficulties. In particular, feedstock selection can have a profound impact on the production process but also on food prices when food crops such as palm oil are diverted to energy. Additionally, the conversion of forests and other critical habitats for biodiesel feedstocks cultivation have the associated damage on biodiversity, loss of soil quality or land fertility, emissions from carbon stock change and land competition, among others. For these and other reasons the use of waste vegetable oils can be an effective way of minimizing some of the negative impacts associated with biodiesel production and at the same time of using a hard to treat residue.

In which concerns biodiesel production the transesterification reaction is very sensitive to the feedstock purity that directly affect the reaction performance.

Contrarily to refined vegetable oils, waste oils and animal fats usually have a lot of impurities, such as FFA and water that negatively affect the reaction yield, reducing the reaction rate by several orders of magnitude, even in small amounts (Canakci, 2007). For this reason some pre-treatment operations are conducted, such as the esterification of FFA to biodiesel (Aranda *et al.*, 2008).

2. Process Design and Simulation

In order to gather the process data needed for the inventory analysis, models of the two biodiesel production process alternatives were developed and simulated using ASPEN Plus®. A combination of NRTL (Non-Random Two Liquid) and UNIQUAC (UNiversal QUAsiChemical) thermodynamic/activity models were used to predict the activity coefficients in the process simulations, due to the presence of highly polar components (Zhang *et al.*, 2003ab; Kiwjaroun *et al.*, 2009). Some thermodynamic properties not available in the component library of the process simulator were estimated by providing the process simulator with the component molecular structure. Additionally, some components not directly available in the process simulator were represented by similar components chosen from the available components database. The triglycerides are represented by triolein ($C_{57}H_{104}O_6$), the FFA present in the waste vegetable oil are represented by oleic acid ($C_{18}H_{34}O_2$), and FAME are represented by oleic acid methyl ester ($C_{19}H_{36}O_2$). For the two process alternatives, product purities were defined to be 96.5% (w/w) for biodiesel, according to the European biodiesel standard (EN 14214) specification for esters content, and 91.0% (w/w) for crude glycerol. Vacuum operation for methanol recovery and products purification was applied to keep the temperature at suitably low levels because of the FAME and glycerol thermal decomposition temperatures (523K for FAME and 423.15K for glycerol).

Figure 1 shows a model, obtained using the process simulator ASPEN Plus®, of the alkali-catalyzed process for biodiesel production from waste vegetable oils, with and without the FFA pre-treatment. The pre-treatment steps are indicated inside the dashed line. Concerning the biodiesel production with the FFA pre-treatment, a waste vegetable oil stream flowrate of 428.50 kg/h (stream 104) is fed to an esterification reactor (R-100) where the FFA are converted to methyl esters. It is assumed a 6% (w/w) of FFA content in the waste vegetable oil (Zhang *et al.*, 2003a). It is used a fresh methanol stream flowrate of 9.50kg/h (stream 101) mixed together with a stream of recycled methanol (110) and the sulfuric acid catalyst (stream 103) as reactants to perform the esterification. The esterification reaction is performed at a temperature of 343.15K, a pressure of 405.3kPa, a methanol to oil molar ratio of 6:1 and with 5% (w/w) of sulfuric acid catalyst in methanol (Lepper and Friesenhagen, 1986; Zhang *et al.*, 2003a; Ma and Hanna, 1999). It is assumed a 95% conversion of FFA's in methyl esters (Zhang *et al.*, 2003a).The esterification products are sent to a glycerol washing column (X-100) to wash out sulfuric acid and water.
The pretreated oil stream (107) resulting from the esterification process is sent to the transesterification reactor (R-200), where a 6:1 molar ratio of methanol to

Designing Eco-Efficient Biodiesel Production Processes from Waste Vegetable Oil

oil is used with 1% (w/w) of sodium hydroxide to perform the reaction (Canacki, 2007; Freedman *et al.*, 1984, Ma and Hanna, 1999).
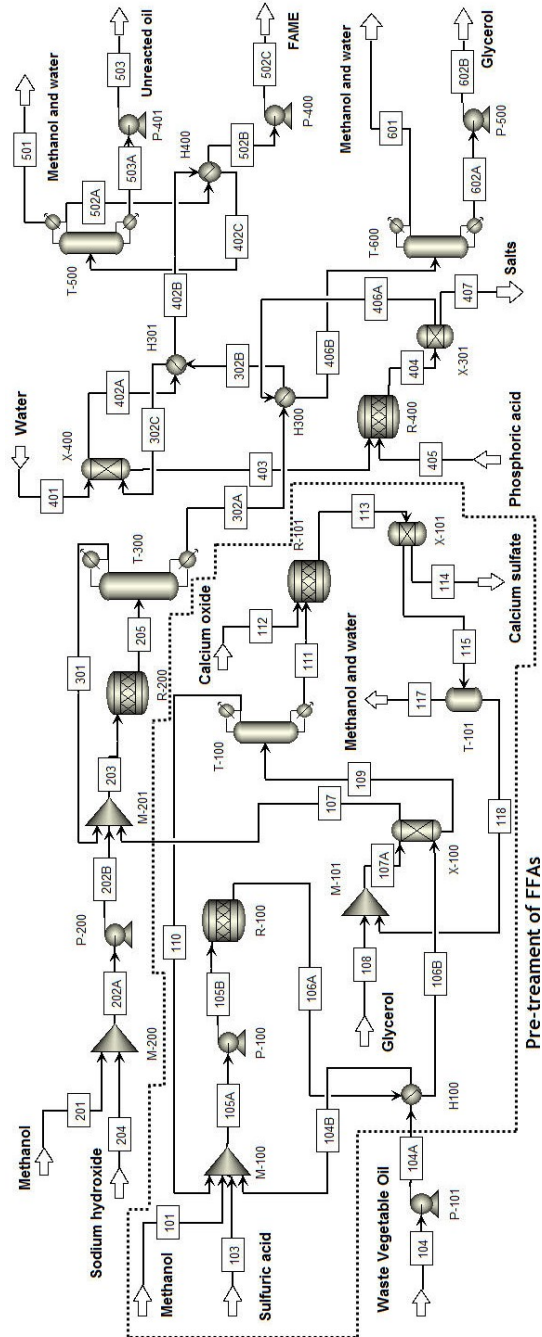

Figure 1 - Alkali-catalyzed process with and without FFA pre-treatment steps, indicated by the dashed line

A fresh methanol stream flowrate of 42.98 kg/h is mixed with anhydrous sodium hydroxide as catalyst and fed to the reactor. Transesterification takes place at 333.15K and 405.3kPa and reaches a 95% conversion of oil to FAME, after 2h (Zhang *et al.*, 2003a; Kiwjaroun *et al.*, 2009). The transesterification reactor products are fed to a vacuum distillation column (T-300), where 94% of methanol is recovered through four theoretical stages. The column bottom stream (302A), mainly containing biodiesel and glycerol, is charged to a washing column (X-400), where water is used to wash biodiesel, providing separation of methanol, soap, glycerol and catalyst from FAME. The top stream leaving the washing column (402A), mainly containing methyl esters and unconverted oil, is then feed to a vacuum distillation column (T-500) in order to separate FAME from water and methanol. FAME is obtained in the bottom stream (502A) of the column overhead condenser, with a mass flowrate of 411.13kg/h. The washing column bottom stream containing glycerol (403) is then fed to a neutralization reactor (R-400) in order to remove the catalyst using phosphoric acid. The resulting salts (Na_3PO_4) are then removed in a gravity separator (X-301) and treated as waste. Glycerol resulting from the neutralization reactor (R-400) is further purified in distillation column at a flowrate of 43.73kg/h.

Concerning the biodiesel production without the FFA pre-treatment that is represented outside the dashed line in Figure 1, the waste vegetable oil stream (104) is mixed with the unreacted oil stream (503), obtained from the bottom stream of distillation column T-500, at a flowrate of 428.5kg/h and is sent to the transesterification reactor. A 6:1 molar ratio of methanol to oil is used with 1% (w/w) of sodium hydroxide. It is also assumed a 6% (w/w) of FFA content in the waste vegetable oil. The transesterification reaction occurs at 333.15K and 405.3kPa and reaches a 78% conversion of oil to FAME, after 2 h, due to the presence of FFA's (Yan *et al.*, 2008). The next steps of the process are quite similar to the process alternative with the FFA pre-treatment. In the vacuum distillation column T-500, the FAME is obtained in the bottom stream (502A) of the column overhead condenser, with a mass flowrate of 330.22kg/h.

3. Potential Environmental Impacts

The Waste Reduction algorithm (WAR) developed by U.S. EPA algorithm was applied in order to evaluate the potential environmental impacts (PEIs) of the two process alternatives. It enables the assessment of the streams' PEIs crossing the system boundaries and uses a database for more than 1600 chemicals PEIs (Smith *et al.*, 2004). The mass flowrate of each component in the process streams is multiplied by its chemical potency to determine its contribution to the PEIs.

The WAR algorithm database values include local impact categories such as human toxicity by dermal/inhalation and ingestion routes (HTTPi and HTTPe), terrestrial toxicity and aquatic toxicity (TTP and ATP). Regional impact categories include photo-chemical oxidation and acidification (PCOP and AP), while global categories include ozone depletion and global warming (ODP and

Designing Eco-Efficient Biodiesel Production Processes from Waste Vegetable Oil

GWP). Each of these categories has scores that have been normalized within the category, while weighting factors are applied between categories. In this work all of the weighting factors are set to 1.0 (equal to each other) and it is assumed that the impacts from the various categories are additive.

Figure 2 shows the PEIs of the alkali-catalyzed process of biodiesel production from waste vegetable oils comparing both process alternatives with and without the FFA pre-treatment. As shown in Figure 2, the process alternative without the FFA pre-treatment has a lower contribution to the PEIs in all the impact categories. Also, the total PEI of this process alternative is 20% lower PEI than the alkali-catalyzed process with FFA pre-treatment.
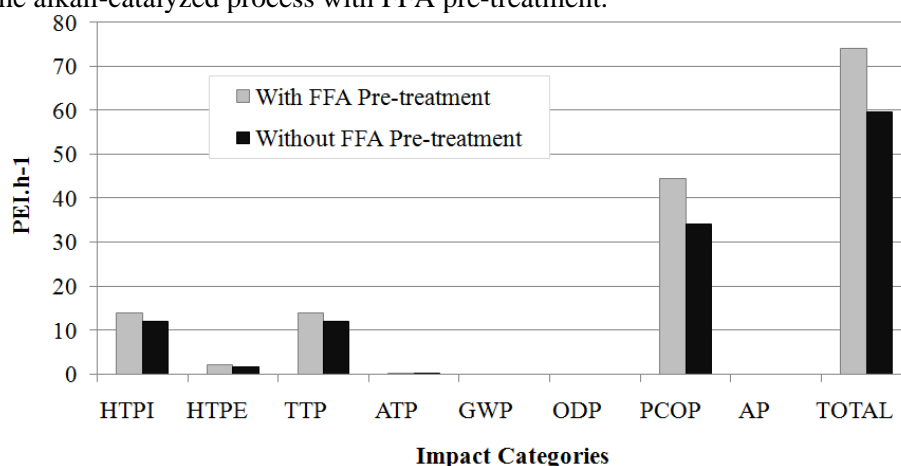

Figure 2 – PEIs of the alkali-catalyzed process of biodiesel production from waste vegetable oils with and without the FFA pre-treatment

4. Economic Analysis

The economic evaluation of both process alternatives is also performed to provide an opportunity to identify the potential relationships between economic potentials and PEIs (Mata *et al.*, 2003). Table 1 shows the economic evaluation of both alkali-catalyzed process alternatives for biodiesel production from waste vegetable oils, with and without the FFA pre-treatment, where the economic indicators determined are the net present value (NPV), the internal rate of return (IRR) and the payback period, calculated for a window of three years.

Table 1. Initial investment, Net present value (NPV), internal rate of return (IRR) and payback period comparing both process alternatives, with and without the FFA pre-treatment

Economic indicators	With FFA pre-treatment	Without FFA pre-treatment
Investment (EUR)	410.366	274.768
NPV (EUR)	906.460	489.556
IRR (%)	102,9	84,2
Payback period (years)	1,15	1,32

As shown in Table 1 the process alternative with the FFA pre-treatment is more advantageous economically. Although its higher investment it has a greater

NPV and IRR and a shorter payback period by comparison with the alternative without the FFA pretreatment. This is because the FFA pre-treatment improves the conversion of triglycerides and fatty acids to biodiesel.

5. Conclusions

This work compares the two alternatives normally used industrially for the alkali-catalyzed transesterification process: with and without a pre-treatment step of FFA. Models of these process alternatives are developed using the process simulator ASPEN Plus® and their PEIs are evaluated WAR algorithm. The economic potential of both process alternatives are also evaluated showing that there are environmental and economic trade-offs between the two designs. The process alternative with the FFA pre-treatment has a higher contribution to the PEIs but is more advantageous economically than the process alternative without the FFA pre-treatment. The comparison using plant data will also be performed as future developments of this work.

References

D.A.G. Aranda, R.T.P.Santos, N.C.O.Tapanes, A.L.D. Ramos, O.A.C. Antunes, 2008, Acid-Catalyzed Homogeneous Esterification Reaction for Biodiesel Production from Palm Fatty Acids, Catal. Lett. 122: 20-25.

M. Canacki, J. Van Gerpen, 1999, Biodiesel production via acid catalysis, Transactions of the ASAE; 42 (5): 1203-12120.

M. Canakci, 2007, The potential of restaurant waste lipids as Biodiesel feedstocks, Bioresource Technology 98(1):183-190.

B. Freedman, E.H. Pryde, T.L. Mounts, 1984, Variables affecting the yields of fatty esters from transesterified vegetable oils, JAOCS 61:1638– 1643.

C. Kiwjaroun, C. Tubtimdee, P. Piumsomboon, 2009, LCA studies comparing biodiesel synthesized by conventional and supercritical methanol methods, J Clean Prod. 17:143-153.

H. Lepper, L. Friesenhagen, 1986, Process for the production of fatty acid esters of short-chain aliphatic alcohols from fats and/or oils containing free fatty acids. US Patent 4608202.

F. Ma, M.A. Hanna, 1999, Biodiesel production: a review, Bioresource Technology 70: 1-15.

T.M. Mata, R.L. Smith, D.M. Young, C.A.V. Costa, 2003, Evaluating the Environmental Friendliness, Economics and Energy Efficiency of Chemical Processes: Heat Integration, Clean Technologies and Environmental Policy 5(3-4), 302-309.

R.L. Smith, T.M. Mata, D.M. Young, H. Cabezas, C.A.V. Costa, 2004, Designing Environmentally Friendly Chemical Processes with Fugitive and Open Emissions, Journal of Cleaner Production 12(2), 125-129.

S. Yan, M. Kim, S.O. Salley, J. Wilson, K.Y.S. Ng, 2008, Biodiesel Production by Simultaneous Transesterification and Esterification, Present at AIChE Meeting, Nov. 20.

Y. Zhang, M.A. Dubé, D.D. McLean, M. Kates, 2003a, Biodiesel production from waste cooking oil: 1 Process Design and Technological Assessment, Bioresource Technology 89:1-16.

Y. Zhang, M.A. Dubé, D.D. McLean, M. Kates, 2003b, Biodiesel production from waste cooking oil: 2. Economic assessment and sensitivity analysis, Bioresource Technology 90:229-40.

Rate-Based Modeling Approach and Simulation for Molecular Distillation of Green Coffee Oil

Melvin A. Durán,[a] Rubens Maciel Filho,[b] Maria R. Wolf Maciel[c]

[a] *Laboratory of Oleochemical, School of Chemical Technology, Technological University of Pereira (UTP), Vereda la Julita, Pereira, postcode 97, Colombia, melvin.duran@utp.edu.co*

[b] *LOPCA-School of Chemical Engineering, State University of Campinas (UNICAMP), Campinas, P.O. Box: 6066, Zip Code: 13083-970, Brazil, maciel@feq.unicamp.br*

[c] *LDPS-School of Chemical Engineering, State University of Campinas (UNICAMP), Campinas, P.O. Box: 6066, Zip Code: 13083-970, Brazil, wolf@feq.unicamp.br*

Abstract

This work describes a non-equilibrium mathematical model and simulation procedures for the fractionation of green coffee oil via molecular distillation. The simulation results were in quantitative agreement with previously reported experimental. Green coffee oil makes up to 18% (w/w) of coffee beans (C*offea arabica*). The main components of the coffee's lipids are triglycerides accounting up to 80% w/w, diterpene fatty acid esters amounting up to 18% w/w. The large amount of diterpene fatty acids renders Green Coffee Oil unsuitable for use as an edible vegetable oil. Fractionation of green coffee oil by molecular distillation offers an avenue to improve the quality of green coffee oil allowing its use in nutritional, cosmetic and pharmaceutical applications. Molecular distillation also provides a viable process to purify valuable products such as diterpene esters which has been reported to exhibit anticarcinogenic properties.

Keywords: rate-based, molecular distillation, Aspen Plus, coffee oil.

1. Introduction

Green and roasted coffee oil has a high price in the market and it is commonly obtained by mechanical cold-pressing and solvent extraction procedures. Crude green coffee oil exhibits a dark green color with a cloudy aspect and slight vegetable odor as well as an excessive amount of diterpenes of the kaurane family, mainly cafestol and kahweol.
Green coffee oil consists mainly of lipid components such as free fatty acids, free sterols, triglycerides, sterol esters, partial glycerides, diterpene fatty acid esters and polar lipids [1]. Diterpenes are receiving significant attention due their demonstrated emollient properties, their ability to increase serum cholesterol and block solar radiation as well as potential anticarcinogenic properties [2]. Diterpenes are present in the unsaponifiable lipid fraction of coffee oil [3]. Cafestol and kahweol are mainly esterified with various fatty acids, mainly palmitic and linoleic acids, hence only a small amount of the diterpenes is present in the free form [4].
Molecular distillation is well known fractionating process usually used for concentrating vitamins, essential fatty acids, antioxidants and minor components from crude vegetable oils [5-8]. Molecular distillation occurs at low temperatures, high

vacuum, and short residence times, hence reducing thermal decomposition and eliminating oxidation of the oil. During molecular distillation vapor molecules can reach the condenser without intermolecular collisions hence vapour-liquid phase equilibrium can not be reached [5, 6, 9].

The objective of this work is to model and simulate a fractionation process for the concentration of diterpenes fatty acid esters from green coffee oil via molecular distillation.

2. Characterization of Green coffee oil

We consider that green coffee oil is formed by three key compounds representing its most abundant groups: triglycerides, diterpenes fatty acid esters, and free fatty acids. These are complex compounds and many of their key physical properties are currently not available in the literature. Only the properties of free fatty acids are currently available in the database of Aspen-Plus® process simulation software. Green coffee oil was modeled as a mixture of triglycerides (93.66%), diterpene fatty acid esters (5.84%), and free fatty acids (0.5%), with an average molar mass of 818.57 kg kmol^{-1}. The triglycerides profile of green coffee oil has been previously studied [10]. These triglycerides are composed of fatty acids (**L,** linoleic acid, C18:2**; Ln**, linolenic acid, C18:3; **O**, oleic acid, C18:1; **P**, palmitic acid, C16:0; **S**, stearic acid, C18:0; **M**, myristic acid, C14:0). Table 1 shows composition of green coffee oil used in this work.

Cafestol is the primary diterpene component in arabica coffe, with kahweol making up to 50% of the cafestol. Our approach considers that diterpene fatty acid esters are formed by four main components: cafestol palmitate, cafestol linoleate, kahweol palmitate and kahweol linoleate.

Table 1. Composition of green coffee oil

Triglycerides	Molar mass (kg/kmol)	Composition % (w/w)
LLL	879.38	7.13
PLLn	852.72	2.45
OLL	881.40	4.49
PLL	854.74	28.59
OLO	882.77	1.57
PLO	856.75	13.28
SLL	882.77	3.91
PLP	830.73	25.64
POP	832.7	5.61
SOS	889.46	0.99
Cafestol palmitate	554.84	2,54
Cafestol linoleate	578.86	1.31
Kahweol palmitate	552.82	1.31
Kahweol linoleate	576.84	0.68
Palmitic acid	256.42	0.35
Linoleic acid	280.44	0.15

3. Computational model

Advances in the theoretical modeling of this molecular distillation have been reported by several authors [5-8, 12-14] with most of the reported models developed only for binary mixtures. Molecular distillation is characterized by direct transfer of molecules from the evaporator to the condenser with no possibility of return of them to evaporator.

Rate-Based Modeling Approach And Simulation For Molecular Distillation Of Green Coffee Oil

Under these circumstances there is no equilibrium between the vapour and the liquid phases and no true equilibrium pressure of the distilling molecules in the space between evaporator and condenser [5-9].

The proposed model comprises the following three steps:

- The creation of a property database with the main compounds of green coffee oil.
- Steady state simulation of the molecular distillation process using Aspen-Plus®.
- Model validation with experimental data.

3.1. Creating Database for Simulation

Physical property data for many of the key components of green coffee oil are not available. Only the properties of free fatty acids (palmitic and linoleic acids) are included in the database of Aspen-Plus®. Many of the physical properties of green coffee oil components can not be determined experimentally due to thermal decomposition of the components at temperatures below their normal boiling point. Aspen-Plus® requires the knowledge of the molecular structure, vapour pressure as a function of temperature, normal boiling point, liquid density, critical temperature, critical pressure, critical volume and acentric factor. Fortunately in cases in which not all of these properties are available Aspen-Plus® can provide accurate estimates using classical group contribution methods.

The estimation of vapour pressure of liquids as a function of temperature was done through the extended Antoine equation with data estimated by the Ceriani and Meirelles group contribution method [11].

3.2. Modeling the molecular distillation process

The non-equilibrium model, (also denoted as rate-based model), was initially presented by Krishnamurthy and Taylor [15] for conventional distillation process and consists of a set of mass and energy balances for vapor and liquid phases, along with rate equations for the evaluation of mass and heat transfer rates. This model use the Maxwell-Stefan equations for description of vapor-liquid mass transfer [15], and it requires information about parameters such as mass and heat transfer coefficients and vapour-liquid interfacial area. The method requires the evaluation of the mass and heat transfer processes for both phases separately. These parameters are usually obtained from semi-empirical correlations.

The following assumptions were made to simplify the rate-based model:

- The molecular distillation process is represented by a distillation column with only one tray and reboiler in steady state.
- Each phase is perfectly mixed in each segment.
- The assumption of phase equilibrium is made only at the vapor-liquid interface. The thermodynamic model UNIQUAC (for liquid phase activity coefficient calculation) is used in this research.
- The finite-flux mass transfer coefficients are assumed to be the same as the low-flux mass-transfer coefficients.
- The heat transfer coefficients are assumed to be constant for all segments.
- The reboiler is treated as equilibrium stage.

Figure 1 shows the rate-based concept for a column segment (a stage). In the rate-based model, thermodynamic equilibrium is assumed only at the vapour-liquid interface. The bulk phases of both vapour and liquid are assumed to be perfectly mixed, and the resistance to mass and heat transfer is located in tow films next to the phase boundary. Mass transfer rates are calculated by Maxwell-Stefan equations.
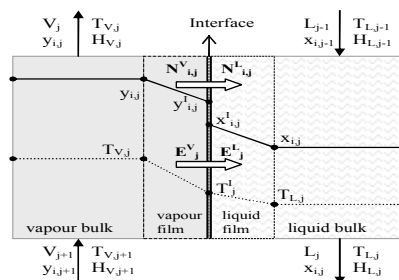
Fig.1. Schematic diagram of rate-based segment.

Where $N_{i,j}$ and E_j are the interfacial mass and heat transfer rate of component i on stage j, where i=1,2,…,c-1. L_j and V_j are the liquid and vapour molar flow rates leaving stage j. $x_{i,j}$ and $y_{i,j}$ are the mole fractions of component i in the liquid and vapour streams leaving stage j. $H_{L,j}$ and $H_{V,j}$ are the liquid and vapour phase enthalpies and $T_{L,j}$ and $T_{V,j}$ are the liquid and vapour phase temperatures. The $N_{i,j}$ are related to the chemical potential gradient in either phase by the Maxwell-Stefan equations [16].

The rate-based model has been implemented into the commercial Aspen-Plus® software package in the separation module RateFrac. The heat transfer coefficient in the vapour phase was calculated by the Chilton-Colburn analogy [17] and the mass transfer coefficient was calculated by the correlations of Scheffe and Weiland [18].

4. Results and discussion

4.1. Estimation of vapour pressure of components of green coffee oil
Equation.1 represents the extended Antoine equation used in this study.

$$\ln P^{sat} = A + \frac{B}{T} + CT + D \ln T \qquad (1)$$

Where P^{sat} is vapour pressure in (kPa), temperature T is in (K), and *A-D* refer to regressed parameters for the extended Antoine equation. The parameters for the extended Antoine vapour pressure equation are summarized in Table 2.

Table 2. Pure component parameters for the extended Antoine equation

Component	A	B	C	D
LLL	23.275	-15762.405	-7.1088792E-06	6.5817175E-03
PLLn	23.344	-15665.966	1.6706963E-08	-1.9010551E-05
OLL	23.292	-15771.513	-1.9035309E-06	2.5597721E-03
PLL	23.302	-15678.897	-6.0311229E-06	6.6242770E-03
OLO	23.300	-15780.231	1.6752738E-08	-1.9061677E-05
PLO	23.343	-15694.472	1.6730396E-08	-1.9040678E-05
SLL	23.301	-15780.231	1.6752738E-08	-1.9061677E-05
PLP	23.278	-15551.269	-5.2015317E-06	6.8921692E-03
POP	23.329	-15573.868	1.6820643E-08	-1.9140037E-05
SOS	23.275	-15801.088	1.6670462E-08	-1.8964202E-05
Cafestol palmitate	20.866	-12184.586	1.9669248E-08	-2.2378719E-05
Cafestol linoleate	21.129	-12520.333	1.9391296E-08	-2.2067425E-05
Kahweol palmitate	20.760	-12048.815	1.9695586E-08	-2.2397744E-05
Kahweol linoleate	21.024	-12386.029	1.9342710E-08	-2.1993897E-05

Rate-Based Modeling Approach And Simulation For Molecular Distillation Of Green Coffee Oil

4.2. Validation of the fractionation process

The rate-based model was compared with experimental data from literature. Table 3 provides a comparison between the cumulative mass percentage of distillate of coffee oil reported in the literature [19] and that predicted by the rate-based model. The cumulative mass percentage of distillate predicted using the rate-based model agrees quantitatively with the experimental data of Khan and Brown [19]. The calculated ARD (average relative deviation) was 2.65% for the rate-based model.

Table 3. Fractionation process by molecular distillation: comparison of experimental data with rate-based model

Fraction	Temperature (K)	Cumulative mass distilled, %	
		Rate-based model	Experimental data [19]
1	423	0.800	0.809
2	448	2.170	2.226
3	483	19.548	20.242
Residue	---------	100	100

In Figure 2, the predicted and reported data for percent cumulative distilled of coffee oil as a function of distillation temperature at 0.0015 mmHg are highlighted.

Most of the unsaponifiable matter is recovered in the third distillation fraction at 483K, together with considerable amounts of triglycerides, while free fatty acids were removed in the first and second fractions. It indicates that molecular distillation process would constitute a satisfactory procedure for separation of diterpene fatty acid esters from green coffee oil. The influence of distillation temperature ranging from 408 K to 490 K on the contents of diterpene fatty acid esters in distillates is shown in Figure 3.
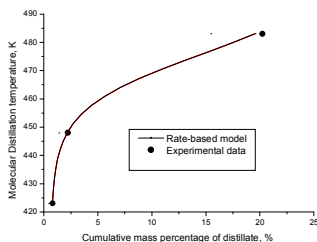

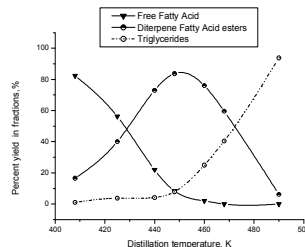Fig 2 Comparison of experimental data with rate-based model at 0.0015 mmHg

Fig. 3. Predicted effect of distillation temperature on diterpene fatty acid esters recovery in fractions (0.0015 mmHg).

The distillation temperature of 453K became a turning point for the changes in diterpenes fatty acid esters content in the distillates. At temperatures lower than 453 K, the diterpenes content in distillates decreased slightly while the content of free fatty acids increased. As shown in Figures 5, at 463 K the yield of distillate is 8.85% (w/w) and, content of diterpene fatty acid esters are about to 60%, what means increasing ten times over that in the raw green coffee oil. When the temperature was above 483 K, the content of large molecules such as triglycerides increased in the distillates and percent yield of diterpene fatty acid esters is less than 20% w/w. This behavior indicates that when mass in the distilled fractions is increased, the content of diterpene fatty acid esters diminish due to increase in the content of triglycerides. The components with larger molecular weights such as triglycerides are more difficult to evaporate; these components consisted of a larger proportion in residues than in distillates.

5. Conclusions

The availability of the rate-based model to represent the non-equilibrium process of molecular distillation for green coffee oil was demonstrated through a quantitative agreement between the experimental and simulated data. Specifically, the rate-based model was able to adequately predict the fractionation of green coffee oil in the temperature range of 408 K to 490 K at pressure of 0.0015 mmHg, which is commonly, used operating conditions for molecular distillation process.

The characterization of green coffee oil mixture, likewise the estimation of pure component vapour pressure and creation of non-databank compounds into Aspen-Plus® is a reliable alternative for predicting the behavior of green coffee oil. Distillation temperature and distillation pressure have important effect on the purification process by molecular distillation process. Molecular distillation is characterized by a direct transfer of molecules from evaporator to condenser without possibility of return to evaporator Due to this fact, the system cannot reach equilibrium state and a non-equilibrium model is needed to correctly simulate the process. For the specific case considered, the rate-based model with the Maxwell–Stefan equations is able to simulate the molecular distillation process of coffee oil.

6. Acknowledgements

References

[1] B. Nikolova-Damyanova, R.E. Velikova, G.N. Jham (1998) Food Res Int 31:479-486
[2] C. Cavin, D. Holzhaeuser, G. Scharf, A. Constable, W.W. Huber, B. Schilter (2002) Food Chem Toxicol 40:1155-1163
[3] I. Kolling-Speer, S. Strohschneider, K. Speer (1999) J High Resolut Chrom 22: 43–46
[4] K. Speer, I. Kölling-Speer (2006) Braz. J. Plant. Physiol 18:201-216
[5] C.B. Batistella, M.R. Wolf Maciel (1996) Comput Chem Eng 20: S19–S24
[6] E.B. Moraes, P.F. Martins, C.B. Batistella, M.E.T. Alvarez, R. Maciel Filho, M.R. Wolf Maciel (2006) Appl Biochem Biotechnol 132:1066-1076
[7] C.B. Batistella, E.B. Moraes, R. Maciel Filho, M.R.Wolf Maciel (2002) Appl Biochem Biotechnol 98:1187–1206
[8] E.B. Moraes, C.B. Batistella, M.E.T. Alvarez, R. Maciel Filho, M.R.Wolf Maciel (2004) Appl Biochem Biotechnol 113:689-709
[9] J. Shi, L. Rodríguez Posada, Y. Kakuda, S.J. Xue (2007). Separ Sci Tech 42:3029-3048
[10] A.G. González, F. Pablos, M.J. Martín, M. León-Camacho, M.S. Valdenebro (2001) Food Chem 73:93–101Z
[11] R. Ceriani, A.J.A. Meirelles (2004) Fluid Phase Equil 215:227-236
[12] A. Bose, H. Palmer (1984) Ind Eng Chem Fund 23:459-465
[13] M. Bhandarkar, J. Ferron (1991) Ind. Eng. Chem. Res 30: 998-1007
[14] J. Cvengros, J. Lutisan, M. Micov (1997) Separ Sci Tech 32:3051-3066
[15] R. Krishnamurthy, R. Taylor (1985) AIChE J 31: 449-456
[16] R. Krishna, J.A. Wesselingh (1997) Chem Eng Sci 52:861–911
[17] C. J. King, 2nd Edition, Separation Processes, McGraw-Hill, New York, (1980), 626
[18] R.D. Scheffe, R.H. Weiland (1987) Ind Eng Chem Res 26:228–236
[19] N.A. Khan, J.B. Brown, (1953) J Am Oil Chem Soc 30:606-609

A reduced model for the freezing step in ice cream manufacture

Bogdan Dorneanu[a], Costin S. Bildea[b], Johan Grievink[b], Peter M. Bongers[c]

aDelft University of Technology,DelftChemTech, Julianalaan 136, 2628BL, Delft, The Netherlands , b.dorneanu@tudelft.nl
bUniversityPolitehnica of Bucharest,Department of Chemical Engineering, St. Gh. Polizu 1-7, 011061, Bucharest, Romania, s_bildea@upb.ro
bUnilever R&D, Discovery Platform Structured Materials & Process Science, O. van Noortlaan 120, 3133AT, Vlaardingen, The Netherlands

Abstract

This contribution deals with the development of a reduced yet complex model, to support process design and operation. The model is computationally effective. The main physical phenomena considered in the model are the axial convective transport of mass, the radial outflow of heat at coolant wall to the refrigerant, the growth of the frozen ice layer, the periodic removal of the ice crystals by scraping and the melting of the ice crystal population in the bulk liquid. Rate equations for the relevant physical phenomena, as well as phase equilibrium conditions and thermodynamic equations of state are also present. The target output variables to meet the product quality specifications are the ice crystals size and the air content. Results of some preliminary steady state simulations are presented.

Keywords: ice cream, freezing, model reduction

1. Introduction

Ice cream is a complex colloidal system comprised, in frozen state, of ice crystals, air bubbles, partially coalesced fat globules, all in discrete phases, surrounded by an unfrozen continuous matrix of sugars, proteins, salts, polysaccharides and water. The freezing of the ice cream is performed in a scraped surface heat exchanger (SSHE). The mix is pumped along with air into the SSHE and the action of the rotor inside the tube blends the air into the matrix. The freezing medium in the jacket, typically ammonia, freezes the water into an ice layer against the wall from which finely dispersed ice crystals are continuously scraped off and incorporated into the ice cream. The quality, the palatability and the yield of the finished product depend on the way the freezing process is being operated [1], making it one of the most important operations in ice cream manufacture. The quality of the final product depends to a large degree on the distribution of the air bubbles and ice crystals within its structure. However, understanding the behavior of the material inside the SSHE is still incomplete [2]. Moreover, direct measurement of these variables inside the freezer barrel is difficult, if not impossible to achieve. For these reasons, a mathematical modeling approach should be applied to predict these quantities to support product design and process operation. In a previous model [3] of the ice cream freezing step, the SSHE was modeled in a simplified way. SSHE was considered as a series of well mixed stages. Heat, mass and momentum equations were developed considering the phases as pseudo-continuous. Details regarding the distribution of the air bubbles and ice crystals within the ice cream structure were not taken into account. In other models, the distribution of the ice crystals

or air bubbles is present only partially. For example, in [4] only the ice crystals distribution is considered. In the most recent model version [5] the freezing was modeled using a bottom-up approach, using a reduced set of coordinates a priori.

In the following sections, a more fundamental physical approach to the development of a reduced model for the ice cream formation is presented. The product is modeled as a non-Newtonian liquid with dispersed ice crystals and air bubbles. The process is modeled in two steps: (a) The fundamental phenomena are conceptually considered, simplifications are introduced in a judicious way, and the variables and the equations are written on a rigorous coordinate set; (b) Having too many coordinates for practical computation times and effort, the coordinate set is reduced. Averaged values are taken for the variables over the domains of the eliminated coordinates. The model includes equations of change for the mass and energy for all the phases and domains considered. Due to the complex rotational movements in axial direction and the non-Newtonian flow behavior of the ice cream, a major short-cut is taken in modeling the momentum conservation: radially symmetric axial convective flow of the bulk liquid driven by a pressure drop over the equipment.

2. Physical aspects

The model considers two types of phases: equipment related phases and process related phases. The equipment related phases are the solid freezer wall + the rotor and the coolant liquid. For the process side, the model takes into account three phases of which two are spatially and size-wise dispersed and the third one is lumped and continuous. The dispersed phases are the ice crystals and the air bubbles, while the pseudo-continuous matrix is a liquid, all sharing the same geometric domain, called the bulk layer (B) (see Figure 1).
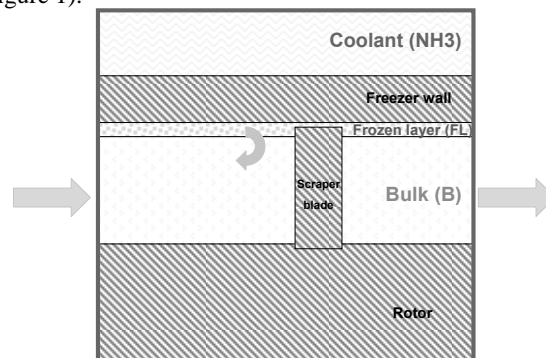

Figure 1. Model domains.

Next to the bulk layer there is a relatively thin frozen ice layer (FL) stuck at the wall of the freezer. Water crystallizes instantaneously on the freezer's wall and the ice is periodically removed into the bulk layer by the movements of the scraper blades. The thickness of the frozen layer in radial direction varies periodically due to growth by freezing in between two successive scrapings. The bulk consists of moving fluid ice cream, in which crystals and bubbles are imbedded. The rotational movement induces flow in radial and angular directions. The entire mixture has a complex non-Newtonian flow. To simplify, this layer is modeled as an axial plug flow. Concerning the ice particles only the melting is considered. No nucleation, attrition or agglomeration of ice is considered. The temperature of the frozen layer is assumed to be lower than the temperature inside the bulk at the same axial position. The rotor and the bulk have the

A reduced model for the freezing step in ice cream manufacture

same temperature locally. The temperature of the ice crystals melting inside the bulk is considered equal to the temperature at the interface with the frozen layer.

3. Model development

We will start by presenting the development of the rigorous, first-principles model and continue with the derivation of the reduced model of the freezing step. The modeling approach is presented in Figure 2.
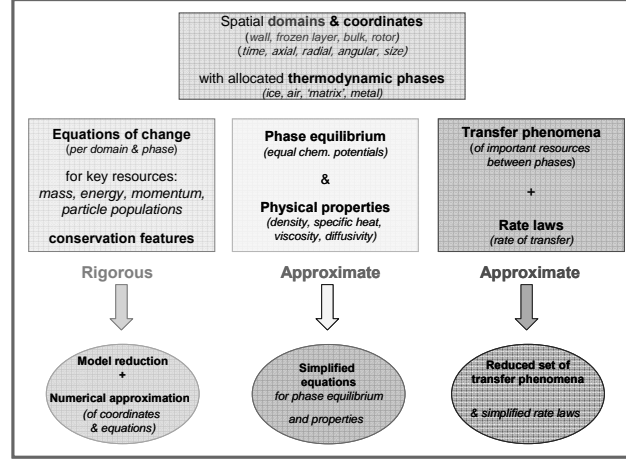

Figure 2. Modeling approach.

3.1. The first-principles model

Cylinder symmetry is imposed on the freezer, covering the external metal wall, the frozen layer and the bulk region. The conservation equations are based on five coordinates: the time, t, the axial coordinate, z, the radial coordinate, r, the angular coordinate, θ, and the internal coordinate of the crystals, $s^{(i)}$.

a) Mass conservation equations

The equations consider the dispersed phase as size distributed population. The matrix is considered as pseudo-continuous system with four components: water, sugar, fat and other components, while the air phase is modeled as a pseudo-continuous phase, too. The mass conservation equations are written for each phase/component and for both the frozen layer and the bulk.

For example, in the case of the size-dispersed phases, the population balance equation is written in the following form:

$$\frac{\partial\left(n_f^{(p)}\right)}{\partial t}+\frac{\partial\left(j_{f,z}^{(p)}\right)}{\partial z}+\frac{1}{r}\cdot\frac{\partial\left(r\cdot j_{f,r}^{(p)}\right)}{\partial r}+\frac{1}{r}\cdot\frac{\partial\left(j_{f,\theta}^{(p)}\right)}{\partial\theta}+\frac{\partial\left(j_{f,s}^{(p)}\right)}{\partial s^{(p)}}=B_f^{(p)}-D_{m,f}^{(p)} \tag{1}$$

Where $n_f^{(p)}$ is the concentration of particles of phase p in domain f, $\left[\dfrac{\#}{m^3}\right]$,

$j_{f,z}^{(p)}, j_{f,r}^{(p)}, j_{f,\theta}^{(p)}, j_{f,s}^{(p)}$ are the fluxes of particles of phase p in domain f in axial, radial,

angular direction and along the internal coordinate, respectively, $\left[\dfrac{\#}{m^2 \cdot s}\right]$, $B_f^{(p)}$ is the

birth rate of particles, $\left[\dfrac{\#}{m^3 \cdot s}\right]$ and $D_{m,f}^{(p)}$ is the death rate of particles, $\left[\dfrac{\#}{m^3 \cdot s}\right]$.

b) Energy conservation equations

Five energy balances are required. These balances are coupled by mass and heat transfer terms. The energy balances will be written for the rotor + bulk fluid phase + the air phase, the melting ice crystals in the fluid bulk, the frozen layer, applying the average thickness over the periodic scraper rotation, the metal wall and the coolant.
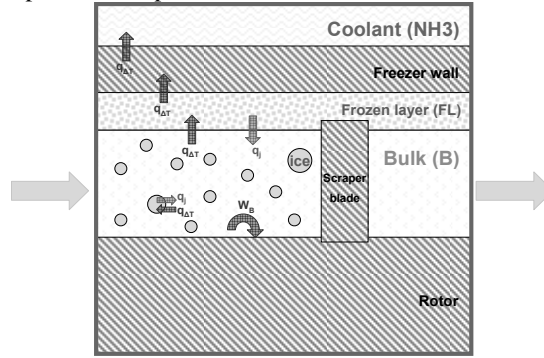

Figure 3. Energy transfer terms.

The general form of the energy conservation is formulated as follows:

$$\frac{\partial\left(\varepsilon_f^{(p)} \cdot c_f^{(p)} \cdot h_f^{(p)}\right)}{\partial t} + \frac{\partial\left(\varepsilon_f^{(p)} \cdot j_{m,f,z}^{(p)} \cdot h_f^{(p)}\right)}{\partial z} + \frac{\partial\left(\varepsilon_f^{(p)} \cdot j_{m,f,r}^{(p)} \cdot h_f^{(p)}\right)}{\partial r} + \frac{1}{r} \cdot \frac{\partial\left(\varepsilon_f^{(p)} \cdot j_{m,f,\theta}^{(p)} \cdot h_f^{(p)}\right)}{\partial \theta}$$
$$+ \left[\frac{\partial\left(q_{z,f}\right)}{\partial z} + \frac{1}{r} \cdot \frac{\partial\left(r \cdot q_{r,f}\right)}{\partial r} + \frac{1}{r} \cdot \frac{\partial\left(q_{\theta,f}\right)}{\partial \theta}\right] = \sum q \tag{2}$$

Where $h_f^{(p)}$ is the specific enthalpy of the phase p in domain f, $\left[\dfrac{J}{kg}\right]$, $q_{z,f}$, $q_{r,f}$, $q_{\theta,f}$ are the conductive heat fluxes in axial, radial and angular direction, respectively, $\left[\dfrac{W}{m^2}\right]$ while the term $\sum q$ represents all the heat transfer terms, $\left[\dfrac{W}{m^3}\right]$.

These transfer terms are presented in Figure 3.

c) Momentum conservation equations

Due to the complex rotational movements in axial and radial direction, and the non-Newtonian flow behavior of the ice cream, the momentum equations become very complex and are not shown here.

Finally, initial and boundary conditions are added to complete these differential equations.

The variables which affect the texture of the ice cream are the air content and the ice crystal size and are used as quality indicators. Since the distributed properties are

A reduced model for the freezing step in ice cream manufacture

difficult to determine experimentally, some more specific lumped measures must be specified. The feed intake, the processing conditions and equipment design parameters are adjustable to meet the target variables.

3.2. The reduced model

a) Conservation equations

The rotation time of the scraper is very short relatively to the time the fluid phase is transported through the freezer in axial direction. Due to the fast rotation of the scraper, the ice cream will exhibit complex spiraling movements and mixing in radial and angular directions with slow net transport in axial direction. This complex hydrodynamics of the ice cream will not be taken into account, assuming the effective mixing time is small. As a consequence, the ice cream is practically well mixed within a cross-sectional area, perpendicular to the axial direction. I.e., all process variables in the bulk region will be averaged over such a cross-sectional area, while they will vary with the axial position and time. Hence the mass conservation equation for the size-dispersed phases:

$$
\frac{\partial\left(\overline{n_f^{(p)}}\right)}{\partial t} + \frac{\partial\left(\overline{j_{f,z}^{(p)}}\right)}{\partial z} + \frac{\partial\left(\overline{j_{f,s}^{(p)}}\right)}{\partial s^{(p)}} = \overline{B_f^{(p)}} - \overline{D_{m,f}^{(p)}} + \overline{R_{t,f}^{(p)}} \cdot \frac{S}{A}
\tag{3}
$$

Where $\overline{R_{t,f,j}^{(p)}}$ is the rate of change of component j in domain f, S, A are the circumference and the cross-sectional area of domain f, respectively.

In a similar way, a reduced energy conservation equation is derived.

For the momentum equations a major short-cut is taken. A radially symmetric axial convective flow of the bulk mixture is assumed, with equal velocities of air, ice and liquid. The momentum conservation equation is written for the whole freezer in the form of a pressure drop equation, linking pressure drop with axial flow velocity.

b) Rate laws

In order to complete the dynamic model, information related to the constitutive equations for the rates of different processes presented above need to be added, as well as the ice-water-sugar phase equilibrium. There is scarce or inexistent information related to some of these rate equations. In the case of some of the rate equations there is an essential uncertainty related to the structure of the equations. For others, the structure of the equation is available, but there is nothing regarding the parameters. Before being able to solve the model, all these information need to be estimated. For this reason, all the rate laws are structured in the same fundamental way: the rate is proportional to one single thermodynamic driving force. This thermodynamic driving force is mechanically enhanced or diminished by the scraper rotation. The relative magnitudes of the rate constants, related to relaxation time constants, must be estimated from empirical data.

4. Results and discussion

The reduced dynamic model of the freezing step obtained in the previous section has a modest number of partial differential equations (PDE's) which have to be solved:12. However, the discretization of two of the three coordinates considered (time, axial position and internal coordinate of the particle) causes a multiplicative increase in the number of ordinary differential equations (ODE's): approximately 12000 ODE's. Algebraic equations for the different rate laws of the model need also to be taken into account. For a start, steady state simulations with the reduced model are performed. The

mass of the crystals is chosen as internal coordinate. The discretization of the internal coordinate assumes that for two classes of particles, $k-1$ and k, the mass ratio is constant. The obtained ODE's are integrated numerically using pre-defined Matlab ode solvers.
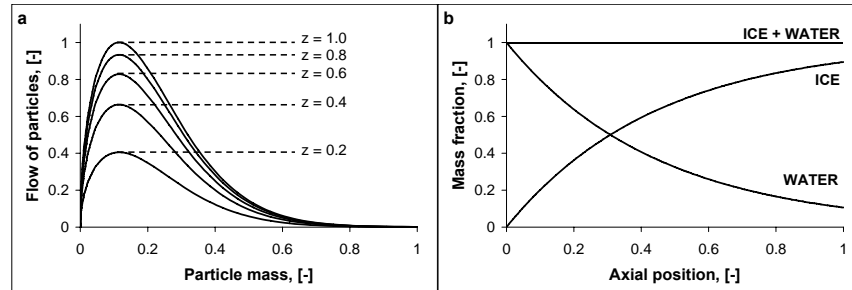

Figure 4. Variation of the (a) mass flow of particles; (b) mass fractions of ice and water along the freezer

Figure 4 presents the simulation results of the mass conservation equations for the ice crystals and the liquid matrix inside the bulk layer. The rate equations parameters are chosen by the user. The results are presented in adimensional form. The model predicts the expected trends in the variation of the number of ice crystals, as well as of the amount of water along the freezer.

5. Conclusion and future work

A reduced, yet complex model of the ice cream freezing step has been formulated, accounting for time, axial position and size coordinate of the particles. For lack of any detailed rate laws, a simple structure of the more important rate laws was postulated: the rate is made proportional to a single major driving force. Steady state simulations of the model were successful, while dynamic simulations form a computational challenge. Parameter estimation and model validation using experimental data is required in order to ensure the fidelity of the model predicted trends. Application to product quality optimization and equipment design is planned.

6. Acknowledgements

References

[1]. W.S. Arbuckle, 1986, Ice cream, AVI Publishing.

[2].B.R. Duffy et al., 2007, A mathematical model of fluid flow in a scraped-surface heat exchanger, Journal of Engineering Mathematics, 57.

[3]. P.M.M. Bongers, 2006, A heat transfer model of a scraped surface heat exchanger for ice cream, Computer Aided Chemical Engineering, 21.

[4]. J. Aldazabal et al., 2006, Deterministic model for ice cream solidification, Computational Materials Science, 38.

[5]. Dorneanu, B., Bildea, C.S., Grievink, J., Bongers, P.M., 2009, A first principles model for the freezing step in ice cream manufacture, Computer Aided Chemical Engineering, 26, 171.

Computer-Aided Lipid Design: phase equilibrium modeling for product design

Moises T. dos Santos,[a,b] Galo A.C.Le Roux,[a] Vincent Gerbaud [b*]

[a] *Universidade de São Paulo, Escola Politécnica, Laboratório de Simulação e Controle de Processos, Av. Prof. Lineu Prestes, São Paulo,5088-900 BRASIL*
[b] *Université de Toulouse, Laboratoire de Génie Chimique (LGC) UMR CNRS INP/UPS, BP 84234. Campus INP-ENSIACET. 4 allée Emile Monso. F-31432 Toulouse cedex 4.*
* Corresponding author:Vincent.Gerbaud@ensiacet.fr

Abstract

The aim of this work is to use phase equilibrium modeling as an auxiliary tool for product design, especially for lipids whose desired final properties are directly related to solid fat content (SFC) and melting behaviour through Solid-liquid equilibrium (SLE) modeling. This has been implemented for triacylglycerols mixtures, the main components of vegetable oils, a renewable raw-material for a wide variety of products. Excess Gibbs energy models were used to model the solid phases. Optimization of Gibbs free energy using Generalized Reduced Gradient was performed aiming to compute the number of molecules in each phase at the whole range of melting. As a result, the computed phase diagram was compared with experimental data from literature as well as a DSC curve. The model was also used to simulate a four-component DSC curve as a predictive tool.

Keywords: vegetable oils, solid-liquid equilibrium, optimization, triacylglycerol.

1. Introduction

Phase equilibrium calculations have been widely applied in chemical process design. However, equilibrium calculations have also a great potential to be used for product design, as many products have their desired properties directly related to phase behavior in multi-component mixtures (paints, rubber, plastic composites, agglomerated powders, extruded products, foams and foods) (Bruin and Jongen, 2003). Regarding the last ones, in fat-based products (lipids domain), the distribution of several triacylglycerols (TAG) molecules between solid and liquid phases directly impacts final products requirements, such as solid fat content (SFC), melting profile, hardness and texture. Besides this, vegetable oils are the main source of TAG molecules used in a large variety of products (food, biofuels, cosmetic, soaps, pharmaceutical and lubricants), and their renewable nature makes them proper to chemical industry sustainability issues. New applications are sought with tailor-made lipids, which are lipids specially designed to match a desired set of properties, and the use of rigorous thermodynamic based models can open opportunities for better predictive knowledge about final properties of such mixtures. However, a general phase equilibrium description of all possible mixtures of TAGs is lacking and experimental data is scarce. Some discussions about general TAG phase equilibrium modeling can be found in literature (Wesdorp, 1990; Won, 1993; Himawan *et al*., 2006), but computational results for multi-component mixtures remain scarce. The goal of this work is contribute to phase equilibrium modeling in such mixtures, using thermodynamic based description of solid and liquid phases and direct minimization of the Gibbs energy, allowing to

calculate the type and amount of triacylglycerols in each phase and solid fraction content (SFC) in a given temperature and overall composition.

2. Problem Modeling

2.1. Multiphase Multicomponent Solid-Liquid Equilibrium (SLE) Modeling

Triacylglycerols (TAG) are made by three fatty acids sterified to a glycerol backbone. Due to their high molecular weight, TAGs tend to crystallize in a solid network with different crystals packing (polymorphisms): unstable α, metastable β' and stable β forms (Sato, 2001) as illustrated in Figure 1.
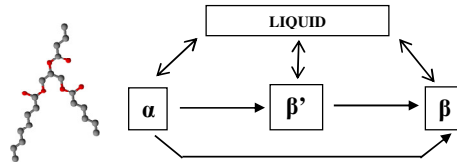

Figure 1: Triacylglycerol structures and state transitions.

In a multi-component system with a liquid phase and at least one solid phase, the condition for thermodynamic equilibrium is that the chemical potential of each component i in liquid phase is equal to that in any other j-solid phase:

$$\mu_i^{liquid} = \mu_i^{solid(j)} \tag{1}$$

or

$$\mu_{i,0}^{liquid} + RT \ln(\gamma_i^{liquid} x_i^{liquid}) = \mu_{i,0}^{solid(j)} + RT \ln(\gamma_i^{solid(j)} x_i^{solid(j)}) \tag{2}$$

For the chemical potential of molecules i in the reference state:

$$d\mu_{i,0}^p = -S_{i,0}^p dT + V_{i,0}^p dP \tag{3}$$

where:

$$\Delta S_{i,0} = \Delta H_{i,0} / T \ \ and \ \ \Delta H_{i,0} = \Delta H_{m,i,0} + \Delta Cp_{i,0}(T - T_{m,i}) \tag{4}$$

The effect of pressure in condensed phases (solid, liquids) can be neglected at pressures not too high. Assuming ΔCpi independent of temperature, after some rearrangements Eq. (2) can be rewritten as:

$$\ln\left(\frac{\gamma_i^{solid(j)} x_i^{solid(j)}}{\gamma_i^{liquid} x_i^{liquid}} \right) = \frac{\Delta H_{m,i}^{solid(j)}}{R} \left(\frac{1}{T} - \frac{1}{T_{m,i}^{solid(j)}} \right) - \frac{\Delta C_{p,i}}{R} \left(\frac{T_{m,i}^{solid(j)} - T}{T} \right) + ...$$

$$... + \frac{\Delta C_{p,i}}{R} \ln \frac{T_{m,i}^{solid(j)}}{T} \tag{5}$$

Computer-Aided Lipid Design: phase equilibrium modeling for product design

Experimental and theoretical backgrounds for lipids allow considering liquid and alpha solid phases as ideal (Wesdorp, 1990; Bruin, 1999; Himawan *et al.*, 2006). Thus, β' and β solid phases need description with an excess Gibbs free-energy model. The Margules equations are used, as this model is well-suited for mixtures whose components have similar molar volumes, shape and chemical nature (Prausnitz et al., 1986) and whose parameters can be predictable by using experimental correlation based on the isomorphism concept (Wesdorp, 90). The intensive Gibbs free energy for a phase p is:

$$g^p = \sum_{i=1}^{nc} x_i^p (\mu_{i,0}{}^p + RT \ln \gamma_i^p x_i^p) \tag{6}$$

Setting the chemical potential in the pure liquid reference state to zero and taking into account that for TAG mixtures ΔCp is small ≈ 0.2 kJ/mol (Wesdorp, 90), Eq.(5) can be simplified and the expressions for the Gibbs free energy become:

$$g^{liquid} = RT \sum_{i=1}^{nc} (x_i^{liquid} \ln x_i^{liquid}) \tag{7}$$

And for p being one of the possible solid phases (α, β' or β):

$$g^p = RT \sum_{i=1}^{nc} x_i^p \left(\frac{\Delta H_{m,i}^p}{R} \left(\frac{1}{T} - \frac{1}{T_{m,i}^p} \right) + \ln \gamma_i^p x_i^p \right) \tag{8}$$

2.2. Direct minimization of Gibbs free energy

Computing phase equilibrium is the solution of a nonlinear problem (NLP) for minimization of the total Gibbs free energy subject to material balance constraints. Therefore, the problem is:

$$\min G(n) = \sum_{i=1}^{nc} \sum_{j=1}^{np} n_i^j \mu_i^j(n) \tag{9}$$

s.t:

$$n_i = \sum_{j=1}^{np} n_i^j \quad i = 1...nc \tag{10}$$

$$0 \le n_i^j \le n_i \quad i = 1...nc; \ j = 1...np \tag{11}$$

where G(n) can be computed using Eq.(7) and Eq.(8).

3. Results and Discussions

3.1. Phase Diagram

The optimization problem was solved using GAMS (v.23). Among three solvers, CONOPT 3 (Generalized Reduced Gradient-GRG algorithm) was selected as this solver gave the best results in terms of CPU time and number of successful solutions obtained. Fig. (2) shows a calculated phase diagram for the binary mixture PPP-POP compared with experimental data from Bruin (99). It can be noted that the model was able to

predict the liquid line with good accuracy, while the solid line shows larger deviations, especially for PPP enriched mixtures.
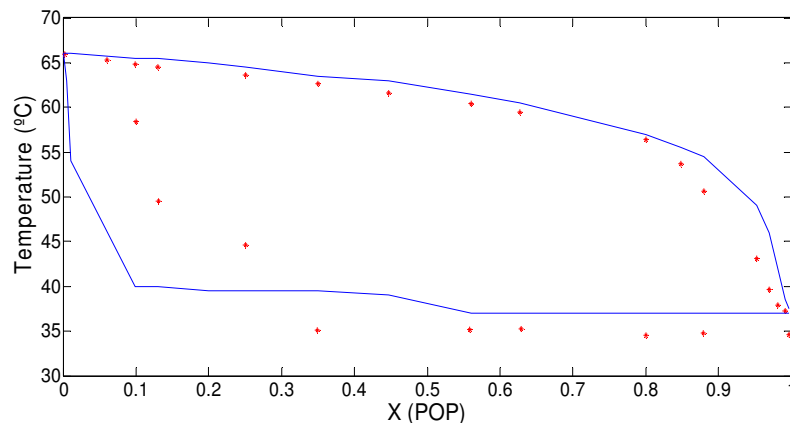

Figure 2: Phase diagram for mixture POP-PPP (experimental points in red).

3.2. Differential Scanning Calorimetry (DSC)

Experimental determination of phase diagrams has drawbacks: time consuming, accuracy sometimes uncertain, little data available, impurities leading to large deviations, start and end melting points difficult to determine and unstable forms not covered (Wesdorp, 90). Besides this, all the intermediary points (solid-liquid mixtures) are not used. To overcome this, Differential Scanning Calorimetry (DSC) can be used. Instead of using just the clear and softening points reported on the phase diagram, all the DSC curve points can be used. They are therefore well-suited for SLE models validation (Takiyama et al., 2002). DSC simulation is based on the following equation:

$$C_p{}^{app} = C_p + \frac{\partial G^E}{\partial T} + \sum_{j=1}^{P} \sum_{i=1}^{N} H_i{}^j \frac{\partial n_j{}^i}{\partial T} \tag{12}$$

Equation (12) shows that the apparent heat capacity (given by DSC measurements) can be calculated by using two derivatives obtained by numerical differentiation requiring two SLE calculations for each temperature in a DSC curve.

For validation of the model, a DSC curve was simulated using the results from the optimization step (number of mols of each molecule in each phase) for the ternary-mixture MPM-SSO-OOO and the results from the model were compared with experimental points from Wesdorp (90). Fig. (3) shows the curve as well as the corresponding simulated melting curve. They are in good quantitative agreement.
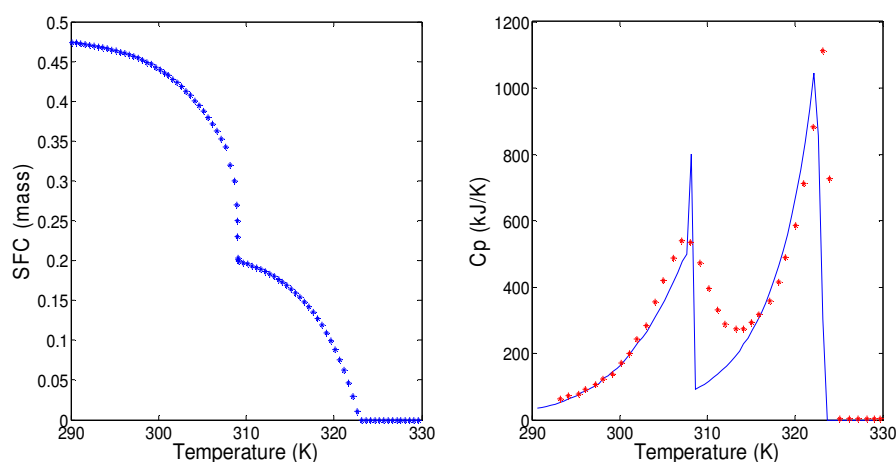
Computer-Aided Lipid Design: phase equilibrium modeling for product design

Figure 3: Simulated SFC vs Temperature and simulated (blue) and experimental (red) DSC for mixture MPM, SSO and OOO.

The model can also be used in a predictive manner. Fig.(4) shows a simulated melting curve and DSC for the mixture PPP, SSS, CCC and OOO. Three peaks are well observed, corresponding to the phase transitions for the molecules CCC, PPP and SSS respectively. The TAG formed by three oleic acids (OOO) in Fig.(3) and Fig.(4) is not observed as a peak, because at the first temperature it is already in liquid phase and it is used just as a liquid medium for make the solid diffusion rates faster.
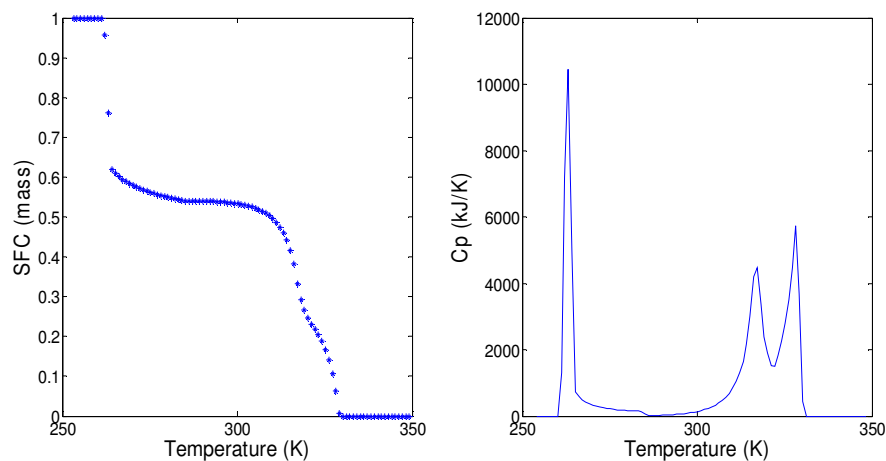

Figure 4: Simulated SFC vs Temperature and simulated DSC curve for mixture PPP, SSS, CCC and OOO.

4. Conclusions

SLE modeling and direct minimization of Gibbs free energy was used to compute a phase diagram and DSC curves for triacylglycerols mixtures. The model was able to

predict the melting behavior of such mixtures, and was evaluated with experimental data from literature. Its predictive capability is also tested. The present modeling has the advantage to be used as an auxiliary tool for product design using glycerol structure, as phase related properties can be evaluated in a pre-experimental step.

5. Acknowledgements

LIST OF SYMBOLS

G: extensive Gibbs free energy

G^E: excess extensive Gibbs free energy

g: intensive Gibbs free energy

nc: number of components

np: number of phases

n: total number of mols

R: gas constant

T: temperature

x_i^p: mol fraction of i in phase p

n_i^p: number of mols of i in phase p

$T_{m,i}^p$: temperature of melting of component i in polymorphism p

$\Delta H_{m,i}^p$: molar enthalpy of melting of component i in polymorphism p

$\Delta Cp,i$: molar heat capacity difference of component i between solid and liquid

C: capric acid (C10:0)

M: myristic acid (C14:0)

O: oleic acid (C18:1)

P: palmitic acid (C16:0)

S: stearic acid (C18:0)

Greek:

γ_i^p: activity coefficient of i in phase p

μ_i^p: chemical potential of i in phase p

$\mu_{i,0}^p$: chemical potential of i in phase p at the reference state

References

Bruin S., Th.R.G. Jongen, 2003, Food Process Engineering: The Last 25 Years and Challenges Ahead, Comprehensive Reviews in Food Science and Food Safety, 2, 42 – 81.

Bruin S., 1999, Phase equilibrium for food product and process design, Fluid Phase Equilibrium, 158–160, p. 657–671.

Himawan C., V.M. Starov, A.G.F. Stapley, 2006, Thermodynamic and kinetic aspects of fat crystallization, Advances in Colloid and Interface Science, 122, 3 – 33.

Prausnitz J.W., Lichtenthaler R.N., Gomes de Azevedo, E., 1998, Molecular thermodynamics of fluid phase equilibria, New York, Prentice-Hall, 3 edition.

Sato K., 2001, Crystallization behaviour of fats and lipids - a review, Chemical Engineering Science, 56, 2255 – 2265.

Takiyama, H., Suzuki, H., Uchida, H., Matsuoka, M. 2002. Determination of solid–liquid phase equilibrium by using measured DSC curves. Fluid Phase Equilibrium, 194–197, p. 1107–1117.

Wesdorp L.H. *et al*., 1990. Liquid – Multiple Solid Phase Equilibria in Fats: Theory and Experiments. In A.G. Marangoni, ed. *Fat Crystal Networks*, New York, 2005.

Won K.W., 1993, Thermodynamic Model of Liquid-Solid Equilibrium for Natural Fats and Oils, Fluid Phase Equilibrium, 82, 261-273.

Experimental investigation and mathematical modeling of water absorption in air-dried chestnuts

Altimari Pietro,[a] Adiletta Giuseppina,[a] Albanese Donatella,[a] Crescitelli Silvestro,[b] Di Matteo Marisa,[a]

[a]*Dipartimento di Ingegneria Chimica Alimentare Università di Salerno, Via Ponte Don Melillo, 84084, Fisciano (SA), Italy*
[b]*Dipartimento di Ingegneria Chimica Università Federico II, piazzale tecchio 80125 Napoli,*

Abstract

Chestnuts are characterized by limited shelf-life because of their high water activity and sugar content. For this reason, air-drying is typically performed to achieve physicochemical and microbiological stability of such fruits. However, undesired structural changes can occur during drying owing to the achievement of high processing temperature. These modifications can affect water permeability ruling out the possibility of uniformly rehydrating the dried fruit.

This contribution provides an experimental and modeling study of the rehydration process of air-dried chestnuts. With the objective of characterizing the effect of the drying temperature on water permeability, chestnuts previously dried at 40, 60 and 80 °C are rehydrated by immersion in water at 90 °C and sorption curves are constructed. Under the explored processing conditions, a purely Fickian diffusion model fails to describe the evolution of the absorbed amount of water due to the occurrence of swelling and starch gelatinization. Therefore, a mathematical model is formulated simultaneously accounting for the effect of diffusion, swelling and starch gelatinization on the transport of water. Parametric estimation of the model parameters is performed, based on the results of rehydration tests, by nonlinear regression techniques. In this way, satisfactory agreement between model predictions and derived experimental data is achieved and valuable information about the influence of the drying pre-processing conditions on water transport are obtained.

Keywords: Chestnuts, rehydration, non-Fickian diffusion.

1. Introduction

Chestnut is a traditional food product in Mediterranean countries and is widely employed for the elaboration of confectionary products. These fruits are characterized by limited shelf-life because of their high water activity and sugar content. For this reason, air-drying is typically performed to reduce water activity through moisture removal so as to achieve physicochemical and microbiological stability (Breisch, 1996). However, further processes can occur during drying owing to the achievement of high processing temperature eventually resulting in undesired modifications of certain characteristics of the material. Shrinkage, color changes and alterations of the porous structure can be, for example, observed depending on the drying temperature. Particularly, changes in the structure of material can affect water permeability ruling out the possibility of uniformly rehydrating the material. This might result in undesired texture properties of the rehydrated fruit. Therefore, it is of great concern to investigate

the effect of air-drying processing conditions on the transport of water during rehydration.

Rehydration is traditionally performed by immersion of the dried fruit in water at ambient temperature. However, this process can take long time and does not guarantee, in general, an homogeneous distribution of water. To overcome these limits, the achievement of larger rehydration temperature or the use of steam have been proposed (Lewicki Piotr, 1998). Under such processing conditions, complex phenomena can take place significantly affecting the transport of water during rehydration. At temperature values greater than 70 ^{o}C, a large growth in the diffusivity of water is, for example, observed due to starch gelatinization when the water concentration reaches a threshold value (Stapley et al., 1998; Cafieri et al., 2008). Moreover, the swelling of the fruit is enhanced as the rehydration temperature is increased. The influence of this phenomenon on the transport of water is evident at large times. Under these conditions, the chestnut slowly absorbs water even at large times due to a growth in the size of the pores although no significant concentration gradient is observed.

In spite of the relevance of such phenomena, no modeling studies accounting for the occurrence of diffusion, swelling and starch gelatinization during the rehydration of dried chestnuts have been performed. Rather, empirical or purely Fickian diffusion kinetics are employed to fit the data of sorption experiments (Moreira et al., 2008).

In this contribution, a mathematical model is formulated enabling to describe the effect of diffusion, swelling and starch gelatinization on the absorption of water in dried chestnuts. Parametric estimation of the model parameters is performed by nonlinear regression techniques based on the results of rehydration tests performed for chestnuts dried at 40, 60, and 80 ^{o}C immersed in water at 90 ^{o}C. In this way, the effect of the drying temperature on the transport of water during rehydration is investigated providing valuable information on how to select processing conditions enabling to achieve desired characteristics of the final product.

The rest of the paper is structured as follows. In section 2, a brief description of the experimental procedure followed to perform drying and rehydration tests is presented. In section 3, the evolution of the amount of water absorbed during rehydration is analyzed and the physical mechanisms responsible for the transport of water during rehydration are described. In section 4, a mathematical model enabling to satisfactorily cover the obtained experimental data is presented. Final remarks end the paper.

2. Materials and methods

Samples of "Palumna Cultivar" chestnuts were obtained from local farms in the Campania region, Italy. Drying was carried out on fruits with external shell and weight ranging from 10 to 11 g, at 40, 60 and 80 ^{o}C in a convection oven (Zanussi FCV/E6L3), with an air speed of 0.5 m/s, so as to reduce the average moisture of chestnuts. The chestnut weight was monitored at suitable time intervals by means of a digital balance (Gibertini E42) and drying was stopped when no appreciable variation in the moisture content was observed. The dried products were then rehydrated by immersion in water at 90 ^{o}C for about 8 h. During the rehydration process, the variations in moisture content were monitored. All drying and rehydration tests were carried out on 10 chestnuts and replicated three times.

3. Experimental analysis of the rehydration process

In order to qualitatively describe the physical mechanisms governing the transport of water in air-dried chestnuts, a preliminary analysis of the results of rehydration

Experimental investigation and mathematical modeling of the rehydration process of air-dried chestnuts

experiments is provided in this section. As representative example, we report in Figure-1, the evolution of the amount of water absorbed during rehydration at 90 $^{\circ}$C by chestnuts previously dried at 80 $^{\circ}$C as function of the square root of time. The choice of employing the square root of time as independent variable is aimed to get information about the transport mechanisms governing the absorption of water. The evolution of the moisture content of materials exposed to a low-molecular solute activity is indeed known to initially exhibit a linear dependence on the square root of time before reaching a constant value (Crank, 1968). Therefore, deviations from such behavior can be assumed to provide indications about the occurrence of further physical mechanisms affecting besides diffusion the transport of water.
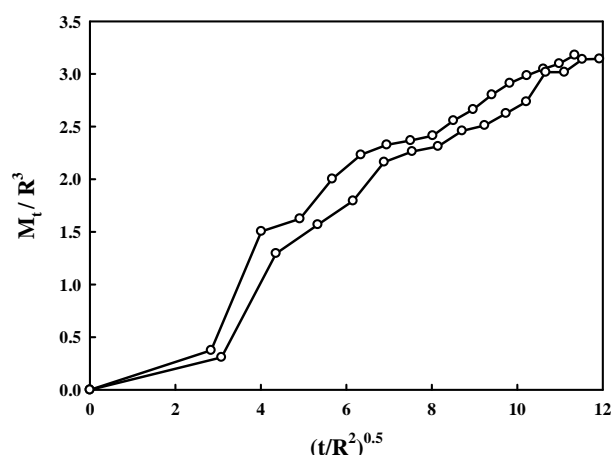

Figure-1. Evolution of the amount of water absorbed by chestnuts dried at 80 $^{\circ}$C as function of the square root of time.

The approach of comparing the evolution of the moisture content as function of the square root of time with the predictions of a purely diffusive model has been widely employed to characterize the physical mechanisms governing the transport of low-molecular solutes in polymeric materials. In this way, a classification of the observed deviations from purely diffusive behavior has been obtained and the phenomena responsible for such deviations have been characterized (Van der Wel and Adan, 1999).

The results of Figure-1 clearly indicate a strong deviation from purely diffusive behavior for the absorption of water in air-dried chestnuts. The evolution of the absorbed amount of water as function of the square root of time is S-shaped at small times. In particular, the absorption rate results initially low and rapidly increases at larger time instants. Such behavior is frequently observed during the absorption of low-molecular solutes in polymeric materials and is commonly attributed to the low establishment of equilibrium at the surface.

The rapid growth in the absorption rate observed in Figure-1 must be also attributed to the gelatinization of the starchy content of the chestnuts which typically occurs at temperatures larger than 70 $^{\circ}$C as the water concentration increases (Attanasio et al., 2004). Starch gelatinization indeed results in a large growth of the diffusivity of water.

Finally, it is important to note that water is slowly absorbed even at large times. In this respect, we stress that rehydration experiments showed that the absorbed amount of water invariably increases till the sample breaks apart. Such phenomenon must be attributed to the swelling of the chestnuts. At large times, water is indeed absorbed due

to the gradual growth in the sizes of the pores even though no significant concentration gradient is observed.

4. Mathematical modeling

According to the analysis presented in the previous section, the phenomena governing the absorption of water are molecular diffusion and starch gelatinization at intermediate times and the relaxation of the chestnut matrix at large times. This behavior is typically observed during the hydration of polymeric materials and is commonly referred to as two-stage sorption. The approach we take to describe such behavior is to consider the absorption process to be composed of two independent contributions: a diffusion part $M_D(t)$ governed by the Fick's law and a structural part $M_R(t)$ resulting from relaxation of the chestnut matrix. This approach has been widely employed to describe the transport of low-molecular solutes in swelling polymeric matrices (van der Wel and Adan, 1999) and has been proved satisfactory as the diffusion time is much smaller than the relaxation time (Wilde and Shopov, 1994). Therefore, the total weight gain at time t is expressed as follows:

$$M(t) = M_D(t) + M_R(t) \tag{1}$$

Here, $M_D(t)$ is computed by integrating the concentration profiles obtained by solving the following diffusive mathematical model:

$$\begin{cases} \dfrac{\partial c}{\partial t} = D\left(\dfrac{\partial^2 c}{\partial r^2} + \dfrac{2}{r}\dfrac{\partial c}{\partial r}\right) \\ c(R,t) = c_s(t), \ \dfrac{\partial c}{\partial r}(0,t) = 0, \ c(R,0) = c_0 \end{cases} \tag{2}$$

while $M_R(t)$ is expressed as follows:

$$M_R(t) = M_{R,\infty}\left(1 - \exp\left(t/\tau_R\right)\right) \tag{3}$$

with τ_R denoting the characteristic time of the relaxation process and $M_{R\infty}$ the amount of water absorbed after infinite time due to the only effect of swelling.

The formulation (1)-(3) accounts for the influence of diffusion and swelling on the transport of water. However, further physical mechanisms must be taken into account, according to the analysis provided in section 2, to describe the evolution of the chestnut moisture content during rehydration. In particular, strong deviations from purely diffusive behavior are initially observed due to the slow establishment of equilibrium at the surface and a rapid growth in the absorption rate is detected at intermediate times as the starch gelatinization process takes place. The approach here followed to describe small times deviations from purely diffusive behavior is to assume that the surface concentration c_s is not constant taking a finite time interval to reach its asymptotic value (Van der Wel and Adan, 1999). Therefore, c_s in (2) is expressed as follows:

$$c_s(t) = c_f\left(1 - \exp\left(t/\tau_s\right)\right) \tag{4}$$

Experimental investigation and mathematical modeling of the rehydration process of air-dried chestnuts

where c_f is the surface concentration of the chestnut under equilibrium conditions and τ_s denotes the characteristic time needed to reach such a value. It must be remarked that c_f does not represent the concentration actually observed in the chestnut at infinitely large times. The asymptotic water concentration observed in the chestnut results, in accordance with the formulation (1)-(3), from the sum of c_f and the concentration $M_{R\infty}/V_c$, V_c being the chestnut volume, absorbed after infinite time due to the only effect of swelling.

In order to account for the rapid growth in the absorption rate due to starch gelatinization, a dependence of the diffusion coefficient on the water concentration is also assumed. In particular, the following step function for the diffusion coefficient is employed:

$$D(c) = \begin{cases} D_1 & c \leq c^* \\ D_2 & c > c^* \end{cases} \tag{5}$$

where D_1 and D_2 denote the diffusion coefficient value before and after gelatinization respectively and c^* is the critical water concentration at which gelatinization occurs.

In order to implement the model (1)-(5), the following unknown parameters must be estimated: $M_{R\infty}$ c_f, c^*, τ_R, τ_s, D_1, D_2. It is worth to remark, in this context, that $M_{R\infty}$ can be expressed as difference between the amount of absorbed water experimentally observed after large times and the amount of absorbed water $c_f V_c$ predicted by solving the model (2), (4), (5). Even with such a simplification, the application of nonlinear regression techniques to the estimation of the unknown parameters results computationally expensive. In this respect, it is important to stress that the use of a step function for the dependence for the diffusion coefficient on the water concentration rules out the possibility to analytically solve the model (2), (4), (5) imposing the recursion to finite difference methods. In these conditions, it is of great importance to provide initial guesses for the unknown parameters close to the optimal solution. To this aim, the diffusion coefficient can be initially assumed to be constant. This approximation allows to analytically solve the resulting model (1)-(4) (Crank, 1980) and, therefore, to easily get an estimate for its unknown parameters. Such parameters can be then used as initial guesses with the complete model (1)-(5). The results of such estimation are reported in Table-1. A comparison between the evolution of the absorbed amount of water predicted by the model (1)-(5) with the estimated parameters and those experimentally observed for chestnuts dried at 40, 60 and 80 ^{o}C is shown in Figure-2.
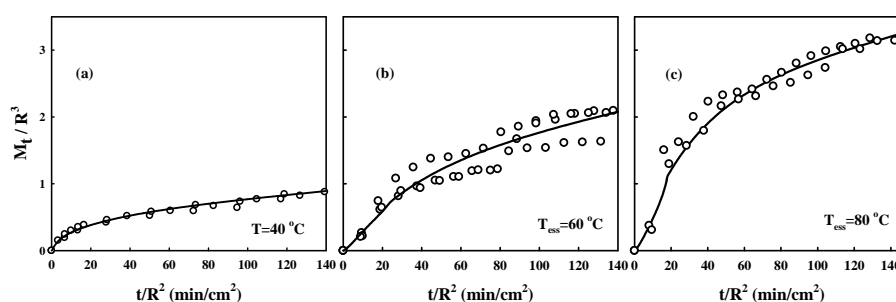

Figure-2. Comparison between the evolution of the amount of absorbed water predicted by the model (1)-(5) with parameter values reported in Table-1 and the results of rehydration tests for chestnuts previously dried at 40, 60, 80 ^{o}C.

Table 1. Estimated model parameters

	40 $^{\circ}$C	60 $^{\circ}$C	80 $^{\circ}$C
D_1 x 10^8 m^2/s	0.98± 0.037	0.96 ± 0.066	0.63 ± 0.05
D_2 x 10^8 m^2/s	4.94 ± 1.85	3.31 ± 0.3	3.32 ± 0.42
c^* g/cm^3	0.075 ± 0.0134	0.0772 ± 0.0045	0.0841 ± 0.0051
c_f g/cm^3	0.0736 ± 0.012	0.29 ± 0.04	0.4826 ± 0.02
τ_S min	1.75 ± 0.37	52.74 ± 9.02	47.72 ± 2.64
τ_R min	642 ± 16.9	371 ± 14.41	188 ± 23.96

It is apparent that the formulated mathematical model satisfactory covers the available experimental data. The presented results suggest a significant dependence of the rehydration behavior on the thermal history of the tested chestnut. A reduction of the diffusion coefficient D_1 and D_2 is observed as the drying temperature is increased. Also, the characteristic time of the relaxation process decreases at larger drying temperature values indicating a growth in the amount of water absorbed by swelling.

5. Conclusions

An experimental and modeling study of the rehydration process of air-dried chestnuts was performed. Chestnuts previously dried at 40, 60 and 80 $^{\circ}$C were rehydrated by immersion in water at 90 $^{\circ}$C and water sorption curves were constructed with the objective of characterizing the influence of the drying temperature. The evolution of the absorbed amount of water was invariably found to exhibit deviation from purely diffusive behavior. These deviations can be mainly imputed to the occurrence of swelling and starch gelatinization. Hence, a mathematical model was formulated accounting for diffusion, starch gelatinization and relaxation of the chestnut matrix. Satisfactory agreement between model predictions and experimental data was achieved and valuable information about the influence of the drying temperature on the extent of the addressed transport mechanisms were derived.

References

G. Attanasio, L. Cinquanta, D. Albanese, M. Di Matteo, (2004), *Food Chemistry*, 88, 583.
H. Breisch, (1996), Chataignes et Marrons. Editions Centre technique Interprofessionnel des fruits et legumes, Paris.
S. Cafieri, S. Chillo, M. Mastromatteo, N. Suriano, M. A. Del Nobile, (2008), *Journal of Cereal Science*, 48, 857.
J. Crank, (1968) "The mathematics of diffusion", Oxford university press, New york.
P. Lewicki (1998). *Journal of Food Engineering*, 36, 81.
R. Moreira, , F. Chenlo, L. Chaguri, L. Fernandes, (2008), *Journal of Food Engineering*, 86, 584.
A. G. Stapley, , P. J. Fryer, L. F. Gladden, (1998), *AIChE Journal*, 44, 1777.
G. K. Van der Wel, O. C. G. Adan (1999), *Progress in organic coatings*, 37, 1.
W. P. Wilde, P. J. Shopov, (1994), *Composite Structures*, 27, 243.

Screen printed biosensors for detection of nitrates in drinking water

Donatella Albanese[a], Marisa Di Matteo[a], Crescitelli Alessio[a]

[a]*Dipartimento di Ingegneria Chimica Alimentare Università di Salerno, Via Ponte Don Melillo, 84084, Fisciano (SA), Italy, mdimatteo@unisa.it*

Abstract

Amperometric screen printed nitrate reductase by Escherichia coli biosensors have been developed to detect nitrate ions. The effects of two different immobilization procedures of NaR and mediator (Azure A and Methyl viologen MV) to increase the performances (response time, sensitivity, stability) of the nitrate sensor have been studied. The best performances were registered by nitrate biosensor developed with a simultaneous cross-linking of enzyme and MV mixed BSA in in satured glutaraldehyde vapour. This biosensor showed the capability to detect nitrate residue levels lower than the European legislation one for a long storage stability (> 36 days) representing an applicable and cheap method for the analysis of nitrate ions in water and food samples.

Keywords: nitrate, water, biosensor, food analysis

1. Introduction

During the past two centuries, the human species has substantially altered the global nitrogen cycle, by means anthropogenic input such as animal farming, urban and agricultural runoff, industrial wastes and sewage effluents. The results of the nitrogen cycle alteration turn out through a significant increase of the nitrate concentration in ground, superficial and ground water causing one of the most prevalent environmental problems on a worldwide scale. Moreover the potential contamination of groundwater through the percolation of nitrates through natural aquifers presents serious risks for human health. The two main risks to health that arise from the ingestion of nitrate are reported as ''blue baby'' syndrome and gastric cancer. In both cases, the principal protagonist is nitrite obtained directly from contaminated water supplies or derived from the reduction of nitrate by the multifarious bacterial colonies that reside within the mouth. The potential hazard for human health linked to daily ingestion of high nitrate amounts has induced the Union European to define 50 mg/l as limit of nitrate content of drinking water [Council Directive 98/83/EEC]. The most frequently used methods for detection of nitrate are based on Ion-Selective Electrodes (ISEs), or Spectroscopic, and liquid chromatographic methods (Moorcroft et al., 2001). These methods are sensitive and reliable, but require an expensive laboratory equipments and trained personnel. The use of biosensors as an analytical method to determine the presence of nitrate ions represents an attractive alternative for research because of their high sensitivity, selectiveness, ease and rapidity of use. Biosensors are based on the intimate contact between a biorecognition element that interacts with the analyte of interest and a transducer element that converts the biorecognition event into a measurable signal. These analytical devices are gaining momentum over classical analysis techniques due to their high selectivity and sensitivity, low instrumentation cost, ease of use and rapidity of the assay.. The nitrate biosensor reported in literature (Ramsay and Wolpert,

1999; Kirstein, et al., 1999; Quan et al., 2005) are based on the immobilization of the Nitrate reductase enzyme, as bio element able to recognize nitrate ions, on glassy electrode. Nitrate is enzymatically reduced to nitrite and the oxidized form of NaR is electrochemically reduced by an electron-transfer mediator. The possibility to miniaturize the nitrate biosensor using as transducer element with small size such as the Screen Printed Electrodes (SCE) represents an important goal in amperometric nitrate biosensors development. In this work the results inherent the development of nitrate biosensor with *Escherichia coli* Nitrate Reductase on screen printed electrodes are reported. In particular the effects of two different immobilization procedures of NaR and mediator to increase the performances (response time, sensitivity, stability) of the nitrate sensor have been studied. To improve the stability the speed and reproducibility of the analysis nitrate biosensor was connected to Flow Injection Analysis (FIA) manifold.

2.Material and methods

2.1 Reagents

Nitrate reductase (E.C. 1.9.6.1) from Escherichia coli (0,1 units/mg solid), Methyl Viologen (MV), Azure A, glutaraldehyde (GA) (50% aqueous solution), potassium nitrate, Bovine Serum Albumin (BSA) were purchased from Sigma (St Louis, USA). All other chemicals were of analytical grade or better.

2.2 Screen Printed Electrodes

Screen printed electrodes (SPE) were produced in three steps, by screen printing different consecutive layers on transparent polyester films. A first layer of a graphite ink was deposed to define the conducting track and the working electrode, the second one was a silver/silver chloride ink used as reference electrode, while the third layer consisted in an insulating ink, UV polymerizable.

2.3 Immobilization.

Nitrate biosensor preparation was made by two different immobilization procedure. In the first one 2 ul of Azure A and MV at 5mM in phosphate buffer (PB) 0.1 M were applied on the working electrode of SPE and drying them in air (method A). Successively the enzymatic membrane was prepared directly by dropping of a mixture BSA (0.1mg/ul): Glutaraldehyde (1.0%): NaR (0.4mg/ul) (20:10:5) on working electrode 5 ul. SPE was stored at room temperature for 1 h and successively at 4°C overnight. For the second immobilization method nitrate biosensors were prepared by dropping two mixture BSA (0.05mg/ul): NaR (0.05mg/ul):glycerol (4%): MV/azure A (0.05mg/ul) on working electrode 5 ul. Enzymatic cross-linking took place in satured glutaraldehyde vapour at 4°C overnight (method B).

2.4 FIA system

The characterization of nitrate biosensors and analysis of water samples were carried out by FIA manifold. PB 0.1 M pH 6.8, deoxigened by sonication for 20 min, was pumped through the FIA system by a peristaltic pump (Miniplus 3, Gilson, France). Nitrate standards at different concentrations were injected into nitrate biosensor by manual sample injection valve (six port loop inject valve, Omnifit, England) at 0.5 ml/min. The current corresponding to the reduction of MV at -800 mV and Azure A at -400 mV was measured by a potentiostat Palm Sens (Palm Instruments BV, Netherland).

Screen printed biosensors for detection of nitrates in drinking water

The components of the flow injection manifold were connected with 1.6mm OD x 0.3 mm ID Polytetrafluoroethylene (PTFE) tubing.

3.Results and discussion

The detection principle of nitrate by nitrate reductase biosensor is based on the redox reactions reported in fig.1. The performance of any biosensor in terms of sensitivity and long term stability is highly depended on both the enzyme loading and immobilization method. Because the enzyme loading was about 0.05mg/μl for both procedure of enzymatic cross-linking, the effect of immobilization method (A; B) and of two electrochemical mediators (MV; Azure A) on performances of biosensors were examined.
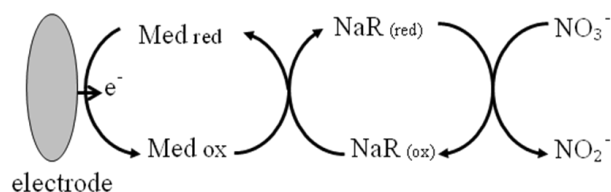

Figure 1. Detection principle of NaR biosensor.

3.1 Nitrate biosensor with Azure A.

Nitrate biosensor prepared with Azure A as electrochemical mediator, for both cross-linking procedure, showed unreproducible signals as reported in fig **2**. This behavior was probably caused by an insufficient immobilization of the mediator that was "eluted" by carried stream.
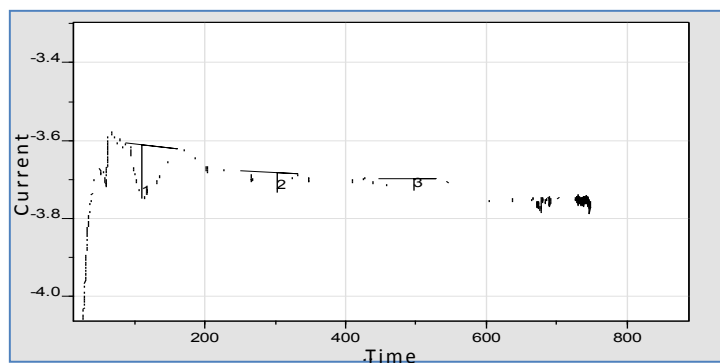

Figure 2. Nitrate biosensor Amperometric signals, developed by nitrate standards (1mM), with azure A as mediators.

3.2 Nitrate biosensor with Methyl Viologen.

Stable and reproducible currents were registered by nitrate biosensors prepared with MV. As far as the immobilization procedure realized by separated dropping of the MV and successively cross-linking of the nitrate reductase with BSA and GA, the results related to the biosensor characterization is reported in table 1. Nitrate biosensor showed good performances with a detection limit (0.5 mM) suitable to detect the maximum

residue level of nitrate content established by the European legislation, even if the storage stability at 4°C was lower than 1 day. After a storage time of 24 hours, the biosensor showed unreproducible amperometric peaks with a detection limit of 2mM. Better performances were registered by nitrate biosensor obtained by method B (table 1).

Table 1. Characterization of Nitrate reductase-MV biosensor prepared by method A and B

Parameter	NaR biosensor Method A	NaRbiosensor Method A
Sensitivity (nA/mM)	22.402	118.33
Correlation coefficient (R^2)	0.999	0.999
precision (RDS%)	3.57	2.98
Equation of calibration curve	y= 22.402x + 25.191	y = 118.334x+6.6155
Linear range (mM)	0.5-10	0.1-10
Detection limit (mM)	0.5	0.1

The data registered during the characterization of this biosensor highlighted that the simultaneous cross-linking of nitrate reductase and MV improved the performances of the biosensor in terms of sensitivity and detection limit of nitrate compounds. In contrast to method A, the simultaneous immobilization of enzyme and mediator (MV) improved the storage stability significantly. At regular interval time storage stability tests were carried out for a storage period of 36 days. During this interval time calibration curves were registered and current values related to nitrate standard injections at 0.5 mM were shown. (figure 3). Currents collected during the tests showed that the lifetime (t $_{L50}$) defined as the storage time necessary for the sensitivity within the linear concentration range, to decrease by a factor of 50%, was more high than 36 days. During this interval time, in fact, nA values decrease about 23 % and 38 % after 4 and 10 preservation days respectively, then stay at constant values of 44-48 nA. The calibration curve obtained at 36 days (fig. 4) showed good biosensor performances too, which exhibits the capability to detect the value of 0.8mM equal to 50mg/L established by the European legislation as maximum residue level of nitrate content in drinking water.

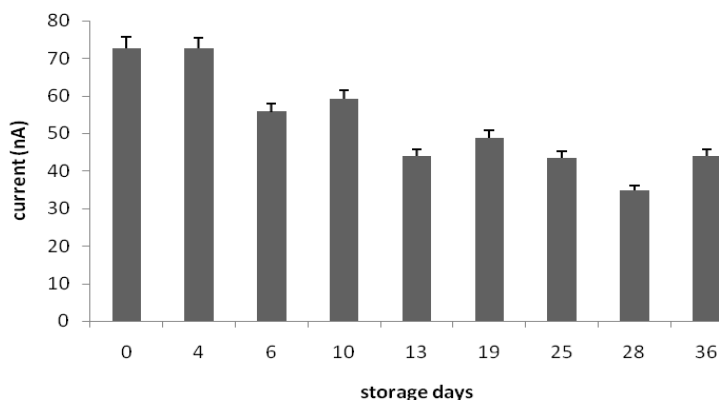*Screen printed biosensors for detection of nitrates in drinking water*

Figure 3. Current values of the nitrate-MV biosensor(method B) registered with nitrate standard at 0.5 mM during storage time.

$y = 52{,}925x + 8{,}5848$
$R^2 = 0{,}9971$
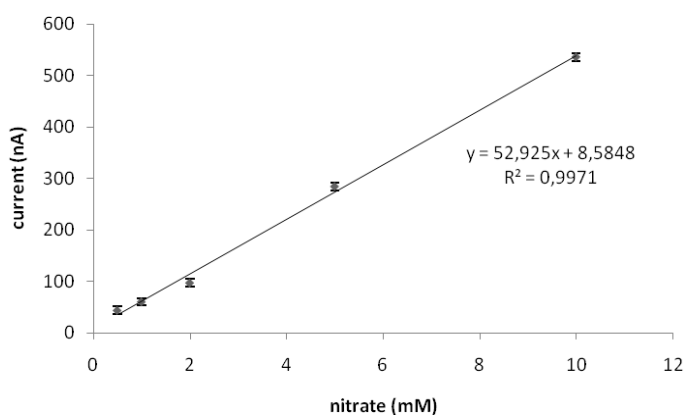
Figure 4. Calibration curve of the nitrate-MV biosensor by method B after 36 day of storage at 4°C.

4. Conclusions

Results suggest that the use of nitrate reductase by *Escherichia coli*, coupled with MV as mediator, can be used to developing nitrate biosensor able to determinate the presence of nitrate ions in water samples. The long storage stability time and the good performances highlighted that biosensor by method B was able to detect nitrate residue levels proposed by the European legislation, thus representing an applicable and cheap method for the analysis of nitrate ions in water and food samples.

5. References

1. Council Directive of 3 November 199812, concerning on the quality of water intended for human consumption (98/83/EEC)
2. Da Silva, S., Cosnier, S., Almeida, M. G., Jos Moura, J.G. (2004). An efficient poly(pyrrole–viologen)-nitrite reductaãse biosensorfor the mediated detection of nitrite. Electrochemistry Communications, 6, (404–408)
3. Kirstein, D., Kirstein, L., Scheller, F., Borcherding, H., Ronnenberg, J., Diekmann, S., & Steinrucke, P. (1999). Amperometric nitrate biosensors on the basis of Pseudomonas stutzeri nitrate reductase. Journal of Electroanalytical Chemistry, 474, 43-51
4. Moorcroft M., Davis J. and Compton R.G. (2001). Detection and determination of nitrate and nitrite: a review. Talanta 54, 785–803
5. Quan, D., Shim, J H., Kim, JD., Park, H S., Cha, GS. and Nam H. (2205) Electrochemical Determination of Nitrate with Nitrate Reductase-Immobilized Electrodes under Ambient Air. Anal. Chem. 77, 4467-4473
6. Ramsay, G. and Wolpert, SM. (1999). Utility of Wiring Nitrate Reductase by Alkylpyrroleviologen-Based Redox Polymers for Electrochemical Biosensor and Bioreactor Applications. Anal. Chem., 71, 504-506

Control System Development for Integrated Bioreactor and Membrane Separation Process

Oscar Andrés Prado-Rubio,[a] Sten Bay Jørgensen,[a] and Gunnar Jonsson,[a]

[a]*Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark, oap@kt.dtu.dk, sbj@kt.dtu.dk, gj@kt.dtu.dk*

Abstract

A bioreactor integrated with an electrically driven membrane separation process (Reverse Electro-Enhanced Dialysis - REED) is under investigation as potential technology for enhancing lactic acid bioproduction. In order reveal the operation of the integrated process to achieve a specific production goal, a methodology for goal driven control system development is expanded to handle periodically operating systems. In this paper the pH control issue is addressed. A sensitivity analysis is used as criterion for the conceptual design of the control structure.

Keywords: Lactic acid production, REED, sensitivity analysis

1. Introduction

Lactic acid is mostly produced using fermentation of carbohydrates by Lactic Acid Bacteria (LAB). Improvements in the design of the lactic acid bioproduction have been studied extensively driven by an increasing number of applications of the fermentation products. The main limitation for this bioprocess is that LAB normally are impaired by product inhibition at a certain concentration level of the main metabolic product. Therefore, the fermentation can be intensified by continuous lactate removal from the cultivation broth during a pH controlled fermentation, this will result in a higher productivity and product yield. An integrated bioreactor and a novel electrically driven membrane separation processes (Reverse Electro-Enhanced Dialysis - REED and Electrodialysis with Bipolar membranes - EDBM) has been recently proposed as a method for *in situ* continuous lactate removal from the fermentation broth, the first extraction stage is depicted in fig. 1 [5]. Additional productivity is achieved by the continuous recycle of biomass and unconsumed substrates, which allow working at higher cell densities and exploitation of the substrate.

The novelty of the process is the innovative electro-membrane separation process employed which selectively extracts the lactate and simultaneously facilitates the pH control in the fermenter. This is possible since the lactate ions are exchanged by hydroxyl ions in the REED module. In addition, the adverse influence of the membrane fouling is diminished by periodically reversing the polarity of the electrical field, ensuring longer operation compared to continuously operated membrane based separations.
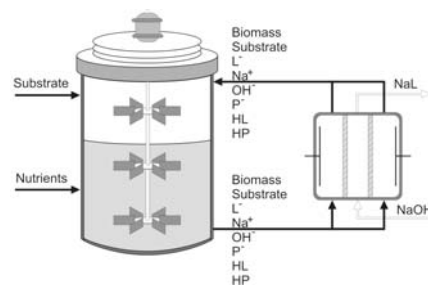

Fig 1. Integrated bioreactor and REED module sketch for intensified lactic acid cultivation

The cleaning effect in the REED module, by alternating the polarity of the potential gradient, has been attributed to an effective destabilization mechanism generated by the hydroxyl flux [5]. The periodic operation of the REED module implies a lost of lactate recovery when the current is reversed. However, the enhanced cleaning enables a higher productivity of the REED module compared to a traditional continuously operating membrane separation process.

The present contribution is part of our efforts on modeling and operational design of a novel process for lactic acid production, where the fermentation and product removal are tightly integrated. To reveal the operation of the integrated system according to different fermentation goals, a systematic procedure for control system design is employed. In this paper, a conceptual pH control structure is proposed based on a systematic but non trivial sensitivity analysis for the periodically operated process.

2. Conceptual control system design methodology

The employed methodology to control structure design is based on ideas for plantwide control design [6]. Existing guidelines for the conceptual control structure design of a goal driven control system needs extension especially when the selection of manipulated variables is not evident for a periodically operated process

2.1. Functionality

The starting point of the control structure design procedure is to define the desired functionality of the process which is the plant operating goal. Interesting industrially relevant primary objective functions are: lactate production as potential feedstock for the production of the biodegradable Poly-Lactic Acid (PLA), probiotic culture production (>10 times higher activity) or extracellular protein production based on genetically modified lactic acid bacteria. To achieve these goals an important subgoal is to achieve pH control in the fermenter. This is selected as a main subgoal and further investigated in this contribution.

2.2. Top- down analysis

The degrees of freedom (DOF) for the integrated system are determined. The DOF which can ensure that the goals and subgoals can be achieved will become actuator variables and define the axes of the operating window for the process. The remaining degrees of freedom are considered as disturbances. In linear systems, the selection of the manipulated variables according to a specific goal can be handled in a relatively simple way based on prior knowledge and guidelines i.e. that large and fast effects on the controlled variables are desired, corresponding to a large steady state gain and relatively small time constants. However, in this integrated periodically operated system it is desirable to define an index that can quantify the mentioned guidelines. As a case study the pH control structure design is chosen.

The selected criterion is the dynamic sensitivity of the controlled variable to changes in the potential manipulated variables. Stationary simulations of the periodically operated system within the operating window are performed and the sensitivity is evaluated. The simulation results additionally provide useful information concerning the type of operating surface for the system and the operational constraints. Simulations have been performed for the fermentation and REED module separately.

2.3. Bottom-up design

Once the conceptual control structure is designed, the controllers can be designed. Usually a multilevel hierarchical structure is employed. It is of particular interest to

Control System Development for Int. Bioreactor and Membrane Separation Process

reveal the potential benefit of multivariable control over a fully decentralized control structure.

2.4. Evaluation of the control structure performance

The implemented control structure is tested through dynamic simulations. Performance indicators provide relevant indicators for the integrated plant monitoring.

3. pH control structure design

3.1. pH model in the REED module

The question to answer at this stage is: what is the most appropriate manipulated variable which can control pH? Previously, a dynamic model was developed to describe simultaneous ion transport across anion exchange membranes in a dialysis cell [4]. Investigations were performed for operation without imposing current density, operation applying an external potential gradient and operation under current reversal conditions [2, 3, 4]. In those contributions, the operating window of the device was explored. Here, we investigate pH changes in the outlet of the REED module as a function of the potential manipulated variables, therefore the pH model is highlighted. The mass balances within the boundary layers and membranes in the REED model can be summarized as follows. A mass balance for component k in phase p is:

$$\frac{\partial C_{k,p}}{\partial t} + \nabla J_{k,p} - R_k = 0 \tag{1}$$

The reaction term (R_k) is used to introduce the acid dissociation into the model. The flux $J_{k,p}$ is estimated using the Nernst-Planck equation for ideal solutions, neglecting convective transport [7]:

$$J_{k,p} = -D_{k,p}\left(\frac{\partial C_{k,p}}{\partial x} + \frac{z_k F C_{k,p}}{RT}\frac{\partial \psi}{\partial x}\right) \tag{2}$$

Where $D_{k,p}$ is the diffusion coefficient, z_k the valence, F is the Faraday number, R is the ideal gas constant, T is temperature and ψ is the electrical potential. The potential gradient can be calculated using the assumption that all current I_d is carried by ions:

$$I_d = \sum_k z_k F J_{k,p} \tag{3}$$

Donnan equilibrium is used to describe the concentration and potential discontinuities at the membrane surface. The bulk channel models are approximated using tanks in series model where in each tank there is mass transport towards the membrane and the dissociation reactions are present. Experimentally, it has been verified that a pH buffer effect is induced by the presence of biomolecules in a fermentation broth. When those components are modeled as highly charged macromolecules, multiple dissociation reactions must be introduced. That increases unnecessarily the complexity of the model. To deal with this situation, the proton acceptor groups in the proteins are considered in terms of equivalents, i.e. the protein concentration is represented as mol of acid equivalents per volume. Therefore, the dissociation of a polyprotic species is simplified to a monoprotic acid reaction [1]. The protein species represents a wide range of components in the fermentation broth from low molecular weight proteins to colloidal material. The system of reactions is given by the following equilibrium expressions:

$$HL + OH^- \underset{k_2}{\overset{k_1}{\rightleftharpoons}} L^- + H_2O$$

$$HP + OH^- \underset{k_4}{\overset{k_3}{\rightleftharpoons}} P^- + H_2O$$
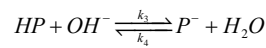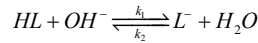
Where the dissociation constants for those reactions (K_d) correspond to the acid dissociation constant (K_a) divided by the ionic product of water (K_w). Using a stoichiometric matrix, the dissociation reaction rates are systematically introduced into the model.

3.2. Degrees of freedom

The potential manipulated variables to control hydroxyl flux towards the REED feed channel are: the feed and dialysate input flow rates (q_{feed} and q_{dia}), the polarity reversal time (t_{rev}), the imposed current density (I_d) and the inlet hydroxyl concentration in the dialysate channel ($C_{OH,dia}^{in}$). A first screening is performed based in the knowledge earned from previous investigations [2, 3, 4]. The first variables discarded are the flow rates, the reason is the lack of information about how the thickness of boundary layers change as a function of the flow conditions. Therefore, the model predictive power during flow rate changes is very limited. The reversal time, or operation time before the polarity of the potential gradient is inversed, is not investigated since it is a design variable that is chosen by trading off lactate recovery and the energy consumption subject to the power source constrains [3]. These choices leave the current density and the inlet base concentration to the dialysate channel as potential manipulated variables.

3.3. pH sensitivity estimation

In order to evaluate the influence of the REED module operation on the pH of the fermenter, a simulation scenario is proposed. Feed channel input concentrations to REED are assumed constant, corresponding to hypothetical fermentation broth at constant pH. Simulations are performed to estimate the pH at the end of the feed channel based on changes of the potential manipulated variables. The dynamic sensitivity of the pH to the potential manipulated variables is approximated using forward finite differences:

$$NS_{I_d} = \frac{\left.|I_d|\right|_{t_o}}{\left.pH\right|_{t_o}} \frac{dpH(t)}{dI_d} \quad \text{where} \quad \frac{dpH(t)}{dI_d} \approx \frac{pH(I_d + \Delta I_d, t) - pH(I_d, t)}{\Delta I_d} \qquad (4)$$

$$NS_{C_b} = \frac{\left.C_{OH,dia}^{in}\right|_{t_o}}{\left.pH\right|_{t_o}} \frac{dpH(t)}{dC_{OH,dia}^{in}} \quad \text{where} \quad \frac{dpH(t)}{dC_{OH,dia}^{in}} \approx \frac{pH(C_{OH,dia}^{in} + \Delta C_{OH,dia}^{in}, t) - pH(C_{OH,dia}^{in}, t)}{\Delta C_{OH,dia}^{in}} \quad (5)$$

Where *NS* is the dimensionless sensitivity. It is selected in order to make a fair comparison between both actuator variables. The scaling values chosen are the initial conditions before the disturbance is applied, i.e. at $t_o = 0$. The simulation procedure to estimate the sensitivity around each operating point involves three steps:

a. The dynamic model is initialized from a known operating point to periodically stationary operation. This point is function of I_d and $C_{OH,dia}^{in}$. The other input variables are fixed. The last simulation point during the period is selected as the initial condition for the following simulations.

b. The model is solved from the new initial conditions and the previous input variables. The simulation results correspond to the stationary operation, and become the reference to estimate the sensitivity. The main variable stored is either $pH(I_d, t)$ or $pH(C_{OH,dia}^{in}, t)$.

c. The model is solved again introducing the disturbance in the potential manipulated variable. The main variable stored is either $pH(I_d + \Delta I_d, t)$ or $pH(C_{OH,dia}^{in} + \Delta C_{OH,dia}^{in}, t)$.

Control System Development for Int. Bioreactor and Membrane Separation Process

4. Results and discussion

4.1. Cultivation broth pH behavior during REED operation

Under constant I_d and $C_{OH,dia}^{in}$ operation, the system is driven by the periodicity and strength of the current. The pH behavior at the outlet of the feed channel during a step change in the operating current of 100 A/m^2 is depicted in fig. 2. The period to period dynamics is visible before a periodic stationary operation is achieved after a few cycles in the open loop simulation. After the polarity is reversed, there is a reduction in the pH due to a temporal flux inversion in the system. Due to the symmetry of the unit, the pH behavior is not a function of the polarity of the electrical field under stationary operation, but of the absolute current density.
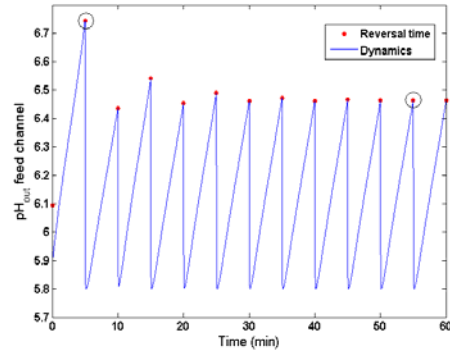

Fig 2. Feed channel pH behavior for t_{rev}=5 *min*, $C_{OH,dia}^{in}$ =50 *mol/m³*. When I_d is changed from *100 A/m²* to *200 A/m²*. The highlighted points indicate the time just before the current is reversed.

4.2. pH Sensitivity

The pH sensitivity is evaluated dynamically, however a comparison is only performed when the derivative has been approximated at the same point in time within subsequent periods. The chosen points are at the time just before the polarity is reversed - shown highlighted in fig. 2. Following the simulation procedure described above, the sensitivity is investigated within a selected operating window of the potential manipulated variables. The sensitivity response is underdamped and overdamped, after introducing disturbances in I_d and $C_{OH,dia}^{in}$ respectively, while the settling time for both responses is comparable. The ratio between the dimensionless sensitivities was investigated at one point during the transient response -the first middle point- and under stationary operation after the transient has settled as shown in fig. 3 and 4.
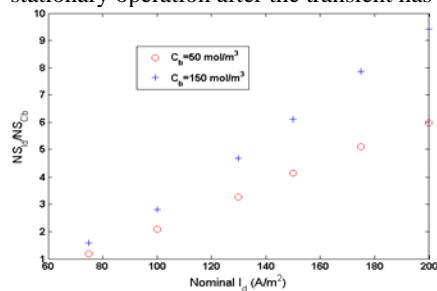
Fig 3. Stationary sensitivity ratio at different operating points
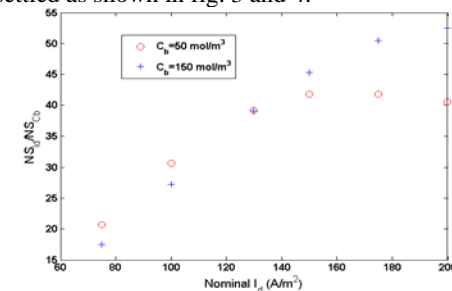
Fig 4. First middle point sensitivity ratio at different operating points

Fig. 3 shows that the stationary normalized sensitivity does not change radically by introducing disturbances either in the current density or the base concentration. At low current densities there are no practical differences. At high current densities, the pH is between 5 and 10 times more sensitive to I_d than to $C_{OH,dia}^{in}$.

On the other hand, larger differences are evident in the first middle point after a step change where normalized pH sensitivity is approx. 20 to 50 times higher for current density than for base concentration changes (fig. 4). These results indicate that I_d is an appropriate manipulated variable to handle the dynamic response; while $C_{OH,dia}^{in}$ can control the stationary behavior to achieve optimal operation. From a practical point of view, both potential manipulated variables have operating constraints. The current density should not be larger than the current saturation value, for this system which is around 280 A/m^2 [2]. The base concentration has a larger operating range, but low concentrations are desired since that prolong the anion exchange membrane life time.

5. Conclusions

This paper illustrates how to perform a quantitative systematic analysis that lead us to select an appropriate manipulated variable to satisfy a specific control goal for a highly non linear and dynamic system, when the selection is neither intuitive nor trivial. As case study, we have selected as goal the pH control at the end of the feed channel of the REED module. A defined operative window was explored using a dynamic model derived from first principles. The model describes the simultaneous transport of multiple ions across anion exchange membranes in a REED cell during lactate recovery from a fermentation broth. The solution of the system of multiregion partial differential equations was approximated numerically [2, 3, 4]. In this contribution, the pH model in the cultivation is highlighted including a pH buffer effect which was been experimentally validated [5]. A dimensionless pH sensitivity, towards I_d and $C_{OH,dia}^{in}$, was used as the quantitative criterion. The sensitivity was evaluated at a specific point within the periodic shifting of the imposed gradient polarity. The results show that pH can be controlled, during stationary operation, by both manipulated variables; especially around low current densities. However, pH has shown a considerably higher sensitivity towards current density in the first middle period point after a step change in the actuator. That behavior leads us to propose that the dynamic pH response can be controlled by manipulating the current density. Additionally, the base concentration in the inlet of the dialysate channel can control the stationary response since the I_d operative window more narrow than for the $C_{OH,dia}^{in}$. Our current investigation is focused on the control implementation.

References

[1] Møllerhøj, M. (2006). Modeling the REED Process. Master's thesis, Technical University of Denmark.

[2] Prado Rubio, O.A.; Jørgensen, S.B. and Jonsson, G. (2009a) Lactic Acid Recovery in Electro-Enhanced Dialysis: Modelling and Validation. *ESCAPE19.*

[3] Prado-Rubio, O.A.; Jørgensen, S.B and Jonsson, G. (2009b). Tool for Optimizing the Design and Operation of Reverse Electro-Enhanced Dialysis of Monoprotic Carboxylic Acids. *PSE09.*

[4] Prado-Rubio, O.A.; Møllerhøj, M.;Jørgensen, S.B and Jonsson, G. (2009c). Modeling Donnan Dialysis Separation for Carboxylic Anion Recovery. *Submitted to CCE Journal.*

[5] Rype, J. (2003). Modelling of Electrically Driven Processes. Ph.D. thesis, Technical University of Denmark.

[6] Skogestad, S. (2002) Plantwide Control: Toward a Systematic Procedure. *ESCAPE12.*

[7] Strathmann, H. (2004). Ion-Exchange Membrane Separation Processes. Membrane Science and Technology Series, 9. Elsevier.

Modeling and optimization of biological sequential batch reactors

Mauren Fuentes, Oscar Iribarren, Miguel Mussati, Nicolás Scenna, Pío Aguirre

INGAR (CONICET-UTN), Avellaneda 3657 (3000) Santa Fe, Argentina
(mfuentes; iribarr; mmussati; nscenna; paguir)@santafe-conicet.gov.ar

Abstract

This work deals with modeling and operation optimization of biological sequential batch reactors (SBR). The SBR is a fill-and-draw biological sludge system for wastewater treatment. In this system, wastewater is added to a single batch reactor, treated to remove undesirable components, and then, discharged. In this paper, a global model of a gas-solid-liquid SBR is presented to investigate and optimize operational strategies. The model can address the differences between aerated and anaerobic systems by assigning adequate parameter values related to the aeration and reaction systems. Fluctuating operation conditions during cycles such as disturbances in the organic loading rate, stirring rate and cycle time, result in strong numerical discontinuities that can be included in the simulation schedules. An existing set of experimental data is used to show a model application based on an anaerobic SBR. A good agreement was obtained between experimental and predicted values. Optimization results are based on minimizing the reaction time/total cycle time ratio subjected to path pH constraints and interior- and end-point constraints related to the pollutant removal efficiency and settling conditions. A decrease of 22% in the total cycle time, i.e. an increase in the organic loading rate from 787 to 985 mg dm^{-3} d^{-1} is reached without modifying the quality of effluent.

Keywords: Dynamic modeling, Optimization, Sequential batch reactors, Wastewater treatment

1. Introduction

The sequencing batch reactor (SBR) is the most promising and viable of the proposed activated sludge modifications for the removal of organic carbon and nutrients [1]. Due to its simplicity and flexibility of operation, it has become increasingly popular for the biological treatment of domestic and industrial wastewater [2].They are uniquely suited for wastewater treatment applications characterized by low and/or intermittent flow conditions.

The most common (aerated) SBR is a fill-and-draw activated sludge system for wastewater treatment. Equalization, aeration, and clarification can all be performed in a single batch reactor. This technology has also been used for anaerobic (non-aerated) digestion of industrial wastewaters [3]. In general, SBR systems have a relatively small footprint; they are useful for areas where the available land is limited. In addition, system cycles can be easily modified, making SBRs extremely flexible to adapt to more restrictive effluent quality standards by public authorities.

The determination of the influent characteristics and effluent requirements, site specific parameters such as temperature, and key design parameters such as nutrient-to-biomass ratio, treatment cycle duration, suspended solids and hydraulic retention time is imperative to establish the operation sequence of the SBR. It allows calculating the

number of cycles per day, number of basins (batches), decanting volume, reactor size, and detention times.

More than one operation strategy is possible for the complete process. For most municipal wastewater applications, treatability studies have not been required to determine the operating sequence because the municipal wastewater flow rate and composition fluctuations are usually predictable, and most designers follow conservative design approaches. However, for industrial wastewater applications, treatability studies have typically been required to determine the optimum operating sequence. Therefore, the aim of the paper is to present a global model of a SBR system, and discuss some aspects on process optimization based on model solutions. The mathematical model has been derived to be used in aerated and anaerobic applications. Due to space restrictions, only one example is presented; an anaerobic application is selected. It is based on the experimental results obtained by Rodrigues et al. [4] operating a bioreactor (New Brunswick Scientific Co. model BIOFLO III) with a working capacity of 5 L. A complete description of the model and other application examples could be presented in an extended version of the paper. Following, a brief explanation on the main SBR model equations and hypotheses is presented.

2. Mathematical model

The overall operation of a SBR is based on five steps: (static, mixed and aerated) fill, react, settle, decant and idle (including wasting). The aerated procedure is simplified in anaerobic reactors. The static or mixed fill depend on kinetics and economy of processes. The wasting consists in pumping out a small amount of sludge at the bottom of the SBR basin, and is more frequent in aerobic applications.

A heterogeneous dynamic model is proposed to describe the three-phase gas-solid-liquid system present in a SBR. The solid phase consists of bioparticles (granules) composed by active and non-active biomass. The liquid phase is composed by the chemical species in solution (substrates, products, enzymes, ions, and water) and (active and non-active) single suspended cells, which are assumed to behave as solutes. The gas phase is formed by the oxygen stream (in aerated systems) and the gaseous products from degradation stages. Eq. (1) represents the mass balance equation for components ϕ_{ik} in the k-phase (liquid, solid, gas).

$$A_c d \frac{\varepsilon_k \phi_{iL} H}{dt} = Q_{k_{in}} \phi_{ik_{in}} - Q_{k_{out}} \phi_{ik} + \lambda V \left(\sum_j R_{ik}^j + \sum_j T_{ik}^j \right), \; k{=}L, S, G \qquad \text{(Eqn. 1)}$$

In this equation the input (fill) and output (discharge) flow rates for liquid and solid phases are calculated as $Q_{k_{in}} = V_{k_f} / t_{k_f}$ and $Q_{k_{out}} = V_{k_d} / t_{k_d}$. As initial condition, the SBR is assumed to be an inoculated system, i.e. $Q_{S_{in}} = 0$ afterwards an eventual sludge load occurs. Variations of volume fractions (phase holdups, ε_k) and reactor height H are assumed to occur only during the fill and decant steps. When the agitator driver turns off during the settling and withdrawal steps, parameter λ takes the value zero. Although, when the driver turns on, the phases are mixed, reaction and mass transfer processes ($\sum_j R_{ik}^j + \sum_j T_{ik}^j$) occur, and parameter λ takes the value one.

The kinetics model involves biochemical (growth-uptake, death, hydrolysis, disaggregating) and physico-chemical (system charge balance for calculating pH, gas-liquid mass transfer) processes which take place in the bioreactor. The model can address the differences between aerated and anaerobic systems by assigning adequate

parameter values related to the aeration process, and selecting the kinetics model to represent the digestion stage [5-7]. Kinetic model parameters and constants are described in the original sources.

2.1. Settling time

As a first modeling approach, a simple relationship is used to analyze granular sludge settling characteristics. It involves the concept of mean settling time and is based on measuring the time required for settling out half of the sludge poured into a graduated cylinder, after homogenized and left to stand [8]. Assuming z^* as the dimensionless vessel height, the solid-liquid interface follows the $1 > z^* > 0.5$ trajectory. During this stage, the sedimentation process mainly depends on granule (density, diameter) and liquid bulk (density, viscosity) characteristics.

The settling time from $z^* = 0.5$ to the static sludge level, i.e. when the final sludge height is reached, involves several factors and it is difficult to obtain a phenomenological expression. An exponential expression ($z^* = \alpha e^{-\beta t}$) is proposed to represent this stage. Finally, the two components of settling time are:

$$t_s = t_{(1 > z^* > 0.5)} + t_{(0.5 > z^* > z^*_{final})} \qquad \text{(Eqn. 2)}$$

Since the global model allows calculating time variation of granule diameter, more sophisticated settling models can be investigated in future works [9]. Homogeneous biomass distribution, constant wet biomass density and spherical geometry are assumed to model the bioparticle.

2.2. Discharge and fill (pumping) times

Eq. (3) can be used for calculating the pumping time, after the batch unit cycle time (T) is estimated [10]. For small scale equipment, parameter γ adopts smaller values than the published for industrial applications. Values of γ have implicit the effect of some factors such as the vessel and pump type, the pressure drop and prevention of adverse hydrodynamic events on the sludge. Therefore, different values may be assumed to calculate either discharge and feed times.

$$t_{pump} \approx \gamma T^{0.625} \begin{cases} t_f = \gamma_f T^{0.625} \\ t_d = \gamma_d T^{0.625} \end{cases} \qquad \text{(Eqn. 3)}$$

The mathematical model resulted in a differential and algebraic equations (DAE) system, and was implemented and solved using the process modeling software tool gPROMS (Process Systems Enterprise Ltd).

3. Example

An anaerobic SBR example ($Q_{g_{in}} = 0$) is selected to illustrate the model application. Kinetics model hypotheses and all terms ($\sum_j R_{ik}^j + \sum_j T_{ik}^j$) related with mass transfer processes, model parameters and constants, are described in Fuentes et al. [7]. The digestion model involves eight groups of microorganisms for consuming pollutants measured as chemical oxygen demand (COD) concentration. A novel aspect in the SBR model is related to modeling the disaggregation of single suspended cells from biomass granules. Hydrodynamics affects biomass processes inside a bioreactor attending to the operational fluid velocity ranges [11]. A parameter (specific rate of granule rupture) has been defined to explain how the process of granulation is affected by the environmental and operational conditions. The rate of granule rupture is modeled as a first order function of the reactor stirring rate.

In the case study described in the following section, the wasting step was not necessary ($Q_{s_{out}} = 0$) and the idle time is zero. Since $Q_{s_{in}} = Q_{s_{out}} = 0$, time variations of the reactor height mainly depends on changes in the liquid volume due to the wastewater feeding and discharge of the treated effluent ($A_c d \dfrac{H}{dt} = Q_{l_{in}} - Q_{l_{out}}$).

3.1. Case study

The experimental results obtained by Rodrigues et al. [4] are here used to calculate the application example. Details related to the anaerobic SBR design, sludge and wastewater characteristics and operational conditions are summarized in Table 1.

Table 1. Details of case study.

Parameter (dimension)	Value
Reactor design	
Working reactor volume (dm^3)	5
Inner diameter (dm)	1.8
Sludge characteristics	
Initial load (dm^{-3})	2
Density (g dm^{-3})	1026
Biomass concentration (gTVS dm^{-3})	22.1
Granule mean diameter (10^2 dm)	2.3
Wastewater characteristics	
Influent COD (mg dm^{-3})	500
Substrate composition (%COD)	Carbohydrate 41%, protein 47%, lipid 12%
Operational conditions	
Temperature (°C)	30±2
Stirring rate (rpm)	50
Volume of treated effluent discharged (dm^3)	2
Reaction time (min)	310
Settling time (min)	30
Discharge time (min)	13
Feed time (min)	7
Total time per cycle (min)	360
Time horizon (days)	17 (68 cycles)

3.2. Simulation results

As observed in Table 2 and Fig. 1, a good agreement between experimental and predicted values of total and soluble COD, volatile suspended solids (VSS) and volatile fatty acids (VFA), as acetic acid (HAc) concentration, is obtained. An experimental value of 6.9 was reported for pH of treated effluent (see Fig. 1b). For a clearer presentation, only one operation day (four cycles) has been represented.

In Fig. 1(a), main hydrodynamic disturbances are depicted. The total volume of medium in the batch is 5L, i.e., the volume of wastewater to be treated and biomass. A volume of 2 L of treated medium is discharged during 13 min at the end of the cycle. The same volume of fresh wastewater is fed into the reactor during 7 min at the start-up of the next cycle. Parameter λ has been added to show the agitation and reaction periods. A mixing police is used during the fill step. An increase in the total COD is predicted for values of the specific rate of granule rupture higher than 1×10^{-4} rpm^{-1} d^{-1}.

Table 2. Experimental, modeled and predicted values of influent and effluent streams.

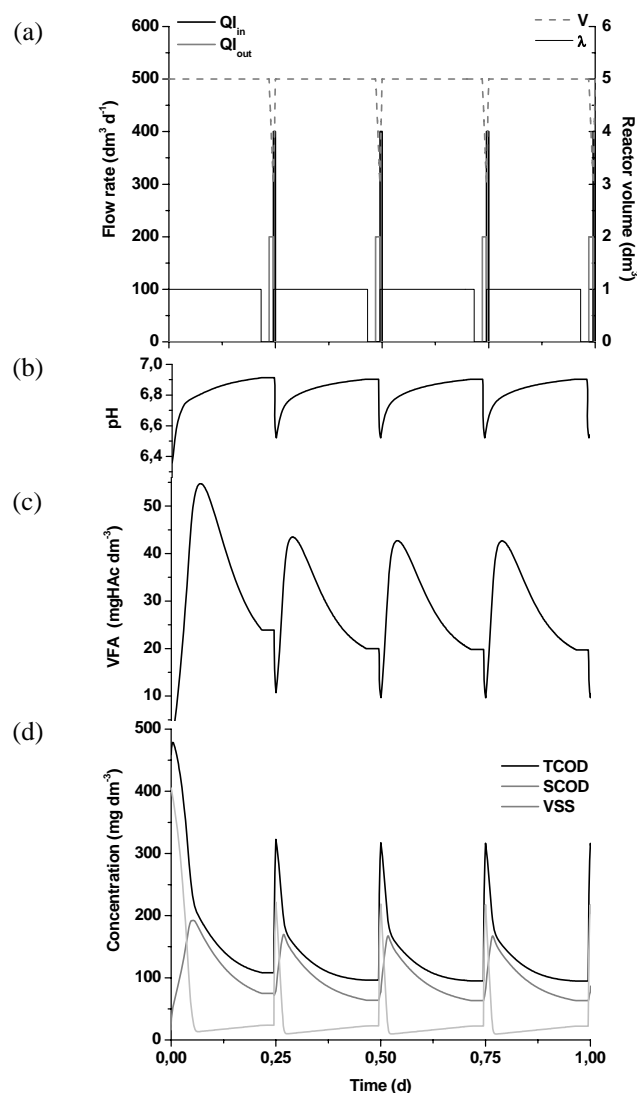| Stream	TCOD (mg dm^{-3})		SCOD (mg dm^{-3})		VFA (mgHAc dm^{-3})		VSS (mg dm^{-3})	
	Exp.	Mod.	Exp.	Mod.	Exp.	Mod.	Exp.	Mod.
Influent	508±40	500	-	50	36±3	36	19±7	20
Effluent	97±11	95	66±3	63	21±1	21	23±18	22

Fig. 1. Simulation results: (a) feed and discharge flow rates, reactor volume, parameter λ (agitation and reaction), (b) pH, (c) VFA, (d) TCOD, SCOD and VSS concentration.

3.3. Process optimization

When dealing with multispecies biofilm models, path constraints need to be imposed to pH to obtain optimal biomass (pollutant removal efficiency) and methane profiles in a single unit [12]. As observed, the model is able to manage strong numerical disturbances to represent stepped cycle strategies. Here, the objective function experienced is minimizing the reaction time/total cycle time (t_r/T) ratio subject to pH path constraints, and interior- and end-point constraints related to COD removal efficiency and settling conditions. A pH variation between 6.2 and 7.2 was allowed. The horizon (cycle) time is unknown. Times t_s, t_f and t_d are calculated by Eq. (2) and Eq. (3), respectively. The reaction time is obtained from the optimum solution. To determine and fix the t_r value, a conditional structure is programmed (if $\lambda=1$, $dt_r/dt = 1$; else for

$\lambda=0$, $dt_r/dt = 0$, $t_r=t_{r\_opt}$). Values of parameter λ, and fill and discharge volume of medium, are the (piecewise constant) control variables. The gOPT tool of gPROMS was used to perform the dynamic optimization. This approach turned robust, i.e. converged and required a low computational time (Total CPU time 13.5 s, 800MHz Pentium IV PC).
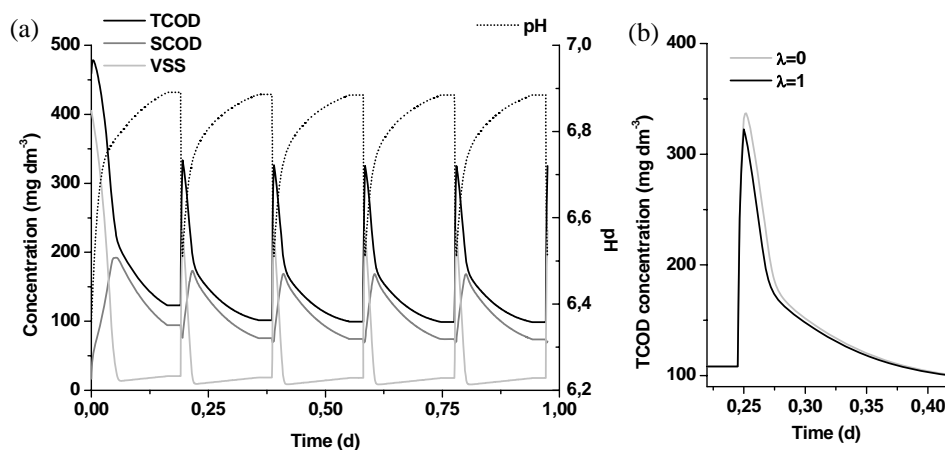

Fig. 2. (a) pH, TCOD, SCOD and VSS optimal profiles; (b) influence of λ in the fill step.

The optimal profiles for total and soluble COD, VSS and pH are depicted in Fig. 2(a). This strategy implies an increase in the number of cycles per day (T_{opt}=280, t_r=233, t_s=30, t_d=11, t_f=6 min), i.e. a decrease of 22% in the total cycle time and an increase in the organic load from 787 to 985 mg dm^{-3} d^{-1} are reached without modifying the quality of effluent. On the other hand, Fig. 2(b) shows the influence of the agitation (and reaction, λ=1) during the fill step. For both conditions, the output values of total COD are the same. Static fill can be used during low-flow periods to save power. In this case, because the mixers remain off, this scenario has an energy-savings component.

As concluding remarks, the model here proposed provides a useful tool for optimization of operational strategies and design of SBR systems. Main process variables such as the number of cycles per day, decanting volume, reactor size, and reaction and detention times can be optimized.

References

[1] N. Artan and D. Orhon, Mechanism and Design of Sequencing Batch Reactors for Nutrient Removal, IWA Publishing, Hove, UK (2005).
[2] S. Mace and J. Mata-Alvarez, Ind. Eng. Chem. Res. 41 (2002) 5539-5553.
[3] N. Bernet et al, Water Res. 34 (1999) 611-619.
[4] J.A.D. Rodrigues et al., Braz. J. Chem. Eng., 21 (2004) 423-434.
[5] W. Gujer et al., Activated Sludge Model N° 3. Water Sci. Technol., 39 (1999) 183-193.
[6] D.J. Batstone et al., Anaerobic Digestion Model No. 1, IWA Publishing, London, UK (2002).
[7] M. Fuentes et al., Lat. Am. App. Res., 37 (2007) 235-242.
[8] L.W. Hulshoff-Pol et al., 3rd Int. Symp. on Microb. Ecology Proceed., Washington (1984) 646-632.
[9] Y. Liu et al., Biotechnol. Adv., 23 (2006) 335-344.
[10] O.A. Iribarren, Batch Chem. Processes Design, Ph. D. Thesis, Univ. of Massachusetts, Amherst (1985).
[11] C. Nicolella et al., J. Biotechnol., 80 (2000) 1-33.
[12] M. Fuentes, Modeling and Dynamic Simulation of Anaerobic Biofilm Fluidized Bed Reactors. Application to the Complex Effluent Treatment, Ph. D. Thesis, Univ. Nat. of Litoral, Argentina (2006).

Combining Genetic Circuit and Microbial Growth Kinetic Models: A Challenge for Biological Modelling

Michalis Koutinas,[a] Alexandros Kiparissides,[a] Ming-Chi Lam,[a,b] Rafael Silva-Rocha,[c] Victor de Lorenzo,[c] Vitor A.P. Martins dos Santos,[b] Efstratios N. Pistikopoulos,[a] Athanasios Mantalaris[a,*]

[a] *Department of Chemical Engineering, Imperial College, London SW7 2AZ, UK*
[b] *Systems and Synthetic Biology Group, Helmholtz Center for Infection Research, Inhoffenstrasse 7, D-38124, Braunschweig, Germany*
[c] *Centro Nacional de Biotecnologia, Consejo Superior de Investigaciones Cientificas, Darwin 3, Cantoblanco, 28049 Madrid, Spain*
[*] *Author to whom correspondence should be addressed: a.mantalaris@imperial.ac.uk*

Abstract

A modelling framework that consists of model building, validation and analysis, leading to model-based design of experiments and to the application of optimisation-based model-predictive control strategies for the development of optimised bioprocesses is presented. An example of this framework is given with the construction and experimental validation of a dynamic mathematical model of the *Ps/Pr* promoters system of the TOL plasmid, which is used for the metabolism of *m*-xylene by *Pseudomonas putida* mt-2. Furthermore, the genetic circuit model is combined with the growth kinetics of the strain in batch cultures, demonstrating how the description of key genetic circuits can facilitate the improvement of existing growth kinetic models that fail to predict unusual growth patterns. Consequently, the dynamic model is combined with global sensitivity analysis, which is used to identify the presence of significant model parameters, constituting a model-based methodology for the formulation of genetic circuit optimization methods.

Keywords: Dynamic modelling, Sensitivity analysis, Genetic circuit, pWW0 (TOL) plasmid, *Pseudomonas putida*.

1. Introduction

Genetic circuits are groups of elements that interact producing certain behaviour [1]. Based on our capability to engineer genetic circuits, fundamental biological processes can be studied systematically and targets can be identified for genetic modification, producing the desired properties. However, the extensive experimentation required to understand the function of genetic circuits, is often limited by the time and cost required. Although the experimental techniques required for the study of genetic circuits are very sophisticated, reliable mathematical models are equally important in reducing substantially the trial-and-error experimentation. In line with this, dynamic modelling can be used for characterisation of the cellular function integrating biological information into predictive models [2]. Furthermore, the molecular and genetic events responsible for the growth kinetics of a microorganism can be extensively influenced by the presence of mixtures of substrates leading to unusual growth patterns, which cannot be accurately predicted from existing models [3]. Thus, a novel approach combining

genetic circuit and growth kinetic models constitutes an improved version of the currently used models for the prediction of microbial growth kinetics.

Complex biological models may include a large number of parameters, which can be difficult to estimate and may incorporate expensive and time consuming experiments. Global sensitivity analysis (GSA) is a tool used to quantify the importance of model parameters and their interactions with regards to the model output [4]. Analyzing the properties of parameters included in genetic circuit models provides the identification of the most significant ones with respect to the output of interest. Thus, the experiments required for circuit optimisation can be aimed at genetic modifications altering these parameters alone, consequently reducing the cost and the number of experiments required. Application of GSA methods to biological systems has been limited to a few examples [5]. In this work, we present a modeling framework that consists of model building, validation and analysis providing a solid basis for genetic circuit optimization. A combined model has been constructed describing the function of the *Ps/Pr* promoters system and the growth kinetics of *P. putida* mt-2. The model's prediction has been compared to that of models accounting for enzymatic interactions. Finally, preliminary model analysis has been performed with the application of Sobol' GSA method [6].

2. Results and discussion

2.1. Genetic circuit model

P. putida mt-2 is equipped with the TOL plasmid (pWW0), specifying a pathway for the catabolism of *m*-xylene. The enzymes required for these reactions are produced by the two gene operons of TOL (upper- and *meta*- operon), while two genes (*xylS* and *xylR*) control the regulation of transcription of the operons. These four transcriptional units are driven by four different promoters (*Pu*, *Pm*, *Ps* and *Pr*). The *Ps/Pr* system has been reconstructed into its various interacting molecular components and has been described as a combination of logic gates (Fig. 1) producing a representation in an analogy to electronic circuitry. Based on the logic model of the *Ps/Pr* system, the Hill functions were used as input functions to the genes and a dynamic mathematical model of the system was generated, as described below.
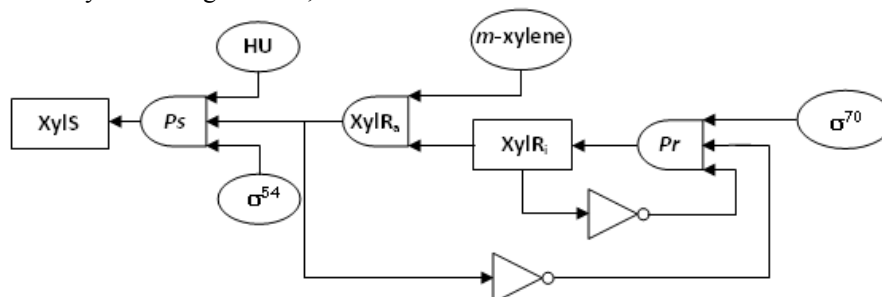

Figure 1. Logic diagram of the *Ps/Pr* system of TOL plasmid pWW0. ⬭ : input; ▭ : output; ◖ : AND; ▷∘ : NOT.

P. putida mt-2 degrades aromatic substrates through a series of events leading to coordinated expression from the upper- and *meta*-cleavage pathways of TOL. The master regulator of the two pathways, the XylR protein, is transcribed by the *xylR* gene from two σ^{70} tandem promoters (*Pr1* and *Pr2*). After binding with *m*-xylene, the inactive dimer form of the XylR protein (XylR$_i$) binds ATP and oligomerizes to form a hexamer. This leads to the formation of the active form of XylR (XylR$_a$), which induces transcription of the *Pu* promoter synthesizing the upper-pathway enzymes. The

Combining Genetic Circuit and Microbial Growth Kinetic Models: A Challenge for Biological Modelling

synthesis of $XylR_i$, as well as the forward and reverse reactions for XylR activation/deactivation are expressed by Eqs. (1-2) (all symbols are defined in Table 1):

$$\frac{dXylR_i}{dt} = \frac{\beta_{XylR_i} \mathrm{Pr}_{TC}}{K_{\mathrm{Pr},XylR_i} + \mathrm{Pr}_{TC}} - r_{XylR} XylR_i Xyl + 3 r_{R,XylR} XylR_a (Xyl_{INI} - Xyl) - \alpha_{XylR_i} XylR_i \tag{1}$$

$$\frac{dXylR_a}{dt} = \frac{1}{3} r_{XylR} XylR_i Xyl - r_{R,XylR} XylR_a (Xyl_{INI} - Xyl) - \alpha_{XylR_a} XylR_a \tag{2}$$

Table 1. List of symbols.

Symbols	Definition
$XylR_i$ / $XylR_a$	concentrations of the inactive and active forms of XylR protein
r_{XylR} / $r_{R,XylR}$	$XylR_i$ oligomerization and $XylR_a$ dissociation constants
Xyl / Xyl_{INI} / Suc	total m-xylene, total m-xylene initial and total succinate concentrations
Pr_{TC} / Ps_{TC}	relative activities of Pr and Ps
t	time
$K_{Pr,XylRi}$ / β_{XylRi}	$XylR_i$ and maximal $XylR_i$ translation rates
α_{XylRi} / α_{XylRa}	degradation/dilution rates due to cellular volume increase for $XylR_i$ and $XylR_a$
β_0 / β_{Ps} / β_{Pr}	basal and maximal expression levels of Ps and Pr
$K_{XylRa,Ps}$ / K_{XylRi} / K_{XylRa}	activation and repression coefficients of Ps and Pr due to $XylR_i$ and/or $XylR_a$
$n_{Pr,i}$ / $n_{Pr,a}$ / $n_{Ps,a}$	Hill coefficients of Pr and Ps due to $XylR_i$ and/or $XylR_a$ binding
$K_{SUC,Pr}$ / $K_{SUC,Ps}$	inhibition constant of succinate on Pr and Ps
α_{Pr} / α_{Ps}	deactivation rates of Pr and Ps
μ_1 / μ_2 / $\mu_{max,1}$ / $\mu_{max,2}$	specific and maximum specific growth rates of biomass on m-xylene and succinate
S_1 / S_2	m-xylene and succinate concentrations
$K_{S,1}$ / $K_{I,1}$ / $K_{S,2}$	m-xylene and succinate saturation and/or inhibition constants
X	biomass concentration
MWt_1 / MWt_2	m-xylene and succinate molecular weight
Y_1 / Y_2	yield coefficient for biomass on m-xylene and succinate
$K_{I,1,2}$ / $K_{I,1-P,2}$	m-xylene inhibition and by-product inhibition on succinate constant

For simplification of the model developed we express both *xylR* promoters as a single *Pr* promoter. The XylR protein activates *Ps* and represses its own synthesis. During the experiments, the *Pr* promoter was slightly repressed in the presence of both substrates as compared to the case when only *m*-xylene was present. Therefore, we presumed that succinate is repressive for *Pr* in the presence of *m*-xylene and that the concentration of σ^{70} is at a constant level. The function of *Pr* promoter activity is expressed by Eq. (3).

$$\frac{d\,\mathrm{Pr}_{TC}}{dt} = \frac{\beta_{\mathrm{Pr}}}{1 + \left(\dfrac{XylR_i}{K_{XylR_i}}\right)^{n_{Pr,i}} + \left(\dfrac{XylR_a}{K_{XylR_a}}\right)^{n_{Pr,a}}} \frac{1}{1 + K_{SUC,\mathrm{Pr}} Suc^2} - \alpha_{\mathrm{Pr}}\,\mathrm{Pr}_{TC} \tag{3}$$

The *xylS* gene is expressed constitutively at a basal expression level β_0 (Eq. 4) but boosted in the presence of *m*-xylene synthesizing the XylS protein and stimulating the induction of the *meta*-pathway. Activation of *Ps* is assisted by the HU protein, which stabilizes the correct architecture of the promoter. The experiments of this study confirmed in that *Ps* is negatively affected in the presence of succinate. Thus, we consider that succinate is repressive for *Ps* promoter and we assume that the concentration of HU and σ^{54} is constant at housekeeping level.

$$\frac{dPs_{TC}}{dt} = \beta_0 + \beta_{Ps} \frac{XylR_a^{\,n_{Ps,a}}}{K_{XylR_a,Ps}^{\,n_{Ps,a}} + XylR_a^{\,n_{Ps,a}}} \frac{1}{1 + \left(\dfrac{Suc}{K_{SUC,Ps}}\right)^2} - \alpha_{Ps} Ps_{TC} \tag{4}$$

2.2. Growth kinetic model

P. putida mt-2 was first cultivated in the presence of succinate and *m*-xylene as single substrates. The biodegradation of 0.9 mM *m*-xylene fed in a batch experiment was modelled assuming substrate inhibition [7] (Eqs. 5-6).

$$\mu_1 = \frac{\mu_{\max,1} S_1}{K_{S,1} + S_1 + \frac{S_1^3}{K_{I,1}^2}} \quad , \quad (5); \quad \frac{dS_1}{dt} = -\frac{1}{MWt_1} \frac{\mu_1}{Y_1} X \, , \quad (6)$$

A 1 h lag-phase occurred following the introduction of *m*-xylene (data not shown). Since the culture was pre-grown in succinate, the lag might be due to the change in substrate requiring the induction of new enzymes for *m*-xylene biodegradation. Thus, we assume that the transition from the lag to the accelerating phase takes place when the activity of *Ps* increased from its basal level by 65-fold, an amount which corresponds to the activation of the TOL pathway and to the induction of its enzymes. The genetic circuit model was used to calculate *Ps* promoter's activity over time, estimating the duration of the lag-phase, and the growth kinetic model was used after the lag-phase.

The growth kinetics of mt-2 was studied in the presence of succinate. The consumption of 13.6 mM succinate fed in a batch experiment was modelled using Eqs. (7-8).

$$\frac{dS_2}{dt} = -\frac{1}{MWt_2} \frac{\mu_2}{Y_2} X \, , \quad (7); \quad \mu_2 = \frac{\mu_{\max,2} S_2}{K_{S,2} + S_2} \, , \quad (8)$$

The strain was cultured in a batch experiment in the presence of 14 mM succinate and 1.04 mM *m*-xylene. Unlike the cases of simultaneous or diauxic growth often observed when a mixture of substrates is available, mt-2 displayed a different growth pattern. Following the initial lag-phase, *m*-xylene degradation started first followed by succinate degradation resulting in two phases were both substrates were utilised individually and one phase were both substrates were utilised simultaneously. *m*-xylene is sensed by *P. putida* mainly as a stressor to be extruded rather than as a nutrient to be metabolised. Consequently, the lag-phase in succinate degradation can be attributed to the presence of the stressor and the duration of the lag on succinate might depend on the time required to inactivate *m*-xylene. In order to consider the inhibitory effects of *m*-xylene and its intermediates on succinate degradation, a new succinate degradation model is suggested (Eq. 9). We assume that a major intermediate in *m*-xylene degradation accumulates over time proportionally to the removal of *m*-xylene. Furthermore, inhibition of growth on succinate due to the presence of *m*-xylene is also considered.

$$\mu_2 = \frac{\mu_{\max,2} S_2}{K_{S,2} + S_2} \frac{K_{I,1,2}}{K_{I,1,2} + S_1} \frac{K_{I,1-P,2}}{K_{I,1-P,2} + (1 - S_1)} \quad (9)$$

Due to the overall repression of *Ps*, we assumed that in the presence of both substrates, *m*-xylene degradation started when the activity of *Ps* increased from its basal level by 14-fold. Also, as mentioned above, the lag-phase in succinate degradation is attributed to the stress effect caused by *m*-xylene. Thus, we assume that growth on succinate starts when the cellular metabolic resources are redistributed towards succinate assimilation indicating the onset of the TOL pathway deinduction, which is expressed by the time point where *Ps* activity starts decreasing from its maximum value. The model parameter values were obtained from the experiments presented above.

The mixed-substrate experiment was also modelled with the SKIP model, which is used when the type of substrate interactions cannot be directly specified, and with cell growth

Combining Genetic Circuit and Microbial Growth Kinetic Models: A Challenge for Biological Modelling

models accounting for competitive, noncompetitive and uncompetitive inhibition. The uncompetitive inhibition and SKIP models satisfactorily described the experimental data, while the competitive and non-competitive inhibition models failed to follow the experimental results (data not shown).

The predictive capability of the model was tested with an independent experiment. The initial succinate concentration was maintained at 14.1 mM, while *m*-xylene concentration was reduced to 0.8 mM. The duration of the lag-phase for each substrate was calculated from the genetic circuit model (Fig. 2) as described above. The combined mathematical model underpredicted the biomass concentration (Fig. 3) and overpredicted to minor extent the *m*-xylene concentration over time (Fig. 4). However, the model closely tracked succinate concentration (Fig. 5) and overall produced a satisfactory description of the experimental data. In contrast, the competitive inhibition and the SKIP model failed to describe the experimental results confirming that only the combined model can be predictive under different experimental conditions.
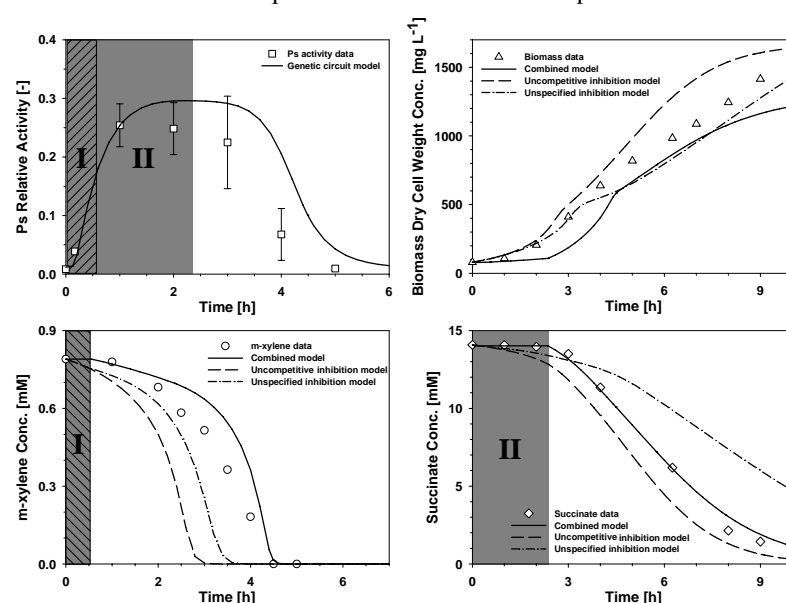
Figures 2-5. Comparison of the combined model and the substrate interaction models prediction. I) lag-phase on *m*-xylene; II) lag-phase on succinate.

2.3. Model analysis

The ability of the Sobol' method to distinguish between individual and total sensitivity index (SI) enables us to identify interacting factors within the system gaining valuable insight into its dynamics. The dimensionality of the sensitivity analysis problem is defined by the number of model parameters; therefore a feasibility constraint regarding the maximum possible number of individually scanned parameters is imposed implicitly in terms of computational time. This constraint is unavoidable due to the - increasing with dimension - number of model evaluations required for the Monte Carlo integrals to converge. Researchers in the field of GSA often resolve to parameter grouping in order to reduce the dimensionality of the problem, thus solving a more tractable version of the original problem. Therefore the parameters of the model have been divided in 16 groups according to their biological function, to make the computation of GSA feasible.

The sensitivities of the parameter groups have been calculated at different time points for $XylR_a$ concentration as the output variable and the results are shown in Fig. 6. The

most significant parameter groups are: i) Gp1 ($\mu_{max,1}$), ii) Gp10 (r_{XylR}, $r_{R,XylR}$), iii) Gp11 (β_{XylRi}, α_{XylRi}, $K_{Pr,XylRi}$) and Gp12 (α_{XylRa}). Therefore, parameters related to XylR$_i$ and XylR$_a$ oligomerization and dissociation respectively, XylR$_i$ translation and degradation, XylR$_a$ degradation and the maximum specific growth rate on *m*-xylene are the most significant. The identification of significant parameters not only enriches our knowledge on the intricate mechanisms that govern cellular behaviour, but may be an indication of which are the most significant parameters to genetically modify towards the production of improved cellular behaviour.
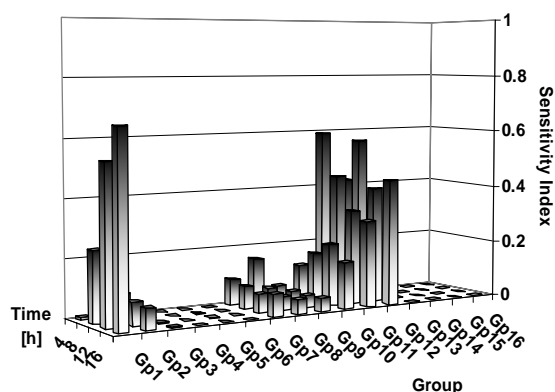

Figure 6. Preliminary calculation of SI for various parameter groups. Indexes were calculated at different time points for XylR$_a$ concentration as the output variable.

3. Conclusions

The mathematical model successfully combines the prediction of a key genetic circuit to the growth kinetics of the microorganism, producing a reliable description of the system. The combination of the constructed model with GSA constitutes a model-based methodology identifying the driving mechanisms of the system, which can be used for hypothesis testing and network optimisation. Thus, the modeling framework presented in this study enables the formulation of genetic circuit optimisation methods, opening a window into the direct re-programming of cellular behaviour and, subsequently, the development of optimised and novel, high-added value biocatalysts.

4. Acknowledgements

References

[1] R. Weiss, S. Basu, S. Hooshangi, A. Kalmbach, D. Karig, R. Mehreja and I. Netravali, Nat. Comput., 2 (2003) 47
[2] F.R. Sidoli, A. Mantalaris and S.P. Asprey, Cytotechnology, 44 (2004) 27
[3] K.F. Reardon, D.C. Mosteller and J.D.B. Rogers, Biotechnol. Bioeng., 69 (2000) 385
[4] A. Kiparissides, S.S. Kucherenko, A. Mantalaris and E.N. Pistikopoulos, Ind. Eng. Chem. Res., 48 (2009) 7168
[5] C. Kontoravdi, S.P. Asprey, E.N. Pistikopoulos and A. Mantalaris, Biotechnol. Prog., 21 (2005) 1128
[6] I.M. Sobol', Math. Comput. Simul., 55 (2001) 271
[7] T. Yano and S. Koga, Biotechnol. Bioeng., 11 (1969) 139

Modeling Heterogeneous Virus Infectiosity for Cells in Bioreactors for Vaccine Production

Juan G. Diaz Ochoa[a], Stefanie Kalweit[a], Andreas Voigt[a], Kai Sundmacher[b]

a Otto von Guericke Universität, Magdeburg 39106, Germany, diaz@mpi-magdeburg.mpg.
b Max Plank Institute for Technical Complex Systems, Magdeburg 39106, Germany.

Abstract

In bioprocesses it is relevant to consider that viruses are inhomogeneous particles infecting the cell population. For instance, the viruses used for the inoculation of the cell population in a bioreactor can have different strains, which can influence the infection spread as well as the release of viral particles. The dependence of the virus strain in the infection process is a known effect; however, little is known about its effect on apoptosis in the cell population, in particular considering virus heterogeneity related to different grades of infectivity in the produced viral particles. We have developed a model, using a kinetic Monte Carlo approach, where viral characteristics are explicitly considered, an aspect that helps to analyze the effect of the heterogeneity of initial and released viral strains on the virus production. With this model we describe the release of influenza A viruses in a cell population in a bioreactor. Using our simulation data we are able to predict certain conditions under which this bioprocess produces a maximum of virus particles with a preselected property.

Keywords: Virus Strains, Monte Carlo Method, Cell Population, Grade of Infectivity

1. Introduction

The modeling of a viral infection process in cell populations *in-vitro* and in bioreactors is a challenging problem from an academic point of view because it helps to understand the dynamics in the spreading of the infection and release of new viral particles by infected cells. Using this information it is possible to understand the interplay of different mechanisms, from the individual cell level (for example individual infection of uninfected cells), to the whole population (accumulation of viral particles with different grade of infectivity) which take place in the infection spreading process. But this problem is also relevant from a technical and biomedical point of view, because by using this information it is possible to optimize the production of viruses used for vaccines. Although theoretical aspects about virus spread are well known (Sidorenko, 2008; Beauchemin, 2005), there is comparatively little knowledge about the effect that different virus strains and its heterogeneous grade of infectivity may have on cell cultures.

One relevant aspect is the consideration of the "quality" of viral particles. In several mathematical models the viral particles are considered as a homogeneous population with a given grade of infectivity. However, experiments on microcarriers and bio-reactors show that there are more than a single class of viral particles, in particular that the viral activity depends not only on physicochemical characteristics of the cultivation media, but also on the individual strain of the viruses (Schulze-Horsel, 2009).

The first intention of this model is to keep the representation of the system as simple as possible. From different possible activity grades for the viral particles we restrict the viral characteristics to two grades or states: *active* or *inactive.* Both states are updated either when the cell releases new viral particles of different classes or when there is a non-reversible transition from the active to the inactive state of the virus. Only active viral particles are able to infect uninfected cells and change their status accordingly.

Due to the higher dimensional population of cells and the transitions based on probabilities we use a stochastic Monte Carlo approach to model and simulate the virus replication process. Our model is based on kinetic Monte Carlo methods (developed by Sidorenko, 2007). In the second part we resume our model and the main steps of our simulations. In the third part we present a compendium of some relevant results obtained with our simulations. In the last part we summarize our results and give a discussion about the virtues and perspectives of the present model.

2. Model

In this model we assume that the total number of cells *C* is constant in time and that the spatial distribution of the cell population is not relevant for the spread of the infection. Uninfected cells degrade with a probability $P_{un\text{-}deg}$ and infected cells undergo apoptosis with a probability P_{apop} (in this case lysis). Additionally, we assume the presence of active and inactive viral particles that are respectively able to infect a cell or leave it uninfected. In figure 1, active viruses are green spheres, whereas inactive viruses are black spheres; uninfected cells are blue spheres and infected cells are represented by red spheres whereas apoptotic or degraded cells are presented by gray spheres.
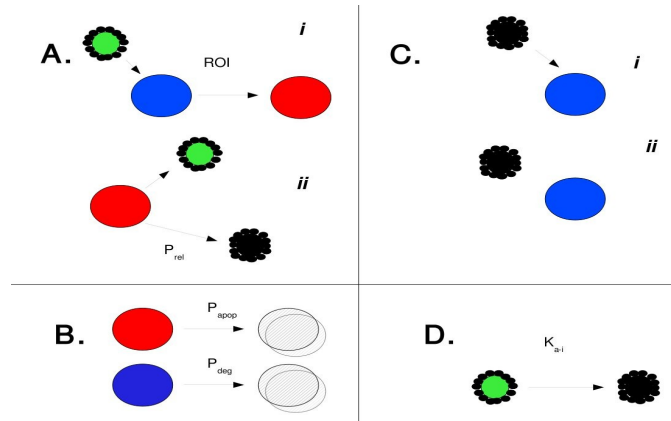
Figure 1. Main steps of simulation

In figure 1 the main steps driving the dynamical evolution of the cell and virus population are depicted: in a first step **i** an *active* viral particle can adhere to an uninfected cell and infect it; here the transition probability is given by a function called risk of infection (ROI), which is defined in the following way

$$ROI(t) = \frac{V_a(t)}{V(t)} \cdot P_{inf} \cdot f\left(\frac{V_a(t)}{C}\right), \tag{1}$$

Modeling Heterogeneous Virus Infectiosity for Cells in Bio-Reactors for Vaccine Production

where $V_a(t)$ is the number of active viruses, $V(t)$ the total number of viruses (active and inactive) and P_{inf} is a constant parameter of infection, and $f(x)$ is a Poisson distribution function that defines the multiplicity of infection, depending on the fraction of active viruses in relation to the total cell population C.

In a subsequent step (figure 1, part **A, ii**) the infected cell internally replicates viruses with a constant probability P_r and releases several new active and inactive viral particles with a constant probability P_{rel}. The degradation and apoptotic transitions are considered in part **B** of the same figure. We consider that only active viruses are able to adhere to uninfected cells; for this reason, inactive viruses cannot infect uninfected cells (figure 1 **C,** steps **i** and **ii**). Finally, active viral particles can have an irreversible transition to an inactive state (inactive viruses are not able to become active again), given by a constant probability K_{a-i} (figure 1, part **D**). The ratio between inactive to active replicated particles is here defined as $R_{ia} = V_i/V_a$. The dependence of the model on K_{a-i} and R_{ia} will be analysed in the third section.

The dynamics for the viral states is implemented using the basic stochastic model implemented by Sidorenko et. al. (Sidorenko, 2007), where the degree of infection of individual cells is included (Müller, 2008). In this model the intracellular Virus Expression -VE- (represented by an individual class J_i for the cell i), contains information on the number of phosphoproteins associated with viral RNA, which probably is involved in the switch from viral protein synthesis to viral genome replication (Flint et. al., 2000). The number J_i, describing this internal state for each individual cell C_i, changes according to virus replication and release by this cell. For virus replication the number increases by 1, $J_i(t+dt)=J_i(t)+1$, for release it decreases by 1, $J_i(t+dt)=J_i(t)-1$.

Table 1: Principal parameters applied in the simulations

Parameter	Value [-]
P_{inf}	0.0299
P_{un-deg}	0.003
P_{ap}	0.05
P_r	0.482
P_{rel}	0.48

3. Results

The production of active viral particles depends on the parameter K_{a-i}, which is a free parameter that must be selected according to the experimental results (in previous models, $K_{a-i}=0$ and $R_{ia}=1$; see Sidorenko, 2008). In figure 2 the total production of viral particles and the number of active viral particles is represented as a function of time (in Monte Carlo steps -MCS-) and K_{a-i}.

For $K_{a-i} < 0.015$ viruses maintain their active state throughout the whole time, i.e. the transition from active to inactive viral states is much slower than the release of new active viral particles. However, in the parameter space there is a region ($K_{a-i} > 0.015$) where the transition from active to inactive states is compensated by the production of new viral particles. Eventually, for $K_{a-i} \to 1$ (not shown in this figure) all the viruses change instantaneously to the inactive state. This would lead to a complete stop of the secondary infection in the bio-process, but has not been observed in any experiment. From experimental data (Schulze-Horsel, 2009) it can be deduced that a transition probability K_{a-i} will be in the range of about 0.05.
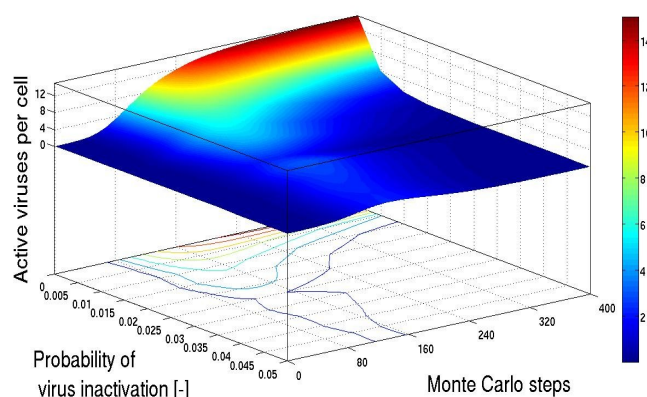
Figure 2. Number of active viral particles as a function of time (MCS*)* and $K_{a-i,}$.

The ratio of the released active respect to inactive viruses depends on particular internal cellular characteristics that can be influenced either by cultivation conditions or by the virus population in the bioreactor. In the previous simulations we assumed a constant ratio for R_{ia} of 0.05 (i.e. more active than inactive viruses are released). This is an ad-hoc assumption that must be analyzed in more detail. In this point we can ask, if there is some critical ratio where the virus replication process will not be sustained any more because no active viruses are replicated by the cells (see figure 3, *-left-*).
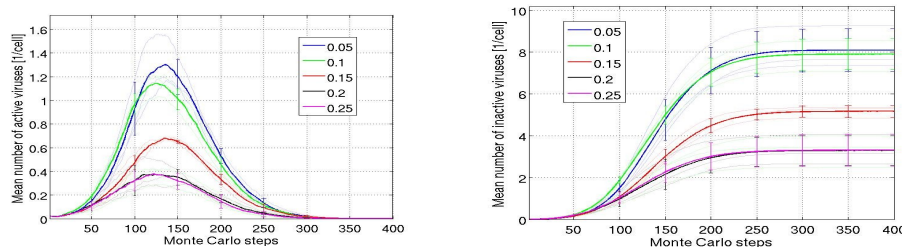

Figure 3. ,Number of active viral particles as a function of time (MCS) and the production ratio of active particles $R_{ia} = V_i/V_a$ *(left)*. The time series of the total number of viruses for the same ratio parameters *(right)*. Each colored line is an averaged data over 10 runs.

We have found that for $R_{ia} < 0.1$ the virus release has no dramatic change respect this ratio. However, above $R_{ia} > 0.1$ there is a continuous decrease in the number of released active viral particles. In general, this ratio has not only influence on the productivity of active viruses, but also on the total productivity of *V* (figure 3, *-right-*). Further experiments would shed some light on the possible value and relevance of $R_{ia.}$

The inclusion of viruses with different grades of infectivity introduces a time delay in the dynamics of the whole process. When this model parameter is considered, the production time can be probably optimized. However, in the fabrication of vaccines the

Modeling Heterogeneous Virus Infectiosity for Cells in Bio-Reactors for Vaccine Production

total amount of viruses to be inoculated for the first infection of the cell population, i.e. the initial multiplicity of infection (MOI = $V_a(0)/C$), is a relevant parameter in the production process. We will therefore study the influence of this parameter in more detail (in this analysis we set $R_{ia}=0.05$ and $K_{a-i}=0.05$).
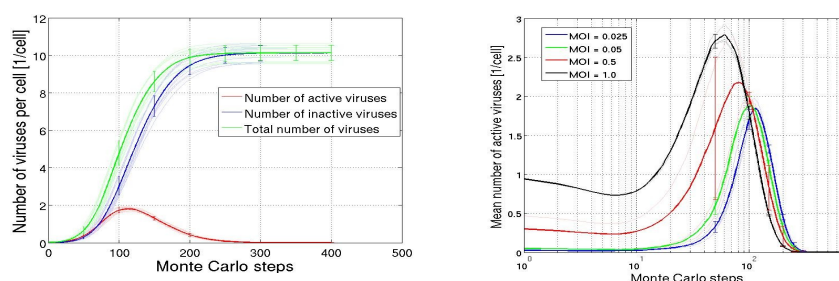

Figure 4. *-left-* Total number of viruses $V[1/C]$ and number of active viral particles $V_a[1/C]$ as function of time for MOI = 0.025. *-right-* $V_a[1/C]$ as function of time (MCS) for different MOI values.

In figure 4 two main results are presented: *-left-* total number of viruses, $V[1/C]$, and the number of inactive and active viruses ($V_i[1/C]$ and $V_a[1/C]$ respectively). In the short to intermediate time regimes (until 50 Monte Carlo Steps) the cell population is infected not only by active viral particles, but its production ability is large enough to produce more active than inactive viral particles. As the population of inactive viruses increases, the apoptosis and degradation of infected and uninfected cells respectively stops the virus production. Eventually, after long time regimes the population of viral particles V converges to the population of inactive particles V_i.

We also make an analysis of the dependence of V_a on the MOI value (figure 4, *-right-*). For low MOI's we observe that there is a time delay in the production of active viral particles. However, in a relative short time the population of infected cells is able to produce a large number of active viral particles, indicating that low MOI values are also optimal for the production of new viruses. In these simulations the number of active viral particles increases proportional to the increase of MOI. In experiments (not shown in this paper; see Schulze-Horsel, 2009) it has been shown that MOI values of 0.05 are optimal with regard to a profitable vaccine production.

4. Conclusion

A new aspect has been introduced into an established Monte Carlo model (Sidorenko, 2008). The production of active and inactive virus particles has been implemented as another population structure with two important parameters, a ratio in the release of active and inactive virus by the cell, R_{ia}, and a transition probability of virus from an active to an inactive state, K_{a-i}. In particular we have shown ways to use the simulation results to validate specific models and assumed phenomena in accordance to their importance on the evolution at hand. A validated model could then be used to optimize certain initial parameters in the production process. Optimizations would be possible for

example to maintain a specific ratio of active and inactive virus particles throughout the whole processing time.

We also explore the dependence of the system on the MOI value. We predict that at low MOI values there is a time delay in the production of active viruses. Hence, the MOI value has influence on the time when the release of new viruses reaches its maximum. This knowledge might be very important for the production of active immunization vaccines. However, we observe that the total concentration of viral particles is proportional to this MOI value. This result is not in line with experimental data (shown in Schulze-Horsel, 2009), where an optimal production is obtained for low MOI values, unless additional mechanisms, as the innate immune response of the cells, are considered. In the last scenario we found that low MOI values could also be efficient in virus production. Our model of virus production with heterogeneous activity simplified here the role that active and inactive virus particles have on the dynamics of the cell population. As an extension of the present work we are currently working on the effect that activity of viral particles could have on the cell cycle.

Acknowledgments

References

C. Beauchemin, J. Samuel, J. Tuszynski, 2005, A simple Cellular Automaton Model for Influenza A Viral Infections. Jour. Theor. Biol., **232**, 223.

J.S. Flint, V.R. Vacaniello, R. Krug, 2000, Principles of Virology: Molecular Biology, Pathogenesis and Control of Animal Viruses. American Society for Microbiology, Washington.

T. Müller, J. Schulze-Horsel, Y. Sidorenko, U. Reichl, A. Kienle, 2008, Population Balance Modeling of Influenza Virus Replication in MDCK cells during Vaccine Production, ESCAPE 18.

J. Schulze-Horsel, M. Schulze, G. Agalaridis, Y. Genzel, U. Reichl, 2009, Interaction Dynamics and Virus-Induced Apoptosis in Cell Culture-based Influenza Vaccine Production-Flow Cytometry and Mathematical Modeling, Vaccine, **27,** 2712.

Y. Sidorenko, J. Schulze-Horsel, A. Voigt, U. Reichl, A. Kienle, 2008, Stochastic Population Balance Modeling of Influenza Virus replication in Vaccine Production Processes, Chem. Eng. Sci., **63**, 157.

Self-organizing maps and learning vector quantization networks as tools to identify vegetable oils and detect adulterations of extra virgin olive oil

José S. Torrecilla[a], Ester Rojo[b], Mercedes Oliet[c], Juan C. Domínguez[d], Francisco Rodríguez[e]

Departamento de Ingeniería Química, Facultad de Ciencias Químicas, Universidad Complutense de Madrid, 28040-Madrid, Spain. (a) Tel.: +34 91 394 42 40, Fax: +34 91 394 42 43, jstorre@quim.ucm.es; (b) erojorec@pci.ucm.es; (c) Tel.: +34 91 394 42 41, moliet@quim.ucm.es; (d) jucdomin@quim.ucm.es; (e) Tel.: +34 91 394 42 46, E-mail address: frsomo@quim.ucm.es.

Abstract

Unsupervised models have been explored for the identification of edible and vegetable oils and to detect adulteration of extra virgin olive oil (EVOO) using the most common chemicals in these oils such as saturated fatty, oleic and linoleic acids. The optimization and validation processes of the models have been carried out using bibliographical sources. A database for developing learning process and internal validation, and six other different databases to perform their external validation has been used. In the worst of the cases, the unsupervised models are able to classify more than the 94 % of samples and detect adulterations of EVOO with promising results. The adulteration of EVOO with corn, soya, sunflower and hazelnut oils can be detected when their oil concentrations are higher than 10, 5, 5 and 10 %, respectively.

Keywords: Adulteration, Competitive neural networks, extra virgin olive oil, edible oil.

1. Introduction

In recent decades, due to the high price of extra virgin olive oil (EVOO), an appreciable incidence of adulteration has been detected. The substitution or adulteration of EVOO with cheaper ingredients is not only an economic fraud but may also on occasions have severe health implications for consumers an example being the Spanish toxic oil syndrome.

Although EVOO quality can be checked by chemical indices and organoleptic assessment, there is no single analytical index to determine the protected denomination of origin, their geographical origin or even the olive fruit variety. That is why, to determine the adulteration of EVOO or classify the vegetable oils, concentrations of chemicals present in the oils and their physicochemical properties should be quantified [1, 2]. These can be determined using a wide number of chemometric tools which are based on techniques such as nuclear magnetic resonance spectroscopy [3], fourier transform raman spectroscopy [4], etc. To extract the most relevant information from those huge databases formed by the characteristics, composition and concentration of chemicals of each edible oil, statistical techniques are required such as principal component analysis (PCA) [5, 6], multivariate regression techniques [7, 3] or non linear algorithms, etc.

Two of the most used competitive neural networks algorithms used are self-organizing maps (SOMs) and learning vector quantization networks (LVQs) models. SOM and LVQ models have been successfully used in different scientific fields [8,9]. To the best of our knowledge, in the vegetable oil field, SOM models have hardly been used and the LVQ model applications are even scarcer [10, 11]. Given the successful results achieved in other scientific fields and the recognition capability of groups with similar characteristics (using SOM models) [12, 13] and that even these groups can be in addition classified in target classes defined by the user (using LVQ models) [12, 13], the combination of linear (PCA) and non linear algorithms (SOM or LVQ models) have been applied here in the edible oil area.

The main objective of this work is the application of a principal component analysis technique, self organizing map and learning vector quantification networks to identify hazelnut, sunflower, corn, soybean, sesame, walnut, rapeseed, almond, palm, groundnut, safflower, coconut, and extra virgin olive oils and detect adulterations of EVOO with corn, soya, sunflower and hazelnut oils, using only their saturated fatty (mainly palmitic and stearic acids), oleic and linoleic acid concentration values.

2. Material and methods

Learning, verification and validation samples. To design and optimize both SOM and LVQ models a database composed of values of the acidity, iodine value, ratio of 1,2-diglycerides to the total diglycerides and the concentrations of total sterols, total diglycerides, 1,2-diglycerides, 1,3-diglycerides, saturated fatty (SFA), oleic, linolenic and linoleic acids determined by analysis of the respective 1H NMR and 31P NMR spectra was used [3]. These properties have been calculated in 192 samples corresponding to 13 types of vegetables oils (hazelnut, sunflower, corn, soybean, sesame, walnut, rapeseed, almond, palm, groundnut, safflower, coconut, and extra virgin olive oils). In order to test the performance of optimized SOM and LVQ models related with the classification of vegetable oils and detection of adulteration of EVOO with seeds oils, another seven bibliographical references (263 samples) have been employed. [14, 15, 3, 10, 16-18].

Principal component analysis is a classical unsupervised technique based on linear algebra. It involves a mathematical procedure, which transforms a number of possibly correlated variables into a smaller number of uncorrelated variables called principal components. Here, the PCA technique has been applied to check and select the most important information from the aforementioned database. Then, using the selected information, two different non linear models were designed and applied. PCA was carried out by SPSS software version 15.0.

Self-organizing map is one of the most interesting topics in the competitive neural network field [12]. SOM models can learn to detect irregularities and correlations in their input and adapt their future responses to that input accordingly, that is, they are able to recognize groups with similar characteristics [12, 13].

Given that self-organizing maps classify input data according to how they are grouped in the input space, along the leaning process, in order to adequately represent all input data, its weights are optimized. As every neuron is represented as a weight vector, during this process, the neurons look for the best place to represent whole input database. The learning process of the SOM involved two steps viz. ordering and tuning phases [13]. In the former, the ordering phase learning rate (OLr) and neighborhood distance (ND) are decreased from both that rate and the maximum ND between two

neurons to the tuning phase learning rate (TLr) and the tuning phase neighborhood distance, respectively. The ordering phase lasts for a given number of steps (named ordering phase steps, OP). The SOM model used in this work was designed using Matlab version 7.01.24704 (R14) [13].

Learning vector quantization networks can classify any set of input vectors, not only linearly separable sets of input vectors. The only requirement is that the competitive layer must have enough neurons, and each class must be assigned enough competitive neurons [19]. LVQ models classify input vectors into target classes by using a competitive layer to find subclasses of input vectors, and then, combining them into the target classes defined by the user. Therefore, LVQ networks consist of two layers viz., unsupervised (competitive) and supervised (linear) layers [13]. The competitive layer learns to classify input vectors in much the same way as the competitive layers of self-organizing maps. The linear layer transforms the competitive layer's classes into target classifications defined by the user. The linear layers have one neuron per class [13]. The LVQ model used in this work was designed using Matlab version 7.01.24704 (R14).

3. Results and discussion

Learning, verification and validation samples
In order to select the most important variables, mutual correlation coefficients between 1,2-diglycerides, 1,3-diglycerides, the ratio of 1,2-diglycerides to total diglycerides, acidity, iodine value, and fatty acid composition were calculated. Correlations coefficients higher than 0.9 was found in three cases viz., 1,3-diglycerides and the ratio of 1,2-diglycerides to total diglycerides and 1,3-diglycerides and between iodine value and acidity and SFA. In the light of these results, to remove the data redundancy, only three of them (one in each pair) were taken into account. In addition, to reveal the underlying information of the database used, the PCA technique has been applied. In addition to these, as SFA (palmitic and stearic acids), oleic and linoleic acids are present in most edible vegetable oils, the concentration values of these compounds have been selected as the most characteristic of the oil samples tested.

The learning, verification and external validation present the same format. These have as many rows as variables necessary to characterize the process (concentrations of SFA, oleic and linoleic acids) and the same number of columns as the number of vectors to describe the system to be studied. Whole database have been distributed randomly into learning (80%) and verification samples.

Self-organizing map optimization
A non linear mapping method has been used to classify the aforementioned thirteen types of oils. The output neurons were arranged in three different topological grids viz. grid, hexagonal and random topologies. In addition, three different methods to calculate the distances were used, viz. link, Euclidean and Manhattan distances [13]. To classify the edible oil samples in the most reliable way possible, both topologies and distances were combined and the best pair was selected. Following the manufacturer's indications, throughout the selection process, all other parameters were maintained constant and fixed by default (OLr, OP, TLr, ND and the dimension of the network was equal to 0.9, 1000, 0.02, 1 and 5 x 8, respectively) [13]. In most of the nine possible combinations, between 10 to 12 % of EVOO and hazelnut oil samples from verification sample were misclassified as hazelnut, sunflower and EVOO oils. The combination of

hexagonal topology and Manhattan distance was the best combination, misclassifying only less than 5 % of the EVOO samples. Therefore, this combination was selected.

Using the optimized topology and distance of the SOM, the dimension of the network was optimized. Networks of sizes ranging from 18 x 18 to 26 x 26 were tried [8]. From 20 x 20 up to 26 x 26 dimensions, the performance of these maps (number of misclassifications) was somewhat similar. Therefore, the 20 x 20 dimension was selected. Then, with the selected topology, distance and the optimized network dimension, the parameters of the SOM model were optimized by a Central Composite Design 2^5 + star experimental design, where the variables analyzed were OLr (from 0.1 to 1), OP (from 500 to 1500), TLr (from 0.01 to 0.03), ND (from 0.5 to 1.5) and the number of epochs in the learning process (from 10000 to 30000 epochs). The response of the experimental design was the number of the incorrect classifications of the oil samples from the verification sample. In order to reach the least number of misclassifications, the optimum parameter values have been fixed at 0.1, 1500, 0.01, 0.5 and 30000 to OLr, OP, TLr, ND and the number of epochs necessary in the learning process, respectively.

Learning vector quantification network optimization

The LVQ model consists of unsupervised and supervised layers (*vide supra*). The former is a competitive network similar to the SOM model. This part, formed by the input and hidden neurons, is fixed at three and 20 x 20, respectively, as has been described below. The supervised layer consists of an output layer with thirteen neurons, one for each oil type. The only LVQ parameter to optimize is the learning rate (Lr). In the optimization process, Lr was tested between 0.001 to 1. Taking into account that the minimum number of misclassifications is required, the best Lr value was 0.01. This is in agreement with literature [13].

Application of optimized SOM and LVQ models to others databases

In order to validate the optimized SOM and LVQ models, the external validation process has been carried out using six bibliographical databases (235 samples) [14, 15, 3, 10, 16-18]. In addition, analytical values from mixtures of EVOO and edible oils (28 samples) have been also used to test the reliability of the aforementioned models in the EVOO adulteration detection [3].

In order to guarantee the reliability of the classifications carried out by these models, the applicability domain has been evaluated selecting the compounds with cross-validated standardized residuals greater than three standard deviation values [20, 21]. In this evaluation applied to validation sample, no response outlier was found. Then, the validation sample was input into the SOM and LVQ models. Only one input data set is not adequately represented by the optimized SOM and two competitive neurons are not used to classify some input data sets. As can be expected, the number of misclassifications in the external validation process is higher than those in the internal validation. Nevertheless, as the misclassification percentage is less than 5%, the tested models are able to classify vegetable oils adequately. Given that the LVQ model is partially based on the SOM model, the results are similar. Although the misclassification percentage is slightly higher (< 5.5%), the LVQ model has the advantage in that the classifications are organized by the user. To recap, the models tested are able to classify all oil types, even the hazelnut oil samples.

Finally, the capacity to detect adulteration of EVOO with seeds oils has been tested [3]. All twenty eight samples are adequately represented by the optimized SOM or LVQ models. In particular, using these methods, the adulteration with corn, soya, sunflower

Self-organizing maps and learning vector quantization networks as tools to identify vegetable oils and detect adulterations of extra virgin olive oil

and hazelnut oils can be detected when their respective concentrations are higher than 10, 5, 5 and 10 %. In the case of hazelnut adulteration, these results are similar to those published in literature [22]. But here, thanks to non linear models applied, the required information is notably less.

4. Conclusion

In this work unsupervised models have been used to identify edible and vegetable oils and to detect adulteration of extra virgin olive oil. These models have been developed using only three of the chemicals present in most vegetable oils and tested using internal and external validation samples. In the latter, the adulteration of EVOO with corn, soya, sunflower and hazelnut oils was detected when their concentration was higher than 10, 5, 5 and 10 %, respectively. Therefore, the models developed are adequate to classify these studied samples in thirteen types of vegetable oils. Although the results could be improved by specifically designed models for the adulteration databases, the results reached here are promising.

References

[1] Commission Regulation (EC) No 1989/2003 of 6 November 2003 amending Regulation (EEC) No 2568/91 on the characteristics of olive oil and olive-pomace oil and on the relevant methods of analysis.

[2] Food and Agriculture organization of the United Nations. Codex alimentarius commission. Codex Standard for named vegetable oils. CODEX Stan 210, 1999 (revision and amendments: 2003, 2005).

[3] G. Vigli, A. Philippidis, A Spyros, P. Dais, 2003, Classification of Edible Oils by Employing 31P and 1H NMR Spectroscopy in combination with multivariate statistical analysis. A proposal for the detection of seed oil adulteration in virgin olive oils. J. Agric. Food Chem. 51, 5715-5722.

[4] V. Baeten, P. Hourant, M. T. Morales, R. Aparicio, 1998, Oil and fat classification by FT-Raman spectroscopy. J. Agric. Food Chem. 46, 2638-2646.

[5] S. Mildner-Szkudlarz, H. H. Jelen, 2008, The potential of different techniques for volatile compounds analysis coupled with PCA for the detection of the adulteration of olive oil with hazelnut oil. Food Chem. 110, 751–761.

[6] V. G. Dourtoglou, T. Dourtoglou, A. Antonopoulos, E. Stefanou, S. Lalas, C. Poulos, 2003, Detection of olive oil adulteration using principal component analysis Applied on Total and Regio FA Content. J. Am. Oil Chem. Soc. 80, 203-208.

[7] F. Peña, S. Cárdenas, M. Gallego, M. Valcárcel, 2005, Direct olive oil authentication: Detection of adulteration of olive oil with hazelnut oil by direct coupling of headspace and mass spectrometry, and multivariate regression techniques. J. Chromatogr. A 1074, 215–221.

[8] A. M. Fonseca, J. L. Biscaya, J. Aires-de-Sousa, A. M. Lobo, 2006, Geographical classification of crude oils by Kohonen self-organizing maps. Anal. Chim. Acta 556, 374–382.

[9] J. Zhang, H. Yang, Application of self-organizing neural networks to classification of plant communities in Pangquangou Nature Reserve, North China. Front. Biol. China. DOI 10.1007/s11515-008-0061-7.

[10] D. Brodnjak-Voncina, Z. C. Kodba, M. Novic, 2005, Multivariate data analysis in classification of vegetable oils characterized by the content of fatty acids. Chemometrics Intell. Lab. Syst. 75, 31– 43.

[11] F. Marini, A. L. Magrì, R. Bucci, A. D. Magrì, 2007, Use of different artificial neural networks to resolve binary blends of monocultivar Italian olive oils. Anal. Chim. Acta 599, 232–240.

[12] T. Kohonen, 1987, Self-Organization and associative Memory, 2nd Edition, Berlin, Springer-Verlag,

[13] H. Demuth, M. Beale, M. Hagan, 2005, MATLAB User's Guide, v 4.0: Neural Network Toolbox. MathWorks Inc., Mass., USA.

[14] D. S. Lee, B. S. Noh, S. Y. Bae, K. Kim, 1998, Characterization of fatty acids composition in vegetable oils by gas chromatography and chemometrics. Anal. Chim. Acta 358, 163-175.

[15] F. Marini, F. Balestrieri, R. Bucci, A. D. Magrì, A. L. Magrì, D. Marini, 2004, Supervised pattern recognition to authenticate Italian extra virgin olive oil varieties. Chemometrics Intell. Lab. Syst. 73, 85– 93.

[16] H. L. Gan, Y. B. Che Man, C. P. Tan, I. NorAini, S. A. H. Nazimah, 2005, Characterisation of vegetable oils by surface acoustic wave sensing electronic nose. Food Chem. 89, 507–518.

[17] J. E. Spangenberg, S. A. Macko, J. Hunziker, 1998, Characterization of olive oil by carbon isotope analysis of individual fatty acids: Implications for authentication. J. Agric. Food Chem. 46, 4179-4184.

[18] P. Dais, A. Spyros, S. Christophoridou, E. Hatzakis, G. Fragaki, A. Agiomyrgianaki, E. Salivaras, G. Siragakis, D. Daskalaki, M. Tasioula-Margari, M. Brenes, 2007 Comparison of analytical methodologies based on 1H and 31P NMR spectroscopy with conventional methods of analysis for the determination of some olive oil constituents. J. Agric. Food Chem. 55, 577-584.

[19] G. R. Lloyd, R. G. Brereton, R. Faria, J. C. Duncan, 2007, Learning vector quantization for multiclass classification: application to characterization of plastics. J. Chem. Inf. Model. 47, 1553-1563.

[20] P. Gramatica, 2007, Principles of QSAR models validation: internal and external. QSAR Comb. Sci. 26, 694-670.

[21] P. Gramatica, E. Giani, E. Papa, 2007, Statistical external validation and consensus modeling: A QSPR case study for Koc prediction. J. Mol. Graph. 25, 755–766.

[22] A. Sayago, D. L. García-González, M. T. Morales, R. Aparicio, 2007, Detection of the presence of refined hazelnut oil in refined olive oil by fluorescence spectroscopy. J. Agric. Food Chem. 55, 2068-2071.

Systematic Model Analysis for Single Cell Protein (SCP) Production in a U-Loop Reactor

Oscar Andrés Prado-Rubio[a], John Bagterp Jørgensen[b], Sten Bay Jørgensen[a]

[a]*Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Lyngby, Denmark, oap@kt.dtu.dk, sbj@kt.dtu.dk*
[b]*DTU Informatics, Technical University of Denmark, DK-2800 Lyngby, Denmark, jbj@imm.dtu.dk*

Abstract

Cultivation of *Methylococcus capsulatus* for SCP production is a highly exothermic process which is limited by oxygen transport from gas to liquid phase when it is conducted in a conventional fermenter. This motivated the design of the U-loop reactor which enhances heat and mass transport through the integration of functions. As a consequence of intensification, the operability of the system become an issue. In order to understand the system behavior and investigate operability, a systematic model analysis is performed. It results in the design of a control structure which facilitates the reactor start-up and enables stable operation.

Keywords: U-loop reactor, Single Cell Protein (SCP) production, Modeling for control

1. Introduction

The rapidly increasing world population generates the challenge of providing necessary food sources. In particular protein supply poses a problem since essential amino acids can not be replaced. One possible solution to this problem is SCP production. Bacteria and yeast are candidates for the synthesis of SCP. Bacteria grow more rapidly and efficiently than yeast on cheap substrates, and they provide a higher content of protein. The *M. capsulatus* has been studied for SCP production due to its high protein content, almost 70% [4]. The cultivation involves several basic process engineering operations, such as stirring and mixing of a multiphase system (gas-liquid-solid), transport of oxygen from the gas bubbles through the liquid phase to the microorganisms, and heat transfer from the liquid phase to the surroundings. When *M. capsulatus* cultivation is conducted in a traditional stirred reactor, intractable heat and mass transfer problems appear at scale up. The U-loop fermenter is a special bioreactor type designed for intensifying mass and energy transport phenomena by enhancing the mixing of the multiphase system and favoring heat transfer [1, 5]. The purpose of this paper is to develop and investigate a dynamic model of the U-loop reactor to reveal the consequences of intensification. Simulations are employed to provide understanding how the reactor must be operated, both during start-up and during disturbances encountered during continuous culture. A systematic model analysis is required since reactor operation is non trivial.

2. Process description

2.1. Microorganism

M. capsulatus may utilize methane or methanol because they can be easily assimilated. The methane is cheaper than methanol. However, the use of methane has mainly three

problems: the potential explosion hazard, low solubility in water and high purity is required. These limitations are avoided using methanol. However, high methanol concentration inhibits growth and must therefore be carefully controlled [4, 5].

Ammonium is conventionally used as N-source, but the microorganism is very sensitive to fluctuations in its concentration that can lead to culture instability. Nitrate as N-source is attractive since it leads to higher growth rate and protein content. Additionally, the nitrate addition as HNO_3 can facilitate the reactor pH control [5].

2.2. Reactor

The U-loop fermenter is a non-conventional reactor where the fermentation broth with a dispersed gas phase circulate in a U shaped loop as indicated in fig. 1 [5]. Through the integration of functions, the transport limitations evidenced at large scale are diminished and the cultivation intensified [8]. SCP fermentation is an exothermic process and large heat loss to surroundings is desired. Therefore, the reactor was designed to enable a larger heat transfer surface than in a conventional reactor. Furthermore, the use of intensified equipment (static mixers along the reactor) enhances mass transfer of oxygen from the gas to liquid phase. Besides, the gas phase is pumped into the reactor through a jet, which enables good mixing and oxygenation [1, 5]. The SCP productivity can be further increased with this design since it allows going from batch/fed-batch cultivations to a continuous production. The reactor is composed of the following main parts: two vertical tubular legs, a propeller pump, a degassing unit at the top and a heat exchanger [7].
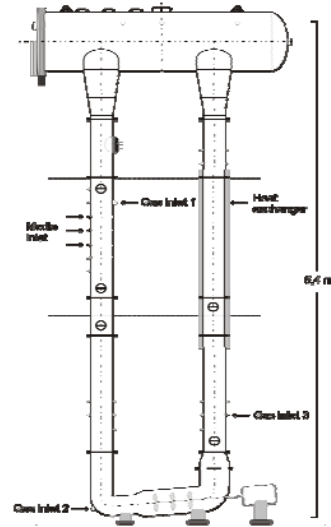

Fig. 1: Sketch of the U-loop pilot plant at DTU

3. Bioreactor model

In the model, methanol is used as C-source, nitrate (HNO_3 and $NaNO_3$) as N-source, and air as O_2-source. The U-loop reactor can be divided into a degassing unit on the top and the U-loop section at the bottom. The degassing unit has large dispersion and is modeled as a CSTR. The U-loop section has low dispersion and is modeled as a plug flow reactor with dispersion [7].

3.1. Stoichiometry and reaction rate

The presence of several metabolic pathways for *M. capsulatus* means that the stoichiometry depends on the actual operating conditions. The overall reaction stoichiometry is shown in the following reaction, using standard biomass composition X= $CH_{1.8}O_{0.5}N_{0.2}$.

$$CH_3OH + Y_{sn}HNO_3 + Y_{so}O_2 \rightarrow Y_{sx}X + Y_{sc}CO_2 + Y_{sw}H_2O$$
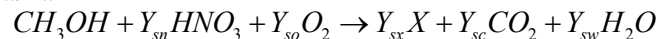
The biomass yield coefficient (Y_{sx}) is determined from experimental data. Biomass reaction rate is calculated using Monod-Haldane rate expressions with parameters based on experiments with methanol and ammonia as substrates [5]. Methanol and oxygen limitations are presently included in the rate expression. The biomass production rate and the yield coefficients are used to calculate the other volumetric production or consumption rates. The specific biomass growth rate is:

$$\mu = \frac{\mu_{max} C_s}{K_s + C_s + (C_s / K_I)^2} \left(\frac{C_o}{K_o + C_o} \right) \tag{1}$$

3.2. Gas-liquid mass transfer

The oxygen mass transfer is kinetically limited. The overall mass transfer coefficient is approximately equal to the liquid film mass transfer coefficient and assumed position independent. In this model, the oxygen limitation condition is included through the coupling between oxygen transport, kinetics and oxygen mass balances in both phases.

3.3. Total volume

Only one gas inlet stream is used, it is located in the upper section of the downward leg. It is assumed that the gas-liquid transport does not significantly alter the total gas volume as oxygen consumption is accompanied by CO_2 production. The liquid stream is composed by methanol, some salts (KH_2PO_4, $NaNO_3$ and $CaCl_2 \cdot 2H_2O$), nitric acid and water. The gas void fraction is estimated from the ratio between gas feed and total volumetric flow rate, which is valid in the case of homogeneous flow [2]. The dilution rate is estimated for the liquid phase.

$$\varepsilon = \frac{\upsilon_g}{\upsilon_l + \upsilon_g} \quad \text{and} \quad D = \frac{\upsilon_l}{(1 - \epsilon)V} \tag{2}$$

3.4. pH model

The addition of nitric acid and its consumption by the microorganism influence the cultivation pH. The quality of a model based pH controller depends on the quality of the pH model. pH is governed by the equilibrium reactions involving KH_2PO_4 and HNO_3. The unknown concentrations are estimated using the equilibrium constants and the electroneutrality condition. The pH is computed by simultaneous solution of these expressions using an eigenvalue approach.

3.5. Spatial discretization

An effective method to approximate the solution of the axially dispersed plug flow mass balance is by using a tank in series model. The model consists of mass balances for the components in the gas and liquid phase as well as an energy balance. Furthermore, phenomena such as reaction, gas-liquid mass transfer, and heat transfer are modeled. The most important assumptions are:

- Liquid phase: constant volume and dilution rate in each section, incompressible liquid. Nitrate stays in liquid phase and N_2 stays in gas phase. The properties for the liquid phase are assumed as water.
- Gas phase: constant molar flow rate. No reaction. Phase is quasi-stationary. All the gas is stripped off in the degassing unit. Ideal gas behavior. The air is saturated in the first reactor section. All the CO_2 formed is immediately transferred to the gas phase. The properties for the gas phase are assumed as dry air.
- Biotic phase: only biomass, carbon dioxide and water are produced and the microorganism does not assimilate nitrogen or carbon dioxide.
- Other: free convection is the dominating heat transfer mechanism, constant pressure in each section, but it changes along the U-loop. The motor efficiency is 80%.

3.6. Mass and energy balances

The mass balance for the specie i in section k includes the mass generated by reaction and the mass transported from/to the gas phase. The general equation is:

$$\frac{dC_{i,k}}{dt} = q_{i,k} + q_{it,k} + \frac{1}{(1-\in_k)V_k}\left(\upsilon_{l,k-1}C_{i,k} + \upsilon_{lf}C_{if} - \upsilon_{l,k}C_{i,k}\right) \tag{3}$$

The energy balance is formulated taking into account the heat added by the heat exchanger, the reaction heat, heat losses, heat removed by evaporation and the work added by the impeller pump. Mass balances for oxygen and carbon dioxide are formulated for the gas phase:

$$\frac{dC_{ig,k}}{dt} = \frac{y_{if}n_{gf,k}}{\in_k V_k} \pm \frac{1-\in_k}{\in_k}\frac{q_{it,k}}{MW_i} + \frac{1}{\in_k V_k}\left(\upsilon_{g,k-1}C_{ig,k-1} - \upsilon_{g,k}C_{ig,k}\right) \tag{4}$$

4. Model analysis and simulation results

4.1. Model analysis strategy
First part of the model analysis is the degrees of freedom (DOF) estimation, from there the potential manipulable variables are identified: the flow rates of methanol, F_1, salts, F_2, nitric acid, F_3, water, F_4, air, v_{gf} and the heat supply rate, Q_{heat}. How those DOF will be distributed, between inputs and disturbances, depends on the control objectives. Secondly, static simulations are used to explore the operative window and to find the optimal biomass productivity. Afterwards, the control problem is defined. Dynamic simulations are used to compare a fedforward and feedback controlled start up.

4.2. Static analysis – optimal operating conditions
The most interesting DOF to investigate at this stage are the methanol input flow rate and the gas input flow rate. The salts and acid flow rate, F_2 and F_3, can be related to the methanol feed flow rate, F_1, to guarantee that the reactor has the amount of nutrients necessary for the growth [1, 2, 5]. A feed forward ratio control strategy is used. The water input flow rate F_4 is fixed and no external heat is provided to the system.
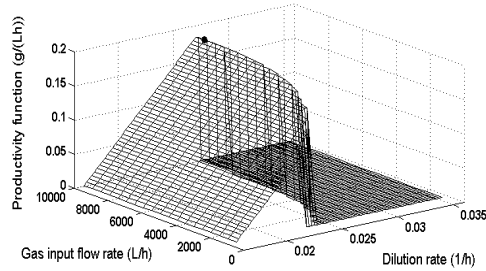

Fig. 2: Static biomass productivity as a function of the dilution rate and gas input flow rate. Maximum biomass productivity is marked

The static biomass productivity response surface is illustrated in fig. 2. The maximum biomass productivity is achieved using v_{gf}=9500 L/h and D=0.027 h^{-1} (F_1=4.31 L/h), approximately. It is well known experimentally, that the U-loop reactor is close to instability at high gas and methanol input flow rates [5]. Therefore, control is necessary to realize maximum productivity of the U-loop reactor.

4.3. Dynamic analysis – start up
Start-up of the U-loop reactor is neither trivial nor intuitive. Experimental experience and simulations were used to develop an open loop start-up procedure which avoids the reactor washout around a known experimental operating point. The gas input flow rate is kept constant and a stepwise input methanol flow rate is used (illustrated in fig. 3d, when the reactor is divided in 4 sections). There, the evolution of important concentrations can be followed. It can be seen that the methanol concentration is kept below the lethal concentration and that sufficient dissolved oxygen is present. The concentration differences between the sections are almost negligible as the system

approaches the behavior of a single CSTR at high recirculation flow rates. Larger differences are observed for the O_2 concentrations in both the liquid and the gas phase. This can be explained by the pressure change along the U-loop and the fact that gas is not recirculated. To illustrate gas phase concentration changes, the U-loop section was divided into a larger number of sections. These results are depicted in fig. 3 (e and f).

| (a) Methanol | (b) Dissolved oxygen | (c) Biomass |

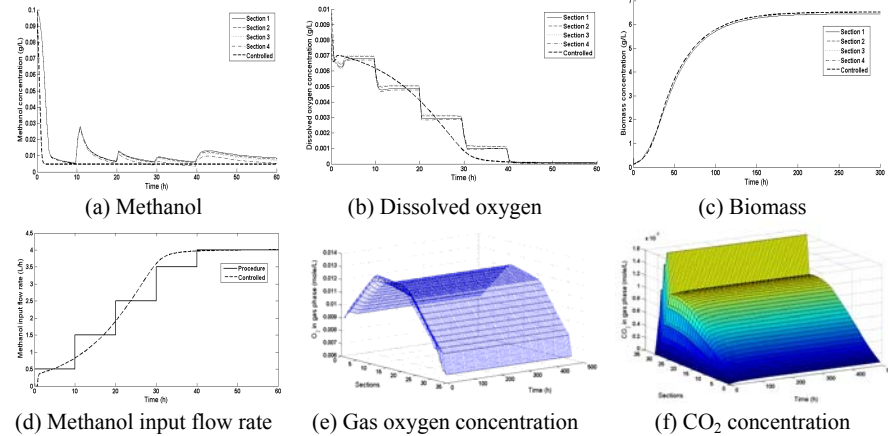| (d) Methanol input flow rate | (e) Gas oxygen concentration | (f) CO_2 concentration |

Fig. 3: System response to the proposed start-up procedure. The controlled start-up uses an inventory based controller at the supervisory layer (the pH buffer effect was neglected for the controller design). Notice that the time scales were adjusted to highlight the periods with non-constant concentrations. Figures (e) and (f) are computed using 31 tanks and no control.

5. Discussion

For SCP production, the yield and specific growth rate are adversely influenced by high methanol concentrations. Thus there is a risk of process instability and further washout as a consequence of disturbances during process operation [3]. A microorganism growth model is essential for successful design and control of the SCP cultivation, since reactor operation is very sensitive to variations in substrate concentrations. In SCP production, the transfer rate of oxygen from the gas phase to the liquid phase is the key limiting factor along the U-loop. The main purpose of the U-loop reactor for SCP production, is to maximize the biomass produced. Therefore, the biomass productivity ($J=C_{x,CSTR}{\cdot}D$ $g/L/h$) is used as a simple objective function. The estimated optimum gas flow rate is twice the value of the operating point at which most of the model parameters were estimated [2]. Therefore the validity of the simulations in this paper remains to be verified experimentally. From the static and dynamic analysis, the control objectives and potential manipulated variables are identified. An initial control structure is proposed based on a hierarchical structure [6]. Methanol and dissolved oxygen must be controlled to avoid biomass growth inhibition and reactor washout. In addition, temperature, pH and nitrate concentration must be regulated to provide the optimal epigenetic conditions for microorganism growth. The top optimization layer defines the optimal static values for substrate concentration and addition rate to be used in the lower control layers. At the supervisory layer, multivariable inventory controllers are under investigation to handle the pH, methanol and nitrate concentrations. For the regulatory layer, inventory and PI controllers regulate temperature and dissolved oxygen. A preliminary simulation of start-up using inventory control at the supervisory layer is

illustrated in fig. 3. The results show that smooth startup is achieved, while gradually approaching oxygen limiting conditions.

6. Conclusions

A dynamic model for cultivation of *M. capsulatus* in an U-loop reactor is derived. The parameters in the model have been estimated based on steady state data from the U-loop pilot plant at the Technical University of Denmark. Qualitatively, the results obtained by the model agree with experimental data [1, 2, 5, 7]. Prediction of the dynamics can be improved by dedicated dynamic experiments designed for parameter estimation. The model can be applied for optimization and control of SCP production in the U-loop reactor. To use a systematic model analysis has shown to be very useful to reveal interesting system behavior, define the control objectives, find the possible manipulated variables and design a control structure. We have used the model to determine a new operating point that improves the biomass productivity with approximately 10%. It remains to verify this improvement in the pilot plant. Furthermore, we have applied the model to develop a startup procedure for SCP production in the U-loop reactor. In a preliminary study of control structures, the model has been used to design a hierarchical control structure employing inventory based controllers supported by PI-control. We are aiming to implement more advanced model based controllers that guarantee an optimizing operation.

7. Nomenclature

C_i: concentration of i (liquid phase), C_{if}: inlet concentration i, C_{ig}: concentration of i (gas phase), D: internal dilution rate, F_i: input flow rate, K_I: inhibitory constant, K_o: limiting dissolved oxygen concentration, K_s: limiting methanol concentration, MW: molecular weight, n_{gf}: molar gas input flow, q_i: reaction rate of the specie i, q_{it}: transport rate of i, V: volume, y_{if}: feed molar fraction, Y_{si}: yield coefficient of i referred to substrate. $\in$: void fraction of gas, μ: specific growth rate, ν_g: total gas flow rate, ν_{gf}: gas feed flow rate, ν_l: circulation flow rate of liquid phase, ν_{lf}: liquid input flow rate.

8. Acknowledgements

9. References

[1] B.R. Andersen, J.B. Jørgensen, and S.B. Jørgensen. U-loop reactor modeling for optimization, part 1: Estimation of heat loss. CAPEC, Technical University of Denmark, 2005.

[2] B.R. Andersen, J.B. Jørgensen, and S.B. Jørgensen. U-loop reactor modeling for optimization, part 2: Mass transfer. CAPEC, Technical University of Denmark, 2005.

[3] I.M. Chu and E. Papoutsakis. Biotechnology and Bioengineering, 29:55–64, 1987.

[4] R.S. Hanson and T.E. Hanson. Microbiological Reviews, 60(2):439–471, 1981.

[5] S. Piper. Continuous Cultures of *Methylococcus capsulatus*. Center of Microbial Biotechnology (Biocentrum)- Technical University of Denmark, 2004. Master's thesis.

[6] S. Skogestad. Plantwide. Journal of Process Control, 10:487–507, 2000.

[7] L. Soland. Characterization of Liquid Mixing and Dispersion in a U-loop Fermentor. Technical University of Denmark, 2005. Master's Thesis.

[8] Stankiewicz, A. and Moulijn, J. Chemical Engineering Progress, 96(1):22-34, 2000

Aglianico wine dealcoholization tests

Loredana Liguori, Gerardina Attanasio, Donatella Albanese, Marisa Di Matteo

Dipartimento di Ingegneria Chimica e Alimentare, Università degli studi di Salerno, Fisciano 84084, Salerno, mdimatteo@unisa.it

Abstract

In the recent years, there has been a growing interest in partial or total dealcoholization of alcoholic beverages. The demand for drinks with low levels of alcohol is attributable both for health reasons and for the desire to reduce alcohol consumption in new generations. This research aims to produce a dealcoholized wine with less than 0.5% alcohol-by-volume from Aglianico. Dealcoholization tests were carried out by plant working on direct osmosis, equipped with a membrane contactor and some temperature sensors and automatic control of the process phases. The fixed alcohol content was achieved with 5 cycles, under mild condition. Chemical-physical and aromatic composition was evaluated in feed (wine) during the dealcoholization process.

Keywords: direct osmosis, wine dealcoholization, chemical-physical composition, aromatic compounds, beverage.

1. Introduction

Wine is one of the most popular alcoholic drinks in the world, which contributes to reducing the risk of cardiovascular diseases, thanks to a plenty of compounds playing a role of great significance from the aspect of human health. However, non-alcoholic and low alcoholic fermented beverages such as wines and beers have become of great interest because they offer traditional beer and wine flavours without certain unhealthy and socially objectionable side effects of alcohol. Furthermore, they have a lower caloric content, and the same intake of natural antidotes of cardiovascular diseases, anthocyanins and phenolic compounds (Takács *et al*., 2007). The technologies for the production of dealcoholized, low- and reduced-alcohol beverages have progressively improved in order to safe and get better their quality. There are several methods disclosed in the art for removing or reducing alcohol in wines. However, each process has its advantages and disadvantages, in terms of process costs and product quality.

Various processes based on the use of high temperature like evaporation and cryoconcentration, employed for the concentration of fruit juices, cause strong alteration or loss of the wine aroma. Wine vacuum distillation promotes a dealcoholized wine with low volatile compounds that could be used for blending with other wines. Only the recovery of the lost volatile fraction during the process by means of distillation will give a desired organoleptic product (Gomez-Plaza *et al*., 1999). Another dealcoholization technique is based on spinning cone column. It is operated at mild operation temperatures (26-35°C) and takes places in two steps: aroma recovery and ethanol removal. After ethanol separation, the aromatic fraction is added back to the wine. It is a long and expensive operation (Diban *et al*., 2008).

Other technologies such as adsorption on zeolites (Diban *et al.*, 2008) and supercritical fluid extraction (Medina *et al*., 1997) are studied in the literature as possible alternatives to reduce the alcoholic content in beverages, but these both methods have some disadvantages and/or are too much onerous like for example for supercritical fluid

extraction or another they change the composition of wine. The membrane processes can be also utilized for the dealcoholization and they allow the ethanol content to be reduced under mild conditions, so the organoleptic features might remain unchanged. There are some experimental works on wine dealcoholization by pervaporation (Takács *et al*., 2007), reverse osmosis (Pilipovik *et al*., 2005; Labanda *et al*., 2009; Catarino *et al*., 2006) and osmotic distillation (Varavuth *et al.*, 2009; Diban *et al.*, 2008). Direct osmosis finds its application in the concentration of fruit juices, vegetable juices, skim milk, coffee extract, etc. and it can be also used for the concentration of pharmaceutical products and beverages dealcoholization. In earlier years, direct osmosis process could not be exploited commercially because of low flux due to thick membranes. With the advent of thin membranes in recent years resulted in increased flux and hence the direct osmosis process has been gaining the importance due to the concentration of heat sensitive liquid foods/natural colours (Ravindra Babu *et al*., 2006). It will be interesting to develop an easy, rapid and profitable plant for beverages dealcoholization based on direct osmosis. The objective of this work was to study the feasibility of applying a pilot plant working on direct osmosis using a membrane contactor to dealcoholize wine in order to produce a non alcoholic beverage similar to the original wine.

2. Experimental

The dealcoholization tests were carried out in a pilot plant equipped with a polypropylene hollow fiber membrane contactor (Liqui-Cel, Extra-Flow 4x28, Celgard X50). Inside the module, hollow fibers are wrapped around a central tube; a baffle for the liquid phase is created in the middle of the contactor so maximizes surface area and improves liquid transfer efficiency. Feed stream (wine) enters the shellside port and travels into the distribution tube. Wine is forced radially over the fibers on each side of the baffle and it exits through the collection tube and the second shellside port. Pure water is introduced to the other side port and it is gradually enriched by volatile compounds, especially ethanol, due to their different concentration in the two streams. An overview on the fluid dynamics of the module is shown in fig.1.
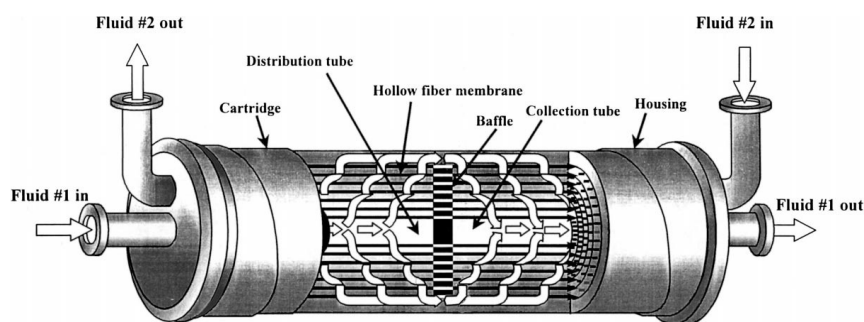
Figure 1: Liqui-Cel Extra-Flow 4x28 Module

From theoretical calculations performed on the module's exchange characteristics and based on operational considerations, the dealcoholization process consists of five cycles: the first two lasting 45 minutes, the others 30 minutes. Pure water flows in counter-current and temperature is monitored during all the cycles' process. At the end of every cycle, samples were taken from each of the two streams and were evaluated for chemical-physical and aromatic compositions.

Alcohol content (% vol.) was evaluated according to AOAC methods (1995). On dealcoholized wine samples pH was determined through the use of pHmeter (Hanna Instrument, mod.122); instead, total acidity determinations, expressed in equivalent of tartaric acid content (g/L), were carried out by titrating 25 ml sample with 0.1N NaOH to a pH endpoint of 7.0. The reducing sugars content was measured with Fehling method, according to AOAC (1995). The total phenols amount in Aglianico wine and in dealcoholized wine samples was determined according to the Folin-Ciocalteu colorimetric method (Singleton et Rossi, 1965). Absorbances were measured at 765 nm, using Perkin Elmer UV/VIS Spectrometer Lambda Bio 40. Total phenols were expressed as gallic acid equivalents (GAE mg/L). Gallic acid standard solutions were prepared at a concentration ranging from 100 to 400 mg/L. The colour measurements of Aglianico and dealcoholized wine samples were carried out using the Glories methods (1986). Colour density and hue were calculated using the following equations (eqn. nr. 1-2) which incorporated corrected values for a 1 cm cuvette, as reported in a previous paper (Cliff M.A. et al.,2007);

$Colour\ density = [(A_{520}-A_{700}) + (A_{420}- A_{700})]$ (Eqn. nr.1)
$Colour\ hue/tint = [(A_{420}-A_{700}) + (A_{520}- A_{700})]$ (Eqn. nr.2)

Another method for colour evaluation utilized L*, a* b* coordinate values in the Hunter system for each wine sample with a CR-300 Chromometer.

The identification and determination of aroma compounds concentration in Aglianico wine and dehalcoholizated samples were made respectively by GC coupled to a mass spectrometer (Trace MS plus, TermoFinnigan, USA) and by GC (HP 6890, Agilent) both equipped with a capillary column (Equity 5, 300 mm x 0.25 mm x 0.25 μm, Supelco). The extraction of the aroma compounds was made according to Cocito et al, 1995. The gascromatografy conditions were: injector worked in split mode and was set at 250°C, one microlitre (1μl) was the injected volume, helium was used as carrier gas at a constant flow of 1.0 ml. The temperature program in the oven was 5 min at 50°C, an increase until to 200°C at a rate of 2°C/min and finally 5 min at 200°C. Mass detector conditions were: electronic impact (EI) mode at 70 eV, source temperature 280°C, scanning rate 1 scan/s, mass acquisition 30-350 amu. The identification was based on comparison of the GC retention times and mass spectra with authentic standards from Sigma-Aldrich when standards were available; for these compounds, calibration curves were calculated in order to the quantification. When the authentic standards were not available, the identification was based on comparison with the spectral data of the Wiley library and the chromathographic data from literature; semi-quantitative analyses of these compounds were done assuming the response factors equal to the 2-octanol, used as internal standard.

3. Results and discussion

The Aglianico wine ethanol content was 12.80% (v/v) and it decreased during the dealcoholization process, in such amount between 62.10 % and 83.67 % in the first three cycles, then alcohol amount reduced gradually in the last cycles until to achieve an alcohol content less than 0,5% (as shown in table 1). pH, total acidity and also reducing sugars content showed very slight variations during the dealcoholization, as reported in table 2. Total phenols amount was high, typical of red wines (Minussi R.C. et al., 2003) and it was nearly unchanged, with a slightly increase at the end of the dealcoholization, probably because of the concentration effect produced for removal of ethanol from the

corresponding red wine, according to research carried out by Belisario-Sanchez YY. et al. (2009). Colour density was similar to Versari et al. (2007) varying around 7 during the dealcoholization process, with a slight decrease in the final product. Tonality values were roughly similar in all cycles and testified the mellow state of wine (table 2). About chromatic parameters, dealcoholized wine with 0.42% alcohol content showed brightness value (L*) higher than Aglianico wine and a degree of redness similar to the original wine. The degree of redness (a*) ranged from 20.23 to 29.07 as in other studies about red wines (Cliff M.A. et al., 2007), like as b* values (table 2).

Table 1: Alcohol content (%v/v), alcohol loss (%), pH, total acidity (mean values ± standard deviation), reducing sugars and total phenols content of wine samples during the dealcoholization process. V_0 is starting wine (Aglianico); V_1, V_2, V_3, V_4, V_5 are wine samples respectively at the end of every cycle.

Samples	Alcohol content (% vol.)	Alcohol loss (%)	pH	T (°C)	Total acidity (tartaric acid g/L)	Reducing sugar (g/100 ml)	Total phenols (Gallic acid mg/L)
V_0	12,8		3,49	20,0	5,59 ± 0,18	0,60	3381,75
V_1	4,85	62,10	3,38	20,1	5,64 ±0,04	0,59	3616,75
V_2	3,50	72,65	3,34	20,0	5,80 ±0,04	0,57	3724,25
V_3	2,09	83,67	3,24	20,1	5,68 ± 0,04	0,54	3604,25
V_4	0,75	94,14	3,21	20,1	5,78 ± 0,03	0,56	3919,25
V_5	0,42	95,71	3,20	20,1	5,64 ± 0,02	0,57	4145,50

Table 2: Chromatic Glories and CIE Lab parameters (mean values ± standard deviation) of dealcoholized wine samples.

Samples	Colour density	Tonality	L*	a*	b*
V_0	7,61 ± 0,04	0,71 ± 0,01	14,82 ± 0,77	27,92 ± 1,52	5,68 ± 1,18
V_1	7,65 ± 0,12	0,69 ± 0,01	14,98 ± 0,04	29,07 ± 0,24	8,51 ± 0,10
V_2	7,73 ± 0,18	0,55 ± 0,02	17,96 ± 0,14	20,23 ± 0,6	2,15 ± 0,45
V_3	7,50 ± 0,16	0,53 ± 0,01	17,91 ± 0,06	20,84 ± 1,16	1,87 ± 0,39
V_4	7,54 ± 0,09	0,63 ± 0,01	19,49 ± 0,12	24,92 ± 0,17	5,24 ± 0,46
V_5	7,58 ± 0,01	0,63 ± 0,00	21,92 ± 0,04	27,26 ± 0,30	9,60 ± 0,19

In order to evaluate the sensory impact owing to the aroma compounds loss by the dealcoholization process, the odour activity values (OAV) of the most significant compounds were evaluated as ratio between aromatic compounds concentration and corresponding olfactory threshold. The OAV allows to estimate the contribution of specific compound to the wine aroma. Each compound was assigned to one or several aroma series, depending on its principal odour descriptors; the solvent, floral, sweet, green, fatty, fruit and balsamic series were chosen for this purpose.

Aglianico wine dealcoholization tests

In figure 2 it was shown the odour activity values for each dealcoholizated cycle. Sweet and solvent aroma series suffered major changes, and V_4 and V_5 dealcoholizated samples were characterized by only solvent aroma serie (Genovese et al., 2007).
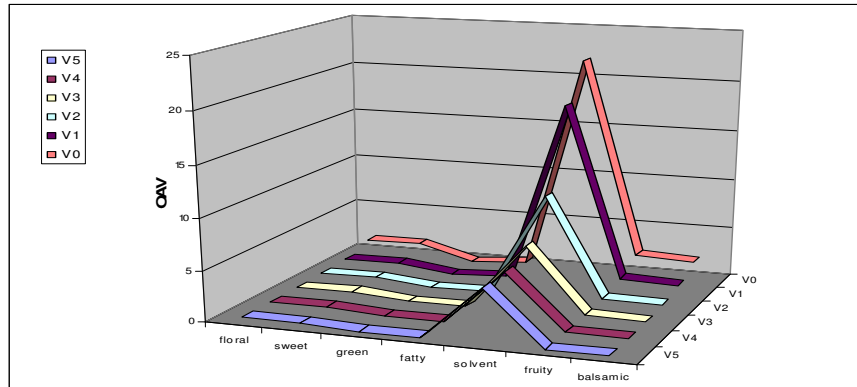

Figure 2: Odour activity values of odorant series of dealcoholized wine samples.

4. Conclusions

The dealcoholization plant working on direct osmosis resulted efficient for removing alcohol from wine until to ethanol concentration lower than 0.5% (v/v). The final dealcoholized product has a chemical composition similar to wine of origin in terms of reducing sugars, total acidity and is rich in phenolic substances, but it is lacking in aroma. So, this should be an obstacle for the development of non alcoholic beverages, which had been reconstituted in the wine aroma for a better tasting drink.

These results show that the membrane technology based on direct osmosis and used to separate the ethanol from wine keeps or increases the amount of beneficial compounds in the dealcoholized wine. It also can be used for a reduction of alcohol degree in wine industry.

References

AOAC, 1995, Official methods of analysis of AOAC International, 16th ed., Arlington

Y.Y.Belisario-Sanchez, A. Taboada-Rodriguez, f. Marin-Iniesta, A. Lopez.gomez, 2009, Dealcoholized wines by spinning cone column distillation: phenolic compounds and antioxidant activity measured by the 1,1-diphenyl-2-picrylhydrazyl method, Journal of Agriculture and food Chemistry, 57, 6770-6778

M. Catarino, A. Mendesa, L. Madeira, A. Ferreira, 2006, Beer dealcoholization by reverse osmosis, Desalination, 200, 397–399

CIE, 1986, Colorimetry, 2nd ed., Vienna: Central Bureau of the Commission Internationale de L'Eclairage

M.A. Cliff, M.C. King, J. Schlosser, 2007, Anthocyanin, phenolic composition, colour measurements and sensory analysis of BC commercial red wines, Food Research International, 40, 92-100

C. Cocito, G. Gaetano, C. Delfino, 1995, Rapid extraction of aroma compounds in must and wine by means of ultrasound, Food Chemistry, 52, 311-320

N. Diban, V. Athes , M. Bes, I. Souchon, 2008, Ethanol and aroma compounds transfer study for partial dealcoholization of wine using membrane contactor, Journal of Membrane Science 311, 136-146

A.Genovese, A.Gambuti, P. Piombino, L. Moio, 2007, Sensory properties and aroma compounds of sweet Fiano wine, Food Chemistry, 103, 1228-1236

Y. Glories, 1984, La couleur des Vins rouges, 2a partie, Connaissance de la Vigne et du vin, 18, 253-271

E. Gómez-Plaza, J.M. López-Nicolás, J.M. López-Roca, A. Martinez-Cutillas, 1999, Dealcoholization of Wine Behaviour of the Aroma Components during the Process, Lebensm.-Wiss. u.-Technol., 32, 384-386

J. Labanda, S. Vichi, J. Llorens, E. Lopez-Tamames, 2009, Membrane separation technology for the reduction of alcoholic degree of a white model wine, Food Science and Technology, article in press.

I. Medina, J. L. Martinez, 1997, Dealcoholisation of Cider by Supercritical Extraction with Carbon Dioxide, Journal of Chemical Technology & Biotechnology, 68, 14-18

R.C. Minussi, M. Rossi, L. Bologna, L.Cordi, D. Rotilio, G.M. Pastore, N. Duran, 2003, Phenolic compounds and total antioxidant potential of commercial wines, Food Chemistry, 82, 409–416

M.V. Pilipovik, C. Riverol, 2005, Assessing dealcoholization systems based on reverse osmosis, Journal of Food Engineering, 69, 437–441

B. Ravindra Babu, N.K. Rastogi, K.S.M.S. Raghavarao, 2006, Effect of process parameters on transmembrane flux during direct osmosis, Journal of Membrane Science, 280, 185–194

V.L. Singleton, J.A. Rossi, 1965, Colorimetry of total phenolics with phosphomolybic-phosphotungstic reagents, American Journal of Enology and Viticulture, 16, 144-158

L.Takács, G. Vataia, K. Korány, 2007, Production of alcohol free wine by pervaporation, Journal of Food Engineering, 78, 118–125

S. Varavuth, R. Jiraratananon , S. Atchariyawut , 2009, Experimental study on dealcoholization of wine by osmotic distillation process, Separation and Purification Technology, 66, 313–321

A. Versari, G.P. Parpinello, A.U. Mattioli, 2007, Characterisation of Colour Components and Polymeric Pigments of Commercial Red Wines by Using Selected UV-Vis Spectrophotometric Methods, South African Journal of Enology and Viticulture, 28, 1, 6-10

Competition of two microbial species in a turbidostat

Andrea Cammarota[a], Michele Miccio[a]

[a]*Dipartimento di Ingegneria Chimica e Alimentare, Università degli Studi di Salerno, Via Ponte Don Melillo, 84084 Fisciano (SA), Italy mmiccio@unisa.it*

Abstract

The competition between two microbial species in a chemostat definitely leads to the disappearance of one of them (Smith and Waltman, 1995). In this paper, a turbidostat is taken into account, i.e., a continuous bioreactor in which the microbial concentration is controlled by manipulating the feed rate, and the competition between two microbial species is modelled by means of Mathematica® and AUTO97 simulation codes. The growth of the two populations is supposed to be described by a simple unstructured model, the growth rate being a function of the limiting substrate concentration for both species. It is shown that there is a range of the set point parameters for which the continuous bioreactor has a steady state in which the microbial species can coexist. The stability of such solutions depends on both the growth rate and the microbial yield.
If the growth kinetics are monotonic with substrate concentration (such as the Monod-like one), no periodic solutions can occur. On the contrary, if the growth rates are not monotonic, the bioreactor can show periodic regimes, too. Such regimes can involve either oscillations of the concentration of a single species (while the other is constantly equal to zero) or oscillations of the concentrations of both the populations.

Keywords: Turbidostat, Mathematical Model, Simulation, Dynamical Analysis.

1. Introduction

The competition among different microbial species in chemostat is a widely investigated process both from experimental (Balagaddé et al., 2005) and mathematical (Smith and Waltman, 1995) standpoint. The general result is that in a chemostat hosting more microbial cultures competing for the same limiting substrate, the only possible stable steady-states are those in which a single species survives, i.e., the one having the fastest growth rate (Smith and Waltman, 1995). Furthermore, the yield coefficients are proved to have no influence on the selection mechanism inside the bioreactor. Such a result has been proven under quite mild hypotheses on kinetic expressions (which is only needed to be C^1 functions $\mu_i(S)$ of the real nonnegative variable S, the limiting substrate concentration, for every i-th species involved in the process). In particular, when two species are supposed to compete, if the growth law of the first is greater than the one of the second for every value of $S \in \,]0;S_0]$ (where S_0 is the substrate input concentration) then the first species is the only one which survives and proliferate for every value of the dilution rate. On the contrary, if the set $]0;S_0]$ can be parted into intervals in each of which the kinetic law of one of the species is greater than the other, then the only surviving population is the one with the highest growth rate in correspondence of the new steady state substrate concentration: in this case, the dilution rate, which is the main operating variable, can be accurately chosen in order to cause the survival of a pre-selected population and the simultaneous disappearance of the other. The only case in which a coexistence of two species may establish is when the dilution

rate assumes a special value for which the substrate concentration is such that $\mu_1(S^*)= \mu_2(S^*)$. Obviously, provided that the growth laws are not equal in a whole neighbourhood of S^*, an infinitesimal disturbance on the dilution rate disrupts such a coexistence.

No systematical studies are in literature on the microbial competition in a turbidostat except for de Leenheer et al. (2003), in which the control law is assumed to be proportional. In this paper, we determine the conditions that guarantee the coexistence of two different microbial species as a function of the main operating variables in the turbidostat: the set point for biomass concentration and the inlet substrate concentration.

2. The model

In order to describe the competition between two microbial species for a common limiting substrate S, we assumed a simple, non-structured, non-segregated kinetic expression for each species. As a first assumption, we suppose that the biomass growth rates $\mu_1(S)$ and $\mu_2(S)$ are both C^1 class nonnegative functions defined on R_0^+ with $\mu_1(0)= \mu_2(0)=0$ and $\mu_1'(0)>0$, $\mu_2'(0)>0$. Furthermore, it is assumed that S=0 is the only zero of both $\mu_1(S)$ and $\mu_2(S)$ in their domain. A last assumption concerning the kinetic laws is that if their plots intersect at a point $S=S_{int}$, they are not tangent in this point (i.e., $\mu_1(S_{int})=\mu_2(S_{int}) \Rightarrow \mu_1'(S_{int}) \neq \mu_2'(S_{int})$).

The dilution rate is the manipulated variable and the biomass concentration is the controlled one. Actually, the concentration of the biomass is detected using a turbidimetric probe; in order to consider the different contribution of the two species to the turbidity of the solution, we define a constant α which is the ratio between the turbidity produced by an unit concentration of species 1 and the one determined by an unit concentration of species 2 (as proposed by de Leenheer et al., 2003). The controller operates according to a traditional PI algorithm described by the following relationship

$$\text{Dil} = \text{Dil}_{sp} + K_p \left(X_1 + \alpha X_2 - X_{sp} + \frac{1}{T_I} \int_0^t (X_1 + \alpha X_2 - X_{sp}) d\theta \right)$$

(1.)

where X_1 and X_2 the biomass concentration of the two species, Dil the dilution rate and X_{sp}, K_p and T_I the feedback controller parameters (set point referred to the biomass concentration of the species 1, proportional gain and integral time). Note that this control algorithm is valid until Dil$\geq$0 as the option Dil<0 in physically meaningless.

In order to study the system dynamics, it is useful to write it in a dimensionless form

$$\begin{cases} \dfrac{dx_1}{d\theta} = (m_1(s) - dil)x_1 \\[2mm] \dfrac{dx_2}{d\theta} = (m_2(s) - dil)x_2 \\[2mm] \dfrac{ds}{d\theta} = dil(s_0 - s) - m_1(s)x_1 - \dfrac{m_2(s)}{r_y}x_2 \\[2mm] \dfrac{d(dil)}{d\theta} = \begin{cases} 0 \quad \text{if} \left(dil = 0 \text{ and } k_p \left[\dfrac{dx_1}{d\theta} + \alpha \dfrac{dx_2}{d\theta} + \dfrac{1}{\tau_I}(x_1 + \alpha x_2 - x_{sp}) \right] \leq 0 \right) \\[3mm] k_p \left[\dfrac{dx_1}{d\theta} + \alpha \dfrac{dx_2}{d\theta} + \dfrac{1}{\tau_I}(x_1 + \alpha x_2 - x_{sp}) \right] \quad \text{otherwise} \end{cases} \end{cases}$$

(2.)

Given a reference substrate concentration S_r and a reference time t_r, the dimensionless quantities are defined as

$$s = \frac{S}{S_r} \quad dil = Dil \cdot t_r \quad \theta = \frac{t}{t_r} \quad k_p = \frac{K_p Y_1 S_r}{t_r^2} \quad \tau_I = \frac{T_I}{t_r} \quad r_y = \frac{Y_2}{Y_1} \quad x_n = \frac{X_n}{Y_1 S_r} \quad m_n(s) = \mu_n(S_r s) \cdot t_r \quad n = 1,2$$

where Y_1 and Y_2 the biomass yields for the two species.

It is a four-dimensional dynamical system whose state vector is (x_1, x_2, s, dil). It can be easily noticed that both the hyperplanes $x_1=0$ and $x_2=0$ are invariant (this corresponds to the physical fact that if a species is not present either in the bioreactor or in the feeding stream, its concentration cannot be nonzero): hence, if a given orbit is a regime for a turbidostat in which either species 1 or 2 is present, it is a regime solution also for the above system, but the stability properties can be different from the ones computed for a single-species turbidostat.

3. Occurrence and stability of the steady states

The physically meaningful region of the state variables, described by

$$A = \left\{ (x_1, x_2, s, dil) \in \mathbb{R}^4 \mid x_1 \geq 0, \ x_2 \geq 0, \quad s \geq 0, \ x_1 + x_2 r_y + s \leq s_0, \quad dil \geq 0 \right\}$$

can be easily proved to be positively invariant; therefore, we can restrict our attention to this subset of the phase space. Furthermore, the parameters x_{sp}, k_p, s_0 and τ_I are assumed to be strictly positive. We restrict our attention to the case in which the controller does not get the windup condition. In order to compute the steady states and the singularities for this system, it is useful to perform the Lyapunov-Schmidt reduction of the equations of the system performed computing the Groebner basis (using the symbolic tools of the software Mathematica®) of the components of the vector field of the dynamical system:

$$L\left(s, x_{sp}, k_p, \tau_I\right) = \begin{cases} \dfrac{k_p}{\tau_I} \, x_{sp} \, \left[x_{sp} - (s_0 - s)\right]\left[x_{sp} - \alpha \, r_y \, (s_0 - s)\right] \left[m_1(s) - m_2(s)\right] m_1(s) \, m_2(s) & \text{if } \alpha r_y \neq 1 \\ x_{sp} \, \left[x_{sp} - (s_0 - s)\right] m_1(s) \, m_2(s) & \text{if } \alpha r_y = 1 \end{cases}$$

In this paper it is assumed that $\alpha r_y \neq 1$; in fact, if it were $\alpha r_y = 1$, it would be virtually impossible to control the populations of the two competing species separately using only the turbidimetric probe. A special case in which such a condition occurs is when the two species are actually the same one ($\alpha=1$ and $r_y=1$).

The steady states can be easily computed, as they are the zeroes of the function L: in particular, the value of the steady-state substrate concentration can be obtained as a function of the parameter x_{sp}. L is the product of five functions and it can be zero only when one of them is zero. Nevertheless, the last two functions $m_1(s)$ and $m_2(s)$ are zero only when s=0 according to the hypotheses listed in the previous paragraph. Such a case occurs only when dil=0 (as it can be easily obtained by the third equation of the dynamical system) and this implies that the bioreactor operates in a batch mode (and, in this case, the reactor is no longer a turbidostat). Hence, the stationary solutions for the system can be obtained when

a. $x_{sp} = (s_0 - s)$: the value of the state variables for this set of steady states can be easily obtained solving the algebraic system given by the vector field set to the null vector and substituting $s_1 = s_0 - x_{sp}$ to every occurrence of s. The results are $x_1 = x_{sp} = s_0 - s_1$, $x_2 = 0$, dil$=m_1(s_1)$ (obviously x_{sp} can be set to any value in the interval $]0; s_0[$). Such solutions indicate that the species 1 is the only one surviving and all the consumed substrate is used for the growth of this microorganism. These steady states belong to the invariant hyperplane described by $x_2=0$. It is interesting to analyse the stability of such regime solutions: in order to discuss this point, the characteristic polynomial for the Jacobian of the dynamical system in these points can be computed. The result can be factored is the following one

$$p_1(\lambda) = \left(\lambda + m_1(s_1)\right)\left(\lambda + m_1(s_1) - m_2(s_1)\right)\left(\lambda^2 + x_{sp}\left(k_p + m_1{}'(s_1)\right)\lambda + \frac{k_p x_{sp}}{\tau_I}\right)$$

The eigenvalue computed from the first factor is always negative; the one computed from the second factor is negative only when $m_1(s_1) > m_2(s_1)$: this means that such steady states can be stable only if the growth rate of the 1st species is greater than the competing one. The third factor has an always positive constant term, so its roots have either positive real parts when $k_p + m_1{}'(s_1) > 0$ or negative real parts if $k_p + m_1{}'(s_1) < 0$. If there are values s_1 which make $k_p + m_1{}'(s_1) = 0$, they are Hopf bifurcation points. As k_p is supposed to be positive, in order to have Hopf bifurcations $m_1(s)$ must have a negative first derivative in its domain: this cannot be true for a Monod kinetics which is strictly monotonic; however, it is possible if, for example, there is a substrate inhibition effect. Hence, the presence of Hopf bifurcations as well as the existence of periodic regimes is connected with both the non monotonicity of the growth law and the value chosen for the proportional gain.

b. $x_{sp} = \alpha r_y(s_0 - s)$: the value of the state variables for this set of steady states can be easily obtained solving the algebraic system given by the vector field set to the null vector and substituting $s_2 = s_0 - x_{sp}/\alpha r_y$ to every occurrence of s. The results are $x_2 = x_{sp}/\alpha = (s_0 - s_2)r_y$, $x_1 = 0$, $dil = m_2(s_2)$ (obviously x_{sp} can be set to any value in the interval $]0; s_0 \, \alpha \, r_y[$). Such solutions indicate that the species 2 is the only one surviving and all the consumed substrate is used for the growth of this microorganism. These steady states belong to the invariant hyperplane described by $x_2 = 0$. Such solutions are symmetric to the previous ones, even though such a symmetry is not perfect because of the different biomass yield and the different turbidity of the two species (i.e., $\alpha r_y \neq 1$). The analysis of the stability of such regime solutions can be performed using the same procedure of the previous case and the results are similar. A first eigenvalue of the Jacobian is $-m_2(s)$ that is always negative (provided that $s > 0$); a second one is negative when $m_1(s_1) < m_2(s_1)$: again, such regimes can be stable only if the growth rate of this 2nd species is greater than the competing one. The last two eigenvalues have positive real parts if $r_y\alpha k_p + m_2{}'(s_2) < 0$ or negative real parts if $r_y\alpha \, k_p + m_2{}'(s_2) > 0$. The values of s_2 which make $r_y\alpha \, k_p + m_2{}'(s_2) = 0$ are Hopf bifurcation points, which can occur only if $m_2(s)$ have a negative first derivative for some values of its domain.

c. $m_1(s_3) = m_2(s_3)$: in order to have a set of steady state solutions corresponding to this factor of the Lyapunov–Schmidt function, there should be at least one intersection point for the two kinetic laws different from $s = 0$. Let us suppose that $0 < s_3 < s_0$ is a value for which $m_1(s_3) = m_2(s_3)$. The other variables for this set of steady states can be easily obtained solving the algebraic system given by the vector field set to the null vector and substituting $m_1(s_3)$ to any occurrence of $m_2(s_3)$. The result is $dil = m_1(s_3)$, $x_1 = [x_{sp} - \alpha r_y(s_0 - s_3)]/(1 - \alpha r_y)$, $x_2 = [r_y(s_0 - s_3 - x_{sp})]/(1 - \alpha r_y)$. In order to have physically meaningful solutions, both x_1 and x_2 must be nonnegative; this occurs if $x_{sp} \in [\min\{s_0 - s_3, \, \alpha r_y(s_0 - s_3)\}, \, \max\{s_0 - s_3, \, \alpha r_y(s_0 - s_3)\}]$: the larger the difference between 1 and αr_y, the wider the interval is (it actually collapses in a point when $\alpha r_y = 1$, but this case is not considered in this paper). These regimes have nonzero concentration for both the species for every value of the set point belonging to the specified interval (except for its extremes): more precisely, the concentration of both x_1 and x_2 changes linearly as a function of x_{sp} (in fact x_1 grows from 0 to $s_0 - s_3$ and x_2 decreases from $r_y(s_0 - s_3)$ to 0 when the set point varies from an extreme to the other of the interval). The stability of these steady-state points cannot be analysed as easily as in the cases a. and b. because of the complex

mathematical expression of the characteristic polynomial of the Jacobian. Anyway, some considerations can be made by studying the sign of the Jacobian determinant (in fact, a necessary condition for the asymptotic stability of a steady-state is that the Jacobian determinant evaluated in this point must be positive).

$$\det J(s_3) = \frac{k_p}{\tau r_y} x_1 x_2 [m_1'(s_3) - m_2'(s_3)](\alpha r_y - 1) \qquad (3.)$$

Hence, this regime set can be stable only if either $m_1'(s_3) > m_2'(s_3)$ and $\alpha r_y > 1$ or $m_1'(s_3) < m_2'(s_3)$ and $\alpha r_y < 1$. This mathematical result lends itself to a physico-chemical explanation, e.g., the steady state is stable if the microbial species showing the higher rate of change of its growing function produces at the same time a lower degree of turbidity per unit consumed substrate.

The singularity points in the set of steady-states are the zeroes of both L and its first partial derivative toward the state variable s: it can be easily proved that such points are the intersections between each of the algebraic manifold defined in a. and b. and the one defined in c: it is also easy to prove that such singularities are transcritical, provided that the non-degeneracy conditions $1 \neq \alpha r_y$, $m_1'(s_3) \neq m_2'(s_3)$, $s_3 \neq s_0$ and $s_3 \neq 0$ are satisfied.

4. Some numerical examples

4.1. Monod growth kinetics

The case in which a Monod growth rate is valid for both the competing species is discussed first. Assuming that the reference substrate concentration and the reference time are respectively the Monod constant and the reciprocal of the maximum growth rate for the first species, the dimensionless kinetic laws are

$$m_1(s) = \frac{s}{1+s} \quad m_2(s) = \frac{r_{\mu max} \cdot s}{r_M + s} \qquad (4.)$$

where r_M is the ratio between the Monod constants and $r_{\mu max}$ is the one between the maximum growth rates of the two species. In the two plots in Fig. 1, the solution diagrams for the state variable s as a function of x_{sp} are reported for the cases:

a) $s_0 = 3$, $r_{\mu max} = 1.5$, $r_M = 2$, $r_y = 4$, $\alpha = 1.5$, $k_p = 1$, $\tau_I = 1$ ($s_3 = 1$, $m_1'(s_3) < m_2'(s_3)$ and $\alpha r_y > 1$)

b) $s_0 = 3$, $r_{\mu max} = 0.8$, $r_M = 0.6$, $r_y = 4$, $\alpha = 1.5$, $k_p = 1$, $\tau_I = 1$ ($s_3 = 1$, $m_1'(s_3) > m_2'(s_3)$ and $\alpha r_y > 1$)
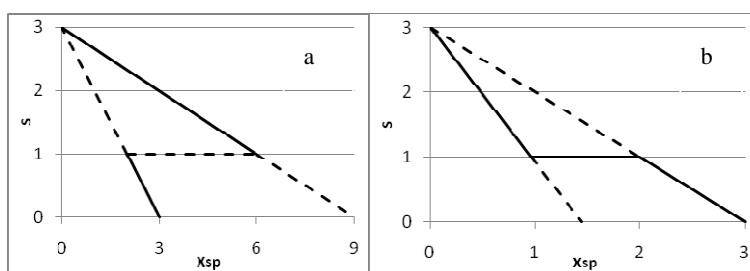

Figure 1: Solution diagrams for a turbidostat with two competing species ($s_0 = 3$, $r_y = 4$, $\alpha = 1.5$, $k_p = 1$, $\tau_I = 1$):   a) $r_{\mu max} = 1.5$, $r_M = 2.0$;   b) $r_{\mu max} = 0.8$, $r_M = 0.6$

4.2. Kinetics with substrate inhibition

If there is a substrate inhibition effect, the kinetic law is not monotonic. In this paper it is supposed that the Haldane kinetics (Bailey and Ollis, 1986) describes the growth rate

of both the competing species. Using a technique similar to the one of the previous subparagraph, the dimensionless form of the kinetic expressions can be computed

$$m_1(s) = \frac{s}{1 + s + k_{I1}s^2} \quad m_2(s) = \frac{r_{\mu\,max} \cdot s}{r_M + s + k_{I2}s^2} \tag{5.}$$

As proved in the paragraph 3, when the proportional gain is low enough, Hopf bifurcations may take place and, therefore, periodic regimes may establish. These bifurcations can occur in one of the invariant hyperplanes ($x_1=0$ and $x_2=0$) and, hence, determine the birth of limit cycles involving the oscillation of the concentration of a single population (the other one being 0). Anyway, it can be shown that also periodic regimes involving both species are possible. In Fig. 2a a solution diagram of the state variable *s* as a function of the parameter x_{sp} is reported. It was obtained using the parametric continuation software AUTO97 for $s_0=3.0$, $r_y=0.4$, $\alpha=0.5$, $r_M=5.0$, $r_{\mu max}=1.25$, $k_{I1}=12.0$, $k_{I2}=0.5$, $k_p=0.01$ and $\tau_I=1.0$. A part of the periodic solutions, plotted with black dots, involves only species 1, while the ones plotted with grey dots involve both of them. These two periodic branches intersect in a transcritical bifurcation point for the limit cycles (for a better readability, the unstable periodic regimes are not plotted).

In Fig. 2b a phase portrait in the plane spanned by x_1 and x_2 of a stable oscillating regime for both species is reported for the same set of the above parameters and $x_{sp}=1.3$.
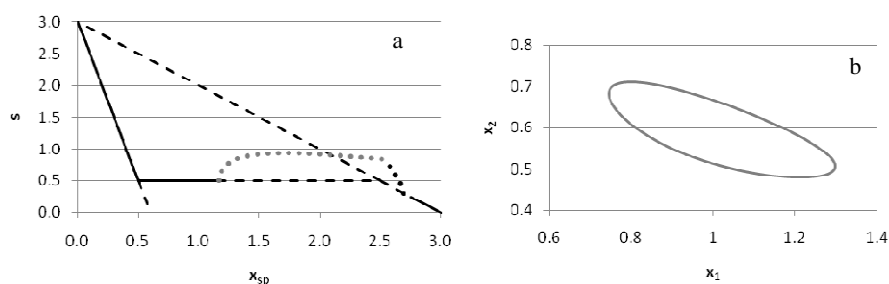

Figure 2: a) Solution diagram of the state variable s as a function of the controller parameter x_{sp} (only the stable periodic regimes are reported) for $s_0=3.0$, $r_y=0.4$, $\alpha=0.5$, $r_M=5.0$, $r_{\mu max}=1.25$, $k_{I1}=12.0$, $k_{I2}=0.5$, $k_p=0.01$ and $\tau_I=1.0$
b) Phase portrait showing the limit cycle obtained for $x_{sp}=1.3$. The period is 49.6.

5. Conclusions

While in a chemostat two different microbial species competing for the same limiting substrate cannot coexist (i.e., the set of the dilution rates which allows such a coexistence has zero as a measure), in a turbidostat such a coexistence is possible, provided that a substrate concentration exists such that the growth rates for the two species are equal and certain simple stability conditions are satisfied. Further, periodic regimes that involve either a single species or both ones may occur simply if the growth kinetics is a non-monotonic function of the substrate concentration.

References

Bailey G.E. and Ollis D.F. (1986) Biochemical Engineering Fondamentals. MacGraw – Hill
Balagaddé F.K., You L., Hansen C.L., Arnold F.H. and Quake S.R. (2005) *Science*, 309, 137-140
De Leenheer, Li, Smith (2003) *Canad. App. Math. Quart.*, 3, 229
Smith M. and Waltman P. (1995) The theory of the chemostat: Dynamics of microbial competition. Cambridge University Press.

A Combination of Parallel Computing and Object-Oriented Programming to Improve Optimizer Robustness and Efficiency

Guido Buzzi-Ferraris, Flavio Manenti

Dipartimento di Chimica, Materiali e Ingegneria Chimica "Giulio Natta"
Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano, ITALY
E-mail: guido.buzziferraris@polimi.it

Abstract

This research activity is mainly aimed at showing potentialities in coupling object-oriented programming with parallel computing. Wide margins of benefits could be obtained in algorithm efficiency and robustness with a relative small programming effort. The case of unconstrained multi-dimensional optimization is proposed as quantitative example.

Keywords: Robust optimization, Multimodality, Narrow valleys, Discontinuous functions, Parallel Computing.

1. Introduction

We are undergoing two silent revolutions that directly involve Process Systems Engineering (PSE) and Computer-Aided Process Engineering (CAPE) communities, besides many other scientific and industrial areas: the object-oriented programming and the parallel computing on personal computer. Both these transformations have been widely discussed in the literature as they significantly modify the numerical analysis as it was conceived since the second part of the previous century and the way to apply numerical methods and algorithms for solving more and more complex problems and multifaceted issues.

Nevertheless, since it is not clearly stated in the current literature, it is worth remarking that the parallel computing is easy to integrate in object-oriented programming and their combination seems particularly appealing as many objects generated by the same class might run simultaneously on different processors or cluster nodes. By thinking parallel and object-oriented both together, it is possible to write by new many algorithms which were not considered for solving numerical problems because of their reduced performances in the procedural programming and sequential computing.

Thus, this research activity specifically deals with the development of very robust optimizers that exploit:

- All features of object-oriented programming (Buzzi-Ferraris, 1994), which allow going beyond the procedural programming and its limitations.
- The shared memory nowadays commonly available on multi-processor machines (distributed memory machines are not considered for the time being, even though the same reasoning here described can be extended to this branch of parallel computing).

Motivation and practical interests in some scientific and industrial areas are briefly reported in Paragraph 2. Basic concepts of coupling parallel computing with object-

oriented programming for improving the optimizer robustness and efficiency are stated in Paragraph 3. Some literature tests involving multidimensionality, strong and weak multimodality, very narrow valleys, and functions that are undefined in some regions are proposed in Paragraph 4.

2. Motivation and Practical Interests

Looking at the increasing spread of multi-processor machines as well as the larger and larger amount of processors available on the common PCs, it is easy to see how the period we are living is very similar to the one of 1970s when the most powerful (and very large-size) machines were gradually replaced by smallest personal computers with reduced computational power, but with a large impact on research activities for their faster spread, reduced costs, reasonably good performances, and especially for their higher slope in innovation. In our opinion, but it is easy to find out some confirmations yet, shared memory machines will have a faster evolution than distributed memory architectures and looking forward this, we preferred to use openMP directives rather than MPI ones in starting exploiting parallel computing, at least for this preliminary research activity.

In any case, the result is practically the same by using one or the other set of directives as efficiency and robustness of many algorithms can be significantly improved so to increase performance solution of a series of industrial issues and to allow moving from reacting to predicting technologies applied to industrial plants (Manenti, 2009) and from those solutions still performed off-line to their on-line application (White, 2001) by preserving the robustness of the selected methods.

To quote some examples typical of process industry and PSE/CAPE communities, improvements in optimizer efficiency and robustness can provide practical benefits in data reconciliation (Arora and Biegler, 2001; Bagajewicz, 2003), in data regression (Buzzi-Ferraris and Manenti, 2009a, 2009b, 2010b; Manenti and Buzzi-Ferraris, 2009), in solving nonlinear algebraic or differential systems (Cuoci *et al.*, 2007; Manenti *et al.*, 2009), or in the supply chain management optimization levels (Dones *et al.*, 2009; Lima *et al.*, 2009; Manenti and Rovaglio, 2008).

3. Exploiting Shared Memory to Improve Efficiency and Robustness

Conventional programs easily fail when some specific families of optimization problems have to be solved. Very robust optimizers are required in these cases:

- When the function is multimodal and the global optimum is required
- The function and/or its derivatives are discontinuous
- The function cannot be approximated by a quadric in correspondence with the optimum
- Very narrow valleys (or steep walls) are present
- The function is undefined in some regions and the domain cannot be analytically described

Although no one can ensure the global optimum is found, a robust algorithm should be effective in tackling all previous situations.

Let us start analyzing the problem of very narrow valleys. From this perspective, the OPTNOV's method (Buzzi-Ferraris, 1967) seems one of the most appealing approach. It is important to realize the reason that makes traditional methods such as Simplex (Nelder and Mead, 1965), Hooke-Jeeves (Hooke and Jeeves, 1961), Rosembrock (Rosenbrock, 1960), Quasi-Newton algorithms and so on ineffective when the function valleys are particularly narrow. For example, Rosembrock's method is based on the

rotation of axes of search so to follow the bottom of the valley. Since one of rotated axes is adopted as search direction, it may occur that moving along it does not bring to any function improvement when the valley is very narrow as shown in Figure 1.
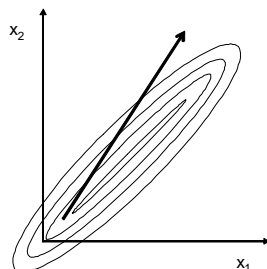

Figure 1. Rosembrock's method fails with very narrow valleys.

To exploit the search direction that inaccurately detects the bottom of the valley, it is necessary to change the point of view. OPTNOV's method is based on some simple ideas that make it particularly robust and efficient in the case of very narrow valleys:

- Whatever optimization algorithm is able to find the bottom of the valley by starting from a point outside the same valley
- The line joining two points on the bottom of the valley is a reasonable valley direction; therefore a point projected along such a direction has good probabilities to be close to the valley
- Nevertheless, this valley direction must not be used as direction of one-dimensional search, rather as a direction which a new point projection must be carried out along
- This new point should not be discarded even though it is worse than the previous one, rather it is the new starting point for the search.
- This search must be performed in the sub-space orthogonal to the valley direction to prevent the problem of having small steps

This philosophy is particularly effective in an object-oriented programming coupled with parallel computing as many reduced optimizations must be carried out starting from distinct points and they can be independently solved each other. Consequently, this philosophy of simultaneously solving different optimization problems by starting from distinct guesses allows rationally facing even the global minimum paradigm.

The concept to build up a program for effectively tackling all aforementioned issues is rather trivial as it is possible to develop an optimizer consisting of N objects, where N is the number of available processors and each of them uses in turn an optimizer reasonably robust.

Hence, two distinct problems must be solved: the first is the selection of points used in the N objects as initial guess and the second is which optimizer to use within each of these N objects; it is worth remarking that even this optimizer must be opportunely robust: to manage possible first- and second-order discontinuities of the function; to overcome possible regions where the same function is undefined; to ensure the global minimum in one-dimensional searches is found; and to efficiently tackle the problem of slightly narrow valleys.

For the sake of clarity, let us call *inner* the optimizer used within each of the N objects and *outer* the one managing the overall optimization problem. The *outer* optimizer only is discussed in this paper.

The problem of searching for the global optimum of the overall problem and to overcome its possible narrow valleys are both tasks of the *outer* optimizer. The following strategy is proposed to manage the N objects: three objects are required for applying OPTNOV's philosophy whereas the remaining $N-3$ objects are selected by using the same techniques employed in optimal experimental design (Buzzi-Ferraris, 1999; Buzzi-Ferraris and Manenti, 2009a, 2010a, 2010b; Manenti and Buzzi-Ferraris, 2009). Therefore, there is the lower bound of using four processors (QUAD CORE machines). The *outer* optimizer collects initial and arrival points of each *inner* object and it selects the two points having the best performances among all those ones collected. If these two points are significantly close, the best third, fourth… is selected in spite of the second to avoid any ill-conditioning while detecting the valley direction.
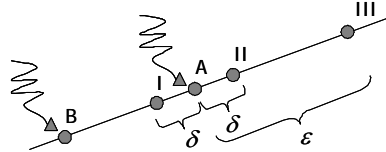

Figure 2. Points A and B are on the bottom of the valley (A is the best one); points I, II, and III are the possible point projections along the valley direction

Distances δ and ε among points can be reduced or expanded according to the results: for example, if the point III brings to a better *inner* optimum, distances are expanded.

Points from the fourth to the $N-th$ are selected so to have the farthest points against all the collected ones. This selection is efficiently carried out by using those techniques adopted and proven for the optimal design of experiments. The following procedure is adopted as stop criterion. At each iteration, the number of points in the neighborhood of the optimum (given a tolerance value) is checked. If such a number is reasonable (according to an assigned value), a possible solution is reached. Theoretically, the number of points should be in the order of magnitude of the optimization problem dimensions, but it is preferable to use smaller numbers when the optimization size is large.

4. Numerical Tests

Many numerical tests were carried out to check the algorithm robustness for problems of different dimensions. Tests of Table 1 are well-known literature functions for:

- "Strong" multimodality issues: all directions are directions of function increase in correspondence with local minima. Both Rastrigin (1) and Haupt (2) functions were adopted and reported in Figure 3.
- "Weak" multimodality issues: the function is constant along at least one direction in correspondence of some local minima. Michalewicz's function (3) was adopted and reported in Figure 3.
- Discontinuities and regions where the function is not defined (4). Figure 4 shows the function against the variable x_1 for the optimal value of x_2.
- Extremely narrow valleys: valleys consisting of steep walls make the search of the minimum a problematic issue for many optimizers. Buzzi-Ferraris's function (5) is adopted and reported in Figure 4.

$$F_{RASTRIGIN} = -110 \cdot n + \sum_{i=1}^{n} \left(x_i^2 - 10\cos\left(2\pi x_i\right) \right) \qquad (1)$$

A Combination of Parallel Computing and Object-Oriented Programming to Improve Optimizer Robustness and Efficiency

$$\begin{cases} F_{HAUPT} = -a, & a > 0 \\ F_{HAUPT} = 0, & a \le 0 \end{cases} \quad \text{where:} \quad a = \prod_{i=1}^{4} \left(\sqrt{x_i} \, \sin\left(2\pi x_i\right) \right) \tag{2}$$

$$F_{MICHALEWICZ} = -\sum_{i=1}^{n} \left(\sin\left(x_i\right) \cdot \left(\sin\left(\frac{i \cdot x_i^2}{\pi} \right) \right)^{2 \cdot n} \right) \tag{3}$$

$$F_{BUZZI-FERRARIS,A} = \sqrt{-945 + x_1 \left[1689 + x_1 \left(-950 + x_1 \left(230 + x_1 \left(-25 + x_1 \right) \right) \right) \right] + e^{-x_1}} + \\ + 10\left| x_2 - 10 x_1 \right| + 10\left| x_1 - 6 \right| \tag{4}$$

$$F_{BUZZI-FERRARIS,B} = \left[x_2 - 10000\left(x_1 - 1 \right)\left(x_1 - 3 \right)\left(x_1 - 5 \right)\left(x_1 - 7 \right)\left(x_1 - 9 \right) \right]^2 + \left(x_1 - 8 \right)^2 \tag{5}$$
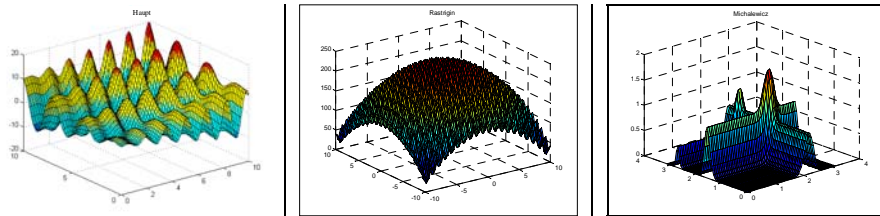

Figure 3. Two-dimensional Haupt (left), Rastrigin (middle), and Michalewicz (right) functions
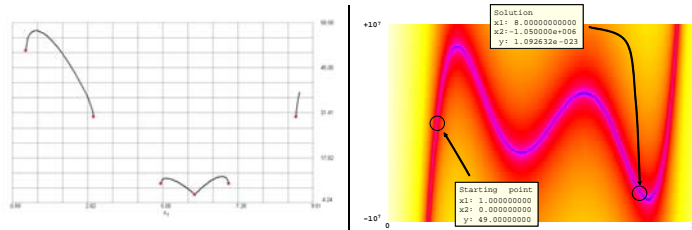
Figure 4. Buzzi-Ferraris's functions (A, left; B, right) to test optimizer robustness

Table 1. Optimization tests

	Starting point	Number of iterations	Optimum value
Rastrigin $n = 2$	$\mathbf{x}_0 = 8.$	2414	-240
Rastrigin $n = 10$	$\mathbf{x}_0 = 8.$	8357	-1200
Haupt	$\mathbf{x}_0 = 0.$	5559	-19.8630560097267
Michalewicz $n = 2$	$\mathbf{x}_0 = 3.$	1683	-1.80130340983985
Michalewicz $n = 10$	$\mathbf{x}_0 = 3.$	20144	-9.660152
Buzzi-Ferraris A	$\mathbf{x}_0 = \{1.;1.\}$	16714	6.708389
Buzzi-Ferraris B	$\mathbf{x}_0 = \{1.;0.\}$	1085	1.009e-018

5. Conclusions and Future Developments

This preliminary research activity shows the way and reports some benefits coming from the interaction of parallel computing and object-oriented programming. Specifically, the example of C++ class for robust optimization that could generate a series of objects so that each of them could run on a specific processor by increasing the same optimizer robustness with a small programming effort is proposed.

References

Arora, N., L.T. Biegler. (2001). Redescending estimators for data reconciliation and parameter estimation. Computers & Chemical Engineering, 25(11-12), 1585-1599.

Bagajewicz, M.J. (2003). Data Reconciliation and Instrumentation Upgrade. Overview and Challenges. FOCAPO 2003. 4th International Conference of Computer-Aided Process Operations, Coral Springs, Florida, 103-116.

Buzzi-Ferraris, G. (1967). Ottimizzazione di funzioni a più variabili. Nota I. Variabili non vincolate. Ing. Chim. It., 3, 101.

Buzzi-Ferraris, G. (1994). Scientific C++. Building Numerical Libraries, the Object-Oriented Way. 2nd Ed., 479pp, Addison-Wesley, Cambridge University Press, ISBN 0-201-63192-X.

Buzzi-Ferraris, G. (1999). Planning of experiments and kinetic analysis. Catalysis Today, 52, 125-132.

Buzzi-Ferraris, G., F. Manenti. (2009a). Kinetic models analysis. Chemical Engineering Science, 64(5), 1061-1074.

Buzzi-Ferraris, G., F. Manenti. (2009b). Outlier Detection. Computers & Chemical Engineering, submitted.

Buzzi-Ferraris, G., F. Manenti. (2010a). Fundamentals and Linear Algebra for the Chemical Engineer Solving Numerical Problems. ISBN: 978-3-527-32552-8, WILEY-VCH, Weinheim.

Buzzi-Ferraris, G., F. Manenti. (2010b). Interpolation and Regression Models for the Chemical Engineer Solving Numerical Problems. ISBN: 978-3-527-32652-5, WILEY-VCH, Weinheim.

Cuoci, A., A. Frassoldati, G. Buzzi-Ferraris, T. Faravelli, E. Ranzi. (2007). The ignition, combustion and flame structure of carbon monoxide/hydrogen mixtures. Note 2: Fluid dynamics and kinetic aspects of syngas combustion. International Journal of Hydrogen Energy, 32(15), 3486-3500.

Dones, I., F. Manenti, H.A. Preisig, G. Buzzi-Ferraris. (2009). Nonlinear Model Predictive Control: a Self-Adaptive Approach. Industrial & Engineering Chemistry Research, to appear.

Hooke, R., T.A. Jeeves. (1961). "Direct Search" Solution of Numerical and Statistical Problems. J. of the Assn. For Computing Machinery, 8, 212-229.

Lima, N.M.N., F. Manenti, R. Maciel Filho, M. Embiruçu, M.R. Wolf Maciel. (2009). Fuzzy Model-Based Predictive Hybrid Control of Polymerization Processes. Industrial & Engineering Chemistry Research, 48(18), 8542–8550.

Manenti, F. (2009). From Reacting to predicting technologies: A novel performance monitoring technique based on detailed dynamic models. Chemical Product and Process Modeling, 4(2).

Manenti, F., G. Buzzi-Ferraris. (2009). Criteria for Outliers Detection in Nonlinear Regression Problems. In J. Jezowski & J. Thullie (Eds.), Computer Aided Chemical Engineering (Vol. 26, pp. 913-917).

Manenti, F., I. Dones, G. Buzzi-Ferraris, H.A. Preisig. (2009). Efficient Numerical Solver of Partially Structured Differential and Algebraic Equation Systems. Industrial & Engineering Chemistry Research, 48(22), 9979-9984.

Manenti, F., M. Rovaglio. (2008). Integrated multilevel optimization in large-scale polyethylene terephthalate plants. Industrial and Engineering Chemistry Research, 47(1), 92-104.

Nelder, J.A., R. Mead. (1965). A simplex method for function minimization. Computer Journal, 7, 308-313.

Rosenbrock, H.H. (1960). An Automatic Method for Finding the Greater or Least Value of a Function. the Computer Journal, 3, 175-184.

White, D.C. (2001). Increased Refinery Productivity through Online Performance Monitoring. Emerson Process Management MCD Technology Division, NPRA Computer Conference.

Polyhedral Results for Discrete-time Production Planning MIP Formulations

Christos T. Maravelias[a], Konstantinos Papalamprou[b]

[a] University of Wisconsin, Madison, WI 53706, USA, maravelias@wisc.edu
*[b] London School of Economics and Political Science, London, WC2A 2AE, UK,
k.papalamprou@lse.ac.uk*

Abstract

We develop a series of polyhedral results for the widely used discrete-time mixed-integer programming (MIP) formulations for production planning and scheduling of continuous processes. We show that for a set of special cases, the incidence matrices of these problems are totally unimodular or network matrices. We also present how these results can be used to facilitate the effective solution of a wide range of practical problems.

Keywords: Mixed-integer programming, polyhedral theory, production planning.

1. Introduction

The increasing product customization and diversification in the chemical industry have led to the installation (or retrofit) of facilities where multiple products compete for limited resources (equipment units and utilities) and which can be operated in multiple modes. The flexibility of these so called *multiproduct* facilities allows for higher resource utilization, lower inventory costs, and better responsiveness to demand fluctuations. Nevertheless, these advantages can only be materialized if the production is planned well, a task which is hard though exactly because of the increased processing flexibility and thus multiplicity of solutions.

To address this challenge, researchers in the area of process systems engineering (PSE) have developed a number of production planning (long-term) and scheduling (short-term) methods, typically mixed-integer programming (MIP) formulations (Maravelias and Sung, 2009). However, most existing attempts to develop effective MIP models in PSE are rather empirical. In particular, the focus has been on the development of smaller formulations (continuous- vs. discrete-time formulations, unit-specific vs. global time points, global- vs. local precedence, etc.) or formulations that appear to be more effective on a subset of (academic) instances.

In this paper, we attempt to rigorously characterize the tightness of a well known and widely used discrete-time formulation, whose polyhedral properties however have not be studied in depth. We present results for the two major constraint sets of the general formulation, and we develop a series of stronger results for special cases. Finally, we show how these results can be used in practice to solve large-scale problems.

It is important to note here that discrete-time formulations have a number of advantages over their continuous-time counterparts: (i) account linearly for inventory and backlog costs; (ii) handle intermediate release and due dates at no additional computational cost; (iii) can be readily modified to model events taking place during the execution of a task. However, the interest in discrete-time models has diminished because it was believed that they lead to intractable formulations. Our results show that discrete-time formulations can in fact be solved effectively if their polyhedral structures are studied.

2. Background

2.1. Polyhedral theory

Totally unimodular matrices form one of the most important classes of matrices for integer linear programming problems. An integral matrix is called *totally unimodular* (TU) if the determinant of each square submatrix of A is equal to 0,+1, or -1. The importance of this class stems mainly from the following characterization of Hoffman and Kruskal (1956):

Theorem 1. The polyhedron $\mathbf{P}(A, \boldsymbol{b}) = \{\boldsymbol{x}: A\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq 0\}$ is integral for all integral vectors $\boldsymbol{b}$ if and only if A is *totally unimodular.*

Theorem 1 implies that the integer programming model $\mathbf{Q} = \max \{\boldsymbol{c}^{\mathrm{T}}\boldsymbol{x}: A\boldsymbol{x} \leq \boldsymbol{b}, \boldsymbol{x} \geq 0, \boldsymbol{x}$ integral$\}$ can be solved by simply dropping the integrality constraints and solving the resulting linear relaxation. A well-known and useful characterization for TU matrices is the following by Ghouila-Houri (1962):

Theorem 2. Let A be a matrix with set of rows M and set of columns L. The following statements are equivalent:
(i)   matrix A is totally unimodular
(ii)   for every $M' \subseteq M$, there exists a partition M_1, M_2 of M such that

$$\left| \sum_{m \in M_1} a_{lm} - \sum_{m \in M_2} a_{lm} \right| \leq 1, \qquad \forall\, l \in L$$

An important subclass of TU matrices is the class of *network matrices* (for a complete overview, see Nemhauser and Wolsey, 1988) which we define as follows: A matrix A with elements in $\{0, +1, -1\}$ is called network if there exists a directed tree T with edges labeled by the rows of A such that, for each column l of A, the following two conditions are satisfied: (i) the non-zeros of l are in rows which correspond to edges inducing a path P_l in T, and (ii) two non-zero entries of l are different if the corresponding edges have opposite direction in P_l.

If matrix A is network, then the integer programming problem **Q** can be solved using efficient methods such as the *network simplex method* which can be up to 200 times faster than general linear programming codes (se Glover et al., 1974).

Furthermore, a full-row rank $m \times n$ matrix A is called *unimodular* if A is integral and each basis of A (i.e. non-singular $m \times m$ submatrix of A) has determinant +1 or -1. Clearly, unimodularity is a generalization of total unimodularity for full-row rank integral matrices. Moreover, it has been shown that the polyhedron $\mathbf{P}(A, \boldsymbol{b})=\{\boldsymbol{x}: A\boldsymbol{x}=\boldsymbol{b}, \boldsymbol{x} \geq 0\}$, where A is an integral full-row rank matrix and $\boldsymbol{b}$ is an integral vector, is integral if and only if A is unimodular (see Schrijver, 1986).

Another generalization of totally unimodular matrices is formed by the class of κ-regular matrices. A rational matrix is called κ-*regular* if for all its non-singular square submatrices R, κR^{-1} is integral. The class of κ-regular matrices possesses important polyhedral properties as shown in the following theorem (Appa and Kotnyek, 2004):

Theorem 3. Let A be an $m{\times}n$ rational matrix. Then the polyhedron $\mathbf{P}(A, \kappa\boldsymbol{b}) = \{\boldsymbol{x}: A\boldsymbol{x} \leq \kappa\boldsymbol{b}, \boldsymbol{x} \geq 0\}$ is integral for each vector $\boldsymbol{b} \in \mathbf{Z}^{\mathrm{m}}$, if and only if A is κ-regular.

2.2. Problem statement

We consider the production planning of multi-stage multi-product continuous processes that involve units that can carry out multiple tasks. Given are a planning horizon η, a set of tasks $i \in \mathbf{I}$ processing units $j \in \mathbf{J}$, and chemicals (states) $k \in \mathbf{K}$. In addition, the process has the following characteristics:

a. A processing unit j can be used to carry out tasks $i \in \mathbf{I}_j$; tasks that can be carried in multiple units are modeled as multiple tasks, each one carried out in only one unit.

b. The production rate of task i is ρ_i; there is no minimum run length.

c. A chemical can be consumed/produced by multiple tasks; the set of tasks consuming/producing chemical k is denoted by $\mathbf{I}_{k-}/\mathbf{I}_{k+}$.

d. Each task consumes/produces one chemical, denoted by $k^C(i)/k^P(i)$.

e. Chemical k is stored in a dedicated tank with capacity ζ_k.

f. Due to continuous processing, a chemical cannot be stored in a processing unit, thus making no-intermediate storage and zero-wait policies identical; they are modeled by setting $\zeta_k = 0$.

g. Raw materials, $k \in \mathbf{K}^{RM}$, and intermediates, $k \in \mathbf{K}^{INT}$, can be delivered and final products, $k \in \mathbf{K}^{FP}$, can be shipped at different time points.

h. There are no utility requirements.

2.3. Mixed-integer programming formulation

We consider a modification of the discrete-time STN formulation of Maravelias (2005). The time horizon is divided into periods (time buckets) $n \in \mathbf{N}=\{1, 2, \ldots, N\}$ of uniform length $\Delta t = \eta/N$, defining $N+1$ time points $\{0, 1, 2, \ldots, N\}$. The optimization variables are: (i) $W_{in} \in \{0, 1\}$: it is equal to 1 if task i is processed during period n; (ii) $B_{in} \geq 0$: extent (amount processed) of task i during period n; and (iii) $S_{sn} \geq 0$: inventory of chemical k at time point n.

The assignment constraint for unit j during period n is expressed via,

$$\sum_{i \in \mathbf{I}_j} W_{in} = 1, \qquad \forall j, n \tag{1}$$

Given the fixed rate ρ_i of task i and the duration Δt of the time period, and assuming that when $W_{in} = 1$ the processing of task i spans period n, we can calculate the amount processed (batchsize) β_i of task i during a period, $\beta_i = \rho_i \Delta t$, and express the extent B_{in} as $B_{it} = \beta_i W_{it}, \forall i,t$. Thus, the material balance constraints can be written as:

$$S_{kn} = S_{k,n-1} + \sum_{i \in \mathbf{I}_{k+}} \beta_i W_{in} - \sum_{i \in \mathbf{I}_{k-}} \beta_i W_{in} + \gamma_{kn}, \qquad \forall k, n > 0 \tag{2}$$

where γ_{kn} is the net delivery of chemical k at time point n; γ_{kn} is positive for deliveries of raw materials and negative for shipments (i.e. demand satisfaction) of final products.

The feasible region of the MIP model **M** we consider in this paper is defined by eqs (1) and (2) with $W_{in} \in \{0,1\} \ \forall i, n$; and $0 \leq S_{kn} \leq \zeta_k, \forall k, n$. The model can be extended to account for unmet or backlogged demand and final product shipments and the corresponding costs (Maravelias and Papalamprou, 2009).

2.4. Previous polyhedral results

The feasible region **P** of the LP-relaxation of model M is $\mathbf{P} = \mathbf{P}^A \cap \mathbf{P}^{MB}$ with

$$\mathbf{P}^A = \left\{ w \in [0,1]^{I \cdot N} : \sum_{i \in \mathbf{I}_j} W_{in} = 1, \quad \forall j, n \right\}$$

$$\mathbf{P}^{MB} = \left\{ (s, w) \in [0, \xi] \times \mathbf{R}_+^{I \cdot N} : S_{kn} - S_{k,n-1} - \sum_{i \in \mathbf{I}_{k+}} \beta_i W_{in} + \sum_{i \in \mathbf{I}_{k-}} \beta_i W_{in} = \gamma_{kn}, \quad \forall k, n > 0 \right\}$$

where w is the $(|\mathbf{I}| \cdot |\mathbf{N}|)$-dimensional vector of W_{in} variables, s is the $(|\mathbf{K}| \cdot |\mathbf{N}|)$-dimensional vector of S_{kn} variables, and ξ is the $(|\mathbf{K}| \cdot |\mathbf{N}|)$-dimensional vector of bounds for S_{kn} variables with $\xi_l = \zeta_k, l \in \{(k-1)n+1, \ldots, kn\}, k \in \mathbf{K}, n \in \mathbf{N}$.

Maravelias and Papalamprou (2009) showed the following for integral data:

Proposition 1. The matrix of the assignment constraints in Eq. (1) is a network matrix.

Proposition 2. The matrix of the material balance constraints in Eq. (2) is lcm_β-regular, where lcm_β is the least common multiple of batchsizes β_i.

Proposition 3. Polyhedron $\mathbf{P}^{\text{MB}}$ is integral if $\Delta t = \text{gcf}_\varepsilon/\text{lcm}_\rho$, where gcf_ε is the greatest common factor of deliveries, γ_{kn}, and storage capacities ζ_k; and lcm_ρ is the least common multiple of production rates ρ_i.

Maravelias and Papalamprou (2009) generalized these results to problems with rational data and showed that variables W_{in} obtain integral values at the vertices of polyhedron $\mathbf{P}^{\text{MB}}$ if Δt is selected appropriately:

Proposition 4. If λ is the least common multiple of the denominators of rational parameters γ_{kn}, ζ_k, and ρ_i, then variables W_{in} obtain integral values at the vertices of $\mathbf{P}^{\text{MB}}$ if $\Delta t = \text{gcf}_{\lambda\varepsilon}/\text{lcm}_{\lambda\rho}$, where $\text{gcf}_{\lambda\varepsilon}$ is the greatest common factor of $\lambda\gamma_k$ and $\lambda\zeta_k$ and $\text{lcm}_{\lambda\rho}$ is the least common multiple of $\lambda\rho_i$.

Maravelias and Papalamprou (2009) then used these results to address large-scale problems effectively and posed the question on whether there exist special cases in which the whole constraint matrix is transformable into a network or totally unimodular matrix. Such special cases are discussed in the next section.

3. New Theoretical Results

In this section, we consider special cases of the general formulation we defined in the previous section and derive results for the incident matrix of all constraints. An outline of a proof for each result is also provided. Finally, note that in all cases we assume that all batchsizes are equal to 1.

3.1. Special cases: totally unimodular matrices

Two totally unimodular special cases are discussed; an example process network associated with Theorems 4 and 5 is given in Figures 1(i) and 1(ii), respectively.

Theorem 4. Let N be a process network such that each unit consumes the same chemical. Then the constraint matrix A associated with N is a totally unimodular matrix.

Proof outline. Let M be the set of rows of A and let M^{AS} and M^{MB} be the subsets of rows corresponding to the assignment and the material balance constraints, respectively. We shall show that, for any $\text{M}' \subseteq \text{M}$ there exists a partition M_1, M_2 of M' such that (1) of Theorem 2 is satisfied and thereby, A is TU. We shall denote by A' the row submatrix of A with row set M'. The partition rules such that (1) of Theorem 2 is satisfied are as follows: (i) all rows $\text{M}' \cap \text{M}^{\text{MB}}$ are included in M_1; (ii) a row of $\text{M}' \cap \text{M}^{\text{AS}}$ is included in M_1 only if there exists a column j of A' such that $\sum_{m \in M'} a_{mj} = +2$. $\square$

Theorem 5. Let N be a process network such that all the tasks of each unit produce the same chemical. Then the constraint matrix A associated with N is a totally unimodular matrix.

Proof outline. If all the rows of A corresponding to the material balance constraints are multiplied by -1 then the matrix so-obtained corresponds to a process network which is as described in Theorem 4 (i.e. each unit of the process network consumes the same chemical). Thus, by Theorem 4 and the fact that TU matrices are closed under column scaling by -1, the result follows. $\square$
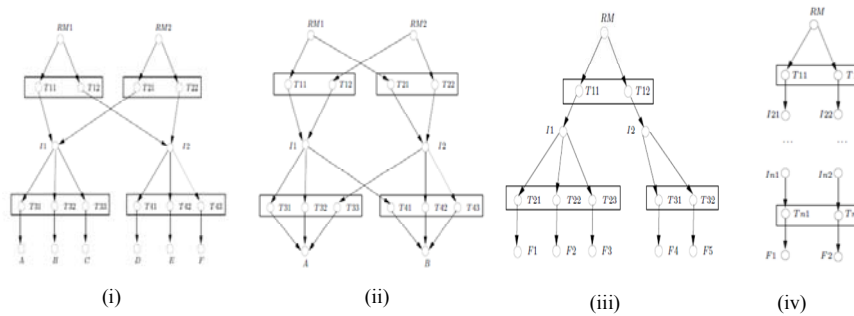
Polyhedral Results for Discrete-time Production Planning MIP Formulations

| (i) | (ii) | (iii) | (iv) |

Figure 1: Process networks with totally unimodular or network constraint matrices.

3.2. Special cases: network matrices

A network case is provided by the following theorem; Figures 1(iii) depicts an example process network of this case.

Theorem 6. Let N be a process network such that the following conditions are satisfied:
(i) all tasks carried in each unit consume the same chemical, and
(ii) each chemical is produced by a single task.
Then the constraint matrix A associated with N is a network matrix.

Proof outline. Let us partition the row set of A into three sets B, C and D, where B={b1,…, bn} is the subset of rows of A which corresponds to the assignment constraints, C={c1,…,cn} consists of the rows of A each of which has a -1 element and D={d1,…,dm} consists of the remaining rows of A (clearly, C and D partition the set of rows which corresponds to the material balance constraints). Matrix A is network since the tree of Figure 2 is associated with A.   □
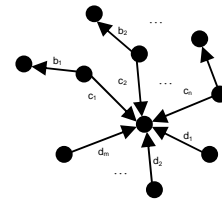

Figure 2

Finally, we would also like to discuss another interesting case, which corresponds to the process network in Figure 1(iv). If A is the associated constraint matrix, then we can show that if B is a basis of A (i.e. a square non-singular submatrix of A) then C= B^{-1}A is a network matrix. By a known result (see Chapter 21 in Schrijver 1986) we have that C is a unimodular matrix and that the polyhedron P is integral. However, the fact that the problem can be transformed to a network one gives us the option to use the network simplex method and speed up the solution process. These facts are summarized in the following theorem which is given here without proof due to space limitations.

Theorem 7. The constraint matrix A associated with the process network of Figure 1(iv) is unimodular. Moreover, for a basis B of A the matrix B^{-1}A is a network matrix.

4. Solution of general classes of problems

The results presented in the previous section were derived for special cases of the production planning problems stated in §2.3. However, they provide insights and lead to solution methods that can be used to address more general classes of problems.

4.1. From network and TU matrices to κ-regular matrices

Assuming all batchsizes are equal to 1, we showed that the incidence matrices of certain problems are totally unimodular or network matrices. If the batchsizes are not equal to 1, then we can show that the corresponding matrices are κ-regular. This is because κ-

regularity is preserved under several matrix operations and because row (or column) multiplication/division weakens/preserves κ-regularity (Appa and Kotnyek, 2004). We can prove κ-regularity by showing that we can obtain the incident matrix of our problem from a 1-regular matrix through matrix operations. Thus, if a given process network satisfies the conditions of one of Theorems 4-7, then its incident matrix is κ-regular (where κ depends on problem data), which means that can be solved using LP technology.

4.2. Extension of proposed results

It is easy to show that more general problems with shipments of raw materials and products, as well as problems with shipments and backlogged demand lead to formulations whose incidence matrices have the properties of Theorems 4-7, if the corresponding conditions are satisfied. This can be shown by generating the matrices of these more general problems from the matrices of the problems discussed in the previous section via operations we know that preserve total unimodularity.

4.3. Embedded structures

In the presence of *complex* side constraints (e.g., minimum processing time requirement, changeover times), the incidence matrix of the problems discussed in §3.1-2 and §4.1-2 do not poses the polyhedral properties (at least, we have not been able to show so). However, the network or TU structure for a subset of constraints (assignment + mass balance) is recognized by powerful commercial MIP solvers resulting in substantial reductions in computational times. In fact, our computational results indicate that the proposed time discretization leads to substantial reduction in computational requirements in problems with sequence-dependent changeover times and minimum run requirements.

5. Acknowledgements

References

G. Appa and B. Kotnyek, 2004, Rational and integral k-regular matrices, *Discrete Mathematics*, 275, 1-15.

F. Glover, D. Karney and D. Klingman, 1974, Implementation and computational comparisons of primal, dual and primal–dual computer codes for minimum cost network flow problem, *Networks*, 4, 191–212.

A. Ghouila-Houri, 1962, Caracterisation des matrices totalement unimodulaires, *Comptes Rendus Hebdomadaires des Seances de l' Academie des Sciences*, 254, 1192-1194.

A.J. Hoffman, J.B. Kruskal, 1956, Integral boundary points of convex polyhedra, In H.W. Kuhn, & A.W. Tucker (Ed.), *Linear inequalities and related systems*, Princeton Univ. Press, 223-246.

C.T. Maravelias, 2005, Mixed time representation for state-task network Models. *Ind. Eng. Chem. Res.*, 44 (24), 9129-9145.

C. T. Maravelias, K. Papalamprou, 2009, Polyhedral results for discrete-time production planning MIP formulations for continuous processes. *Comput. Chem. Eng.*, 33(11), 1890-1904.

C. T. Maravelias, C. Sung, 2009, Integration of production planning and scheduling: overview, challenges and opportunities, *Comput. Chem. Eng.*, 33 (12), 1919-1930.

G.L. Nemhauser and L.A. Wolsey, 1989, *Integer and Combinatorial Optimization*, John Willey and Sons, Inc., New York.

A. Schrijver, 1986, *Theory of linear and integer programming*, Wiley, Chichester.

A dynamic screening algorithm for multiple objective simulated annealing optimization

Eftychia C. Marcoulaki, Ioannis A. Papazoglou

System Reliability and Industrial Safety Laboratory, National Center for Scientific Research "DEMOKRITOS", PO Box 60228, 15310 Aghia Paraskevi, Athens, Greece, emarcoulaki@ipta.demokritos.gr

Abstract

This work proposes new multiple objective optimization (MOO) technology, using a Monte Carlo-based algorithm stemmed from simulated annealing (SA). Since the expected result in MOO tasks is usually a set of Pareto-optimal solutions, the optimization problem states assumed here are themselves sets of solutions. The stochastic search follows a series of reversible state transitions at constant probability, to enjoy convergence properties of stationary Markov processes. The proposed technology is tested against the optimal design of a process system involving equipment placed in a serial/parallel arrangement, with three optimization objectives: the system cost, reliability and weight.

Keywords: multiple objective optimization, simulated annealing, redundancy apportionment problem

1. Introduction

Real process design problems involve the co-consideration of many objectives, like revenue, energy, dependability. Such problems can be formulated mathematically and solved as MOO tasks. To avoid assumptions on the relative importance of the objectives, the final result is usually set of solutions, the Pareto-optimal front (POF). The area of MOO is dominated by evolutionary tools (Coello Coello, 2006), usually in the form of genetic algorithms (GA) where the optimal solution set is the final population. Different genetic tools propose alternative schemes for selection and fitness assignment, and may apply non-dominated sorting, elitistic biases, solution archives, diversity promotion etc to guide the search (Deb, 2002). The population notion takes the attention away from individual solutions and allows optimization of the solutions set as a whole.

Simulated Annealing is a popular and robust meta-heuristic method for single objective optimization. Previous efforts to extend SA to MOO applications investigated alternative strategies for the propagation of the search and the generation of final solution sets. Nam and Park (2000) considered various energy functions to depict the relative benefit of moves in a multidimensional plane and applied multiple runs to generate the POF. Though certain acceptance functions exhibit asymptotic convergence to solutions in the POF, there was no proof for uniform convergence (Villalobos-Arias et al., 2006). To improve the diversity of the final solutions, many researchers used an archive of all the non-dominated solutions in the run. The archive was updated according to each candidate solution generated on the fly. Suppapitnar et al. (2000) used a combined acceptance probability calculated as the product of probability functions defined for each objective. Teghem et al. (2000) proposed using a scalarizing function to guide the random search. Following advancements in GAs, Suman (2004) used an energy function based on the strength Patero fitness assignment procedure. Smith et al. (2008) compared the candidate to the current solution according to the cardinalities of their dominant subsets in the

archive. Bandyopadhyay et al. (2008) used a non-Boltzmann probability function and devised different strategies to quatify the amount of domination, according to domination cases. The archive was initially refined using hill-climbing, and clustered during the run to reduce the computational effort of its management.

In Monte Carlo algorithms, like the standard SA, the final problem state is part of a statistical distribution that remains independent of the algorithm initialization strategies (Geman and Geman, 1984). The convergence property can serve as a guarantee for the quality of the final state and this is particularly useful in optimization. This work proposes a MOO algorithm which aims to maintain the benefits from the mathematical background and the subsequent robustness of the original SA, while addressing the challenges rising from working with a set of solutions rather than a single solution.

2. Background

Without loss of generality, a multi-objective optimization problem assumes the form:

$$\text{minimize} \quad \boldsymbol{F}(\boldsymbol{x}) = \left[f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), ..., f_n(\boldsymbol{x}) \right]$$

$$\text{subject to} \quad \boldsymbol{G}(\boldsymbol{x}) = \left[g_1(\boldsymbol{x}), g_2(\boldsymbol{x}), ..., g_m(\boldsymbol{x}) \right] \leq \boldsymbol{O}_m$$

where $\boldsymbol{x}$ is the variable vector; $\boldsymbol{F}$ is the vector of n objective functions f_i, $i = 1,..n$; $\boldsymbol{G}$ is the vector of the m problem constraints g_j, $j = 1,..m$; $\boldsymbol{O}_m$ is the m-dimensional null vector. Let $\boldsymbol{D}$ denote the domain of variable vectors $\boldsymbol{x}$. MOO algorithms generate a search through $\boldsymbol{D}$ comparing among different solutions to find those featuring the minimal $\boldsymbol{F}$ values. In single objective optimization, solution $\boldsymbol{x}_1$ is better than $\boldsymbol{x}_2$ iff $f(\boldsymbol{x}_1) \leq f(\boldsymbol{x}_2)$. In multi-objective domains, the comparison can be done according to Pareto dominance, so the relation "$\boldsymbol{x}_1$ dominates over $\boldsymbol{x}_2$" (or "$\boldsymbol{x}_2$ is dominated by $\boldsymbol{x}_1$") is written as

$$\boldsymbol{x}_1 \triangleleft \boldsymbol{x}_2 \Leftrightarrow \left(f_i(\boldsymbol{x}_1) \leq f_i(\boldsymbol{x}_2), \forall i \text{ and } \exists i : f_i(\boldsymbol{x}_1) < f_i(\boldsymbol{x}_2) \right), i = 1, 2, ..., n$$

The "domination" relation leaves out the option of two solutions being mutually non-dominating. The POF is the set containing all the solutions $\boldsymbol{x} \in \boldsymbol{D}$ which are non-dominated by any other member of $\boldsymbol{D}$. The aim of MOO technologies is to generate a set of solutions that appears adequately representative of the POF. This includes

(i) guiding the search towards the optimal front, and
(ii) maintaining diversity in the solutions visited during the search and in those stored in the final set of solutions (Villalobos-Arias et al., 2006; Deb, 2002).

This work attempts to address these two challenges using a Monte Carlo–based algorithm stemmed from SA. The standard single objective SA algorithm starts from a random initial state $\boldsymbol{x}_0$, performs a series of random state transitions, and converges to a final state. Each state is an instance, a solution, of the system under consideration. Each transition ξ consists of (a) a random perturbation from a current state $\boldsymbol{x}_\xi$ to a candidate new state $\boldsymbol{x}^*$, and (b) the acceptance ($\boldsymbol{x}_{\xi+1} = \boldsymbol{x}^*$) or rejection ($\boldsymbol{x}_{\xi+1} = \boldsymbol{x}_\xi$) of the perturbation. Acceptance relies on the stochastic improvement that the perturbation brings to the objective function value. The probabilistic biases towards better solutions increase according to a control variable, the effective temperature t, which decreases during the search according to an appropriate cooling schedule. The state transitions performed between subsequent reductions of t constitute a Markov chain. The chain converges to a stationary distribution of problem state probabilities, provided that the state transitions are reversible and their probabilities are constant. As the effective temperature drops, the series of stationary chains converges stochastically to the global optimum (Geman and Geman, 1984; Aarts and van Laarhoven, 1985).

A dynamic screening algorithm for multiple objective simulated annealing optimization

3. Proposed methodology

Since the expected result in MOO tasks is usually a set of Pareto-optimal solutions, this work suggests that the problem states are themselves sets of solutions. So the states assumed here have the form $X = \{x_1, x_2, ..., x_{|X|}\}$ where $x_i \in D$ and $|X|$ denotes the cardinality of X. To enjoy the convergence properties of SA, the state transitions should:

  (i) allow deterioration in terms of the Pareto-dominance of the solutions in the state
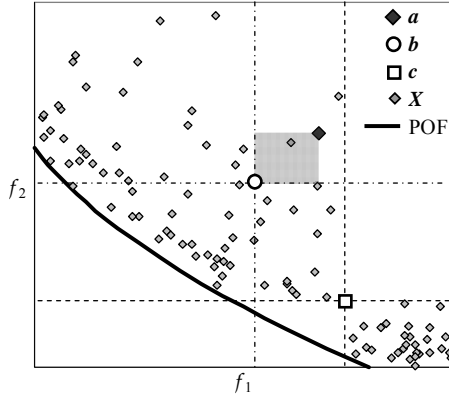 (ii) have constant probability of occurrence during the same temperature interval.

Figure 1: Set X and solutions a, b and c

The energy function $E(X)$ assumed here is a measure of the total "area" laying between the solutions in state X and the POF. Let the set X and two solutions $a \in X$ and $b \in D$ (Figure 1a). First consider that a is comparable to b, and there is a choice to move from X to $X' = X + \{b\} - \{a\}$. If set X is sufficiently big, the difference in the total area between the sets X, X' and the POF is negligible, so these differences can safely be estimated at microscopic level. Now consider that $c \in D$ is not comparable to any member of X, and the choice between the states X and $X' = X + \{c\}$. In this case, the inclusion of c in the new state is always beneficial since it increases the diversity within the state. To achieve a simple implementation and fast execution, the energy formulae used here become:

$$\Delta \varepsilon = E(X') - E(X) = \begin{cases} (-1)^n \cdot \prod_{i=1}^{n} (f_i^b - f_i^a) & , \quad a \lhd b \vee b \lhd a \\ 0 & , \quad \text{otherwise} \end{cases}$$

Given an initial temperature t_0 and a minimum state cardinality N, the proposed algorithm proceeds as follows

---

Start with a randomly chosen state $X = \{x_1, x_2, ..., x_N\}, x_i \in D$; let $\varepsilon = 0, \delta E_{t_0} = \{\varepsilon\}$

for ($t = t_0$; until *search termination criteria are met*; $t = g(t, \delta E_t), \varepsilon = 0, \delta E_t = \{\varepsilon\}$)

    for (k=0; while *chain propagation criteria are met*; k++)

        select randomly $Y \subseteq X$; apply the perturbation $h : y = h(Y), y \in D$

        let $Z \subseteq X : \{z \in X : z \lhd y \vee z \lhd y\}$

        if ($Z \neq \varnothing$) then

            select randomly $z \in Z$ (giving priority to $z \in Z \cap Y$)

            let $\Delta \varepsilon = (-1)^{n-1} \cdot \prod_{i=1}^{n} (f_i^y - f_i^z)$

            if $p \leq \exp(-\Delta \varepsilon / t)$ let $X = X - \{z\} + \{y\}$; $\varepsilon = \varepsilon + \Delta \varepsilon$

        else let $X = X + \{z\}$

        update the set of energies $\delta E_t = \delta E_t + \{\varepsilon\}$

    end k–loop: *proceed to the next chain node*

    apply the ordering $r : \vec{n} = r(X)$; let $X = \{x \in X : n(x) \leq N\}$

end t–loop: *proceed to the next temperature*

---

where ε denotes the energy gain of the new state compared to the fist state of the current chain; $\delta \boldsymbol{E}_t$ is the set of ε's encountered at temperature t; g is the cooling schedule function; h is a perturbation function to generate a new solution from a set of current ones.

In the present implementation of the algorithm, the termination and propagation criteria are similar to those used in Marcoulaki and Kokossis (1999). The initial annealing temperature is estimated automatically based on the statistics of a random initial guess. The employed cooling schedule is from Aarts and van Laarhoven (1985) using $\gamma = 0.02$. The state ordering is according to the non-dominated sorting approach of Deb et al. (2002). The h perturbation used in the application example assumes the form of a randomly assigned crossover/mutation hybrid. Due to lack of space, the implementation choices, as well as the benefits of the proposed energy function are not explained in detail.

4. Application example

The proposed algorithm is tested against the problem of designing a complex configuration of process units. The system is composed of several subsystems placed in series, each performing a different operation, and all of them needed to be operational for the overall system to operate. Each subsystem can contain various types of equipment, able to perform exactly the same operation, but featuring different characteristics. The design problem consists in appointing the optimal combinations of component types and redundancies per subsystem, according to several conflicting objectives.

For the application considered here, the characteristics of each subsystem equipment type are its reliability, cost and weight. Similarly, the problem objectives are the overall system reliability, cost and equipment weight, according to the following formulations

$$\text{maximize} \quad R(\boldsymbol{x}) = \prod_{n=1}^{N} R_i(\boldsymbol{x}) = \prod_{n=1}^{N} \left(1 - \prod_{k=1}^{K_n} \left(1 - r_{n,k}\right)^{x_{n,k}}\right)$$

$$\text{minimize} \quad \begin{cases} C(\boldsymbol{x}) = \sum_{n=1}^{N} C_n(\boldsymbol{x}) = \sum_{n=1}^{N} \sum_{k=1}^{K_n} x_{n,k} \cdot c_{n,k} \\ W(\boldsymbol{x}) = \sum_{n=1}^{N} W_n(\boldsymbol{x}) = \sum_{n=1}^{N} \sum_{k=1}^{K_n} x_{n,k} \cdot w_{n,k} \end{cases}$$

$$\text{subject to} \quad \begin{cases} 0 \leq x_{n,k} \leq m_{n,k} \\ \sum_{k=1}^{K_n} x_{n,k} \geq 1, \forall n \end{cases}$$

where R, C and W denote the reliability, the cost and the weight of the overall system (or subsystem n), respectively. The number of subsystems is N, and the total number of available equipment types in subsystem n is K_n. The variable matrix $\boldsymbol{x} = \{x_{n,k}\}$ stores the number of equipment components of type k in subsystem n. The $r_{n,k}$, $c_{n,k}$ and $w_{n,k}$ denote the reliability, cost and weight of equipment type k in subsystem n, respectively.

The optimization experiments considered here assume a system of 14 subsystems. Of these subsystems, 6 may have up to 4 equipment types and 8 may have up to 3. The maximum redundancy of each equipment type is set to 6, so the total number of system configurations equals 4.24×10^{39}. Data for the component characteristics are taken from Zhao et al., (2007), who treated a simpler problem version using ant colonies.

In this particular problem, the cost and the weight usually increase as the reliability increases, though this not always the case, even among equipment components of the same subsystem. A set of 10^5 random solutions (see Figure 2) indicates a very high solution density as the reliability approaches unity and the solution density drops steeply as the reliability decreases. The solution with minimal reliability has $[R, C, W] = [0.190, 36, 77]$. However, the probability of randomly generating solutions with R between 0.8–0.9, 0.7–0.8 and below 0.7 is less than 2×10^{-2}, 9×10^{-4} and 10^{-5}, respectively.
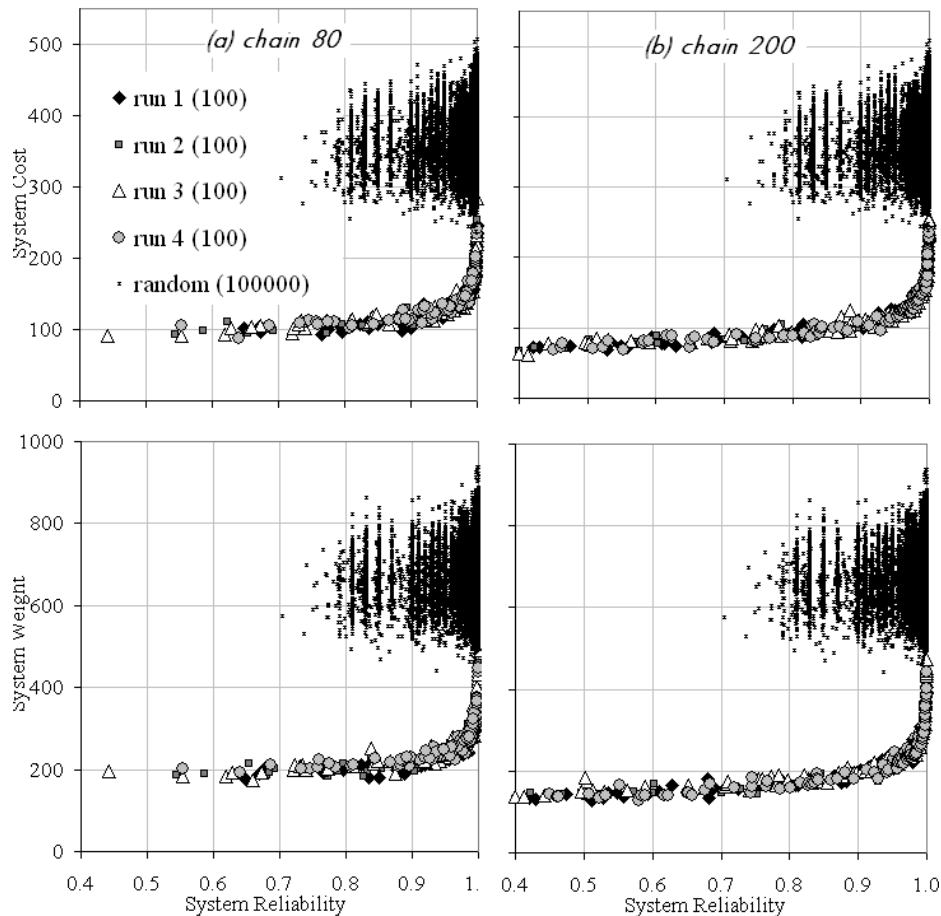
Fig. 2: Optimization results for the redundancy appointment problem.

In the optimization runs, the minimum problem state cardinality is set to 100 configurations. Figure 2 presents final results for the three objectives. Two stationary chain lengths are considered, a short chain with 80 iterations (Figure 2a) and a longer with 200 iterations (Figure 2b). The figure shows the convergences of 4 different stochastic experiments per chain length. The average number of problem simulations for lengths 80 and 200 ranges between $7.9 \times 10^4 – 8.2 \times 10^4$ and $6.5 \times 10^5 – 6.7 \times 10^5$, respectively. The results show that at longer chains the final fronts move to lower cost and weight, and span over a wider range of reliabilities. The three objective function distributions are well approached even with the short chain length. The final fronts agree with Zhao et *al.*, (2007) who maximized only the reliability putting constraints on cost and weight.

5. Summary and conclusions

This paper proposes a new simple and flexible MOO algorithm based on the original SA aiming to benefit from the convergence properties of Markov processes. Rather than assuming individual solutions tested against an archive of non-dominated solutions, the state of the optimization problem considered here is itself a set of solutions. Stochastic moves are applied to each problem state in the form of perturbations on the solutions comprising the state. The generated candidate new set is compared to the current set

using the annealing criterion, and the acceptance of a proposed modification depends on the stochastic benefit gained from modifying the current set of solutions. The benefit evaluation is based here on the difference of the total "areas" between the members of the state and the true Pareto-optimal front. An approximation is proposed to avoid computationally intensive procedures, and resorts to a micro-scale comparison between the current and the candidate states. The problem states are allowed to expand at constant annealing temperature, but their size is reduced according to dominance and clustering criteria when the statistical cooling occurs. The dynamic memory demands of the proposed stochastic search and the effort spent on manipulating the solutions comprising the problem state are low, compared to other MOO algorithms.

The new algorithm exhibits good convergence properties, as explained in the description of the proposed search scheme and demonstrated through the computational experiments for the design of a fictitious process system using three conflicting objectives. Despite the combinatorial scale of the design application and its intrinsic tendency towards certain solutions, the proposed algorithm is found efficient in guiding the search, and the final state covers uniformly most of the true optimal front. This is very important since there is no proof that the convergence guarantees of the standard single-objective SA are inherited in its extensions to treat multi-objective problems.

A major advantage of SA-based algorithms is in treating constraints by embedding them in the problem representation and manipulating the perturbation strategies. So, much faster performances can be achieved by applying intelligent perturbations, tailored to the peculiarities of each application. Current work focuses on further testing the new algorithm, and applying it on process synthesis problems involving constraints.

References

E.H.L. Aarts and P.G.M. van Laarhoven, 1985, Statistical cooling: A general approach to combinatorial optimization problems, Philips Journal of Research, 40, 193-226

S. Bandyopadhyay, S. Saha, U. Maulik and K. Deb, 2008, A simulated annealing-based multiobjective optimization algorithm: AMOSA, IEEE Transactions on Evolutionary Computation, 12 (3), 269–283

C.A Coello Coello, 2006, Evolutionary multi-objective optimization: a historical view of the field, IEEE Computational Intelligence Magazine, 1 (1), 28-36

K. Deb, A. Pratap, S. Agarwal and T. Meyarivan, 2002, A fast and elististic multiobjective genetic algorithm: NSGA-II, IEEE Transactions on Evolutionary Computation, 6 (2), 182-197

S. Geman and D. Geman, 1984, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, 721-741

E.C. Marcoulaki and A.C. Kokossis, 1999, Scoping and screening complex reaction networks using stochastic optimization, AIChE Journal, 45 (9), 1977-1991

D. Nam and C. Park, 2000, Multiobjective simulated annealing: a comparative study to evolutionary algorithms, Internation Journal of Fuzzy Systems, 2 (2), 87-97

K.I. Smith, R.M. Everson and J.E. Fieldsend, 2008, Dominance-based multi-objective simulated annealing, IEEE Transactions on Evolutionary Computation, 12 (3), 323-342

B. Suman, 2004, Study of simulated annealing based algorithms for multiobjective optimization of a constrained problem, Computers and Chemical Engineering, 28, 1849-1871

J. Teghem, D. Tuyttens and E. L. Ulungu, An interactive heuristic method for multi-objective combinatorial optimization, Computers and Operations Research, 27 (7-8), 621-634

M. Villalobos-Arias, C.A. Coello Coello and O. Hernandez-Lerma, 2006, Asymptotic convergence of a simulated annealing algorithm for multiobjective optimization problems, Mathematical Methods of Operations Research, 64, 353-362

J.-H. Zhao, Z. Liu and M.-T. Dao, 2007, Reliability optimization using multiobjective ant colony system approaches, Realiability Engineering and System Safety, 92, 109-120

Parameter Estimation in Kinetic Models for Large Scale Metabolic Networks with Advanced Mathematical Programming Techniques

Jimena Di Maggio[a], Juan C. Diaz Ricci[b], M. Soledad Diaz[a]

[a]*Planta Piloto de Ingeniería Química-PLAPIQUI (UNS-CONICET),Camino la Carrindanga km 7, Bahía Blanca 8000, Argentina, sdiaz@plapiqui.edu.ar*
[b]*Instituto Superior de Investigaciones Biológicas-INSIBIO (UNT-CONICET), Chacabuco 461, San Miguel de Tucumán 4000, Argentina, juan@fbqf.unt.edu.ar*

Abstract

In this work, we formulate a parameter estimation problem for a large-scale dynamic metabolic network. The DAE system represents the dynamic model for the Embden-Meyerhof-Parnas pathway, the phosphotransferase system and the pentose-phosphate pathway of *Escherichia coli* K-12 W3110 (Chassagnole et al., 2002), with modifications on several enzyme kinetics and the addition of fermentation reactions. Model parameters have been estimated based on recently published experimental data for this strain. Most sensitive parameters have been ranked by performing global sensitivity analysis on the dynamic metabolic network (Di Maggio et al., 2009a,b). Eleven kinetic parameters, including maximum reaction rates, inhibition and half-saturation constants, have been estimated with good agreement with available experimental data.

Keywords: dynamic metabolic network, dynamic optimization, control vector parametrization

1. Introduction

Intracellular and extracellular metabolite concentrations can now be measured, as well as protein levels and activities. The advances on experimental techniques and the consequent increase on the amount of accessible data on the dynamics of functioning cells pave the way to building dynamic models for metabolic networks, which in turn can predict microbial behavior and constitute important tools in metabolic engineering. Dynamic models provide time profiles for the concentration of metabolites involved in the metabolic network under study. They comprise a nonlinear differential algebraic system of equations, which arise from mass balances for metabolites and have a large number of kinetic parameters that require tuning for a specific growth condition.

Chassagnole et al. (2002) applied simulated annealing to estimate all the parameters, around one hundred, in a model describing a large-scale metabolic network for *Escherichia coli*. Ceric and Kurtanjek (2006) compared three different estrategies, Nelder and Mead (1965) optimization, Simulated annealing and Differential evolution, to estimate kinetic parameters of a dynamic model of central carbon metabolism of *Escherichia coli*.

In this work, we propose dynamic model extensions for the phosphotransferase system, the Embden-Meyerhof-Parnas pathway, pentose-phosphate pathway, as well as fermentation reactions of *Escherichia coli* K-12 W3110 (Chassagnole et al., 2002) to formulate a parameter estimation problem subject to this differential algebraic system,

within a parameter optimization framework in gPROMS. Numerical results provide maximum reaction rates and half-saturation and inhibition constants for several enzymes from the main metabolic pathways.

2. Mathematical modeling

2.1. Dynamic model of metabolic network

In this work, we have introduced several modifications on the dynamic model representing the Embden-Meyerhof-Parnas pathway, the pentose-phosphate pathway and the phosphotransferase system of *Escherichia coli* K-12 W3110 by Chassagnole et al. (2002). Within the glycolysis pathway, the phosphofructokinase (PFK) kinetics was taken from Diaz Ricci (1996). We have added fermentation reactions, including acetate, formate, lactate, ethanol and succinate production. Kinetic expressions for fermentation pathways for lactate, ethanol and acetate were taken from Hoefnagel et al. (2002). Succinate pathway was modeled with Michaelis-Menten kinetics. In the formate synthesis pathway, the kinetics for pyruvate-formate lyase (PFL) was taken from Knappe et al. (1974). The resulting model comprises twenty five differential equations that represent dynamic mass balances of extracellular glucose, intracellular metabolites and fermentation products, thirty seven kinetic rate expressions and seven additional algebraic equations for co-metabolites. Equation 1 shows mass balance for extracellular glucose; Equations 2 and 3 show general expressions for mass balances on intracellular metabolites and fermentation products, respectively. As it can be seen from Equations 4 to 9 kinetic expressions have a large number of parameters that must be estimated from experimental data. However, not all kinetic parameters are feasible candidates for tuning: In this sense maximum reaction rates, which are related to enzyme concentration and inhibition and half-saturation constants could be estimated and, among them, the most influential parameters have been determined through global sensitivity analysis.

$$\frac{dC_{glc}^{ext}}{dt} = D(C_{glc}^{a\lim} - C_{glc}^{ext}) + f^{pulso} - \frac{C_X r_{PTS}}{\rho_X} \tag{1}$$

$$\frac{dC_i}{dt} = \sum_{k \in S_{enz}} v_{ik} r_k - \mu C_i \quad i = 1,..,NC_i \tag{2}$$

$$\frac{dC_i}{dt} = r_j \left(\frac{C_x}{\rho} \right) - DC_i \quad i = 1,..,NC_e \tag{3}$$

where

NC_i=20, number of intracellular metabolites (g6p, f6p, fdp, dhap, gap, pgp, 3pg, 2pg, pep, pyr, g1p, 6pg, ribu5p, xyl5p, rib5p, sed7p, e4p, accoa, acetaldehyde, acetil-P)

NC_e=5, number of fermentation products (acetate, formate, lactate, ethanol and succinate)

$$r_{PFK} = \frac{r_{PFK}^{\max} \left[\frac{eC_{f6p}}{K_{rf6p}} \left(1 + \frac{eC_{f6p}}{K_{rf6p}}\right)^{n_{PFK}-1} \left(1 + \frac{C_{adp}}{K_{radp}}\right)^{n_{PFK}} + L\theta ce' \frac{C_{f6p}}{K_{rf6p}} \left(1 + ce' \frac{C_{f6p}}{K_{rf6p}}\right)^{n_{PFK}-1} \right]}{\left(1 + \frac{eC_{f6p}}{K_{rf6p}}\right)^{n_{PFK}} \left(1 + \frac{C_{adp}}{K_{radp}}\right)^{n_{PFK}} + L\left(1 + ce' \frac{C_{f6p}}{K_{rf6p}}\right)^{C_{pep}/K_{ipep}}} \tag{4}$$

Parameter Estimation in Kinetic Models for Large Scale Metabolic Networks with
 Advanced Mathematical Programming Techniques

$$L = L_0 \left[\frac{\left(1+\frac{C_{pep}}{K_{tpep}}\right)\left(1+\frac{C_{adp}}{K_{tadp}}\right)}{\left(1+\frac{C_{pep}}{K_{rpep}}\right)\left(1+\frac{C_{adp}}{K_{radp}}\right)} \right]^{n_{PFK}} \tag{5}$$

$$r_{PFL} = \frac{r_{PFLfor}^{max} C_{pyr} C_{CoA}}{\left(C_{pyr} C_{coA} + C_{pyr} K_{mB} + C_{coA} K_{mA}\right)} - \frac{r_{PFLrev}^{max} C_{formate} C_{AcCoA}}{\left(C_{formate} C_{AcCoA} + C_{formate} K_{mD} + C_{AcCoA} K_{mC}\right)} \tag{6}$$

$$r_{PDH} = \frac{r_{PDH}^{max} c_{pyr}^{n_{PDH}}}{K_{PDH,pyr} + c_{pyr}^{n_{PDH}}} \tag{7}$$

$$r_{PGDH} = \frac{r_{PGDH}^{max} c_{6pg} c_{nadp}}{\left(c_{6pg} + K_{PGDH,6pg}\right)\left(c_{nadp} + K_{G6PDH,nadp}\left(1+\frac{c_{nadph}}{K_{PGDH,nadph,inh}}\right)\left(1+\frac{c_{ATP}}{K_{PGDH,ATP,g6pinh}}\right)\right)} \tag{8}$$

$$r_{PTS} = \frac{r_{PTS}^{max} C_{glc}^{extracellular} \frac{c_{pep}}{c_{pyr}}}{\left(K_{PTS,a1} + K_{PTS,a2} \frac{C_{pep}}{C_{pyr}} + K_{PTS,a3} C_{glc}^{extracellular} + C_{glc}^{extracellular} \frac{C_{pep}}{C_{pyr}}\right)\left(1+\frac{C_{g6p}^{n_{PTS,g6p}}}{K_{PTS,g6p}}\right)} \tag{9}$$

2.2. Experimental data

Experimental data correspond to an *Escherichia coli* K-12 culture perturbed with a glucose pulse (Degenring et al., 2004) throughout a time horizon of twenty seconds. Temporal profiles for glucose-6-phosphate (g6p), fructose-6-phosphate (f6p), fructose-1,6-diphosphate (fdp), pyruvate (pyr) and phosphoenolpyruvate (pep) were used for parameter estimation problem.

2.3. Parameter estimation problem

The parameter estimation problem has been formulated in g-PROMS (g-PROMS, 2007) as a Maximum Likelihood parameter estimation problem with constant variance. The problem has been formulated as follows:

$$\frac{N}{2} ln(2\pi) + min \sum_{i=1}^{NM} \sum_{j=1}^{NT} \left[ln(\sigma_{ij}^2) + \frac{\left(C_{ij}^M - C_{ij}\right)^2}{\sigma_{ij}^2} \right] \tag{10}$$

s.t.

DAE model
$C_i(0) = C_i^0$
$\boldsymbol{p}^L \leq \boldsymbol{p} \leq \boldsymbol{p}^U$

where the summation in the objective function is over *NM* measured state variables and *NT* data points for each measured variable; σ_{ij} is the variance of the *j*th measurement of variable *i*, which is determined by the measured variable's variance model whose elements correspond to variances of the measured variables. Vector $\boldsymbol{p}$ corresponds to estimated parameters.

3. Numerical results

Eleven model parameters have been identified as the most influential ones through a previous global sensitivity analysis applied to the DAE system representing the metabolic network for *E. coli* K-12 W3110 (Di Maggio et al., 2009[a,b]). These parameters stand for maximum reaction rates and half-saturation and inhibition constants for some enzymes that participate in the metabolic network.

Figure 1 shows dynamic sensitivity indices for fructose-6-phosphate (f6p) concentration; it can be noted that the most influential parameter within the first twenty seconds (time horizon for the parameter estimation), is $K_{PTS,f6p}$, which represents by-product inhibition constant for the phosphotransferase system. Regarding pyruvate (pyr) concentration (Fig. 2), it must be pointed out that the most influential parameter throughout the entire time horizon is r^{max}_{PDH}, the maximum reaction rate for pyruvate dehydrogenase (PDH), enzyme for which pyruvate is the substrate. Taking into account most important parameters for the entire set of intracellular metabolite concentrations, we have determined the set of eleven parameters for estimation that is shown in Table 1.
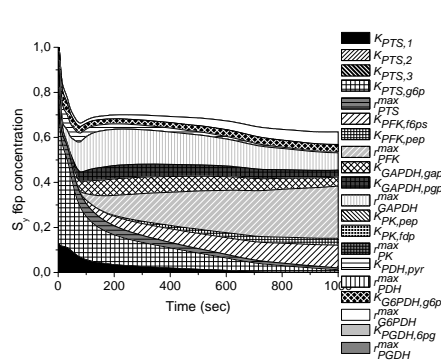

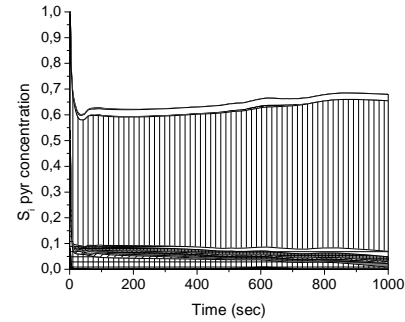Figure 1. Global sensitivity indices for f6p concentration

Figure 2. Global sensitivity indices for pyr concentration

Table 1. Estimated parameters: nominal and calibrated values

Parameter	Description	Nominal value	Calibrated value
$K_{PTS,g6p}$	Inhibition constant (mM)	1.5	1.362
$K_{r,f6p}$	Half-saturation constant (mM)	0.00156	0.00163
r_{PTS}^{max}	Maximum reaction rate (mM/sec)	1.2	1.381
$K_{GAPDH,gap}$	Half-saturation constant (mM)	0.569	0.652
$K_{GAPDH,pgp}$	Inhibition constant (mM)	0.000013	0.000012
r_{GAPDH}^{max}	Maximum reaction rate (mM/sec)	1132.390	1079.976
r_{PDH}^{max}	Maximum reaction rate (mM/sec)	13.244	12.442
$K_{G6PDH,g6p}$	Half-saturation constant (mM)	12.5	11.898
r_{G6PDH}^{max}	Maximum reaction rate (mM/sec)	0.469	0.544
$K_{PGDH6pg}$	Half-saturation constant (mM)	31.25	25
r_{PGDH}^{max}	Maximum reaction rate (mM/sec)	15.613	18.736

Parameter Estimation in Kinetic Models for Large Scale Metabolic Networks with Advanced Mathematical Programming Techniques
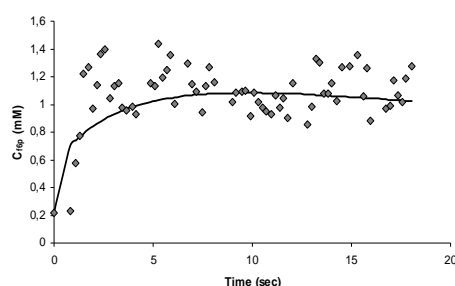

Figure 3. Fructose-6-phosphate concentration profile. Experimental data (Degenring et al., 2004) and simulation results
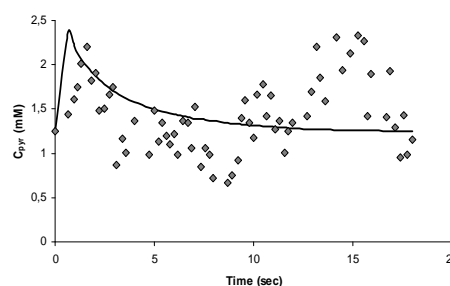

Figure 4. Pyruvate concentration profile. Experimental data (Degenring et al., 2004) and simulation results
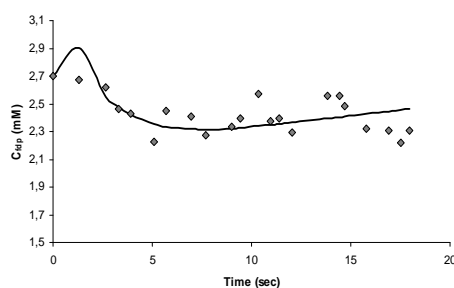

Figure 5. Fructose-1,6-diphosphate concentration profile. Experimental data (Degenring et al., 2004) and simulation results

Table 1 shows nominal values taken from the literature and calibrated values for the eleven estimated parameters. Experimental data and simulated profiles for fructose-6-phosphate (f6p), fructose-1,6-diphosphate (fdp) and pyruvate (pyr) concentrations are shown in Figs. 3 to 5. A good fit between experimental data and predicted profiles for can be seen for both f6p and fdp concentrations, while in the case of pyruvate concentration, it can be noted that experimental data are rather disperse and the best fit we could determine is shown in Fig. 4.

4. Conclusions

We have formulated and solved a dynamic parameter estimation problem for a large-scale metabolic network for the phosphotransferase system, glycolysis, pentose-phosphate pathway and fermentation pathways. Eleven kinetic parameters have been estimated, which had been previously identified as the most influential ones through a global sensitivity analysis. To our knowledge, it is the first time the parameter estimation problem for a large-scale metabolic network has been addressed with advanced mathematical programming techniques (gPROMS, PSE Enterprise, 2007). A satisfactory fit was observed between the predicted profiles and the experimental data, not only the recently reported but also previously published ones (Chassagnole et al., 2002), not reported in this work. Current research includes the determination of additional experimental data for this system.

5. Acknowledgements

References

S. Ceric, Z. Kurtanjek, 2006, Model Identification, Parameter Estimation and Dynamic Flux Analysis of E. coli Central Metabolism, Chemical and Biochemical Engineering Quarterly, 20, 3, 243-253.

C. Chassagnole, N. Noisommit-Ricci, J. Schmid, K. Mauch, M. Reuss, 2002, Dynamic Modeling of the Central Carbon Metabolism of *Escherichia coli*, Biotechnology and Bioengineering, 79, 1, 54-73

D. Degenring, C. Froemel, G. Dikta, R. Takors, 2004, Sensitivity analysis for the reduction of complex metabolism models, Journal of Process Control, 14, 729-745.

J. C. Diaz Ricci, 1996, Influence of Phosphoenolpyruvate on the Dynamic Behavior of Phosphofructokinase of *Escherichia coli*, Journal of Theoretical Biology, 178, 2, 145-150.

J. Di Maggio, J. Diaz Ricci, M.S. Diaz, 2009[a], Global Sensitivity Analysis in Dynamic Metabolic Networks, Computer Aided Chemical Engineering, 26, 1075-1080.

J. Di Maggio, J. Diaz Ricci, M.S. Diaz, 2009[b], Global Sensitivity Analysis in Dynamic Metabolic Networks, Computers and Chemical Engineering, under review.

V. Estrada, E. Parodi, M.S. Diaz, 2009, Determination of biogeochemical parameters in eutrophication models with simultaneous dynamic optimization approaches, Computers and Chemical Engineering, 33, 1760-1769.

g-PROMS, PSE Enterprise, 2007, www.psenterprise.com.

M. Hoefnagel, M. Starrenburg, D. Martens, J. Hugenholtz, M. Kleerebezem, I. Van Swam, R. Bongers, H. Westerhoff, J. Snoep, 2002, Metabolic engineering of lactic acid bacteria, the combined approach: kinetic modelling, metabolic control and experimental analysis, Microbiology, 148, 1003-1013.

J. Knappe, H. Blaschkowski, P. Gröbner, T. Schmitt, 1974, Pyruvate Formate-Lyase of *Escherichia coli*: the Acetyl-Enzyme Intermediate, European Journal of Biochemistry, 50, 1, 253-263.

J. A. Nelder and R. Mead, "A simplex method for function minimization", Computer Journal, 1965, 7, 308-313

A Decomposition Approach for Parameter Estimation of System of Ordinary Differential Equations

Vivek Dua

Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, London WC1E 7JE, E-mail: v.dua@ucl.ac.uk

Abstract

The parameter estimation problem for ordinary differential equations (ODE) is decomposed into two sub-problems. The first sub-problem corresponds to developing an Artificial Neural Network (ANN) model for the given data. The second subproblem is formulated as a parameter estimation problem where the differential terms in the model are obtained by analytically differentiating the ANN model. The second subproblem corresponds to a Linear Program (LP) or a Nonlinear Program (NLP) involving only algebraic variables, for which reliable solvers are available. The solution of the second subproblem does not involve integration of differential equations and therefore the solution is much easier to obtain than using the traditional parameter estimation techniques.

Keywords: Dynamic Optimization, Global Optimization, Artificial Neural Networks

1. Introduction

Advanced design and optimization relies on developing predictive models that can accurately represent the various phenomena taking place in a given system. Good quality models can be developed by using the experimentally observed data to compute the optimal values of the parameters involved in the model proposed for the system. Parameter estimation for models involving differential equations corresponds to a dynamic optimization problem [1], which in general has multiple locally optimal solutions [2-5]. The solution techniques for global optimization can be broadly classified as based upon deterministic and non-deterministic methods. Deterministic methods can guarantee that the solution obtained is globally optimal within a certain pre-specified tolerance [6] whereas the non-determinstic methods do not provide any such guarantee. However global optimization techniques are computationally demanding. The main objective of this work is to develop an algorithm which can speed-up the solution times, reduce solver failures, and increase possibility of obtaining the globally optimal solution. The rest of the paper is organized as follows. In the next section problem formulation and a new solution technique are presented. Section 3 presents an illustrative example where the proposed solution approach is tested and concluding remarks are given in section 4.

2. Problem Statement and Proposed Solution Approach

2.1. Problem Definition

Parameter estimation for a system of ordinary differential equations (ODE) is usually formulated as the following optimization problem:

Problem P1:

$$\varepsilon_1 = \min_{\boldsymbol{\theta}, \mathbf{z}(t)} \sum_{i \in I} \sum_{j \in J} \left(\hat{z}_j(t_i) - z_j(t_i) \right)^2 \qquad (1)$$

$$\text{subject to}: \quad \frac{dz_j(t)}{dt} = f_j(\mathbf{z}(t), \boldsymbol{\theta}, t), \quad j \in J \qquad (2)$$

$$z_j(t = 0) = z_0, \quad j \in J \qquad (3)$$

$$t \in [t_o, t_f] \qquad (4)$$

where **z** is the *J* dimensional vector of state variables in the given ODE system, $\hat{z}(t_i)$ are the experimentally observed values of the state variables at time points t_i, and $\boldsymbol{\theta}$ is the vector of parameters that must be estimated such that the error, ε_1, between the observed values and the model predictions of state variables is minimized.

2.2. Decomposition Approach for Parameter Estimation

In this work, the parameter estimation problem is decomposed into two subproblems. The first subproblem uses the data, $\hat{z}(t_i)$, to obtain an approximate model based upon Artificial Neural Networks (ANN). The second subproblem uses the ANN approximation of the model to obtain the derivatives of state variables, representing the left hand side (LHS) of equation (2) and to also evaluate the right hand side (RHS) of equation (2). An optimization problem is formulated so as to minimize the sum of the square of the difference between the LHS and the RHS of equation (2). This optimization problem is much simpler than the original parameter estimation problem, (1)-(4), and involves only $\boldsymbol{\theta}$, the algebraic variables and does not involve **z**, the differential variables, as described next.

2.2.1. ANN Model

The observed data is used to construct an ANN model, the inputs are $\mathbf{x}(t_i) = [t_i]$ and the outputs are $\hat{z}_j(t_i)$. The ANN model for one hidden layer is given as follows [7]:

$$\mathbf{z}^{ANN}(t) = \mathbf{w}_2.\mathbf{h}(t) + \mathbf{b}_2 \qquad (5)$$

$$\mathbf{h}(t) = \tanh(\mathbf{a}(t)) \qquad (6)$$

$$\mathbf{a}(t) = \mathbf{w}_1.\mathbf{x}(t) + \mathbf{b}_1 \qquad (7)$$

where $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{b}_1$ and $\mathbf{b}_2$ are the constant weight matrices and bias vectors respectively. Note that $\mathbf{h}(t)$ is the output vector of the hidden layer which is obtained by nonlinear tanh transformation of $\mathbf{a}(t)$, where $\mathbf{a}(t)$ is obtained by a linear combination of the input vector $\mathbf{x}(t)$. The output vector of the network, $\mathbf{z}^{ANN}(t)$, is obtained by a linear combination $\mathbf{h}(t)$. In this work, tanh nonlinear transformation was used, other transformations are also available and can be used. The numerical values of $\mathbf{w}_1$, $\mathbf{w}_2$, $\mathbf{b}_1$ and $\mathbf{b}_2$ are obtained by minimizing the sum of the squares of the difference between $z_j^{ANN}(t_i)$ and $\hat{z}_j(t_i)$:

Problem P2:

$$\varepsilon_2 = \min_{\mathbf{z}^{ANN}(t),\mathbf{a}_1,\mathbf{a}_2,\mathbf{b}_1,\mathbf{b}_2,\mathbf{w}_1,\mathbf{w}_2} \sum_{i \in I} \sum_{j \in J} \left(\hat{z}_j(t_i) - z_j^{ANN}(t_i) \right)^2 \tag{8}$$

subject to :   equations $(5) - (7)$

This problem can be solved efficiently by using various algorithms from the open literature.

2.2.2. Parameter Estimation
Differentiating (5) with respect to $\mathbf{x}$ we obtain:

$$\frac{d\mathbf{z}^{ANN}(t)}{d\mathbf{x}(t)} = \mathbf{w}_1.\mathbf{w}_2.\text{sech}^2(\mathbf{b}_1 + \mathbf{w}_1.\mathbf{x}(t)) \tag{9}$$

The right hand side of equation can be simplified to $\mathbf{w}_1.\mathbf{w}_2.(1-\mathbf{h}(t)^2)$. Note that the first column on the left hand side of equation (11) represents $\dfrac{d\mathbf{z}^{ANN}(t)}{dt}$. Parameter estimation is formulated as the following nonlinear programming (NLP) problem:

Problem P3:

$$\varepsilon_3 = \min_{\boldsymbol{\theta}} \sum_{k \in K} \sum_{j \in J} \left(\frac{dz_j^{ANN}(t_k)}{dt} - f_j(\mathbf{z}^{ANN}(t_k),\boldsymbol{\theta},t_k) \right)^2 \tag{10}$$

In this problem $\dfrac{dz_j^{ANN}(t)}{dt}$ is obtained from equation (9) and $f_j(\mathbf{z}^{ANN}(t),\boldsymbol{\theta},t)$ is evaluated by using equation (5). Note that the values of the differential terms, $\dfrac{dz_j^{ANN}(t)}{dt}$, are available and therefore (10) does note require an integration scheme. The NLP given by (10) is much simpler to solve than a traditional parameter estimation problem, P1, involving integration of differential equations.

In problem P3, the time points are given by t_k, which are necessarily not the same as the original time points, t_i, i.e., one can also use the ANN model to estimate the model predictions at time points where experimental data is not available. One can therefore also use a reduced data set for solving the parameter estimation problem, as shown in Figure 1.
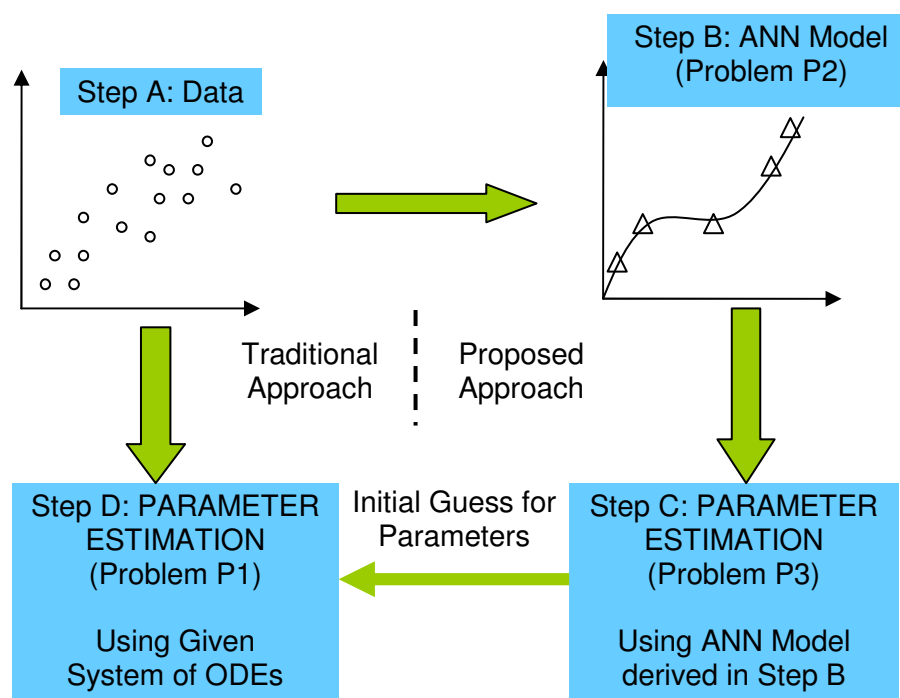

Figure 1: Traditional versus proposed approach for parameter estimation. The original data set (given by circles in step A) is used to obtain the ANN model as shown in step B. A reduced number of data points (given by triangles in step B) and the differentials of the ANN model at the reduced number of data points are used to formulate and solve the parameter estimation problem in step C

It may be noted that original problem P1 minimizes the error between the observed and model predicted values of state variables, whereas problem P3 minimizes the error between the derivatives of the state variables. The proposed decomposition based parameter estimation methodology is summarized in Figure 1. This methodology was tested on several parameter estimation problems and the solution was obtained reliably and in a few iterations only; one such example is discussed in the next section.

3. Illustrative Example

Consider the following first-order irreversible chain reactions [2-3]:

$$A \quad \overset{k_1}{\rightarrow} \quad B \quad \overset{k_2}{\rightarrow} \quad C$$
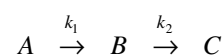
where the model is given by:

A Decomposition Approach for Parameter Estimation

$$\frac{dz_1}{dt} = -\theta_1 z_1 \tag{11}$$

$$\frac{dz_2}{dt} = \theta_1 z_1 - \theta_2 z_2 \tag{12}$$

where z_1 and z_2 are concentrations of *A* and *B* respectively and θ_1 and θ_2 are the rate constant parameters, k_1 and k_2 respectively, to be estimated. The simulated data, $\hat{z}_1$ and $\hat{z}_2$, was generated at $t = t_i$ for $k_1 = 5$, $k_2 = 1$ and $z(t=0) = [1\ 0]$, and is given in Table 1. Solution of problem P2 for this example gives $\varepsilon_2 = 4.95E\text{-}7$ for three nodes in the hidden layer of the ANN. Note that problem P3 for this example is given by a simple quadratic program with quadratic objective function and linear constraints in $\mathbf{\theta}$. Solving problem P3 gives $\varepsilon_3 = 7.437E\text{-}4$, $k_1 = 5.017$ and $k_2 = 1.001$, which is quite close to the values of k_1 and k_2 used for obtaining the data. A comparison of the data, ANN model and model predictions is shown in Figures 2 and 3.

Table 1. Data for illustrative example

| | | | | Approximate solution from ANN model | | | |
| | | Observed Data | | Evaluation of ANN model equations (5) – (7) for t_i | | Evaluation of differential of the ANN model equation (9) for t_i | |
i	t_i	$\hat{z}_1$	$\hat{z}_2$	$\hat{z}_1^{ANN}$	$\hat{z}_2^{ANN}$	$\dfrac{dz_1^{ANN}}{dt}$	$\dfrac{dz_2^{ANN}}{dt}$
0	0	1	0	0.998	0.002	-4.641	4.577
1	0.1	0.606	0.373	0.611	0.367	-3.106	2.746
2	0.2	0.368	0.564	0.366	0.566	-1.880	1.315
3	0.3	0.223	0.647	0.220	0.650	-1.104	0.448
4	0.4	0.135	0.669	0.134	0.669	-0.656	-0.017
5	0.5	0.082	0.656	0.083	0.654	-0.400	-0.251
6	0.6	0.050	0.624	0.051	0.622	-0.250	-0.365
7	0.7	0.030	0.583	0.031	0.583	-0.158	-0.418
8	0.8	0.018	0.539	0.018	0.540	-0.097	-0.440
9	0.9	0.011	0.494	0.011	0.496	-0.055	-0.445
10	1.0	0.007	0.451	0.007	0.451	-0.025	-0.440

4. Concluding Remarks

The main advantage of using the proposed decomposition approach is that the error between the data and model predictions is minimized in the first step and the parameter estimation for a reduced data set is carried out in the second step. ANN are well known for their ability to handle large data sets and characterize highly nonlinear functions

very effectively. This ability together with the differentiability properties of ANN makes the proposed approach a very useful tool for parameter estimation of problems involving differential equations and large data sets.
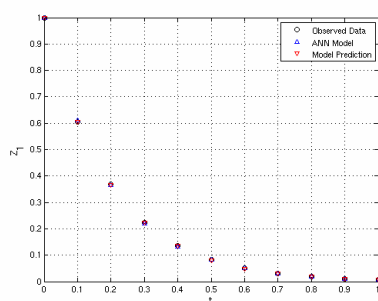

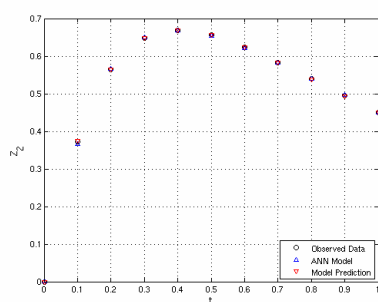Figure 2: Concentration of A (z_1) as a function of time (t)

Figure 3: Concentration of B (z_2) as a function of time (t)

5. Acknowledgements

References

[1] V.S. Vassiliadis, R.W.H. Sargent and C.C. Pantelides. Solution of a class of multistage dynamic optimization problems 1. Problems without Path Constraints. Ind. Eng. Chem. Res. 33 (1994) 2111.

[2] W.R. Esposito and C.A. Floudas. Global optimization for the parameter estimation of differential-algebraic systems. Ind. Eng. Chem. Res. 39 (2000) 1291.

[3] S. Katare, A. Bhan, J.M. Caruthers, W.N. Delgass, W.N. and V. Venkatasubramanian. A hybrid genetic algorithm for efficient parameter estimation of large kinetic models. Computers and Chemical Engineering. 28 (2004) 2569.

[4] C.G. Moles, P. Mendes and J.R. Banga. Parameter estimation in biochemical pathways: A comparison of global optimization methods. Genome Res. 13 (2003) 2467.

[5] I. Papamichail and C.S. Adjiman. Global optimization of dynamic systems. Computers & Chemical Engineering. 28 (2004) 403.

[6] V. Dua, K.P. Papalexandri and E.N. Pistikopoulos. Global optimization issues in multiparametric continuous and mixed-integer optimization problems. Journal of Global Optimization. 30 (2004) 59.

[7] V. Dua. A mixed-integer programming approach for optimal configuration of artificial neural networks, Chemical Engineering Research and Design (2010) 88, 55.

Parameter Estimation using Multi-Wavelets

Heinz A Preisig

Dept of Chemical Engineering, NTNU, Trondheim, Norway
heinz.preisig@chemeng.ntnu.no

Abstract

Model identification is a common operation in connection with model-based operation activities, such as design, control, scheduling and planning. Many methods have been suggested indicating the existence of basic problems with matching plants to mathematical models. Four issues are discussed here (i) the mathematical structure of the identification problem, (ii) high frequency contents in excitation signals, (iii) presence of parasitic stochastic signal components and (iv) the effect of neglected parasitic dynamics.

Keywords: Process identification, observers, model-based operations

1. The Stage

Models are a base ingredient to nearly all activities related to design and operation of plants making modelling a core business for chemical engineers. The basic physical principles are an important source of models. These mechanistic models often have some shady parts that do not quite describe the behaviour of the respective part of the plant accurately but often are rather crude approximations. Thus models must normally be fitted thereby adjusting the odd parameter so as to match the model's behaviour to the one of the plant making system and model identification an everyday activity. Identification has three major ingredients: (i) input/output data, (ii) a model set, and (iii) a criterion. For (i): the data are generated by experiments. For (ii): If the model set consists of one parameterised model, then one talks about a model identification or parameter estimation problem. In the case where the nature of the model changes, one uses the term system identification and obviously system identification can be combined with model identification. For (iii): The criterion provides the measure for the gap between the model and the plant under the conditions imposed by the input signal. Any identification method will minimize the gap by choosing the best model from the set, which is done with respect to structure in the case of system identification and with respect to parameters in the case of model identification or parameter estimation.

Process identification is seen as an art and the continuous stream of publications suggesting improved methods and procedures are a vivid indication for the existence of many practical problems. The objective of this paper is to shed some light on the source of these problems suggesting a view angle that provides an excellent insight. The issues being touched are: problems with (i) the mathematical structure of the problem, (ii) high frequency contents in excitation signals, (iii) presence of parasitic stochastic signal components and (iv) the effect of neglected parasitic dynamics. The work was triggered by a presentation in a PhD defense [0] on yet another approach to identification in an industrial environment as the core subject. The used toy system will serve us as a discussion case, but first two facts.

2. Two Facts

Fact 1: Methods suitable for the identification of dynamic systems make intrinsic use of differentiated signals. For example in the celebrated Kalman filter and its variations it is the derivative of the state that is computed internally. Increasing the gain in the feedback of the output error also is approaching a differentiation as the gain is being increased. The associated subject is high-gain observers.

Fact 2: Multi-wavelets all approach in the limit the differential operator [1] . Several types of multi-wavelets are discussed in the literature. Maletinsky [2] derived B-spline functions from a system-theoretical argument formulating an identification problem, which was extended in [3] [4] [5] . The resulting functions are identical to what is being derived from the a signal processing point of view in [1] extending Shannon's sampling theory to non-ideal sampling systems[1]. In identification the multi-wavelet functions are being used as filters, which at the same time also provides integral measures of the derivative of the signal with respect to time. With the idea being anchored in various places, a number of properties have been discovered: recursive construction [5] [1] and B-spline literature, filtering properties [4] , satisfy Riesz condition [1] , bi-orthogonality [3] [4] [1] to mention the main ones.[2] The signal processing operation being applied can be represented as a convolution of the signal with the wavelet's base function. Loosely adopting the notation of [4] [5] we define

$$\tilde{y}^{(i)} := \frac{d^i}{dt^i} \int_0^{nT} y(\tau)\, \Phi(t - \tau)\, d\tau = \int_0^{nT} y(\tau)\, \Phi^{(i)}(t - \tau)\, d\tau$$

with the equivalence being the consequence of the wavelets having a compact base. This enables us to simply replace the derivative in a model with the modulated derivative being $\tilde{y}^{(i)}$. The time in the above equations is the relative time and the parameter n and T is the order of the wavelet and the characteristic time, respectively.

3. The Case

A relatively well-known toy example will be taken as a demonstration object [0] . It is a gas reactor consisting of a container, shape not essential, assumed to be ideally mixed and empty at the beginning and filled in an instance with a gas A, latter being a rather unrealistic condition. The molecules of A are spontaneously reacting to a single molecule B. The kinetics is of second order in the partial pressure of A. Isothermal conditions are being assumed, which makes the reaction "constant" k indeed to be constant. In a first step a physical model is assembled, which is translated into system notation by defining the component masses as the state, the component mass inflow as the input and the total pressure as the output:

	Physical representation	Systems representation
Component mass balances	$\underline{\dot{\mathbf{n}}} := \underline{\hat{\mathbf{n}}} + \underline{\mathbf{s}}\,\tilde{n}$	$\underline{\dot{\mathbf{x}}} = \underline{\mathbf{F}}\,\underline{\mathbf{u}} + \underline{\mathbf{s}}\,z(\underline{\mathbf{x}})$
Production rate	$\tilde{n} := V\,\tilde{r}$	$z(x) := a(k)\,x_1^2$
Reaction rate	$\tilde{r} := \frac{k}{|\nu_A|}\,p_A^2$	$a(k) := k\,\alpha\,\beta^2$
Partial pressure	$\underline{\mathbf{p}} := \underline{\mathbf{n}}\,\frac{R\,T}{V}$	$y := \beta\,\underline{\mathbf{e}}^T\,\underline{\mathbf{x}}$

The total pressure is the sum of the two partial pressures. The vector $\underline{\mathbf{e}}^T$ is a row vector of ones, thus summing up the component masses, with $\beta := \frac{R\,T}{V}$ and $\alpha := \frac{V}{|\nu_A|}$.

---

[1] *This implies a link between sampling theory and the theory of linear dynamic systems.*
[2] *For details the reader is referred to the cited literature.*

Parameter Estimation Using Multi-Wavelets

4. The Issues

As identification is such a central activity in the application of models in the process industry, four ingredients in the procedure are put into the limelight of this exposition:
1) the structure of the problem yielding the information of what information is required and what underlaying mathematical problem needs to be solved. 2) The experiment that is to be performed in order to extract the desired information. 3) The effect stochastic signal components have on the outcome and finally 4) the effect neglected dynamics have on the estimated parameters.

4.1. The Structure

The classification is done with respect to a number of attributes, the main ones being listed in a non-precedence order: a) linear, non-linear in state and/or parameters, b) stochastic or deterministic, c) static or dynamic, d) discrete or continuous. The wavelet approach brings all these structural components to the light in a jiffy.

In a first instance let us not be concerned with the stochastic component, the noise issue, but focus on the structure instead using the fact that the wavelet approach provides measures for the derivatives. The first observation is that the state is not directly measurable, but that the two state variables, namely the partial pressures must be reconstructed from the total pressure measurement using the model. The derivation follows the classical observability derivation by differentiating the observation:

$$y := \beta \, \underline{e}^T \, \underline{x} \quad \Rightarrow \quad \dot{y} := \beta \, \underline{e}^T \, \underline{\dot{x}} := \beta \, \underline{e}^T \, (\underline{s} \, r + \underline{\underline{F}} u)$$

Isolating what can be measured and defining a new variable for this part:

$$\underline{w} := \dot{y} - b \, \underline{e}^T \underline{\underline{F}}, \qquad u := \beta \, \underline{e}^T \underline{s} r \quad \text{and} \qquad \underline{\dot{w}} := y^{(2)} - b \, \underline{e}^T \underline{\underline{F}}, \qquad \dot{u} := \beta \, \underline{e}^T \underline{s} \dot{r}$$

from which one gets

$$\frac{w}{\dot{w}} := \frac{r}{\dot{r}} \quad := \frac{a(k) \, x_1^2}{a(k) \, 2 \, x_1 \, \dot{x}_1} \quad := \frac{x_1}{2 \, \dot{x}_1}$$

to be solved for the reaction rate and the sought partial pressure:

$$r := \frac{w}{b \, \underline{e}^T \underline{s}} \quad \text{and} \quad x_1 := 2 \, \frac{w}{\dot{w}} \, (s_1 \, r + f_1 \, u)$$

So far the problem is linear and it is only now that the nonlinearity hits.

$$r := a(k) \, x_1^2 \quad \therefore \quad a(k) := r \, x_1^{-2} \quad \therefore \quad k := \frac{a(k)}{\alpha \, b^2}$$

The concept can be generalized by isolating the reaction part from the network representation [6] [7] :

Plant model	compute reaction rates
$\underline{\dot{x}} := \underline{\underline{F}} \, \underline{\hat{x}} + \underline{S} \, \underline{r}$	$\underline{S} \, \underline{r} := \underline{\dot{x}} - \underline{\underline{F}} \, \underline{\hat{x}}$
$\underline{\hat{x}} :: \text{measured}$	$\underline{\underline{S}} :: \text{invertible for simplicity}$
$\underline{r} := V \, \underline{\underline{K}} \, \underline{g}(\underline{x})$	$\underline{r} := \underline{\underline{S}}^{-1} \, (\underline{\dot{x}} - \underline{\underline{F}} \, \underline{\hat{x}})$

which then can be regularized and used as the data input for the estimation of the reaction constants. Sine the rates are a linear function of the rate constants, the problem is a linear one [8] . This idea, whilst not new, has been become quite popular through the work of the Aachen group. See for example [9] .

Conclusions: @ Starting with the assumption that derivative information is available, which it is in approximate form through wavelet-modulation, the structure of the combined observer and parameter estimation problem stands out clearly. @ What is being required in terms of information from the plant and the model becomes apparent.

@ The incremental identification method for reactive systems becomes an easy to derive concept.

4.2. The Excitation Signal

The excitation signal is one of the main ingredients to process identification. Its nature has a profound effect on the results. The most commonly used signal is the step. From a frequency point of view, it injects all frequencies: The high frequencies are most active at the beginning of the step and loose significance as time passes with the low frequencies gradually taking over. The use of high frequency inputs has several undesirable effects. For one the models are in most cases not designed to actually describe the behaviour of the plant on that frequency level, they are simply not detailed enough to capture essentially infinitely fast dynamics at their boundary. Secondly, as one uses at least intrinsically differential information of signals, the use of non-smooth inputs is not a good idea: the differential operator is unbounded. And yes one could use filtering, and must use filtering to reduce the effect of the stochastic components, it is nevertheless not very wise to use such an input. Figure 1[3] shows process' responses to a step at time 0 with pure A and a pulse of A injected at time 10 of length 1 all sampled at three different rates.
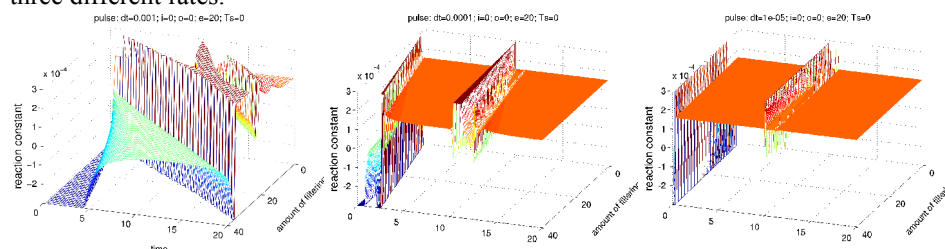

Figure 1: *Demo plant without noise, three different sampling rates. Excitation is a step at time zero and a 1 time-unit pulse at time 10. The flat piece reflects the true value of the parameter, the reaction constant. Surface is the estimated reaction constant as a function of the length of the wavelet filter thus varying the "amount of filtering".*

It is obvious that the high-frequency parts occurring at the very beginning and the pulse up an down causes problems for the identification.
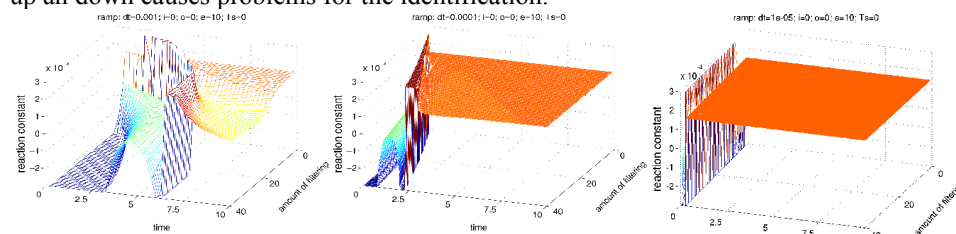

Figure 2: *Demo plant without noise, the same three different sampling rates. Excitation a ramp starting at time 0 all the way to time 10.*

Figure 2 shows the situation when changing to a ramp input (time 0 - 10), which is still not smooth, but imposes much less intense high frequencies.

Conclusions: @ The input signal's frequency contents should match the fidelity of the model with respect to its ability to describe the plant's behaviour on this level of excitation on its boundary. @ The signal should be chosen such that it is sufficiently differentiable where the level of differentiation can be determined by analysing the

---

[3]*The figures show: **up** the identified reaction constant, **front** time in samples, **back** filtering (time resolution) increasing to the back, or amount of filtering increasing to the front.*

Parameter Estimation Using Multi-Wavelets

structure of the combined observation and identification problem. @ More restrictive: the input signal should be band limited.

4.3. The Stochastic Signal Components

Having stochastic components superimposed on the process signals is the source of another problem namely the information contents in the signal. The stochastic component, often termed "noise" is a parasitic phenomena. To quantify the information contents of the signal one often uses the signal to noise ratio, which is the ratio of the power of the two signal components. Alternatively one may use the ratio of the expected signal value and the variance or standard deviation or more sophisticated a entropy measure. Input signals like a step or a long pulse, which are commonly used in chemical engineering, have not only the problem of the high frequencies being most active in the vicinity of where the jump occurred, but also that after some time the process adapts to the change. The input does not inject any new information after the step has occurred, and the output is closing in on the new stationary point, given the process is stable. Thus the information contents reduces, whilst the noise component remains usually about constant. Figure 3 shows the a comparison between the step-pulse and the ramp input having imposed "noise" both on the input and the output. Clearly the pulse response shows that more and more filtering is required as one moves away from the step-event time. Reason being that the signal noise ration decreases with time.
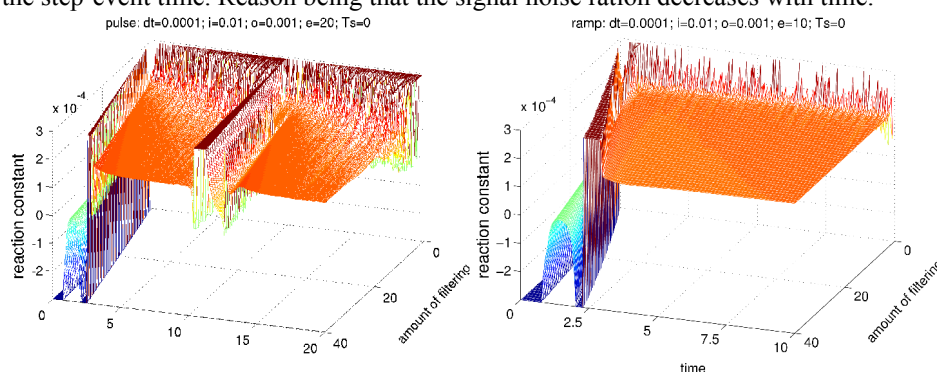

Figure 3: *Demo plant with noise both on input and output. Left a pulse input and on the right a ramp input. The ramp experiment is half the length of the pulse experiment.*

Clearly the ramp provides the better estimates under the same conditions. The obvious problem being that one has to cope with the practical effects associated with an increasing input, but then there are certainly alternatives to ramps, specifically one could use the wavelets basic functions as inputs.

4.4. The Effect of Neglected Dynamics

The model being used for the description of the process is never perfect, simply because the model IS NOT the plant. Beyond that though, one also neglects more or less knowingly dynamic components associated with controlling and observing the process. Very commonly one forgets about the dynamics of the sensor though neglecting such dynamics clearly has an effect on the estimated parameters during dynamic periods. The figures qualitatively show that the dynamic behaviour of the identified parameter changes during the highly dynamic period but dies away as the process approaches equilibrium (pulse input). The effect is "permanent" for a continuously dynamic input, as the ramp is. There the value of the identified parameter is affected also in the relative constant phase of the parameter identification.
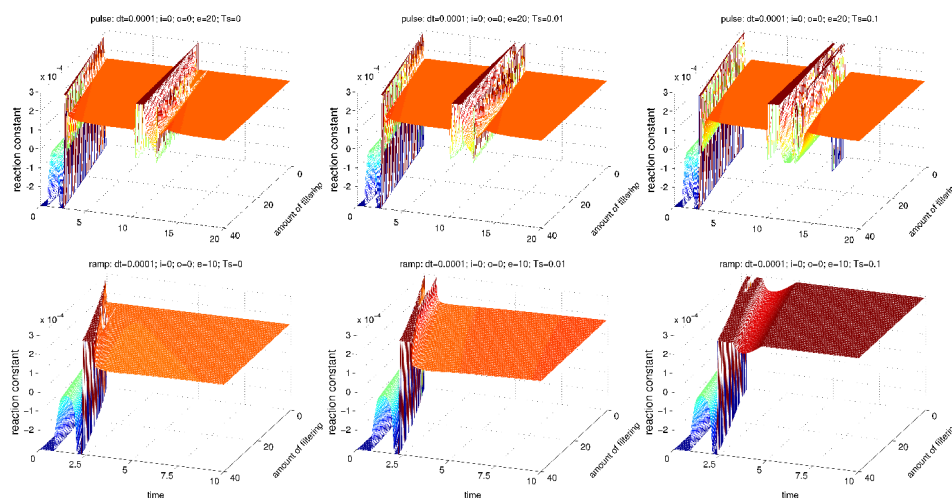
Figure 4: *Parasitic dynamic increases from the left to the right, starting with none. No parasitic stochastic components. The parameter value clearly changes in the case of the ramp as input. In the dynamic domain, the behaviour of the constant is smoothened.*

Conclusions: The presence of parasitic unmodelled dynamics affects the estimate in the dynamic domain.

5. Final Remarks

Changing the point of view and focusing on the structure of the problem provides problem-essential insight. Acknowledging the fact that one uses filtered, but differential signal information in model identification and also in often needed state reconstruction is the key to the analysis, which not only reveals the structure of the estimation problem but also provides guidelines to what input signals are suitable: The input signal must be smooth to the degree of the used time-differentiated signals and it must be band limited. The modern sampling theory is closing in on identification. Wavelets being for three decades rejected by this community have a broad theoretical basis relevant to identification, sampling and signal reconstruction.

References

[0]   Kolaas, Steinar, **2008**, Estimation in nonlinear constrained system with severe disturbances; PhD Thesis, NTNU, Trondheim, Norway, 2008:261.

[1]   Unser, M., **2000**, Sampling -- 50 Years after Shannon, *Proc. of the IEEE, 88*, 569-587

[2]   Maletinsky, M. **1978**, Identification of Continuous Dynamical Systems with Spline-Type Modulating Function Method, *ETH, Zuerich, Switzerland, Diss ETH Nr 6206*

[3]   *Preisig, H. A., **1984**, On the Identification of Structurally Simple Dynamic Models for the Energy Distribution in Stirred-Tank Reactor Equipment, ETH-Zuerich, CH, Diss 7616*

[4]   *Preisig, H. A. & Rippin, D. W. T. ,**1993**, Theory and Application of the Modulating Function Method-- Part I Review and Theory of the Method and Theory of the Spline-Type Modulating Function Method, Comp & Chem Eng, 17, 1-16*

[5]   *ibid-Part II Algebraic Representation of Maletinsky's Spline-Type Mod. Functions, 17-28*

[6]   *Preisig, H A, **2010**, Constructing and maintaining proper process models, in press, Comp & Chem Eng*

[7]   *----, **2004**, A Topology Approach to Modelling, SIMS 45, 413-420*

[8]   *----, **1989**, On-line observation of the composition in non-isothermal batch reactors with non-linear reactions, ACC 89, 1549*

[9]   *Kahrs, O and Marquardt, W, 2005, Incremental identification of hybrid process models, Comp & Chem Eng, 32(4-5), 694-705*

Adaptive Data Reconciliation Coupling C++ and PRO/II and On-line Application by the Field

Flavio Manenti,[a] Stefano Signor,[b] Maria Grazia Grottoli,[b] Paolo Fabbri[b]

[a]*Politecnico di Milano, CMIC dept. "Giulio Natta", Piazza Leonardo da Vinci 32, 20133 Milano, ITALY, E-mail: flavio.manenti@polimi.it*
[b]*Chemprod Srl, Automation & Process Solutions, Via Stradella 3, 20129 Milano, ITALY, E-mail: mariagrazia.grottoli@chemprod.it*

Abstract

The paper discusses the possibility of coupling object-oriented programming languages such as Visual C++ and the commercial simulators such as PRO/II, UniSim, and AspenHysys to get a generalized framework for robust and reliable data reconciliation of sulfur recovery units by taking to a series of benefits from their interaction: (i) simulation software provides a high-level and field-proven degree of detail in simulating processes; (ii) they usually allow inferring measures even when the instrumentation presents serious lacks by the field; and (iii) the programming languages allow using specific numerical libraries to solve large-scale, nonlinear, constrained optimization problems by making them feasible even for the on-line industrial application. The proposed approach was validated on a large-scale Sulfur Recovery Unit (technology by Technip-KTI SpA) operating in an oil refinery placed in Italy.

Keywords: Adaptive Data Reconciliation, Parameter Estimation, Sulfur Recovery

1. Introduction

As the market competitiveness and the more stringent environmental regulations are both forcing process industries to apply predictive control and dynamic optimization solutions (Bauer and Craig, 2008; Lima *et al.*, 2009; Manenti and Rovaglio, 2008) for managing at best unit operations and production scheduling and planning in real-time (Busch *et al.*, 2007; Manenti and Manca, 2009; Varma *et al.*, 2007), people are looking forward to switch from their current off-line technology to the on-line application of data reconciliation (Manenti, 2009). Facing the simplicity of solving the theoretical problem, as it is trivially accomplished by iteratively solving the problem at some specific time intervals usually evenly spaced, some overwhelming problems arise especially in ensuring performing solutions of complex and nonlinear systems and/or of large-scale optimization problems by accounting for the process unit details and the evolution of the plant at the same time. On this subject, the research activity here proposed is the result of an academic and industrial collaboration between Politecnico di Milano and Chemprod Srl aimed at developing a reliable and generalized tool for data reconciliation and performance monitoring able to tackle the aforementioned issues.

The approach mainly consists of three steps: (i) the development of a detailed simulation to characterize at best the process in study; (ii) the introduction of adaptive parameters to generalize the simulation by making it available for all those processes based on the same technology but even to account for the plant evolution (performances, fouling factors…); and (iii) the coupling of the simulation to the most performing numerical libraries to solve both the adaptive parameter estimation and the on-line data reconciliation by taking care of possible gross errors affecting the raw data.

The architecture of the adaptive data reconciliation here proposed is explained in Paragraph 2 together with methodologies and numerical solvers adopted. Paragraph 3 describes the industrial application selected to validate it: a large-scale Sulfur Recovery Unit (SRU) operating in an oil refinery placed in Italy. At last, Paragraph 4 shows the validation by the field of our approach.

2. Architecture of the Adaptive Data Reconciliation

As mentioned above, the approach we are going to propose involves a series of steps and issues such as detailed process simulation, parameter estimation, gross error detection, and data reconciliation taking to a multifaceted optimization problem (qualitatively reported in Figure 1).
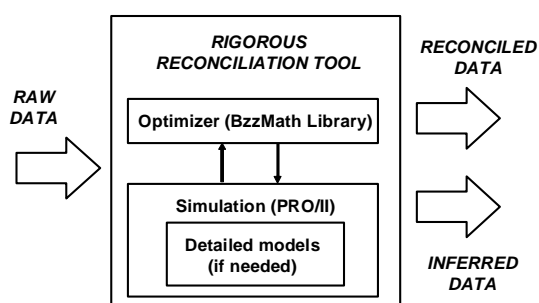

Figure 1: Architecture of the fully integrated approach for adaptive data reconciliation

2.1. Process Simulation: the Need of a Detailed Model

One of the hardest tasks in R&D activities of engineers and scientists is to apply their appealing theoretical solutions by the field. It is not a coincidence that even though data reconciliation was widely studied in these last decades and many authors have described its benefits (Bagajewicz, 2003; Narasimhan and Jordache, 2000; Romagnoli and Sanchez, 1999), it is not yet applied everywhere for different reasons. Some of the most important reasons, which are the ones affecting the industrial case in study here adopted (Paragraph 3), are that many measures cannot be acquired on-line by the field (*i.e.*, the molar composition in a thermal furnace) and that some instrumentation devices may either be too much expensive or require long time to get a response. As a consequence, many processes are generally characterized by a lack of instrumentation that takes to an inadequate redundancy to properly carry out data reconciliation. It is strange to say it nowadays, but this lack of information is sometimes so relevant that both field and control-room operators cannot know exactly what the current plant condition is. This is the case of those processes that are not directly economically appealing even though their presence is essential in the overall plant. SRUs are a typical example as they do not directly increase the net profit margin of the plant because of the low sulfur market price, but they are necessary to match the more and more stringent environmental regulations. From this perspective, there is the need to develop a detailed model to effectively infer the missing measures and information and hence to make not only feasible, but also reliable and robust the data reconciliation (Signor *et al.*, 2010). An appealing solution is represented by the use of commercial simulators as they may take to a series of benefits:

- The mathematical model can cover each level of detail according to the model libraries proposed by the commercial simulator. The degree of detail should be a good compromise between process characterization and computational effort.

*Adaptive Data Reconciliation Coupling C++ and PRO/II and On-line Application
by the Field*

- Some consolidated solutions, especially dictated by the practical experience and already implemented in process simulators, can be successfully involved in the solution of data reconciliation problems.
- Using a commercial simulator that is adopted even for process design, the data reconciliation may have a feedback on both the instrumentation and the same process design (plant debottlenecking, revamping…).
- At last, engineering companies and production sites usually have the licenses of process simulators (reduced license duties).

2.2. Adaptive Parameter Estimation

Unfortunately, facing the need of having a detailed process simulation to infer missing measures to make the data reconciliation feasible even where it is currently not, the problem of obtaining a simulation strictly related to the selected plant arises. This brings to two serious consequences: first of all, it makes the framework of being specifically designed for a single industrial application rather than for many of them; second, as each operating process is subject to grade/load changes, temporary and persistent perturbations, and performance degradations (fouling, cleanliness factor…), a detailed simulation could be unable to properly follow the dynamic evolution of the process and it must be continuously re-tuned to match the process behaviour. There is the need of generalizing the framework by preserving its level of detail, robustness, and reliability, but also making it more flexible and able to adapt itself to the current plant operating condition. To do so, a parametric simulation shall be developed by selecting the most important adaptive parameters according to the kind of process. An adaptive parameter estimation could be accomplished by means of error-in-variable method (EVM) already discussed in the literature (Arora and Biegler, 2001; Weiss *et al.*, 1996). This method has a mathematical formulation similar to the data reconciliation, where the difference resides in the numerical problem size as more steady-state conditions (SSC) of the process are considered and the degrees of freedom of the optimization problem are not only reconciled variables $\mathbf{x}_i$, but even the same adaptive parameters $\boldsymbol{\theta}$:

$$\min_{\mathbf{x}_i, \boldsymbol{\theta}} \sum_{i=1}^{SSC} (\mathbf{m}_i - \mathbf{x}_i)^{\mathrm{T}} \mathbf{Q}^{-1} (\mathbf{m}_i - \mathbf{x}_i) \quad s.t.: \quad f(\mathbf{x}_i, \boldsymbol{\theta}) = 0 \tag{1}$$

$\mathbf{Q}$ is the variance and covariance matrix obtained by accounting for the set of SSC; $\mathbf{m}_i$ is set of measures acquired at the $i-th$ SSC; $f(\mathbf{x}_i, \boldsymbol{\theta}) = 0$ represents equality constraints of process model. The problem size is enlarged against the basis data reconciliation, which has $N_\mathbf{x}$ degrees of freedom ($N_\mathbf{x}$ is the amount of reconciled variables $\mathbf{x}_i$); EVM approach has $N_\mathbf{x} \cdot SSC + N_\boldsymbol{\theta}$ degrees of freedom, where $N_\boldsymbol{\theta}$ is the amount of adaptive process parameters (Signor *et al.*, 2010).

2.3. On-line Data Reconciliation and Gross Error Detection

After having estimated the adaptive parameters, the data reconciliation problem can be solved. During these last years the attention has been mainly focused on the detection of gross errors affecting raw data coming from the field with the need of identifying their existence, location, and size and many statistical tests were developed to solve this issue. Unfortunately, one reason that makes the gross error detection a multifaceted problem is the fact that gross errors have not a general definition that is well-accepted by everyone and this often causes equivocations and errors in speaking about them. The present paper assumes the recent literature definition (Buzzi-Ferraris and Manenti, 2009a, 2009b, 2010b): *"Whatever experimental point that generates a significant increase in the mean square error for a specific model is called outlier for this model."*

3. Industrial Case in Study: Sulfur Recovery Unit (Claus Process)

The current need of matching environmental regulations on sulfur compound emissions renewed industrial interest in this process and their typical lacks of instrumentation make Claus process the ideal field for validating the adaptive data reconciliation.

Task of Claus processes is to recover elemental sulfur from hydrogen sulfide and, more generally, from by-products gases originated by physical and chemical gas treatment units in refineries, natural gas processing, and gasification plants to quote a few.

It consists of a thermal reaction furnace, a waste heat boiler, and a series of catalytic (Claus) reactors. The overall reaction characterizing the process (2) takes place via complex kinetic mechanisms in both the zones of the thermal reactor furnace and in the catalytic reactors as well. In the thermal furnace one third of hydrogen sulfide is oxidized to sulfur dioxide using air (or enriched air). Temperatures are usually in the order of 900-1200 °C. The oxidizing reaction (3) is exothermic and without any thermodynamic restriction. The two thirds of un-reacted hydrogen sulfide reacts with the sulfur dioxide to produce elemental sulfur through the Claus reaction (4), which takes place at high temperatures in the thermal furnace with an endothermic contribution as well as at low temperatures in the catalytic converters with an exothermic contribution.

$$2H_2S + O_2 \rightarrow S_2 + 2H_2O \tag{2}$$

$$H_2S + 3/2O_2 \rightarrow SO_2 + H_2O \tag{3}$$

$$2H_2S + SO_2 \ \square \ \ 3/2S_2 + 2H_2O \tag{4}$$

A secondary but evenly important task is to oxidize ammonia and other impurities usually included in an acid gas refinery stream. Moreover, some side reactions occurring in the thermal furnace may lower the efficiency in sulfur recovery by generating COS and CS_2. Also these mechanisms are modelled to accurately characterize the process. Many existing data reconciliation packages not based on detailed process modeling easily fail in effectively reconciling raw data as the process moves away from its conventional operating point: a series of empirical relations were added to the model by making it more reliable even at non-conventional conditions.

4. Application to Technip-KTI Technology and Preliminary Results

C++ programming language and BzzMath library (Buzzi-Ferraris, 2009; Buzzi-Ferraris and Manenti, 2010a, 2010b) are adopted as numerical kernel for the optimization problems of EVM method in estimating parameters, gross error detection, and data reconciliation. Process simulator PRO/II (Simsci-Esscor, 2002) is adopted to set up the detailed process simulation of the SRU in study. Lid and Skogestad (Lid and Skogestad, 2008) underlined that optimization problems are better solved by simultaneous approaches, whereas the commercial packages adopt a sequential solver. Nevertheless, the larger diffusion in process industry of PRO/II, UniSim, and AspenHysys pushed us towards the possibility to integrate them as black-box in solving the reconciliation problem. The largest optimization problem solved here deals with the EVM approach for estimating adaptive parameters, consisting of 105 degrees of freedom (with 6 SSC). The computational effort required to get a reasonable solution on an INTEL CORE 2 QUAD CPU, 2.83 GHz, 3 GB of RAM, OS MS WINDOWS 2003, and MS VISUAL STUDIO 6.0 compiler is no more than 6 h by making feasible not only a continuous re-

Adaptive Data Reconciliation Coupling C++ and PRO/II and On-line Application
by the Field

tuning of adaptive parameters so to have a simulation that ever matches the current plant conditions, but even the on-line feasibility of less expensive data reconciliation solution. A period of three months was adopted to validate the C++-PRO/II tool on-line. Two test runs only are proposed in Table 1.

Table 1. Preliminary results of the adaptive data reconciliation for some relevant states. Test Run #1 and #2 were acquired at two different times of the plant run length.

State	*UoM*	*Test Run #1*		*Test Run #2*	
		Raw Data	*Reconciled*	*Raw Data*	*Reconciled*
Acid gas flow	kmol/h	64.99	65.33	64.34	64.13
Acid gas temp.	°C	55.10	55.34	55.10	56.01
Sour water flow	kmol/h	35.05	35.95	36.23	35.28
Sour water temp.	°C	84.67	84.74	85.00	85.40
Combustion air flow	kmol/h	181.18	175.21	184.10	186.32
Combustion air temp.	°C	228.73	221.07	229.02	226.36
Waste heat boiler temp.	°C	305.29	312.41	305.48	313.54
First Claus temp.	°C	229.25	233.28	229.50	239.30
Second Claus temp.	°C	215.07	214.01	215.75	214.25
Waste heat boiler steam	Kg/h	5210.92	4841.00	5289.65	5218.00
First Claus outlet temp.	°C	291.53	298.38	291.85	299.45
Second Claus outlet temp.	°C	221.93	223.4	222.48	223.33
H_2S at the tail gas	ppm	5060.00	5111.13	5350.00	5524.83
SO_2 at the tail gas	ppm	1680.00	1719.53	1510.00	1512.68
Furnace temperature	°C	1395.00	1352.64	1390.00	1392.88

The adaptive data reconciliation tool is able to provide a reliable and coherent picture of the plant. Let us consider the case of Test Run #1. The reconciled value for the steam flow generated in the waste heat boiler is significantly lower than the field measure. This is joined to the fact that the reconciled value for the thermal furnace temperature is about 1350°C rather than the measured value of 1395°C. This is also confirmed by the smaller combustion air entering the furnace and its reduced temperature. Then, the amount of ppm for H_2S and SO_2 is larger at the tail gas by indicating a temporary reduced efficiency of the combustion process. From this perspective, we showed even the reliable on-line inferentiation of some measures that are currently acquired off-line. The approach has given encouraging results even when the plant moved away from the conventional operating condition.

5. Conclusions

The paper proposed a comprehensive approach to face the data reconciliation of sulfur recovery units. As these processes are usually characterized by serious lacks in measures and instrumentation, a detailed process simulation is needed to infer the missing data. Some adaptive parameters are then introduced to make more flexible the process simulation so to allow a better characterization of the same process as it evolves changing its own performances as well as to make the framework generalized to solve the data reconciliation on sulfur recovery units having similar layouts (Technip-KTI SpA technology). The error-in-variable method is adopted to estimate parameters by showing the on-line feasibility of the approach. Robustness of the overall approach is ensured by classes for outlier detection of *BzzMath* library adopted as numerical kernel.

6. Acknowledgements

References

Arora, N., Biegler, L.T. (2001). Redescending estimators for data reconciliation and parameter estimation. Computers & Chemical Engineering, 25(11-12), 1585-1599.

Bagajewicz, M.J. (2003). Data Reconciliation and Instrumentation Upgrade. Overview and Challenges. FOCAPO 2003. 4th International Conference of Computer-Aided Process Operations, Coral Springs, Florida, 103-116.

Bauer, M., Craig, I.K. (2008). Economic Assessment of Advanced Process Control - A Survey and Framework. Journal of Process Control, 18, 2-18.

Busch, J., Oldenburg, J., Santos, M., Cruse, A., Marquardt, W. (2007). Dynamic Predictive Scheduling of Operational Strategies for Continuous Processes Using Mixed-logic Dynamic Optimization. Computers & Chemical Engineering, 31, 574-587.

Buzzi-Ferraris, G. (2009). BzzMath: Numerical library in C++. Politecnico di Milano, http://chem.polimi.it/homes/gbuzzi.

Buzzi-Ferraris, G., Manenti, F. (2009a). Kinetic models analysis. Chemical Engineering Science, 64(5), 1061-1074.

Buzzi-Ferraris, G., Manenti, F. (2009b). Outlier Detection. Computers & Chemical Engineering, submitted.

Buzzi-Ferraris, G., Manenti, F. (2010a). Fundamentals and Linear Algebra for the Chemical Engineer Solving Numerical Problems. ISBN: 978-3-527-32552-8, WILEY-VCH, Germany.

Buzzi-Ferraris, G., Manenti, F. (2010b). Interpolation and Regression Models for the Chemical Engineer Solving Numerical Problems. ISBN: 978-3-527-32652-5, WILEY-VCH, Germany.

Lid, T., Skogestad, S. (2008). Scaled steady state models for effective on-line applications. Computers & Chemical Engineering, 32(4-5), 990-999.

Lima, N.M.N., Manenti, F., Maciel Filho, R., Embiruçu, M., Wolf Maciel, M.R. (2009). Fuzzy Model-Based Predictive Hybrid Control of Polymerization Processes. Industrial & Engineering Chemistry Research, 48(18), 8542–8550.

Manenti, F. (2009). From Reacting to predicting technologies: A novel performance monitoring technique based on detailed dynamic models. Chemical Product and Process Modeling, 4(2).

Manenti, F., Manca, D. (2009). Transients modeling for enterprise-wide optimization: Generalized framework and industrial case study. Chemical Engineering Research and Design, 87(8), 1028-1036.

Manenti, F., Rovaglio, M. (2008). Integrated multilevel optimization in large-scale poly(ethylene terephthalate) plants. Industrial & Engineering Chemistry Research, 47(1), 92-104.

Narasimhan, S., Jordache, C. (2000). Data Reconciliation & Gross Error Detection, An Intelligent Use of Process Data. Gulf Publishing Co., Houston, TX, USA.

Romagnoli, J.A., Sanchez, M.C. (1999). Data Processing & Reconciliation for Chemical Process Operations. Process Systems Engineering Ed. (Vol. 2).

Signor, S., F. Manenti, M.G. Grottoli, P. Fabbri, S. Pierucci. (2010). Sulfur Recovery Units: Adaptive Simulation and Model Validation on Industrial Plant. Industrial & Engineering Chemistry Research, to appear.

Simsci-Esscor. (2002). PRO/II, User Guide. Lake Forest, CA, USA, www.simsci-esscor.com.

Varma, V.A., Reklaitis, G.V., Blau, G.E., Pekny, J.F. (2007). Enterprise-wide modeling & optimization - An overview of emerging research challenges and opportunities. Computers & Chemical Engineering, 31(5-6), 692-711.

Weiss, G.H., Romagnoli, J.A., Islam, K.A. (1996). Data Reconciliation - An Industrial Case Study. Computers & Chemical Engineering, 20(12), 1441-1449.

A computer-aided framework for regression and multi-scale modelling needs in innovative product-process engineering

Martina Heitzig, Gürkan Sin, Peter Glarborg, Rafiqul Gani

Department of Chemical and Biochemical Engineering, Technical University of Denmark, Soltofts Plads, Building 227, 2800 Kgs. Lyngby, Denmark, mat@kt.dtu.dk

Abstract

Model-based computer aided product-process engineering requires models of different types, forms and application mode. A computer-aided modelling framework capable of handling the modelling needs for product-process design and analysis has been developed. In this paper, the systematic and efficient development of multi-scale models, their interconnections, analysis, parameter regression and solution through the modelling framework is presented. The application of the framework is highlighted through a case study related to the batch uptake (process operation) of lysosyme by sepharose (product).

Keywords: modelling framework, product-process design, multi-scale modelling

1. Introduction

Models are playing roles of increasing importance in design and analysis of chemicals/bio-chemicals based products and the processes that manufacture them because of the increasing use of computer-aided methods and tools. The advantage of using these model-based methods and tools is that they have the potential to reduce the number of experiments, which can be expensive and time consuming, while at the same time, lead to truly innovative solutions. The idea here is to use the model-based systems to either find promising candidates for final verification by experiments, or, evaluate and analyze existing products to identify opportunities for improvement. To this end, the concept of a virtual lab for product-process design [1] is interesting as it allows the product designer to perform virtual experiments to find, select and/or evaluate products and the important processing steps that define their production. A modelling framework (ICAS-MoT) [2] has already been established in the virtual lab for product-process design. However, to increase its application range, it needs to be extended in terms of model library, model types and forms. Since a large number of product-process design problems require the use of multiscale modelling options, this is the feature that has been developed in this work and presented in this paper. Whenever a required model is missing in the model library of the virtual lab the user needs to develop and provide it. As these models may be complex and require multiple time and/or lengths scales, their development and application for product-process design is not trivial. Therefore, the modelling framework can contribute by reducing the time and resources needed for model development and application, thereby reducing the overall time for product-process development. A case study related to the batch uptake of lysozyme by sepharose where models on multiple length scales need to be generated, integrated and solved, has been used to validate the multi-scale modelling part of the framework.

2. Systematic Modelling Framework

The work-flow for virtual product-process design has been presented earlier [1]. Within this design work-flow, the specific work-flow for the development and identification of multiscale models has now been added. Figure 1 shows the workflows for these two new options: model development (left of Fig. 1) and model identification (right of Fig. 1). For a priori validated models, only the model development option is needed. When the model also needs to be identified, the two options need to be linked (as shown in Fig. 1). The various modelling tools required for the different steps of the two workflows have been implemented in the ICAS-MoT.
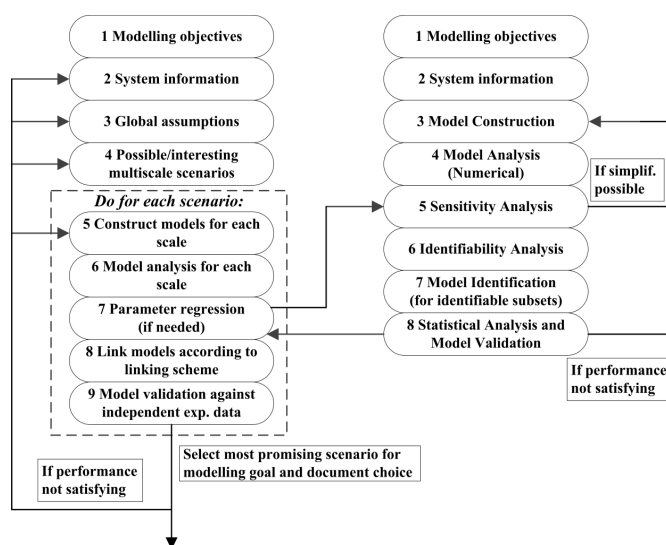

Figure 1: Work-flow for multi-scale modelling (left) and model identification (right)

The first step for multi-scale model development is to define the modelling objective. In the second step, corresponding to the modelling objective, available information on the system and the occurring phenomena is collected. With this information in place, the main assumptions which hold true for the different scenarios that need to be covered for the specific design problem are specified (step 3). In step 4 various modelling scenarios of interest are listed and models of the different scales are generated. In general, it is a good practice to start with the simplest scenario and to gradually increase the model complexity by adding new scales. Steps 5-9 need to be repeated for each multi-scale scenario listed in step 4. First the model assumptions are formulated for each scale and based on that the model equations are derived or retrieved from library. The next step after obtaining the model equations is to perform a numerical analysis of the models on each scale. The ICAS-MoT modelling tool allows the user to specify the variable types and based on this, it performs a degree of freedom analysis and generates the incidence matrix and the corresponding optimal order to form an efficient solution strategy. If there are parameters that need to be regressed, it is done in step 7. At this point the regression work-flow will take over. In step 8, models of different scales generated and/or retrieved from the appropriate model library are combined according to the established work-flow. Figure 2 shows the features of the modelling toolbox, ICAS-MoT to assist the user in the above described steps of the methodology.

A computer-aided framework for regression and mulit-scale modelling needs in innovative product-process engineering
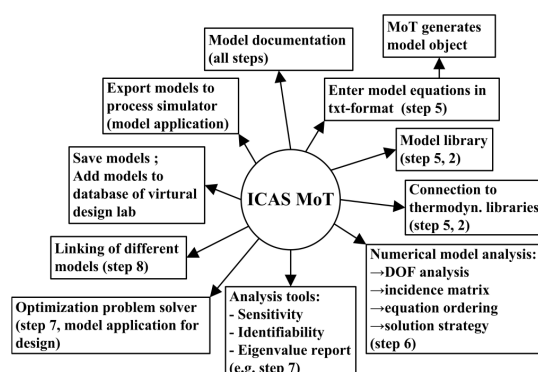

Figure 2. Overview over ICAS-MoT features to support development of multi-scale models.

3. Application: Case Study

This case study involves the batch uptake of lysozyme by sepharose beats. Models of different scales need to be developed so that product behavior (sepharose particles) can be monitored in batch processing step. This case study has been chosen because through it the developed modelling features can be highlighted together with the evaluation of the product behavior and it also provides modelling challenges.

Step 1 Modelling objectives: The objective here is to study the effect of changes in the bulk protein concentration of lysozyme due to protein uptake by sepharose in a batch (process) operation, and also to predict the uptake profiles inside the sepharose.

Step 2 System information: The lysozyme uptake is to be described applying a parallel diffusion approach [3]. Experimental data [4-6] are available for model validation and/or model identification.

Step 3 Principal assumptions: The fluid in the bulk is assumed to be ideally mixed. The size of the particles of the stationary phase is equal and their form is spherical.

Step 4 Multi-scale scenarios: Models for four scenarios (see Figure 3) have been generated and investigated. The shaded boxes illustrate how models of different scales are combined. In Fig 3, c_P^{bulk} is the protein bulk concentration and q_{ads} is the total amount of protein adsorbed by all particles. $c(t,r)$ and $q(t,r)$ describe the protein uptake profile in the particle pores and on the pore surfaces, respectively.

Scenario 1 is the simplest scenario. Here, only the macro scale, which means the overall mass balance for the uptake of proteins are to be considered. From that follows that the amount of protein uptake in each time step by one particle is set to a constant value. The macro scale model is extended by including a meso-scale sub-model, which introduces a parallel diffusion model for the rate of protein uptake (Fig. 4).

For the macro-meso scenario (see Fig. 4a) the values for the pore and surface diffusivities D_p and D_s are assumed to be constant and specified by the user. Depending on the assumption for the protein bulk concentration c_P^{bulk}, models for two scenarios having a macro and a meso scale have been constructed: In scenario 2a, c_P^{bulk} is constant, while in scenario 2b c_P^{bulk} is assumed to be not constant. The change of the overall diffusivity with respect to protein bulk concentration is taken into account by applying a correlation [3] which results in the introduction of a micro-scale model (scenarios 3a-3b) to calculate the surface diffusivity D_s with respect to protein bulk concentration on the adsorbent surface (as shown in Fig. 4b). The meso scale describes, like shown above, how fast the protein enters and penetrates the particle based on a

diffusion model. The micro scale model provides a deeper insight about the protein-wall interaction considering properties of the protein, the solid phase material and the protein adsorption energetics.
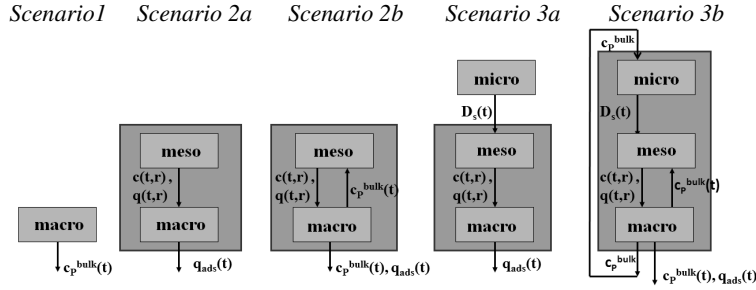

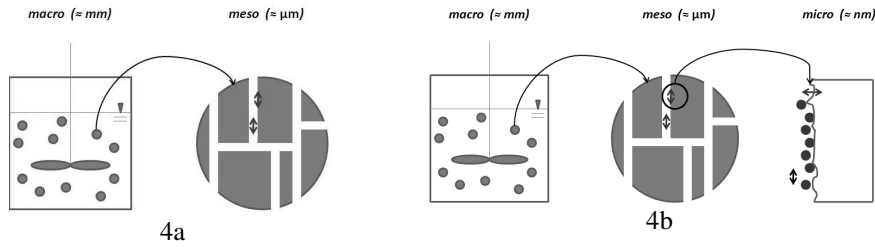Figure 3. Linking schemes for described multi-scale scenarios

Figure 4. Schematic sketches of macro-meso (4a) and macro-meso-micro scales (4b).

Introducing a micro scale model for D_s also increases the flexibility of the model with respect to ability to investigate new scenarios. Now the model can be applied for changing salt concentrations (e.g. salt gradient elution) by adding a correlation for the adsorption isotherm with respect to salt concentration. Again, two scenarios depending on the protein bulk concentrations are distinguished: 3a) assume c_P^{bulk} constant; and 3b) c_P^{bulk} not constant. Due to the dependency of the micro scale model on c_P^{bulk} the micro scale needs to be solved within the integration loop for scenario 3b.

Step 5 Construct models: The model equations are introduced to ICAS-Mot in a text format and translated by ICAS-MoT into a model object. In this paper, the model equations are given below only for scenario 3b. Additional details on the models of each type, including detailed model analysis, etc., can be obtained from the corresponding author. Micro scale model [3] construction (Figure 4b, right) is given below. Eq. (1) gives the quotient of the effective diffusivity and the pore diffusivity. Note that here instead of calculating and communicating the surface diffusivity this quotient is determined.

$$D_a D_p = 1 + 0.35 \cdot \left(\frac{q_0}{\varepsilon_p \cdot c_p^{bulk} \cdot \phi_p \cdot h} \right) \left(\frac{\phi_c}{k_{retention} \cdot \kappa} \right)^{0.56} \tag{1}$$

With:

$$k_{retention} = 0.0017 \cdot IS^{-5.75} \tag{2}$$

$$\kappa = 3.29 \cdot IS^{0.5} \tag{3}$$

A computer-aided framework for regression and mulit-scale modelling needs in innovative product-process engineering

$$q_0 = \frac{q_{mon} \cdot b \cdot c_p^{bulk}}{1 + b \cdot c_p^{bulk}} \tag{4}$$

The meso scale model [3] construction (Figure 4b, middle) is given below, where the partial differential equation describing the parallel diffusion has been discretized and the resulting system of ODEs is given in Eqs. (5) and (6):

$$\frac{dY(j)}{d\tau_p} = f(\Delta\rho, \rho(j-1), \rho(j), \rho(j+1), Y(j-1), Y(j), Y(j+1), Y(j+2), R_{sep}, \alpha, \beta)$$

for $j = 1, ..., 48$ \hfill (5)

$$\frac{dY(49)}{d\tau_p} = f(\Delta\rho, \rho(48), \rho(49), \rho(50), Y(48), Y(49), Y(50), Y, R_{sep}, \alpha, \beta) \tag{6}$$

with: $\alpha = \dfrac{q_0}{\varepsilon_p \cdot c_p^{bulk}}$ \hfill (7)

$$\beta = (D_a D_p - 1)/0.6 \tag{8}$$

$$R_{sep} = \frac{1}{1 + b \cdot c_p^{bulk}} \tag{9}$$

$$\rho(0) = 0 \ , \ \rho(50) = 1 \ , \ \Delta\rho = \rho(50)/50 \ , \ \rho(1) = 0.5 \cdot \Delta\rho \tag{10,11,12,13}$$

$$\rho(j) = \rho(j-1) + \Delta\rho \ , \ j = 2, ..., 50 \tag{14}$$

$$Y(50) = 1 \tag{15}$$

$$q(j) = Y(j) \cdot q_0 \ , \ j = 1, ..., 50 \tag{16}$$

$$c(j) = \frac{q(j)}{b \cdot (q_{mon} - q(j))} \ , \ j = 1, ..., 50 \tag{17}$$

The macro scale model construction (Figure 4b, right) is given below. Eq. (18) determines the concentration of protein uptake by all particles while Eq. (19) gives the protein bulk concentration.

$$q_{ads} = \frac{m_{ads}}{0.5 V_{suspension}} \tag{18}$$

$$c_p^{bulk} = c_p^{bulk, start} - \frac{m_{ads}}{V_{bulk}} \tag{19}$$

With:

$$m_{ads} = \frac{0.5 V_{suspension}}{V_{particle}} (m_{ads1} + m_{ads2}) \tag{20}$$

$$V_{particle} = \frac{4}{3}\pi R^3 \ , \ V_{bulk} = V_{liq} + 0.5 V_{suspension} \tag{21,22}$$

$$m_{ads1} = \sum_{j=1}^{50} (q(j) \cdot V_{cell}(j)) \tag{23}$$

$$m_{ads2} = \sum_{j=1}^{50} (c(j) \cdot V_{cell}(j) \cdot \varepsilon_p) \tag{24}$$

$$V_{cell}(j) = \frac{4\pi}{3}\left(\left((\rho(j) - 0.5\Delta\rho) \cdot R + \Delta\rho \cdot R\right)^3 - \left((\rho(j) - 0.5\Delta\rho) \cdot R\right)^3\right), \ j = 1, ..., 50 \tag{25}$$

Note that the macro scale model equations (Eqs. 18-25) are common to all modelling scenarios. Depending on which scenario is selected, additional scale-models (meso or micro) are added. The remaining steps of the methodology are not shown in detail in this paper. In the text below, the simulation results (model validation) are discussed very briefly for scenario 3b, which was carried out with ICAS-MoT. For scenario 3b, the results show that the batch uptakes on the macro scale can be predicted quite well. However, there are some deviations in the predicted uptake profiles which are for the meso-scale (see Fig. 5 scenario 3b) because the external mass transfer resistance has been neglected [3]. If multi-scale scenario 2b is applied instead of 3b, the predictions are slightly improved (see Fig. 5 scenario 2b). This shows that the most complex scenario is not always the most accurate. However, increasing the complexity will lead to more flexibility of the model with respect to changing conditions.
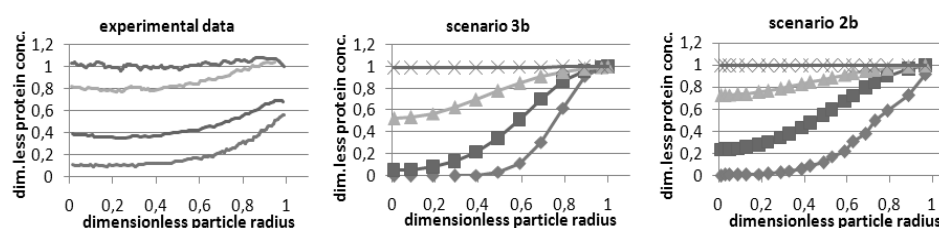

Figure 5. Uptake profile of lysozyme in sepharose particle (for time=2 min, 5 min, 11 min, 26 min, respectively; upper line corresponds to 26 min).

4. Conclusions

Systematic multiscale modelling and identification options have been added to a modelling framework linked to a model-based virtual lab for product-process design. The work-flow for the new options have been tested with an interesting product-process evaluation case study and the results from the multi-scale modelling option have been presented. The case study highlights the important contribution that model-based systems can make in product-process design. Some aspects for further improvement in the model (such as adding a mass transfer model to improve the model performance) and the modelling framework have also been identified. Current and future work is extending the virtual lab and the ICAS-MoT with more models so that interesting product-process design problems can be solved.

5. Acknowledgements

References

[1] R. Morales-Rodríguez, R. Gani, Computer Aided Chemical Engineering, 24 (2007) 207
[2] A. M. Sales-Cruz, Development of a Computer Aided Modelling System for Bio and Chemical Process and Product Design, PhD Thesis, Technical University of Denmark, 2006.
[3] A.M. Lenhoff, Langmuir, 24(2008) 5991
[4] S.R. Dziennik, E.B. Belcher, G.A. Barker, A.M. Lenhoff , Biotechnol. Bioeng., 91 (2005) 139
[5] C. Chang, A.M. Lenhoff, J Chromatogr. A, 827(1998) 281
[6] S.R. Dziennik, E.B. Belcher, G.A. Barker, M.J. DeBergalis, S.E. Fernandez, A.M. Lenhoff, Proc. Nat. Acad. Sci., 100 (2003) 420

Structural Decomposition of Process Models Described by Higher Index DAE Systems

Adrien Leitold,[a] Miklós Gerzson[b]

[a]*Dept. of Mathematics, University of Pannonia, P.O.B. 158, Veszprém, H-8201 Hungary, leitolda@almos.vein.hu*
[a]*Dept. of Electrical Engineering and Information Systems, University of Pannonia, P.O.B. 158, Veszprém, H-8201 Hungary, gerzson@almos.vein.hu*

Abstract

A graph-theoretical method for the structural analysis of dynamic lumped process models described by differential and algebraic equations (DAEs) is applied in this paper in order to determine the most important solvability properties (degree of freedom, structural solvability, model decomposition, dynamic degree of freedom, differential index, e.g.) of these models by using the so-called dynamic representation graph. The structure of the dynamic representation graph is suitable for the determination of the mentioned solvability properties. The aim of our work is to show a decomposition procedure in case of higher index models. This method shows the subpart which causes the higher index and the position of this part in the hierarchy of structurally solvable submodels. Using this procedure we can answer the generally not simple question that in case of large process models which submodel or submodels cause the higher index and what the positions of these submodels are in relation to the other, structurally solvable submodels in the model.

Keywords: process models, DAE-models, structural analysis, differential index, model decomposition

1. Introduction

The structural analysis of dynamic lumped process models forms an important step in the model building procedure [1] and it is used for the determination of the solvability properties of the model, too. Effective graph-theoretical methods have been proposed in the literature [2, 3] based on the analysis tools developed by [4], for the determination of the most important solvability property of lumped dynamic models: the differential index. The properties of the dynamic representation graph of process models described by semi-explicit DAE-systems have also been analysed there in case of index 1 and higher index models.

2. Basic Notions

2.1. Structural Solvability

As a first step, we consider a system of linear or non-linear algebraic equations in its so called *standard form* [4]:

$$y_i = f_i(x, u) \quad i = 1, \ldots, M, \qquad u_k = g_k(x, u), \quad k = 1, \ldots, K \tag{1}$$

where x_j $(j = 1, \ldots, N)$ and u_k $(k = 1, \ldots, K)$ are unknowns, y_i $(i = 1, \ldots, M)$ are known parameters, f_i $(i = 1, \ldots, M)$ and g_k $(k = 1, \ldots, K)$ are assumed to be sufficiently smooth

real-valued functions. The system of equations above is structurally solvable, if the Jacobian matrix $J(x, u)$ referring to the above model is non-singular.

Consider a system of equations in standard form. We construct a directed graph to represent the structure of the set of equations in the following way. The vertex-set corresponding to unknowns and parameters is partitioned as $X \cup U \cup Y$, where $X = \{x_1, …, x_N\}$, $U = \{u_1, …, u_K\}$ and $Y = \{y_1, …, y_M\}$. The functional dependence described by an equation is expressed by arcs coming into y_i or u_k respectively from those x_j and u_l, which appear on its right-hand side. This graph is called the *representation graph* of the system of equations.

A *Menger-type linking* from X to Y is a set of pair-wise vertex-disjoint directed paths from a vertex in X to a vertex in Y. The size of a linking is the number of directed paths from X to Y contained in the linking. In case $|X| = |Y|$, ($M = N$), a linking of size $|X|$ is called a *complete linking*. The graphical condition of the structural solvability is then the following [4]:

Linkage theorem: Assume that the non-vanishing elements of partial derivatives f_i and g_k in the standard form model are algebraically independent over the rational number field Q. *Then the model is structurally solvable if and only if there exists a Menger-type complete linking from X to Y on the representation graph.*

From the viewpoint of structural solvability several types of model decompositions can be defined and constructed using representation graph. The goal of these decompositions is to obtain the sub-systems of a model and to determine a successful way for solving the model equations. The most important decomposition method is the *L*-decomposition [4] which disintegrates the system into structurally solvable subsystems with hierarchical structure among them. We can adapt the graphical techniques to DAE-systems, as well. A system of equations including also differential equations can be represented by a *dynamic graph* [2]. The structural analysis based on graph theoretical technique is carried out in steps performed sequentially:

1. Rewrite the model into standard form.
2. Construct the dynamic representation graph.
3. Assign the types to the vertices in the representation graph according to the model specification (the used labels are <S>, <S*>, <G>).
4. Construct the reduced graph.

The detailed description of these steps can be found in [2].

2.2. Differential Index

Dynamic process models can be described by *semi-explicit DAEs* as follows:

$$z_1' = f(z_1, z_2, t), \quad z_1(t_0) = z_{10} \qquad\qquad 0 = g(z_1, z_2, t) \qquad\qquad (2)$$

The most important structural computational property of DAE models is the differential index [5]. By definition [6] the differential index of the semi-explicit DAE (Eqn. (2)) is one if one differentiation is sufficient to express z_2' as a continuous function of z_1, z_2 and t. One differentiation is sufficient if and only if the Jacobian matrix g_{z_2} is non-singular.

In our earlier work we have proved that the differential index of the models investigated in [2] *is equal to* 1 *if and only if there exists a Menger-type complete linking on the reduced graph.*

If the differential index of the investigated model is greater than 1 then there is no Menger-type complete linking on the representation graph. The properties of a representation graph of a dynamic model, which has differential index >1 are as follows.

1. The fact that the initial values of differential variables cannot be chosen independently results in an *overspecified* part on the graph.
2. Non-singularity of g_{z_2} results in an *underspecified* part on the graph. In this part those algebraic variables appear, which cannot be calculated from algebraic equations and those derivative variables, which we want to calculate from them.

We have also proposed an algorithm using the structure of the representation graph for determination of the differential index of the underlying model. The main steps of this algorithm are the following:

1. Let us form the following variable sets.

I_0 is the set of the differential variables belonging to the overspecified subgraph,

D_0 is the set of the derivative variables referring to the differential variables of set I_0,

I_1 is the set of differential variables from which directed paths lead to the derivative variables in the set D_0,

D_1 is the set of derivative variables referring to the differential variables of set I_1, … ,

I_k is the set of differential variables from which directed paths lead to the derivative variables in the set D_{k-1},

D_k is the set of derivative variables referring to the differential variables of set I_k, …

2. Let n be the smallest natural number for which the set D_n contains some derivative variables of the underspecified subgraph. Then the differential index of the model is

$v_d = n+2$

3. Structural analysis of simple process models using their representation graphs

Small sized real process models having easily understandable representation graphs, in general, contain small implicit parts. According to literature [4] the size of the reduced graph is about one tenth of the representation graph. Hence, in case of small sized real process models, the demonstration of the *L*-decomposition and the hierarchy of *L*-components is a difficult problem. This is the reason why we use a formal model in this section.

Consider the following DAE-system given in its standard form:

$$x_1 = \int x_1' \qquad x_1' = f_1(x_1, x_6) \qquad x_2 = \int x_2' \qquad x_2' = f_2(x_2, x_9, x_{10})$$
$$s_1 = f_3(x_4, x_5), \ s_1 = 0 \qquad s_2 = f_4(x_7, x_8), \ s_2 = 0 \qquad s_3 = f_5(x_3, x_{10}), \ s_3 = 0 \qquad (3)$$
$$x_4 = f_6(x_1, x_3) \qquad x_6 = f_7(x_2, x_7)$$

3.1. Specification 1

Given: x_5, x_7, x_9 as a function of time and $x_1(t_0)$, $x_2(t_0)$;

To be calculated: x_1, x_2, x_3, x_4, x_6, x_8, x_{10} as a function of time.

For the sake of simplicity, we illustrate only one static graph from the dynamic graph as a representation graph of the models.

The representation graph of the model is shown in Fig. 1.a. Labels assigned to the vertices correspond to the *Specification* 1. The reduced graph is shown in Fig. 1.b. It can be established that there is a Menger-type complete linking on the reduced graph, so the differential index of the model with the given specification is equal to 1. The *L*-decomposition can be performed on the reduced graph using the Menger-type complete linking and consequently the hierarchy of *L*-components can be determined (see Fig. 1.c).
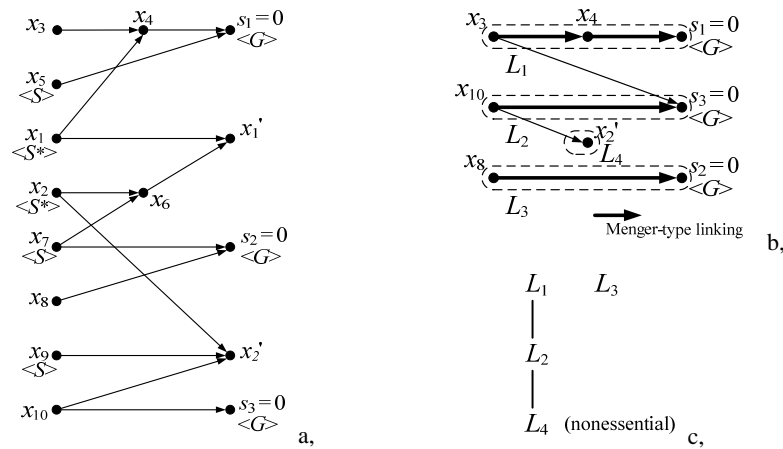
Fig. 1. Structural analysis in case of *Specification* 1

3.2. Specification 2

As the next step, let us investigate the same model with the following specification.

Given: x_3, x_5, x_7 as a function of time and $x_1(t_0)$, $x_2(t_0)$;

To be calculated: x_1, x_2, x_4, x_6, x_8, x_9, x_{10} as a function of time.

The representation graph according to *Specification* 2 can be seen in Fig. 2.a. There is an overspecified and an underspecified subgraph on the representation graph indicating that the differential index of the model is higher than 1.

The differential index can be calculated based on the structure of the representation graph. The sets of variables are the following:

$$I_0=\{x_1\}, \qquad D_0=\{x_1'\}, \qquad I_1=\{x_2\}, \qquad D_1=\{x_2'\} \tag{4}$$

The vertex referring to the derivative variable x_2' can be found in the underspecified subgraph, therefore $n = 1$ and $v_d = n + 2 = 3$.

4. Model decomposition in case of higher index model

We propose the following extension of the *L*-decomposition for the process models having higher than 1 differential index.

Step **1.**: Construct the reduced graph according to following rules:

– Omit the underspecified and the overspecified subgraphs and all arcs starting from their vertices from the representation graph.

– Omit all vertices labelled primarily, secondarily, etc. by type *<S>* and all arcs starting from them.

Step **2.**: Fix the Menger-type complete linking on the reduced graph, then perform the *L*-decomposition.

Step **3.**: Determination of the component hierarchy: Define the position of overspecified and underspecified subgraph relative to the *L*-components based on the original representation graph.

As an example for this procedure, let us see the reduced graph of the previous index 3 model which was constructed using *Step* 1. (Fig. 2.b). Performing the *L*-decomposition we get two *L*-components that are independent from each other. The relation of these components to the overspecified and underspecified subgraphs can be seen in Fig. 2.c. Since there are arcs leading from the overspecified subgraph to the vertex s_3 of L_2 and

from vertex x_{10} of L_2 to vertex x_2' of the underspecified subgraph, the component L_2 can be found between the overspecified (O) and the underspecified (U) subgraphs, and L_1 component is independent from them.
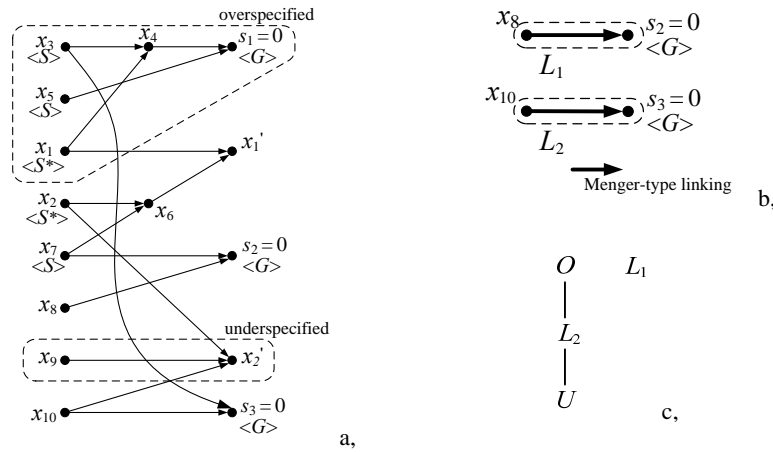

Fig. 2. Structural analysis in case of *Specification* 2

5. Case study

As an example for model decomposition in case of higher index model suppose a simple system consisting of two simple tanks in parallel connection (see Fig. 3.) The inlet flow is forced by pumps and the variables p_x refer to the values of the pressure while variables F_y refer to the values of the flow rates according to the Fig. 3.

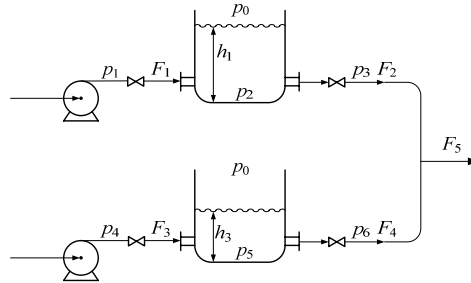Fig. 3. Tanks in parallel connection

The change of the liquid levels in the tanks can be described with the following standard form DAE-system:

$$h_1 = \int h_1' \qquad h_1' = (F_1 - F_2)/A_1 \qquad h_2 = \int h_2' \qquad h_2' = (F_3 - F_4)/A_2 \qquad (5)$$

$$F_1 = c_{v1}(p_2 - p_1)^{1/2} \qquad F_2 = c_{v2}(p_3 - p_2)^{1/2} \qquad F_3 = c_{v3}(p_5 - p_4)^{1/2} \qquad F_4 = c_{v4}(p_6 - p_5)^{1/2}$$

$$F_5 = F_2 + F_4, \qquad s_1 = p_2 - p_0 - \rho g h_1, \quad s_1 = 0 \qquad \qquad s_2 = p_5 - p_0 - \rho g h_2, \quad s_2 = 0$$

Consider the following specification:

 Given: p_0, p_1, p_2, p_4, p_6, as a function of time and $h_1(t_0), h_2(t_0)$;

 To be calculated: $F_1, F_2, F_3, F_4, F_5, p_3, p_5, h_1, h_2$ as a function of time;

 Constants: $A_1, A_2, c_{v1}, c_{v2}, c_{v3}, c_{v4}, \rho, g$.

The representation graph belonging to the model of tank system with the above specification can be seen in Fig. 4.a. There are an overspecified and an underspecified

subgraph referring to the higher index of the model. The differential index can be determined based on the structure of the graph: $I_0 = \{h_1\}$, $D_0 = \{h_1'\}$. Since variable h_1' can be found in underspecified subgraph, therefore $n = 0$ and $v_d = n + 2 = 2$.

The reduced graph (*Step* 1) of the model (Eqns. 5) can be seen in Fig. 4.b. Performing the *L*-decomposition on the reduced graph (*Step* 2) an *essential L-component* (L_1) [2] and other four *nonessential L-components* (L_2, ..., L_5) [2] can be determined. The hierarchy of the *L*-components, the overspecified (*O*) and the underspecified (*U*) subgraphs (*Step* 3) is depicted on Fig. 4.c.
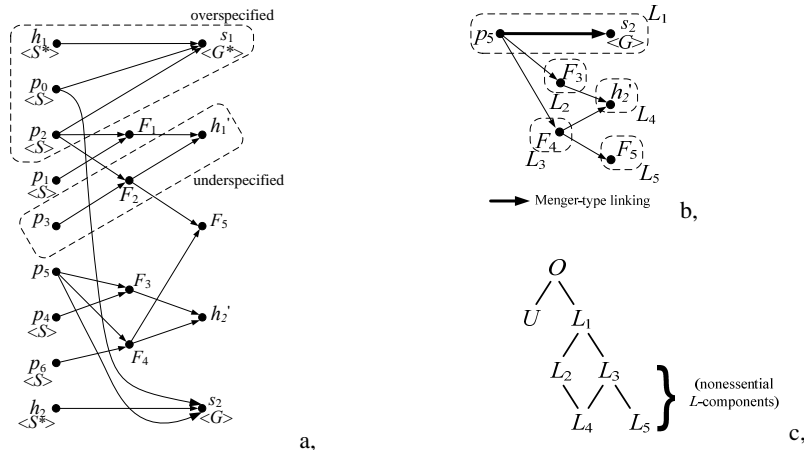

Fig. 4. Structural analysis of simple tank system

6. Conclusion

We propose a decomposition procedure for process models described by higher index DAE-systems. The method is suitable for identifying the parts of large size real models causing the higher index problem and for determining of the relative position of these parts to the other structurally solvable model parts. The introduced decomposition method can be used in case of matrix representation of the representation graph, so the procedure can be performed by an appropriate software tool.

References

[1] K.M. Hangos and I.T. Cameron, Process Modelling and Model Analysis, Academic Press, London, 2001.
[2] A. Leitold and K.M. Hangos, Computers chem. Engng., 25 (2001) 1633
[3] A. Leitold and K.M. Hangos, Hung. J. of Ind. Chem., 30 (2002) 61
[4] K. Murota, Systems Analysis by Graphs and Matroids, Springer Verlag, Berlin 1987.
[5] C. W. Gear, and L. R. Petzold, SIAM J. Numer. Anal., 21 (1984) 716
[6] K.E. Brenan, S.L. Campbell and L.R. Petzold, Numerical Solution of Initial Value Problems in Differential – Algebraic Equations, North-Holland, New York, 1989.

Speeding up a multiobjective genetic algorithm with constraints through artificial neuronal networks

Claudia Gutiérrez-Antonio[a], Abel Briones-Ramírez[b, c]

[a] CIATEQ, A.C., Av. del Retablo 150 Col. Fovissste, Querétaro, Querétaro, 76150, México, claudia.gutierrez@ciateq.mx
[b] Exxerpro Solutions, Av. del Sol 1B Interior 4B El Sol, Querétaro, Querétaro, 76113, México, abel.briones@exxerpro.com
[c] Instituto Tecnológico de Aguascalientes, Av. López Mateos # 1801 Ote. Fracc. Bonagens, Aguascalientes, Aguascalientes, 20256, México.

Abstract

Evolutionary algorithms have been recognized to be well suited for multiobjective optimization [1]; their principal disadvantage is the large number of evaluations of objective function required [2], which is accentuated when those are computationally expensive. In this work, we propose the use of artificial neuronal networks, ANN, to speed up a multiobjective genetic algorithm with constraints, with base on the work of Gaspar-Cunha [3]. The neuronal network generates an approximated function of the original objective function, which are switched during the optimization; so, we reduce the evaluations of the original objective function and the computational time. The use of approximated functions created by the ANN allows reaching the optimal zone rapidly. Results show a significant reduction in the number of evaluations of the objective function, as well as in computational time, required to reaching the Pareto front.

Keywords: Artificial neuronal networks, multiobjective genetic algorithms, dividing wall column, speeding up
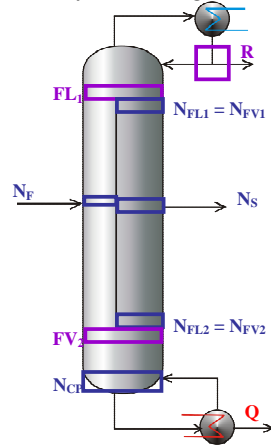
1. Introduction

Evolutionary algorithms have been recognized to be well suited for multiobjective optimization, because of their capability to evolve a set of non dominated solutions distributed along the Pareto front [1]. However, the principal disadvantage of the genetic algorithms, and its variants, is the large amount of computational time that is often required for multiobjective optimization of industrial operations [2]. This has led to the development of strategies to reduce the computational time involved, without sacrifice the complexity of the original problem or increase computational resources, which sometimes is not possible.

Several attempts to speed up evolutionary algorithms were presented previously [4-7], which used modified operators of genetic algorithms, principally mutation and crossover. Nevertheless, these modifications have limited contributions to reduce computational time, when evaluations of objective function are computationally expensive (more than 80% of the total required time). For these cases, we propose the use of artificial neuronal networks, ANN, to speed up multiobjective genetic algorithms with constraints. As study case we select the optimal design of dividing wall columns, considering the complete set of MESH equations. To our knowledge, this is the first

Chemical Engineering application where ANN are using to speed up multiobjective genetic algorithms; until now, the combination of these tools was focused to use genetic algorithms to speed up ANN [8-11]. Results show a significant reduction in the number of evaluations of the objective function, as well as in computational time, required to reaching the Pareto front.

2. Selected problem with an objective function computationally expensive

Nowadays, dividing wall column, DWC, is the most applied thermally coupled distillation column, since it consumes until 30% less energy with respect to conventional sequences [12]. There are three variables in competition in a DWC: number of stages in both sides of the wall, named here as prefractionator, N_{PF}, and main column, N_{CP}, and the heat duty, Q, Figure 1.

These variables are in competition because we can not decrease indefinitely the number of stages without increasing the heat duty, and vice versa. Optimal design of a DWC implies determining 5 integer and 5 continuous variables, Figure 1, subject to meet purities and recoveries in each product stream: distillate, side stream and bottom. Then, for a ternary mixture we have 3 objectives, 6 constraints and 10 variables. This mixed integer nonlinear problem is expressed as:

Figure 1. Dividing wall column

$$Min(Q,N_i) = f(R, F_k, N_{in,k}, N_{out,k}, N_F, N_S, N_i, Q)$$

subject to

$$\vec{y}_k \geq \vec{x}_k$$

(1)

Where R is the reflux ratio, F_k is the interconnection flow k, N_F is the feed stage of the sequence, N_S is the side stream stage, $N_{in,k}$ and $N_{out,k}$ are the number of stage where comes and leaves the interconnection flow k. In addition, the constraints are the required and obtained purities and recoveries, vectors x_k and y_k respectively, along with the equality in the number of stages in prefractionator and in the section between the interconnection stages in main column.

The optimal design of DWC is obtained with a multiobjective genetic algorithm with constraints presented previously [13]. The evaluation of objectives is made through a rigorous simulation in Aspen Plus, which is coupled to the design tool. As was reported before, from the total time required in the optimization almost the 95% is used by simulations performed in Aspen Plus [13]. So, this case is very appropriate to test the proposed approach, which is explained next. For more detailed information about this algorithm and its link to Aspen Plus the reader is referred to the previous work [13].

Speeding up a multiobjective genetic algorithm with constraints through artificial neuronal networks

3. Speeding up approach using artificial neuronal networks

The proposed approach is applied to a multiobjective genetic algorithm with constraints coupled to Aspen Plus [13]. The link of the algorithm with Aspen Plus allows evaluating the objective and constraint functions considering the complete model of the distillation column; this evaluation consumes almost the 95% of the total computing time, hence it is computationally expensive. One alternative to reduce the computational time is using a reduced model of the distillation column, considering constant relative volatilities for instance; this simplifies considerably the complexity of the model, but, at the same time, limits the application of the procedure to ideal o nearly ideal mixtures, which is not recommended.

3.1. Methodology

In order to reduce the computational time required to obtain the Pareto front, without using a reduced model of the distillation column, we propose the use of ANN. The ANN are used to create approximate objective and constraint functions, based on training of individuals evaluated with Aspen Plus; these approximate functions are computationally less expensive, in the order of milliseconds in comparison with the time used with Aspen Plus that varies from a few seconds to several minutes.

The integration of the ANN with the optimization strategy is described next, and it is schematically show in Figure 2. During all generations, the multiobjective genetic algorithm generates, reproduces, and mutates all individuals, both existing and new; also, it calculates the dominances of all of them. The difference lies in the evaluation of the objective and constraints function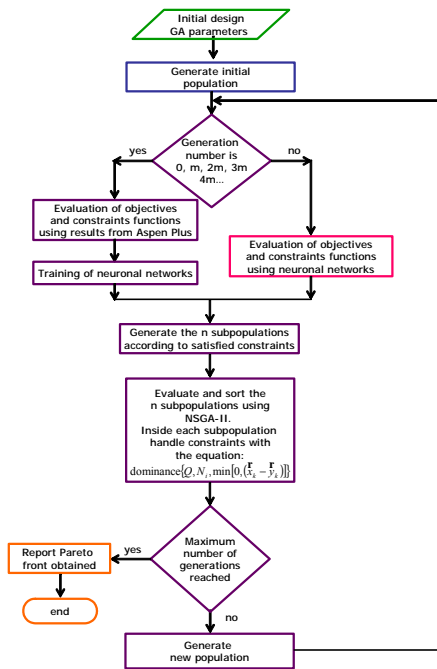s. In the initial generation, 0, the evaluation of objectives and constraints is made with Aspen Plus. The obtained results are used to create the individuals of the first generation, but also to train the ANN; in this way, we have an approximate function to evaluate objectives and constraints. From generations 1 to Am-1 all individuals are evaluated with approximated functions, as constraints as objectives; where A is an integer parameter which values can be 0, 1, 2, …, B, and m is the frequency of use of the original objective function, which values can be 2, 3, 4, …, n.

In the generation Am, evaluation of the objectives and constraints is made with Aspen Plus; the resulting data are used to retrain the ANN. So, the evaluation of objectives and constraints using Aspen Plus is made in generations Am. In other generations, the approximated functions are used to evaluate constraints and objectives, and they change every m generations, when ANN are retraining.

Figure 2. Flow diagram of the proposed approach

In brief, the use of approximated and original functions for evaluation of objectives and constraints is switched each m generations; where m is a frequency parameter which values can be 2, 3, 4, 5, 6 … n. In this way, we reduce the computational time, but we are still solving the original problem. The parameter m determines the frequency of use of the original objective function, both for the optimization procedure as for the training of the ANN. The value of this parameter is resulting from a tuning process; however, it is worth mentioning that for the minimal value of the parameter m, 2, at least the number of evaluations of original objective function is reduced in 50%.

In the selection of the ANN it is important to note that there is formal proof that the simplest neuronal network, multilayer perceptron with two hidden layers, is theoretically sufficient to model any problem [14]. However, previous analysis and performance validation must be made before selecting the ANN for each problem.

3.2. Cases of study

In order to test the new approach, we optimize the DWC for two mixtures using a multiobjective genetic algorithm with constraints [13], and also the new approach that considers the same algorithm along with the ANN. It is worth to mention that the proposed approach can be implemented in any multiobjective or simple algorithm (stochastic or deterministic), where evaluations of the objective function are computationally expensive. In this work, the neuronal network employed is the most simple of all: perceptron with two layers and eight cells in each layer. The functions included are sigmoid and linear, while the training is not supervised of 30 cycles each one. Nevertheless, the kind of ANN and the specific parameters can be varied for different applications.

Selected mixtures for this work are: n-pentane – n-hexane – n-heptane, M1, and methanol – n-butanol – methyl formate, M2. The thermal condition of feed is saturated liquid, while 98% is the value assigned to the recoveries and purities of key components in each product stream. In order to make a comparison of performance of the optimization the initial designs, thermodynamic model, Aspen Plus convergence options, mutation and crossover operators are the same for both mixtures and procedures. Also, for both mixtures we use 50 generations of 1000 individuals each one, as for the original multiobjective algorithm as for the proposed approach with ANN; the parameter m is fixed as 5, resulting from a tuning process. The DWC is simulated in the Radfrac module of Aspen Plus, and the runs are performed in a Xeon 5410 workstation at 2.33 Ghz with 8 GB in RAM.

3.3. Discussion of results

Pareto fronts of the DWC for the two selected mixtures were obtained using the multiobjective genetic algorithm with constraints, GAMOC, and the new approach, GAMOC-ANN. A comparative analysis of the evolution of the Pareto fronts is presented next; it is worth of mention that further analysis about the optimal designs themselves that integrated the Pareto front is not the objective of this work.

The Pareto fronts integrate the three variables in competition defined before: number of stages in both sides of the shell (prefractionator and main column) and the heat duty. The resulting designs represent the best trade-off between these variables in competition. The evolution of Pareto fronts of the DWC generated using GAMOC and GAMOC-ANN for mixture M1 is presented in Figures 3 and 4, respectively. In these figures, each series of dots of different color represent a different generation, beginning

Speeding up a multiobjective genetic algorithm with constraints through artificial neuronal networks

at generation 1 with the red circles. From these figures we observe how the Pareto front evolve as the optimization is in process; however, the followed trajectory is completely different in both figures.
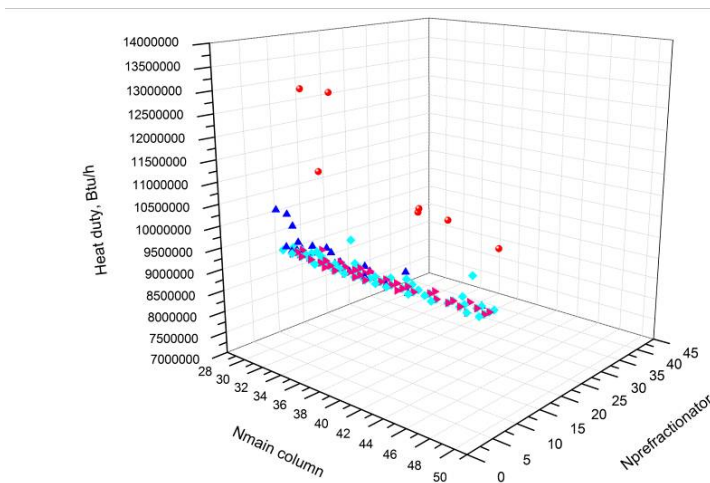

Figure 3. Pareto front of DWC obtained with GAMOC, for mixture M1
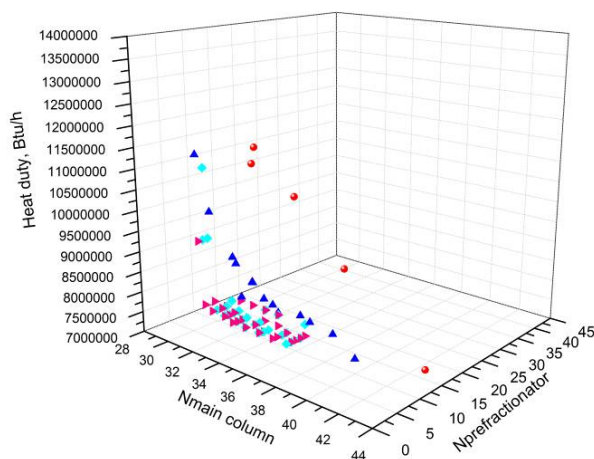

Figure 4. Pareto front of DWC obtained with GAMOC-ANN, for mixture M1.

From figures, it is clear that the use of the ANN allows reaching the Pareto front more quickly, while for this number of generations the Pareto front is inclusive not reached by GAMOC. Using GAMOC-NN were required just 11, 000 evaluations of the original objective function, in comparison with the 50, 000 needed by GAMOC. This represents a significant reduction of 78% in the number of evaluations of the objective function; also a proportional decreasing in the computational time is observed. The frequent retraining of the neuronal network allows refining the estimation of the objectives and

constraint as they approach to the optimal zone. Another important characteristic is that in spite of the use of approximate functions, the Pareto front maintained its diversity when GAMOC-ANN is used. A similar behavior is observed in the Pareto fronts of DWC for mixture M2.

In general, the proposed approach allows reaching the vicinity of the optimal zone, where Pareto front is located, very quickly, in comparison with the use of the optimization strategy alone. However, we observe that once that the vicinity of this zone has been reached, a decreasing in the number of individuals that integrate the Pareto front is observed. This fact suggests that the proposed approach is very useful to reach the optimal zone, but inside is not so efficient. In spite of more test must be done, this findings indicate that the proposed approach should be use just to reach the optimal zone, and inside it the original objective function instead.

4. Concluding remarks

A simple and easy procedure to speed up simple or multiobjective algorithms, stochastic or deterministic, has been presented. The proposed approach considers the use of artificial neuronal networks to generate approximate functions for objective and constraint functions. These approximate functions are several magnitude orders less computationally expensive than the original one. Results show a significant decreasing in the number of evaluations of the original objective function required to reach the Pareto front; as consequence, a decreasing in the computational time is observed also.

It is important to mention that this approach can be used in any simple or multiobjective algorithm (stochastic or deterministic), with or without constraints, where the evaluation of the original objective function is computationally expensive. In spite of the selection of the ANN must be analyzed and validated for each case, results show that even the simpler neuronal network has a good performance, even in a very complex mixed integer highly nonlinear problem of optimal design.

5. Acknowledgements

References

1. D. Sarkar et al, 2005, Chem. Eng. Sci., 60, 2, 481-492.
2. R. B. Kasat et al, 2003, Comp. & Chem. Eng., 27, 12, 1785-1800.
3. A. Gaspar-Cunha et al, 2005, IJCSS, 6, 1, 18-36.
4. K. Rasheed et al, 2000, Proceedings of GECCO, 628-635.
5. B. Suman et al, 2004, Comp. & Chem. Eng., 28, 9, 1849-1871.
6. B. Doerr et al, 2007, Evolutionary Computation, 15, 4, 401-410.
7. R. A. Tikidji-Hamburyan, 2008, Proceed. of 17th Annual CNS Meeting, 9, Suppl 1, P90.
8. D. Whitley, 1995, Genetic Algorithms in Engineering and Computer Science, Chap. 11, 15 pp.
9. B-T Zhang et al, 1993, Complex Systems, 7(3), 199-220.
10. K. F. Man et al, 1996, IEEE Tranactions on Industrial Electronics, 43(5), 519-534.
11. H. Kitano, 1990, AAAI-90 Proceedings, 789-795.
12. Z. Fonyó et al, 2001, Comp. & Chem. Eng., 25, 119-140.
13. C. Gutiérrez-Antonio et al, 2009, Comp. & Chem. Eng., 33, 2, 454-464.
14. V. Kurková, 1992, Neural Networks, 5(3), 501-506.

Constraint Satisfaction Problem for Case-Based Reasoning Adaptation: Application in Process Design.

Eduardo Roldán[a], Stéphane Negny[a], Jean Marc Le Lann[a], Guillermo Cortés[b]

[a]*Laboratoire de Genie Chimique UMR 5503, INPT-ENSIACET, 4 allée Emile Monso, BP 44362 – 31432 Toulouse, France; eduardo.roldan@ensiacet.fr; stephane.negny@ensiacet.fr; jeanmarc.lelann@ensiacet.fr*
[b]*Postgrade Department,ITO,Orizaba, Ver, 31400, Mexico; gcortes@itorizaba.edu.mx*

Abstract

Process design represents a complex domain where past experiences are often used to solve new design problems. Currently in the Artificial Intelligence field and more particularly in Knowledge Management some techniques arise in order to accelerate the entire design process. Among these methods Case Based Reasoning is emerging as an efficient approach to deal with this challenge, allowing to store previous design cases to help for solving new problems and reducing considerably the design time. However existing cases must be adapted to new problem situation. In this article we deal with the case adaptation problem and we propose an improvement of CBR using a Constraints Satisfaction Problem approach applied to the early phase of process design.

Keywords: Case Based Reasoning, Constraint Satisfaction, Adaptation.

1. Introduction

Rapid product development cycle requires methodologies to efficiently explore the design space to build products with reduced costs, improved functionality and quality to compete in global markets. Consequently, design task has had an enormous improvement thanks to new computed aided methods. Nowadays, there are many techniques in the artificial intelligent field that scopes to account for these challenges during the design phase. However the problem complexity often represents a barrier to perform it. Knowledge based systems are emerging as a viable way to require design experiences in order to produce useful solutions. Among the knowledge management approaches, Case Based Reasoning (CBR) and Constraint Satisfaction Problem (CSP) are the most appropriated for the design task.

In CBR, knowledge is stored in a case base. Each case represents the description of a past problem (source) and its associated solution (Sol(srce)). Thanks to the similarity and distance measurement the most similar cases can be retrieved, with adaptation if necessary, and proposed as a solution (Sol(target)) to the new problem (target). This approach has the advantage to build an efficient tool for supporting the design phase and to maintain it easily. On the other hand it is necessary to provide an important number of cases to have significant results. Moreover the similarity measure function is often complex and it does not guarantee to have the most relevant case for the resolution because the most similar is not always the most relevant.

The CSP approach requires a considerable effort to extract and interpret knowledge. Here knowledge is modeled in a constraint form such as logical relations, mathematical

expressions or even domains of validity... The new problem is subject to the knowledge model via certain number of variables. Then a reasoning process is led through constraints to reduce the variables domain validation. All the possible solutions can then be reached. The major disadvantage of this approach is that it requires a tremendous effort to identify and formalize knowledge and to build the reasoning model. However, the systems managing constraints have the advantage to quickly provide original solutions or to establish when a problem does not have a solution.

Due to their complementary, we propose to couple these two approaches in this work. Case Based Reasoning arises as an efficient experiences capitalizing technique which retrieves past experiences with the purpose to use them as a solution for the new problem. Nevertheless these retrieved solutions can not be applied directly to the new problem, particularly in domains such as design, due to the complexity and the problems requirements. From this point, the adaptation process takes a fundamental role in a CBR system. Consequently we introduce a methodology for formalizing the adaptation process and using a constraint satisfaction technique. We also propose an interactive process between the expert and the adaptation phase in order to acquire its knowledge. Viewing adaptation as a constraint satisfaction problem provides several advantages (Purvis, 1998) for instance it gives an efficient method to be applied in many domains, it fits very well with CBR language because it takes into account variables, values and constraints, it provides also a particular way to decompose a large problem in to smaller problems and allows combination of them. In the other hand, the expert interaction will reduce the knowledge engineering effort allowing him to correct the solution and making improvements directly to the adaptation process.

This paper presents our methodology enclosing these terms and addresses the issues of the adaptation process in the domain of process design. In the next section we give an overview of adaptation process in CBR, the representation model utilized by the proposed methodology and its procedures are explains in section 3. Finally the methodology is demonstrated trough an application example in process design.

2. Adaptation

Case adaptation is the process where a retrieved solution can be transformed into an appropriate one for the current problem. Several authors avoid this phase and prefer to develop the part of retrieval by considering that the case abundance will compensate adaptation task. However other authors (like us) consider adaptation as a crucial part of CBR systems because it confers to them their quality of problem solvers. Moreover, our goal is to propose a tool to the preliminary design stage, where the number of past experiences is limited and adaptation is therefore decisive. In the current CBR bibliography, it exists at least four traditional strategies to perform the adaptation task (Pal and Shiu, 2004) and (Wilke, 1998):

- Reinstantiation is the simplest form of adaptation; the new problem solution is copied directly from the retrieved case and used without modification. This kind of adaptation is also known as "null adaptation".
- Substitution replaces parts of the old solution attributes which are unacceptable because they are in conflict with the new problem conditions.
- Transformation is used when no suitable substitution is possible. A solution will be derived while being based on the constraints and the characteristics from the desired solution. These constraints specify which properties the solution would have or should not have.
- Multi-cases adaptation combines several cases to find the new solution.

*Constraint Satisfaction Problem For Case Based Reasoning Adaptation: Application
In Process Design*

It should be noticed that these strategies are complementary according to similarity degree between the source case and the target problem. Carbonell (Carbonell, 1983) pointed out two approaches to perform adaptation viewing this as an analogy; by knowing that the target solution is to source solution the same as target problem is to source problem then by knowing the source problem, source solution and the new problem is possible to calculate the target solution. Two modes emerge for successful adaptation the *transformational mode* that modifies the solution of the recovered case directly and the *derivational mode* which modifies the method describing how the problem was solved.
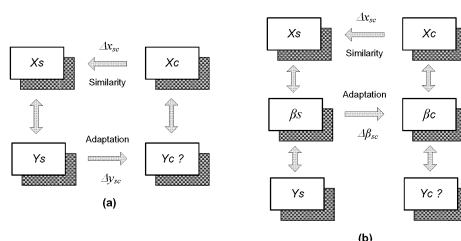

Figure 1. Transformational (a) and Derivational (b) adaptation modes.

Based on the derivational mode explained above (Figure 1), we can formalize the adaptation task as follows: a case can be represented as $C_s\{X_s;\ Y_s\}$ where X_s and Y_s are vectors of attributes describing respectively the problem and the solution part, a new case is described in the same way as $C_c\{X_c;\ Y_c\}$. The objective is to find Y_c. From this point, adaptation can be performed by following the next steps (Cordier, 2007):

1. With $(X_s,\ Y_s) \rightarrow \beta_s$: method to reach Y_s starting from X_s
2. $(X_s,\ X_c) \rightarrow \Delta X_{sc}$
3. ΔX_{sc} with adaptation knowledge $\rightarrow \Delta\beta_{sc}$
4. $(\beta_s,\ \Delta\beta_{sc}) \rightarrow \beta_c$
5. $(X_c,\ \beta_c) \rightarrow Y_c$

The next section describes our methodology and explains how we can acquire the adaptation knowledge and modify the adaptation method ($\Delta\beta_{sc}$) to elaborate β_c (steps 3 and 4).

3. Methodology

A few research on adaptation formalization involving a constraint solving approach is available in the current CBR bibliography, in (Purvis, 1998) the multi-case approach is taken and also solves the local inconsistencies between cases to produce a global solution to the new problem. Its adaptation mechanism uses a CSP algorithm and basically tries to eliminate the costly feasibility testing. However the principal drawback of the CSP method (i.e. the important effort of knowledge extraction and formalization) is still present. Adaptation is maybe the most important CBR step since it adds intelligence to what would be simple pattern matchers or even making CBR as a simply storage and retrieval system. To be satisfactory and effective, the adaptation phase needs some additional adaptation knowledge. This new knowledge can be extracted from expert experiences and coded manually in the CBR system as rules, semantic trees or decision tables. However, this extraction usually demands too much time. Another

alternative consists of making use of the case base by extracting adaptation knowledge using machine-learning techniques in order to generate specific heuristics. Nevertheless they require a base containing many cases in order to avoid inaccuracies and approximations on heuristics. Unfortunately, in preliminary design, it is unusual to have such vast case base. Moreover the adaptation knowledge management is often difficult and demands a great effort of engineering. Under these conditions, it would be more effective to have an "online" subprocess to acquire expert knowledge and check the solution during each problem solving episode.
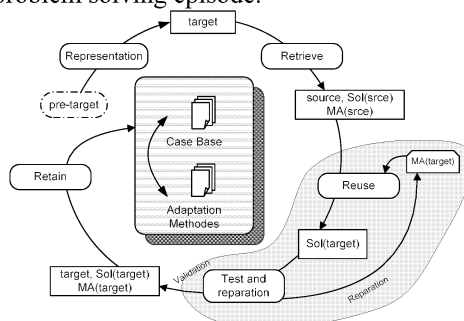

Figure 2. CBR process with the interactive expert loop.

This "online" subprocess (figure 2) relies in the expert interactions with the CBR system, in particular when an adaptation failure occurs and it is necessary to repair the solution. When the expert corrects the suggested resolution method, the knowledge is updated and added to the base. The advantage of this approach is a reduced effort of knowledge engineering by an expert specific request. After adaptation phase, the solution is proposed to the expert, here two situations are possible: 1) satisfactory: the new case is stored as well as the adaptation method; 2) unsatisfactory: the loop with the expert is activated and the adaptation method is modified by the expert until he obtains a satisfactory solution.
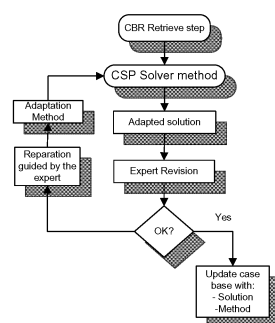

Figure 3. Adaptation methodology.

To exploit expert knowledge we chose to formalize it through constraints. Then for each new adaptation, reasoning is driven through these constraints thanks to Constraint Satisfaction Problem methods. Moreover, it will be possible to distinguish and to gather the general domain and the target problem specific requirements as constraints on a problem. The adaptation methodology consists in the following steps, figure 3 represents it graphically:

1. Retrieve case satisfying mandatory constraints.
2. Select an initial set of parameters covering the target constraints.
3. Solve the constraints with a CSP solver.

Constraint Satisfaction Problem For Case Based Reasoning Adaptation: Application
In Process Design

4. Establish whether the system is over-constrained (too many constraints and thus no solution) under-constrained (not enough constraints and thus too many solutions) or has just one solution.

5. If the system is over-constrained, the selection of an initial set of parameters (step 2) may be replayed, when the system is under-constrained, apply a filter.

6. Display solution to the expert. If its satisfactory go to step 7 else modified the adaptation method and return to step 3.

7. Store new solution and the adaptation method if relevant for the CBR system.

Our methodology tries to exploit the maximum benefit from the user interface and by providing a good initial solution methods based on the already solved cases.

4. Application

In order to validate our approach, we have chosen the design of a new compression station environment for gas supply. The problem is to configure a compression station that better fits with the system requirements.

Component	Methane
Flow rate (kg/s)	130
Temperature (K)	190.60
Pressure (Bar)	46.00
Pipeline: Diameter (m)	0.787
Length (m)	60
Fuel supply for compressors (kg/s)	0.850

Table 1. Attributes and values for compression station needs.

According to our methodology is necessary to find the existing cases that better match the mandatory technical constraints of the new problem. Table 1 lists the main attributes and values required for the system. Next, our CBR system displays the two most similar retrieved cases from the base, Table 2, including features of the compressors.

Features	*Case 1*	*Case 2*	*Proposed Model*
Component	Methane	Methane, Ethane	C1=comp[ethane]
Flow rate (kg/s)	140	140	C2=flow[140]
Pressure (Bar)	41	50	C3=pressure[<50]
Pipeline: Diameter (m)	0.650	0.800	C4=diam[0.8]
Length (m)	55	70	C5=length[<60]
Fuel consumption (kg/s)	0.751	0.900	C6=fuel[0.850]
Compressor	2 (serial)	2 (serial)	*C7=power[user_120]
Powerful (HP)	100	120	*C8=weight[user_320]
Rotational speed (rpm)	2000	2500	
Weight (kg)	300	380	*User constraints

Table2. Compressor retrieved cases and proposed model.

After retrieval and before resolution, the first proposed model is composed of constraints C1 to C6 as stated by the system needs (Table 1), but the designer adds two additional constraints; C7 and C8 according to compressor specific requirements. However, the application of the proposed model with the current problem inputs does

not lead to a satisfactory solution. Therefore, the interactive expert loop is activated. After several iterations, the user modifies the technical constraints, for instance C1 and C2. Furthermore he adds specific problem requirements trough constraints C9 and C10, Table 3. The added constraint C9 was introduced in according to the answer of whether the compressors should be connected in series or in parallel. Because it depends on expert knowledge i.e. back-up strategies, operational strategies or future expansion. Finally the new case results in a consistent final solution for the new problem as seen in Table 3.

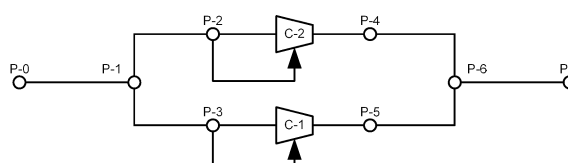| Final Model |
| --- |
| **C1=comp[methane]** |
| **C2=flow[130]** |
| C3=pressure[46.00] |
| C4=diam[0.787] |
| C5=length[60] |
| C6=fuel[0.850] |
| C7=power[user_120] |
| C8=weight[user_320] |
| ***C9= type[user_parallel]*** |
| ***C10=rpm[user_2000]*** |

Table3. Final model with its schema solution.

In this example the synergy between the CBR and CSP via the interactive expert loop, helps to identify the necessary problem variables and provide trough the retrieved cases the initial solution and constraints for the new problem.

5. Conclusions

In this paper several ideas were introduced to present a general formalization of the adaptation phase for CBR process. First, adaptation is viewed as a main step in CBR systems for designing a solution of the target problem, based on the relations existing between a source case and a target problem. From this perspective we assumed adaptation as a Constraint Satisfaction Problem in order to achieve a better performance of the CBR system. In addition this article aims at proposing an interaction with the expert to improve adaptation. Moreover an example application is presented and may be used as a reference for further applications. Also CBR adaptation can be guided by domain knowledge which can be directly provided by the human interaction or extracted from the system itself for designing a solution.

Future work on this methodology will involve the use of machine learning approach in order to extract new knowledge from the case base.

References

A. Cordier, B.Fuchs, J. Lieber and A. Mille, 2007, Acquisition interactive des connaissances d'adaptation integré aux sessions de RAPC, 15th CBR Workshop, pp. 71-84

J. G. Carbonell, 1983, Learning by analogy : Formulating and generalizing plans from past experience, Machine Learning, An Artificial Intelligence Approach, chapter 5, pp. 137-161

Purvis and P. Pu, 1998, COMPOSER: A CBR system for Engineering Design, Robotica, Vol. 16, Issue 3, pp 285-295

S. Pal and S. Shiu, 2004, Foundations of Soft Case-Based Reasoning , Wiley-Interscience, pp. 136-160

W. Wilke, B. Smith, P. Cunningham, 1998, Using Configuration Techniques for Adaptation, Lecture Notes In Computer Science; Vol. 1400, pp. 139-168

Performance monitoring of an industrial boiler: classification of relevant variables with Random Forests

Matthieu Sainlez [a], Georges Heyen [b]

*aCRISIA, Haute-Ecole Robert Schuman, Chemin de Weyler 2, B-6700 Arlon, Belgium,
E-mail: matthieu.sainlez@hers.be*
*bLASSC, Université de Liège (ULg), Sart Tilman B6A, B-4000 Liège, Belgium,
E-Mail: G.Heyen@ulg.ac.be*

Abstract

A data mining methodology, the random forests, is applied to analyze pollutant emission from the recovery boiler of a Kraft pulping process. Starting from a large database of raw process data, the goal is to identify the input variables that explain the most output variations.

Keywords: data mining, random forests, Kraft pulping process, recovery boiler, atmospheric pollutants.

1. Introduction

Data Mining refers to extracting useful knowledge from large amounts of data. Starting from large databases, the main objective is to find interesting latent patterns [1].
Particularly, a random forest [2, 3] is an ensemble of unpruned classification or regression trees, induced from bootstrap samples of the training data, using random feature selection in the CART induction process. Prediction is made by aggregating the predictions of the ensemble. Internal estimates are also used to measure variable importance [2].
Within the framework of a Kraft pulp mill, we analyze the emissions of the recovery boiler, and particularly the nitrogen oxide emission. This kind of boiler acts both as a high-pressure steam boiler and as a chemical reactor with reductive and oxidative zones. Significant perspectives already exist to reduce atmospheric pollutants, and the identification of the most important variables is an interesting byproduct of random forests.

2. The Kraft pulping process

The Kraft process [4] is an alkaline process to produce chemical pulp. A pulp mill can be divided in two main areas: fiber line and chemical recovery loop.
Cellulose fibers are dissociated from lignin by cooking the chips in a solution of sodium hydroxide (NaOH) and sodium sulfide (Na_2S), called white liquor. The residual black liquor is washed from the pulp and treated to recover the cooking chemicals. The black liquor is concentrated and burned in a recovery furnace to yield an inorganic smelt of sodium carbonate (Na_2CO_3) and Na_2S. The smelt is dissolved to form green liquor, which is treated to recycle the calcium carbonate and to regenerate the white liquor.

3. Random Forests methodology

In this paper, we consider a regression problem in which we are trying to predict the value of a continuous variable: pollutant emission of nitrogen oxide (NO_x).

We have a training set $\mathbf{z} = \{(x_1, y_1), \dots, (x_n, y_n)\}$ where each $x_i = \left(x_{i1}, x_{i2}, \dots, x_{ip}\right)^T$ is the ith measurement vector of p input attributes, y_i is the continuous outcome. We fit a model to $\mathbf{z}$, obtaining the prediction $\hat{f}(x)$ at input x.

Bagging is a general strategy for improving predictor accuracy [1]. The bagging algorithm creates an ensemble of models (by bootstrap sampling) for a learning scheme where each model gives an equally-weighted prediction.

Particularly, random forests are a combination of tree predictors such that each binary tree depends on the values of a random vector sampled independently and with the same distribution for all trees in the forest [2].

3.1. From binary trees to random forests

The binary tree [5, 6] is a widely used framework in data mining; this concept can be applied both to classification or regression problems. Basically, it's a sequence of binary decisions applied to the input variables; each non-terminal node contains a decision involving the comparison of an attribute with a given threshold, which then leads to another node or to a leaf (a terminal node). The root node contains the whole dataset which is recursively splits into two branches at each node. A greedy algorithm selects the attribute and threshold that maximizes a given fitness measure.

A particular tree framework called CART (for "*Classification and Regression Trees*") maximizes the Gini index that selects the split with the lowest impurity at each node; CART was introduced in 1984 (Breiman et al., [7]).

Generally, the resulting tree is easily interpretable (giving a set of decision rules), it works with both numerical and categorical data, and it's a non parametric method (no a priori assumption is made). Unfortunately, trees are sensitive to small changes in the learning sample (Breiman, [8]). Moreover, unstable trees can be stabilized via an ensemble method: we average the predictions of a set of individual models (see for example, [14]).

Practically, we have a single training data set, and so we have to find a way to introduce variability between the different models: we use bootstrap data samples [5]. A bootstrap replicate is a random subset of the original dataset, of the same length, taken with replacement [9].

We generate m bootstrap samples and then use each to train a separate copy of a predictive model. This procedure is known as bootstrap aggregating or *bagging* [8].

The aim of aggregating is to create an improved model. We take the average value of each prediction for a given test sample.

For each bootstrap sample $\mathbf{z}_i$ ($i = 1, \dots, m$), we grow a CART tree T_i and we aggregate the ensemble $\{T_i\}_i^m$ (see Figure 1). For a giving prediction $\hat{f}_i(x)$, the bagging estimate is the average of predictions over the m trees.

Bagging is very helpful for reducing variance and, for prediction, Breiman [8] proved theoretically that a bagged predictor will always have improved accuracy over a single predictor.
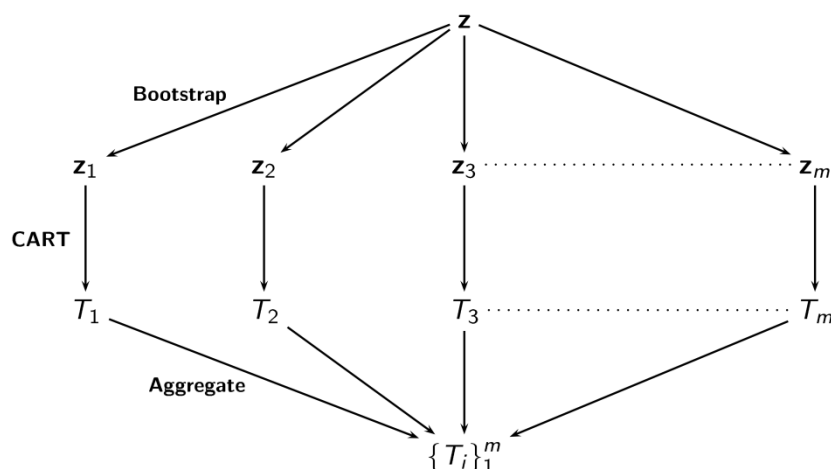
Performance monitoring and optimization of industrial processes

Figure 1 : Bagging : bootstrap aggregating

3.2. Random Forests algorithm

A random forest (Breiman, [2]) is an ensemble of unpruned classification or regression trees, induced from bootstrap samples of the training data, using random feature selection in the CART induction process.

In the random forest methodology [2, 6, 10], a second source of diversity is introduced during the growing of each tree. For each node, the method selects a small random subset of k attributes (from the p input attributes) and uses only this subset to search for the best split. This random selection of features at each node decreases the correlation between the trees in the forest thus decreasing the forest error rate.

We fit each tree on bootstrap sample and we select threshold and attribute at each node from a subset of attributes; the algorithm is described below (Hastie et al., [10]):

- For $i = 1, \dots, m$:
    a) Draw a bootstrap sample $\mathbf{z}_i$ of size n from the original sample $\mathbf{z}$.
    b) Grow a random forest tree T_i to the bootstrapped data, by recursively repeating the following steps for each terminal node of the tree, until the minimum node size n_{min} is reached.
    i) Select k variables at random from the p variables ($k \leq p$).
    ii) Pick the best variables/split-point among the k.
    iii) Split the node into two daughter nodes.
- Output the ensemble of trees $\{T_i\}_i^m$, the prediction at a new point x is given by:

$$\hat{f}_{RF}^m(x) = \frac{1}{m} \sum_{i=1}^{m} T_i(x)$$

In this study, we take $(k, n_{min}) = (\left\lfloor \frac{p}{3} \right\rfloor, 5)$; these are classical values for regression [10]. An analysis of complexity and prediction score helps for selecting appropriate m.

4. Industrial case study

In this paper, we analyze nitrogen oxide (NO_x) emission from a Kraft recovery boiler; the main objective is to find explanatory attributes for predicting NO_x pollutant emissions (we focused this paper on the attributes selection scheme).

The recovery boiler furnace can be considered as consisting of three distinct zones [4]: a drying zone where the black liquor is fired, a reduction zone at the bottom, and the oxidation zone in the turbulent upper section. Air for combustion is introduced from the bottom upward as primary, secondary, tertiary and quaternary air (at different velocities to ensure complete mixing).

The original database is a $(n \times p) = (65509 \times 56)$ matrix. The p attributes are mainly physical flow rates, pressures, and temperatures of black liquor, fuel, air…

We use a Matlab R13 implementation of Breiman's random forest algorithm for regression ([11] , based on Breiman and Cutler's original Fortran code version 3.3).

4.1. A data mining approach for modeling

Firstly, the data need to be preprocessed to make it appropriate for the study [12]. Then, the given original data set is partitioned into two independent sets [1], a training set (70% of the data) and a test set (the remaining 30%). The training set is used to derive the model, whose accuracy is estimated with the test set (see Figure 2).
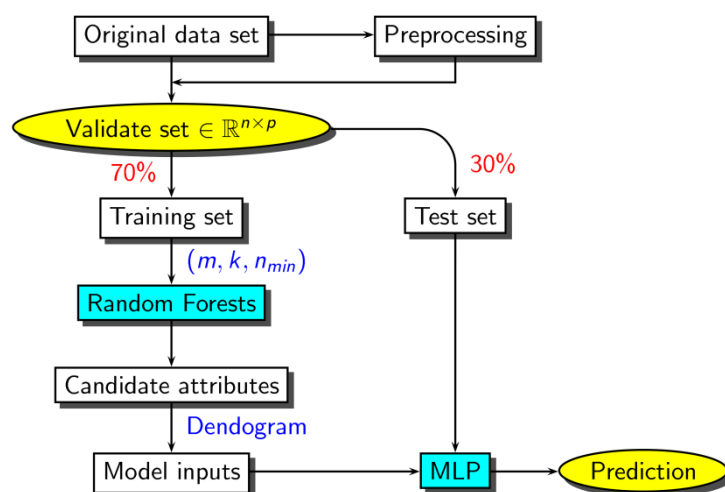

Figure 2: A data mining approach for prediction.

We use random forests to rank input attributes according to their importance measure. Candidate attributes correlations are also analyzed (e.g., with a dendogram) to avoid redundancies. The resulting attributes are the inputs of the model.

A successful model is the feed-forward neural network [5], known as multilayer perceptron (MLP). In the end, the model relevance is assessed by its performance for predicting new observations.

Performance monitoring and optimization of industrial processes

4.2. Attributes selection scheme

In many data mining applications, only a few input variables have substantial influence on the response. It is often useful to learn the relative importance or contribution of each input variable in predicting the response.

Random Forests use the out-of-bag (OOB) samples to a variable importance measure. On average, 37% of the samples will not be present in a given bootstrap replicate [6, 9]: they are called OOB sets. When a tree in the forest is grown, the OOB samples are passed down the tree, and the prediction accuracy is recorded.

Then, one at a time, each attribute values are randomly permuted in the OOB samples, and the accuracy is again computed. The decrease in accuracy as a result of this permuting is averaged over all trees, and is used as a measure of the importance of a variable in the random forest [10, 13].

4.3. Results

Attributes are ranked according to this importance measure (expressed as a percentage of the overall importance, see Figure 3). For a fixed number of trees m, a variable with a larger importance score relative to other variables indicates that the variable is important for regression. This hierarchy presents only the first 25 relevant attributes.
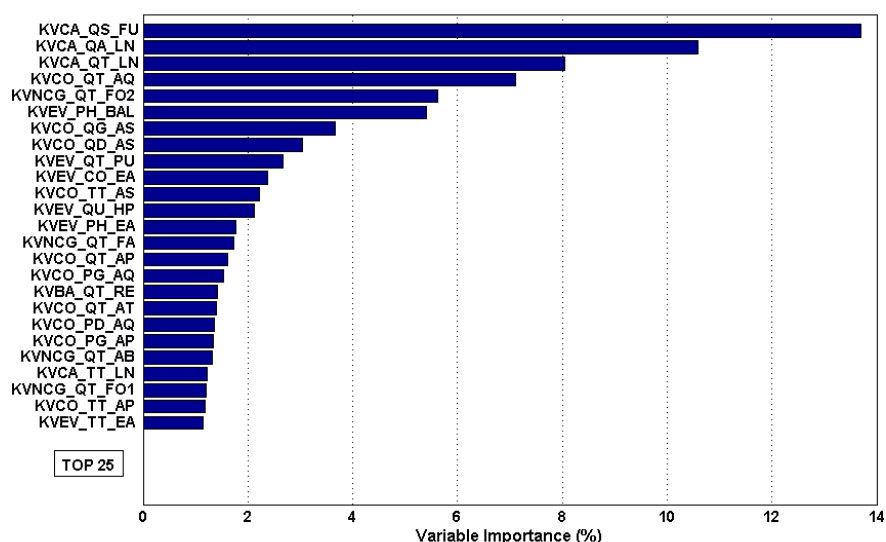

Figure 3: Random Forest attributes score importance - $(m, k, n_{min}) = (80, 18, 5)$

After a breakdown, during a boiler start up or consequent to high variation in steam demand: the furnace is brought to the right temperature by burning heavy fuel. We can observe that, logically, the fuel rate introduces at the lower of the furnace (QS-FU) is the more relevant attribute for predicting NO_x emissions.

The entering black liquor flow rate (QA-LN) is a part of the total black liquor flow rate (QT-LN) which is circulated in a loop around the liquor guns; these variables are highly correlated. Then, quaternary air (QT-AQ) was especially designed to control NO_x emissions and non-condensable gas (NCG-QT) coming from the process are incinerate into the furnace.

5. Conclusions

This paper briefly traced the application of random forests to a pulp mill atmospheric pollutant. We talked only about the attributes selection scheme; these selected attributes can also be used for prediction with a neural network (e.g., a multilayer perceptron).

Random Forests handle very large database and its internal variable importance measure is very helpful for understanding complex interactions between attributes and discovering latent patterns. This method is easy to use and quite fast, requiring only a little tuning on parameters.

Acknowledgements

References

[1]   Jiawei Han, Micheline Kamber. Data Mining : Concepts and Techniques.
        Morgan Kaufmann Publishers, 2006

[2]   Leo Breiman, Random Forests. Machine Learning, 45: 5-32, 2001.

[3]   Yongheng Zhao, Yanxia Zhang, Comparison of decision tree methods for finding active objects. Advances in Space Research, 41: 1955–1959, 2008.

[4]   Gary A. Smook, Handbook for Pulp & Paper Technologists - Third Edition.
        Angus Wilde Publications, 2002.

[5]   Christopher M. Bishop, Pattern Recognition and Machine Learning. Springer, 2006

[6]   P.M. Granitto a, F. Gasperi, F. Biasioli, E. Trainotti, and C. Furlanello. Modern data mining tools in descriptive sensory analysis: A case study with a Random forest approach. Food Quality and Preference, 18: 681–689, 2007.

[7]   Breiman, L., Friedman, J., Olshen, R. and Stone, C. Classification and Regression Trees, Wadsworth, 1984.

[8]   Leo Breiman, Bagging Predictors. Machine Learning, 24:123-140, 1996.

[9]   B. Efron, Estimating the error rate of a prediction rule: some improvements on cross-validation. Journal of the American Statistical Association, 78: 316–331., 1983.

[10]   Trevor Hastie, Robert Tibshirani, Jerome Friedman. The Elements of Statistical Learning: Data Mining, Inference, and Prediction - Second Edition.
        Springer Series in Statistics, 2009.

[11]   Ting Wang, Package random forests for Matlab R13,
        available at http://lib.stat.cmu.edu/matlab/

[12]   Sigurdur Olafsson, Xiaonan Li, and Shuning Wu. Operations research and data mining. European Journal of Operational Research, 187: 1429–1448, 2008.

[13]   Kellie J. Archer, Ryan V. Kimes. Empirical characterization of random forest variable importance measures.
        Computational Statistics & Data Analysis, 52: 2249 – 2260, 2008.

[14]   Pierre Geurts, Damien Ernst, and Louis Wehenkel. Extremely Randomized Trees.
        Machine Learning, 36: 3-42, 2006.

An off-line Model Reduction-based Technique for On-line Linear MPC Applications for Nonlinear Large- Scale Distributed Systems

Weiguo Xie and Constantinos Theodoropoulos

School of Chemical Engineering and Analytical Science, University of Manchester, Manchester, M60 1QD, United Kingdom, E-mail: k.theodoropoulos@manchester.ac.uk

Abstract

Linear Model Predictive Control (MPC) has been effectively applied for many process systems. However, linear MPC is often inappropriate for controlling nonlinear large-scale systems. To overcome this, model reduction methodology has been exploited to enable the efficient application of linear MPC for nonlinear distributed-parameter systems. An implementation of the proper orthogonal decomposition method combined with a finite element Galerkin projection is first used to extract accurate non-linear low-order models from the large-scale ones. Then a Trajectory Piecewise-Linear method is developed to construct a piecewise linear representation of the reduced nonlinear model. Linear MPC, based on quadratic programming, can then be efficiently performed on the resulting system. The stabilisation of the oscillatory behaviour of a tubular reactor with recycle is used as an illustrative example to demonstrate our methodology.

Keywords: model reduction, Model Predictive Control, Distributed Systems, Proper Orthogonal Decomposition, Trajectory Piecewise-Linear

1. Introduction

Model Predictive Control (MPC) is widely used for many processes. Over the last two decades, linear MPC has become a popular and effective advanced control strategy. However, it often leads to poor performance for non-linear systems except near the point at which the model was identified [1]. Nonlinear large-scale system models need expensive computations, restricting the application of MPC. Therefore, nonlinear MPC is mostly used in batch systems and linear MPC in continuous operations. Model reduction techniques have been used with nonlinear MPC for distributed systems (e.g. [2]). The main purpose of this work is to combine an off-line model reduction technique with linear MPC for nonlinear large-scale systems. The proper orthogonal decomposition (POD) method [3] combined with a finite element (FEM) Galerkin projection is first used to extract non-linear reduced models for a "representative" range of parameters. The low-order models are then linearised with a Trajectory Piecewise-Linear (TPWL) method. The resulting reduced piece-wise linear system is then efficiently controlled through linear MPC. A particular appeal of this technique is that despite the nonlinearity of the original system, a quadratic objective function is always extracted for the MPC formulation. Furthermore, the POD-FEM-based reduced model is nonlinear only in one dimension, time. Hence, TPWL linearisation is effectively reduced to a set of 1-D linearisations in time. Moreover, since the POD-FEM-TPWL procedure is performed off-line all the on-line MPC computations are computationally inexpensive. This methodology is especially promising for multi-parametric MPC (e.g. [4]). The developed technique is illustrated using the tubular reactor system which

exhibits complex nonlinear dynamics [5]. We believe that this technique can have significant impact in the applicability of linear MPC to nonlinear large-scale systems.

2. Model reduction/linear MPC methodology

2.1. Proper orthogonal decomposition

The mathematical theory behind POD is the spectral theory of compact, self-adjoint operators employing the Karhunen-Loeve decomposition theorem [6]. It has been used for the model reduction of parabolic PDEs in conjunction with nonlinear controllers [7], and for the optimization [8] and control [9] of transport-reaction processes. POD is very efficient from a data compression point of view among all linear decompositions, which can contain the most "energy" in an average sense [3]. The energy in a given mode can be measured by the magnitude of the eigenvalue corresponding to that mode. The method of snapshots [10] is often used to determine the corresponding *empirical* eigenfunctions. The procedure for POD includes (i) collecting data from the dynamic model or from experiments for an appropriate range of parameters (ii) constructing the two-point correlation matrix of the collected responses, (iii) calculating a few, $l \ll n$, global basis functions, which capture most of the system's energy through singular value decomposition (SVD) of this matrix (iv) expressing the state variables $x(y,t)$ of the system (y denoting spatial coordinates) as linear combinations of these basis functions $\phi(y)$ and of some time coefficients $a(t)$, $\sum_{i=1}^{l} a_i(t)\phi_i(x) + \overline{x}$ and (v) calculating these time coefficients using a Galerkin projection of the original model onto the computed basis functions, producing a low-order model consisting of l equations.

2.2. Trajectory piecewise-linear method

Rewieński, [11] presented trajectory piecewise-linear methods especially the weighted method. It has been noted that it is costly to generate TPWL models of nonlinear large-scale dynamic systems by simulation of the original system. Here we propose to use the POD reduced systems instead to perform the required simulations. Furthermore, an automated procedure to obtain the optimal linearization horizons has been developed. Piecewise linear interpolations are built for $i \in [1, n-1]$, $z \in [x_i, x_{i+1}]$.

$$L_i(z) = a_i + b_i(z - x_i) \tag{1}$$

where the coefficients are defined by $\alpha_i = y_i$, and $b_i = (y_{i+1} - y_i)/(x_{i+1} - x_i)$.

To obtain optimal horizons for the TPWL method, the mean value theorem [12] is used:

$$f(z) = L(z) + \frac{f^{(2)}(\eta)}{2}(z - x_i)(z - x_{i+1}) \tag{2}$$

$\eta \in [x_i, x_{i+1}]$, and $L(z)$ is the linear interpolation for $f(z)$. If the 2nd derivative of f is bounded by M_2 and $\overline{h}$ is the length of the longest subinterval, then for $z \in [\alpha, \beta]$.

$$\left| f(z) - L(z) \right| \le \frac{M_2 \overline{h}^2}{8} \tag{3}$$

We can use this error bound to get the smallest integer that satisfies the above inequality. If $L(x)$ is based on the uniform partition $\alpha = x_1 < x_2 < \cdots < x_n = \beta$, where $x_i = \alpha + (i-1)(\beta-\alpha)/(n-1)$. To ensure that the error between L and f is less than or equal to a given positive tolerance δ, we insist that

An off-line Model Reduction-based Technique for On-line Linear MPC Applications for Nonlinear Large Scale Distributed Systems

$$|f(x) - L(z)| \le \frac{M_2 \bar{h}^2}{8} = \frac{M_2}{8} (\frac{\beta - \alpha}{n-1})^2 \le \delta \tag{4}$$

From this we conclude that n must satisfy

$$n \ge 1 + (\beta - \alpha)\sqrt{\frac{M_2}{8\delta}} \tag{5}$$

Here n is the number of horizons we used in TPWL. However, at highly nonlinear cases, such a "static" TPWL produces a large number of horizons leading to increased computational time. Hence, an adaptive version of TPWL has been developed. The subinterval $[xL, xR]$ is acceptable if $\left| f(\frac{xL + xR}{2}) - \frac{f(xL) + f(xR)}{2} \right| \le \delta$ or if $xR - xL \le h_{min}$, where, $h_{min} > 0$ are refinement parameters [12]. In addition, a partition $x_1 < \cdots < x_n$ is acceptable if each subinterval is acceptable.

2.3. MPC
Linear MPC is often formulated as a state-space model with linear discrete time [13].

$$x(k+1) = Ax(k) + Bu(k), \ x(0) = x_0 \tag{6}$$

Where $x(k) \in \Re^n$ and $u(k) \in \Re^m$ denote the state and control inputs. Receding horizon methods are performed by open-loop optimization with objective function:

$$J_{(p,m)}(x_0) = \min_{u(\bullet)} [x^T(p)P_0 x(p) + \sum_{i=0}^{p-1} x^T(i)Qx(i) + \sum_{i=0}^{m-1} u^T(i)Qu(i)] \tag{7}$$

where, $(p \ge m)$, p denoting the length of the prediction horizon or output horizon, and m the length of the control horizon or input horizon.

3. Case study

3.1. Tubular reactor case
The tubular reactor considered here is described by the following set of partial differential equations [5] defined on a spatial domain $z \in [0,1]$:

$$C_t = -\frac{\partial C}{\partial z} + \frac{1}{Pe_C}\frac{\partial^2 C}{\partial z^2} - f(C,T) \quad T_t = -\frac{\partial T}{\partial z} + \frac{1}{Pe_T}\frac{\partial^2 T}{\partial z^2} + B_T f(C,T) + \beta_T(T_c - T) \tag{8}$$

where, C and T are concentration and temperature, respectively. T_C corresponds to temperature of the cooling medium and $f(C,T) = B_C C \exp(\frac{\gamma T}{1+T})$ is the reaction term.

The parameters used are: Pe_C=7.0, Pe_T=7.0, B_C=0.1, B_T=2.5, γ=10.0 and β_T=2.0. Part of reactant at the output stream is to be recycled to the feed stream at a ratio r, the boundary conditions for concentration and temperature at $z = 0$ become [14]:

$$\frac{\partial C}{\partial z} = -Pe_C[(1-r)(1+C_0) + rC(t,1) - C(t,0)] \text{ and } \frac{\partial T}{\partial z} = -Pe_T[(1-r)(1+T_0) + rT(t,1) - T(t,0)] \tag{9}$$

The boundary conditions at $z = 1$ are $dC/dz = 0$ and $dT/dz = 0$. The reactor exhibits oscillations at $C_0 = T_0 = T_c = 0$ for r=0.5 [15]. The model was discretised using FEM in 16 quadratic elements. Results from these simulations are shown in Fig. 1.

3.2. Control objective

It can be seen that the tubular reactor shows stable behaviour for *r*=0 (Fig. 1a) while it undergoes sustained oscillations for *r*=0.5(Fig. 1b).
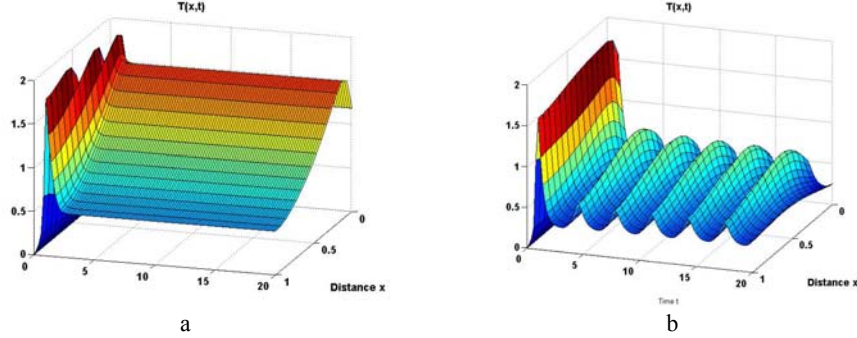

a   b

Figure 1: Temperature profiles of tubular reactor (a) $r = 0$ (b) $r = 0.5$

Our control objective was to stabilize the reactor with *r*=0.5 to behave like the system with *r*=0 by introducing a number of jacket temperature zones (actuators). The objective function is as follows:

$$J = \min_{du} \left(T(t) - T_{ref}(t) \right)^T Q \left(T(t) - T_{ref}(t) \right) + DU^T RDU \qquad (10)$$

where, $T_{ref}(t)$ is the reference state (r=0) and DU is the control on the actuators.

3.3. Data sampling

A method using Heaviside step functions (whose value is 0 or 1) has been applied to facilitate sampling. For 8 actuators we have $2^8 = 256$ states. Taking 11 samples over the range of temperature (-0.999,1) and concentration [0,1], we have $256 * 11 = 2816$ samples. The full-scale FEM model was used for this and the sampling time was 15s.

3.4. POD model reduction

Global basis functions were computed based on the 2816 samples collected. *l*=8, eigenfunctions for concentration and temperature capture 99.5% of the system's energy. The FEM model was then projected onto the POD eigenfunctions to produce the reduced model. The comparison between full and reduced model for the dynamics of the middle and output points for $r = 0.5$ is shown in figure 2a (concentration) and 2b (temperature). The reduced model can accurately predict the complex reactor dynamics.

a   b
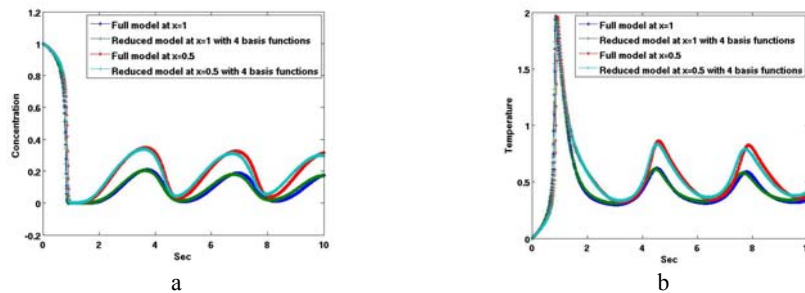
Figure 2: Comparison between (a) concentration (b) temperature predictions of full model and reduced model at the middle and output points for tubular reactor with $r = 0.5$.

3.5. TPWL method for time coefficients

Both static and adaptive TPWL were used to linearise the time coefficients. Figure 3 shows the adaptive TPWL segments for the temperature time coefficients for $\delta = 0.001$

An off-line Model Reduction-based Technique for On-line Linear MPC Applications for Nonlinear Large Scale Distributed Systems

and $r = 0$. Adaptive TPWL produces 198 time segments, much less compared to the 600 static TPWL intervals.
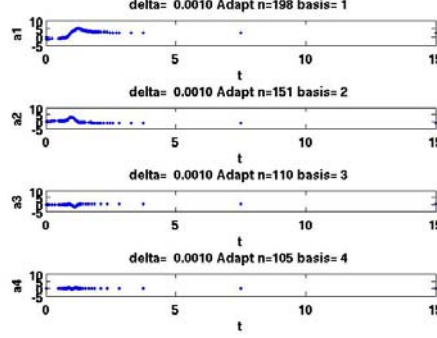

Figure 3: TPWL for time coefficients on temperature of tubular reactor using POD method

3.6. Control law for Linear MPC

The POD method was applied on the control objective (Eq. 10) resulting in a quadratic function (equation 11) due to the linear POD representation of the state variables.

$$J = \min_{du}\left((\sum_{k=1}^{m} \alpha_{k\_T}(t)\varpi_{k\_T}(x) + \overline{T_{16}}) - T_{ref}(t) \right)^{T} Q\left((\sum_{k=1}^{m} \alpha_{k\_T}(t)\varpi_{k\_T}(x) + \overline{T_{16}}) - T_{ref}(t) \right) + DU^{T}RDU \quad (11)$$

POD on the nonlinear constraints resulted in a reduced set of nonlinear constraints as a function of time coefficients $\alpha(t)$. TPWL was then applied on these time coefficients, a 1-dimensional linearization only, to obtain piece-wise linear constraints:

$$\alpha(t+1) = L_1\alpha(t) + B_1 U(t)$$
$$\alpha(t + p/t_n) = L_p\alpha(t + (p-1)/t_n) + B_p U(t + (p-1)/t_n)$$
$$y(t) = H\alpha(t) + \overline{T_{16}} \quad (12)$$

where, $\alpha(t)$ includes time coefficients for concentration and temperature, and

$$H = [\underbrace{0,0,\cdots 0}_{m}, \varpi_{1\_T}(z_{16}), \varpi_{2\_T}(z_{16}), \cdots \varpi_{m\_T}(z_{16})]^{T} \quad (13)$$

Therefore, we obtain a quadratic objective function subject to piece-wise linear constraints. The control law can be obtained explicitly from [16]:

$$DU = (G_{y1}^{T}QG_{y1} + rI)^{-1}G_{y1}^{T}Q[Y_{ref} - G_1\alpha(t) - G_{u1}U(t-1)] \quad (14)$$

where,

$$G_{y1} = \begin{bmatrix} HB_1 & 0 & \cdots & 0 \\ HB_2 + HL_2B_1 & HB_2 & 0 & 0 \\ \vdots & \vdots & \cdots & 0 \\ HB_n + HL_nB_{n-1} + \cdots + HL_nL_{n-1}\cdots L_2B_1 & HB_n + HL_nB_{n-1} + \cdots + HL_nL_{n-1}\cdots L_3B_2 & \cdots & HB_n \end{bmatrix}, G_{u1} = \begin{bmatrix} HB_1 \\ HB_2 + HL_2B_1 \\ \vdots \\ HB_n + HL_nB_{n-1} + \cdots + HL_nL_{n-1}\cdots L_2B_1 \end{bmatrix}, G_1 = \begin{bmatrix} HL_1 \\ HL_2L_1 \\ \vdots \\ HL_nL_{n-1}\cdots L_1 \end{bmatrix}$$

Then, control output variables can be calculated by

$$Y = G_1\alpha(t) + G_{y1}DU(t) + G_{u1}U(t-1) + \overline{T_{16}} \quad (15)$$

In Fig. 4, results of the linear MPC for 8 actuators are shown. Fig. 4a shows the control law for the 8 actuators (zones). The control output (dashed line) and the reference profile (solid line) are shown in Fig. 4b. As it can be seen the reactor is effectively

stabilized, while the on-line computation time is less than 2 minutes. It should be noted that since this is a model predictive controller there is no bias in the control output.
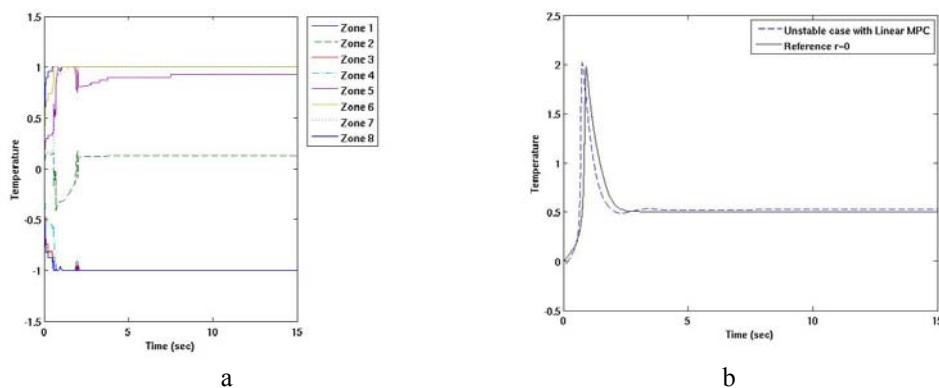

Figure 4: Linear MPC results (a) control law for 8 actuators (b) control and reference profile.

4. Conclusions

We have developed a POD-FEM-TPWL technique, which enables the use of linear MPC for highly non-linear systems. Results of the tubular reactor case showed that the POD accurately predicts the system dynamics. Our adaptive TPWL method can obtain an optimal number of linear segments. Linear MPC effectively stabilises the linearised reduced system predicting the appropriate dynamic temperature profiles for 8 jacket temperature zones, with low computational cost.

5. Acknowledgements

References

[1] Tenny MJ, Wright SJ, Rawlings JB (2004) Comput. Optimization Appl., 28, 87-121.
[2] Li MH, Christofides PD (2008) Computers & Chemical Engineering 32, 2123-2135.
[3] Holmes P, Lumley JL, Berkooz G (1996) Turbulence, Coherent Structures, Dynamical Systems and Symmetry Cambridge University Press.
[4] Pistikopoulos, EN (2009) AICHE Journal, 55, 1918-1925.
[5] Jensen KF, Ray WH, (1982) Chemical Engineering Science 37, 199–222.
[6] Wong E (1971). Stochastic Process in Informaion and Dynamical Systems. McGraw-Hill.
[7] Baker J, Christofides PD (2000) Int J Control 73, 439–456.
[8] Bendersky E, Christofides PD (2000) Chem Eng Sci 55, 4349–4366.
[9] Shvartsman SY, Theodoropoulos C, Rico-Martinez R, Kevrekidis IG, Titi ES, Mountziaris TJ (2000) Journal of Process Control, 10, 177-184.
[10] Sirovich L, (1987), Quart. Appl. Math. 45, 561-590.
[11] Rewieński M, White J (2003) IEEE Trans. Computer-aided Design of Integrated Circuits and Systems, 22, 155-170.
[12] Van Loan CF, (1997). Introduction to scientific computing, Prentice-Hall, Inc.
[13] Morari M, Lee JH (1999) Computers and Chemical Engineering, 23, 667-682.
[14] Antoniades C, Christofides PD (2001) Chem Eng Sci, 56, 4517–4535.
[15] Alonso AA, Frouzakis CE, Kevrekidis IG (2004), AIChE Journal, 50, 1438-1452.
[16] Huang S, Tan KK, Lee TH, (2002). Applied predictive control, Springer-Verlag, London.

Integrated Approach for Enhanced Teaching and Learning towards Richer Problem Solving Experience

Sonia Zheleva,[a] Toshko Zhelev[b]

[a]*Faculty of Engineering & IT, Griffith College Limerick., Limerick, Ireland,*
sonia.zheleva@gmail.com
[b]*Department of Chemical & Environmental Sciences, University of Limerick, Limerick,*
Ireland, toshko.zhelev@ul.ie

Abstract

The utilisation of a game environment for problem solving as part of the concept of three-fold, theory-simulation-experiment laboratory, enriched with advanced mobile wireless technology components for information access, is the focus of the research reported in this paper. The resource design is being motivated by the goal of incorporating both advanced gaming and communications infrastructure to realise a novel multilevel educational experience. The paper illustrates how expertise in two key technology fields – gaming and telecommunications – is combined and integrated with discipline expertise and sound pedagogy for educational content generation, and by this means, how new pathways in the natural knowledge and experience-gaining process may be created. An important principle in the research has been to exploit play and variety of contemporary ICT support for e-Learning and problem solving facilitation, including digital games, wireless mobile technology and learning management systems such as Moodle or Sakai-Sulis. A second principle is to seek an extended access to these new e-Learning paradigms beyond the physical educational campus, through their incorporation into virtual campuses.

1. Introduction

One of the goals in education in present times is to provide an easier access and flexible learning opportunities through use of technology. Technology enhanced learning opens remarkable new avenues for learning and skills development. Learning technologies can be a vital tool to enrich what Higher Education Institutions (HEIs) do best - opening the world of new ideas and helping individuals to develop their learning, critical and creative thinking skills (Anderson, 1997). Countries and HEIs willing to take advantage of these new opportunities, must be proactive in launching meaningful reforms and innovations, and embrace changes. Technology should be used to nourish, transform and enrich the strong relationship between teacher and learner and to promote more active and student-centered learning. Today, people need to be provided with knowledge to be competent, and with incentives to be motivated in acquiring this knowledge.

One reason today's educators "are not more successful at educating children and workforce, despite no lack of effort on their part, is because ***they are working hard to educate a new generation in old ways***, using tools that ceased to be effective" (Prensky, 2001). Thus, the immediate task in front of today's educators is to develop methodologies that speak the language of this highly technological generation (the Net Gen), to "stop telling", but to invent new teaching and learning ways in order to provide

education on some of the driest and boring subject matter imaginable. The latest explorations of multimedia potential interactivity (Aldrich *et al.*, 1998), influenced by the constructivist philosophy of learning resulted in re-discovery of **play** as the most fundamental concept of human instruction. Centuries ago Plato, the Greek philosopher, stated that *"...children should not be kept to their study by compulsion but by play"*. To this, Abraham Maslow has added that *"Almost all creativity involves purposeful play"* . And what better "remedy" to the problem of boredom in the classroom than the use of e-learning through multimedia game technology with its potential interactivity, vivid images, 3-dimensional graphics, and audio! Using the play as the most fundamental concept of human instruction with today's digital game technology, there are ample opportunities to build a unique and stimulating virtual reality in order to improve learners understanding on abstract concepts that otherwise are difficult to grasp. Learners can achieve this through immersion in compelling story lines where fantasy, curiosity, challenge and control, as intrinsic motivations for learning, are met (Castellan, 1993). However, what makes good instruction is not just the medium, it is the methods that guide the way the medium is used (Clark, 1995). Thus, for games to benefit educational practice and learning they need to combine fun elements with aspects of sound pedagogy, instructional and system design that include motivational and interactive learning components (Gagne *et al*, 1988).

Lately, a number of Virtual Environments (VEs) has been developed for educational purposes, which are particularly useful when the learning domain is complex, abstract in nature and difficult to master, and when the virtual features of the learning environment are critical to the success of the learning process. VEs should provide a close physical resemblance to the real world, immediate feedback, and strong sense of learner's presence. Through use of story line in a particular game concept, analogies, metaphor, simulation, and avatars, learners may play an active role in creation of their own collaborative virtual environment, that can be even more educational, entertaining and "real" than the real world.

2. Designed by Students, Aimed for Students

Motivation is the most important aspect of the learning, since learning is not just finding and memorising facts - it should be fun! And what better opportunity for this, but to utilise the creativity of post-graduate students in the development of exciting educational tools to facilitate the learning process of their fellows, namely, undergraduate students! Two such projects for system modeling have been developed at CS&IT department of Durban University of Technology, South Africa (Zheleva, 2001). The main reason for modeling is to deal with systems that are too complex to understand directly. Models reduce complexity by separating a small number of important things to deal with at a time. Since models omit non-essential details, they are easier to manipulate in comparison to the original entity. And this is possible because abstraction is a fundamental human capability that permits us to deal with complexity (Shlaer and Mellor, 1988). Both project developments were successfully integrated into the Systems Analysis and Design module and achieved a dual goal: firstly, to educate postgraduate students in the use of interactive multimedia and authoring tools for educational development, and the underlying principles of Instructional Design, to guide and enable them through an apprenticeship to develop a good quality multimedia courseware on a complex engineering content; and secondly, to implement these developments (at no developmental cost to the department) in the

undergraduate modules as part of tutorial sessions, in order to stimulate learning and foster better understanding of the content.

A digital game, as part of a PhD study, has also been designed (Zheleva *et al*., 2002a, Zheleva *et al*., 2002b). It targeted the creation of a unique virtual environment and investigated the usefulness of digital games in the process of teaching and learning on a specific engineering domain for process integration and environment protection, and the usefulness of virtual learning spaces as educationally viable tools in general. The project attempted to demonstrate that by combining digital game technologies with a sound educational pedagogy and knowledge management, it is possible to build a unique and stimulating virtual reality that may improve learners understanding on abstract and complex engineering concepts, and motivate them act intelligently in challenging situations. The project demonstrated that by instilling best practice academics can introduce engaging, experience-centred, authentic and multi-sensorial learning activities for which the new multimedia game-based technologies provide ample opportunities.

Two other projects have been developed by final year BSc: Computer Systems students at CSIS department, University of Limerick, Ireland. They aimed at harnessing the power of modern technology to create and evaluate virtual learning spaces for environment protection as an appropriate educational tool to teach on complex and abstract engineering content. Based on modern educational and design principles, the projects targeted virtual learning environments from both educational and multimedia game technology perspectives. They resulted at development of an educational product to demonstrate the ability of educators to foster critical skills in learners and learners' abilities of becoming resourceful industrial developers and process system engineers. The projects intended to nurture expertise and gain experience in advanced software development related to game and streaming technologies, and to create an environment for training under- and postgraduate students in the development and evaluation of interactive game learning environments as well as provision of future entrepreneurial opportunities for them.

The results of the evaluation of these projects and the recent developments in technologies showed huge potential of boosting students' learning efficiency, and called for further extension of the discussed teaching and learning concept towards development of students abilities for active problem-solving and troubleshooting.

3. The New Concept (Work in Progress)

The work presented above attempted to provide an overview on some of the modern pedagogical theories and practices. From the many cited in the literature examples, including the discussed above, it is evident that interactive multimedia and game environments can be powerful educational tools, and today's educators should embrace the changes in technology, and provide meaningful reforms and innovations in their teaching methods. One way ahead might be the use of combined conventional methods with technology integrated teaching and learning. As Chris Dede argues, HEIs should plan for "neo-millennial" learning styles that include "*fluency in multiple media and simulation-based virtual settings*", and "*induce learning based on collectively seeking, sieving, and synthesizing experiences, rather than individually locating and absorbing information from a single best source*" (Dede, 2005). This type of active learning, based on both real and simulated experiences, begins with direct participation and then infuses guidance and frequent opportunities for reflection. Such an approach could enhance the

efficiency of knowledge transfer towards building problem solving abilities in engineering education and enrich the engineering students' learning experience.

The new concept is based on integrating recent technological innovations and involving different cross-disciplinary cross-institutional expert groups, with a view to deepening students' educational experiences and broadening instructors' capacity to guide and reinforce meaningful learning towards active problem solving. Its goal is the utilisation of advanced methodologies for teaching and learning based on the original concept of a Three-Fold Laboratory (theory-simulation-experiment), enriched with virtual environments and mobile wireless technology components (mobile wireless devices (MWDs)) for media convergence and information access improvement. This approach combines two key technology fields – digital gaming and telecommunications, to generate discipline specific educational content based on the expertise of senior educators in the field. An important feature of the proposed concept is first, simulation of real problem-solving scenarios, and then building new pathways to enhance the process of experience-gaining learning. Behind this resource development is the concept of integrating alternate reality games (ARG), which incorporate digital gaming and simulations, with robust mobile communication infrastructure for a novel multilevel educational experience. Through this approach, learners can gain knowledge by participating in immersive virtual environments and simulations, where they collaboratively identify problems, form and test hypotheses, and deduce evidence-based conclusions about underlying causes. They use location-aware handheld computers with GPS technology, allowing them to physically move through a real location while collecting place-dependent simulated field data, interviewing virtual characters or real persons, and collaboratively investigating simulated scenarios. Within the realm of an alternate reality game, digital games are used to deliver understanding about particular content area, and learners are provided with opportunities for reflection on and discussions about the content in spaces that are external to the game. Simulations are regarded as dynamic systems with which learners can test theories about how systems work and are affected by manipulating certain factors, and how certain principles of dynamic systems can be observed and played out. Such systems include their own internal assessment measures that can be used to assess students understanding of both micro and macro elements. The educational games used in the larger alternate reality game may be designed to take advantage of the spare time in student's life – the time before or after classes, going to and from classes, etc. Thus, educators do not use class time to play the games, but utilise it to discuss the data coming from the games outside the class. Using wireless PDAs to play the game and GPS services to integrate real world experiences with the virtual experiences of the game, and sending data back to a central server from where the professor may access it, this approach embed students in realistic real world scenarios for which alternate reality games are most suited.

4.1 Some Design Aspects and Components

Before the start of classes, learners sign on each semester to a website where they can post information about their interests and set up learning teams with their peers. There, they find the narrative of a scenario with clues to follow in an alternate reality game. The scenarios with clues and puzzles are designed by a supervision team of educators. These clues may be anywhere – websites, libraries and databases, on/off campus real locations or in virtual worlds, they can be printed materials or recorded telephone messages, provided by real persons, or virtual characters in a digital game. Learners may find GPS coordinates as clues that send them to a real field sites (e.g. waste-water treatment or pharmaceutical plants), where interviewing lead engineers may provide

them with significant information to solve the problem on hand. Or, the clue may point to a digital game on the website where a virtual character is holding the knowledge needed to solve a puzzle. Learners can meet and talk about their strategies, may post their findings and experiences on blogs, and have discussions in online forums.

The developing team of educators has to pre-design a clue-based algorithm for the purpose of problem-solving navigation. The general structure of each problem-solving task is planned to be within the expertise of a particular lecturer from the supervision team of educators. The problem solving process should follow the deductive "what-if" algorithm cycles thus leading to a deeper knowledge acquisition. The guided process of discovery and overcoming obstacles would utilise various possible means of information gathering in the process of achieving final solutions, based on logical deduction and induction, and variety of knowledge-gaining methods. The process will allow for accommodation of uncertainties and non-standard solutions.

The on-line multidisciplinary supervision and hints generation is another specific component of the concept. A "controlling/supervising station" will play a role of guiding navigation hub, where real world support will be provided mainly through visualisation, introduction of a set of hints, and some limiting conditions generation.

The envisaged real-time projects/workshops will facilitate students' dynamics of decision making in finding correct solutions for the problems on hand, which in essence means a time constraint for the project delivery, decision making deadlines and corresponding rate/efficiency of solutions provided in accordance to specified requirements. In practical terms such projects are more often associated with real-life troubleshooting type of problem solving. Such an educational component could be achieved through short-term supervised tasks within a short delivery spans (a day).

The last specific side of the concept is the multidimensionality of resources and methodologies to be utilised/explored, and the agility of their applications. The teaching/learning components and resources will include basic college level retrospectives (bringing old knowledge back to work), past college experience, lecture notes of current courses, library resources, games for teaching and learning, multimedia courseware and simulations, and direct communication between team members.

The composition of the problem and the process of problem-solving assistance incorporate utilisation of an InfoStation-based multi-agent system for the provision of intelligent mobile services within the University campus area. This system will pay particular attention to the interactions of the various entities in providing the e-learning services to students and educators in the 'best possible way' through flexible adaptation to the mix of current user preferences, mobile device capabilities, and wireless access network constraints.

The design of the learning process in such environment is to support achievement of a 'threshold-level knowledge' and guided transition (invisibly supervised by the lecturer in charge) to the next higher level of problem understanding and subsequent hypothesis/solution generation and testing. These levels include information gathering, followed by analysis, problems identification and formulation, and natural approaches towards problem solving. This teaching/learning experience can have both real and virtual nature as a multidimensional game environment with integrated elements of advanced mobile wireless technology and GPS services. An important principle in the concept is to maintain and exploit better the modern ICT support for e-learning (e.g. teaching and learning management systems such as Moodle or Sakai-Sulis).

4. Conclusions

Ideas on learning technologies suitable for HEI under- and postgraduate programmes covering interactive multimedia and simulations, to virtual environments and digital games, to alternate reality games have been discussed. Addressed were evidences that such systems that incorporate **play** can be effective in providing inspiration and stimulation for learning and in fostering the desire to explore advanced knowledge in specialist fields at tertiary educational level. Captured in this research is the approach of using alternate reality game incorporating digital games, simulations, the virtual laboratory concept known as 'theory-simulation-experiment' and the integration of advanced mobile wireless technology, as an e-Learning technological support in the educational process. The attempt to integrate recent technological innovations suggested by different cross-disciplinary cross-institutional expert groups, with a view to reinvent play in a gaming environment based on this 'theory-simulation-experiment' concept has realistic potential to achieve innovations in deepening the students' educational experiences and instructors' capacity to better transfer skills and knowledge to learners in solving complex educational tasks. It is speculated that while both these benefits go well beyond what has been the case heretofore, the initial investment to achieve it will be well rewarded, economically, through the economies of scale in its uptake and usage.

References

F. Aldrich, Y. Roger and M. Scaife M. (1998). Getting to grips with "Interactivity": Helping teachers assess the educational value of CD-ROMs. British Journal of Educational Technology, 29 (4), pp. 321-332

T. Anderson (1997). Integrating Lectures and Electronic Course Material. Innovations in Educational and Training International, 34, pp. 3-10.

N. J. Jr. Castellan (1993). Evaluating information technology in teaching and learning, Behaviour Research Methods, Instruments and Computers, 25 (2), pp.233-237.

R. C. Clark (1995). Authorware, multimedia and instructional methods. Macromedia Inc. San Francisco, CA.

C. Dede (2005). Planning for Neo-Millennial Learning Styles: Implications for Investment in Technology and Faculty, in Educating the Net Generation, D.G. & J. L. Oblinger, eds. (Boulder, Colo.: EDUCAUSE, 2005), available at http://www.educause.edu/ir/library/pdf/pub71010.pdf

R. M. Gagne, L. J. Briggs and W. W. Wagner (1988). Principles of Instructional Design. 3rd ed., Holt, Rinehart & Washington, New York.

M. Prensky (2001) "Digital Natives, Digital Immigrants, Part II: Do They Really Think Differently?" On the Horizon, Vol.9, pp. 15-24, available at http://www.markprensky.com/writing/digital.pdf

T. Shlaer and D. Mellor (1988). Object-Oriented System Analysis: Modeling the World in Data, Englewood Cliffs, New Jersey, Yourdon Press.

S. R. Zheleva (2001) On the Use of Interactive Multimedia as a Teaching and Learning Tool in a Tertiary Institution. REBOC - Faculty of Commerce Research Bulletin, DIT, SA

S. R. Zheleva, T. K. Zhelev and A. Amory (2002a). Virtual Learning Spaces: Africa as Technology Creator. ELSEVIER, AIDIC Conference Series, Vol. 5, pp. 353-360, ISBN 0390-2358.

S. R. Zheleva, A. Amory and T. K. Zhelev (2002b). Using Educational Adventure-style Game as a Teaching Method in a Specific Engineering Domain. Waste Management & the Environment, WIT Press, ISBN 1-85312-919-4.

Accurate and efficient solution of distributed dynamical system models

Andrea Balzano[a], Stefania Tronci[b], Roberto Baratti[b]

[a]*Dipartimento di Ingegneria del Territorio, Università di Cagliari, Piazza d'Armi, 09123 Cagliari, Italy, E-mail: balzano@unica.it*
[b]*Dipartimento di Ingegneria Chimica e Materiali, Università di Cagliari, Piazza d'Armi, 09123 Cagliari, Italy, E-mail: baratti@dicm.unica.it; tronci@dicm.unica.it*

Abstract

Efficient and accurate methods for numerical simulation of time evolution of distributed dynamical systems are presented and discussed. As case example, stochastic models based on the Fokker-Planck equation are considered, describing the evolution of the probability density function (PDF) of the state variables. Unlike most of the methods presented so far, the proposed methods fulfill the normalization and positivity conditions for the PDF even for very small values of the model diffusivities. Hence, the methods appear to be suitable for simulating chemical distributed process systems.

Keywords: Dynamical system, distributed parameter model, Fokker-Planck equation, Numerical methods, uncertainty

1. Introduction

The prediction of time evolution of state variables in distributed parameter models of dynamic systems is a typical problem in the field of chemical engineering. Depending on the variable profiles and on the system dimension, computations can be quite demanding (e.g., highly efficient chromatographic column), as a consequence of efficient and robust integration schemes being required to obtain an accurate solution of the governing partial differential equation (PDE). Among several dynamical systems ruled by a PDE, one is related to the behavior of stochastic models, as recently presented in a study on the effect of model uncertainties in an isothermal CSTR with Langmuir kinetics (Tronci et al., 2009). This approach, based on the Fokker Plank equation (FPE), allowed for a global solution of the underlying problem to be obtained, without neglecting system non-linearity, and revealed to be useful for model-based applications when the uncertainties effects are important. The results obtained suggested to extend the study of stochastic chemical process modeling to multi-dimensional systems (e.g. non-isothermal reactor), in order to increase the possibilities of application of the method. However, analytical solutions of the multi-dimensional FPE are available for a limited number of problems. On the other hand, the numerical integration of the FPE can be quite demanding, as demonstrated by several works recently published on this topic (e.g., Kumar et al. 2009). Beside the fundamental difficulty related to the increasing computational effort with increasing system dimension, the following requirements have also to be satisfied: (i) positive solution; (ii) normality constraints; (iii) unboundedness of the domain of the PDF, which implies that any numerical method has to assume a large enough domain to contain the support of the PDF. Actually, issues (i-ii) are most closely related to an accurate solution of PDF advection in the phase space due to the drift terms, related to the deterministic part of the underlying process. This last aspect is addressed in the present paper, where three

explicit, finite difference schemes are proposed as solvers for the FPE for systems with two state variables with uncorrelated noises. Numerical tests are presented, and solution accuracy and computational efficiency of the three methods are evaluated.

2. The Fokker-Planck equation

The time evolution of state variables in distributed parameter dynamic systems with uncertainties can be modeled by the following Ito system:

$$d\mathbf{x} = \mathbf{f}(\mathbf{x}, t)\,dt + \mathbf{G}(\mathbf{x}, t)\,d\mathbf{W} \tag{1}$$

where $\mathbf{x}$ is a vector of n state variables, $\mathbf{f}(\mathbf{x}, t) \equiv f_i$ is a drift vector representing the deterministic part of the process, $\mathbf{G}(\mathbf{x}, t) \equiv g_{ij}$ is a diffusion matrix, $d\mathbf{W} = \mathbf{R}(t)\,dt$ is an increment vector of a Wiener process with correlation function matrix $\mathbf{Q}\delta(t_1 - t_2)$, and $\mathbf{R}$ is a white noise. The process is said to be nonlinear if the drift is a nonlinear function of the state variables.

The solution of such model is expressed in terms of probability density function (PDF) of the state variable vector, $p(\mathbf{x})$. This can be determined numerically by methods that are all fundamentally related to the well-known Monte Carlo method. However, an alternative is based on the Fokker-Planck equation (FPE), which in the case of additive noise, $\mathbf{G}(\mathbf{x}, t) = \mathbf{G}(t)$, reads:

$$\frac{\partial p}{\partial t} + \frac{\partial f_i p}{\partial x_i} - D_{ij}\frac{\partial^2 p}{\partial x_i x_j} = 0 \ , \tag{2}$$

where use of Einstein's summation rule is assumed. Eqn. (2) is actually an advection-diffusion equation for the PDF p in the phase space, implying conservation of the unit PDF integral over the domain, with symmetric diffusivity tensor $\mathbf{D} = 1/2\,\mathbf{GQG}^{\mathrm{T}} \equiv D_{ij}$. It is to be noted that, even in the case of nonlinear process, the FPE is linear.

3. Description of the numerical models

Three explicit methods for the numerical solution of the 2D FPE have been compared as to accuracy of results and computational efficiency. The analysis is restricted to the case of diagonal diffusion matrix, i.e. the noises affecting the state variables are uncorrelated. The schemes mainly differ as to the handling of the advective terms, related to the drift velocity. To simplify the notation, we let $f_1 = u$, $x_1 = x$; $f_2 = v$, $x_2 = y$, and $D_{11} = D_{xx}$, $D_{22} = D_{yy}$. All three schemes stem from a finite volume integral discretisation of Eqn. (2), resulting in probability fluxes being defined at the boundary of control volumes. The methods are implemented on the staggered Arakawa-C Cartesian grid shown in Figure 1, with PDF values defined in cell centers and drift velocity defined on cell sides as normal components, by which the probability fluxes are computed in an optimal way.

3.1. MOSQUITO (MOS)

The MOSQUITO scheme (Balzano, 1999) was originally proposed to model 2D advective mass transport. It is implemented on the upwind biased stencil associated to a control volume shown in Figure 1a, reflecting the physical mechanism of pure advective transport in a velocity field. For positive u and v the numerical scheme reads:

$$p_{i,j}^{n+1} = p_{i,j}^n - \tau_x(F_{i+1/2,j}^n - F_{i-1/2,j}^n) - \tau_y(G_{i,j+1/2}^n - G_{i,j-1/2}^n) \tag{3}$$

where $\tau_x = \Delta t/\Delta x$, $\tau_y = \Delta t/\Delta y$; F and G are fluxes across the control volume faces, implying conservation of the unit PDF integral with elapsing time, and:

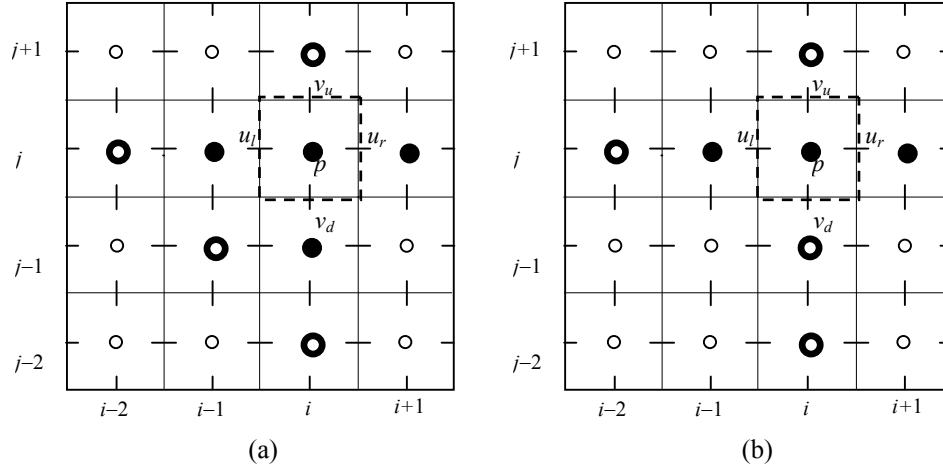*Accurate and effective solution of distributed dynamical system models*

Figure 1. Arakawa-C grid with control volume and computational stencils of the advection schemes for u>0 and v>0: (a) MOSQUITO; (b) split QUICKEST scheme. Full points involved in flux $F_r = F_{i+1/2,j}$; full and bold empty points involved in scheme for the whole control volume.

$$F_{i+1/2,j} = u_{i+1/2,j}\left(a_d p_{i+1,j}^n + a_u p_{i,j}^n + a_{uu} p_{i-1,j}^n + a_{uy} p_{i,j-1}^n\right) - D_{xx}\frac{p_{i+1,j}^n - p_{i,j}^n}{\Delta x} \tag{4}$$

with the diffusion terms handled with the second-order accurate in space scheme, and:

$$a_d = \tfrac{1}{6}(2 - 3c_{xr} + c_{xr}^2) \qquad a_{uu} = \tfrac{1}{6}(c_{xr}^2 - 1)$$
$$a_u = \tfrac{1}{6}(5 + 3c_{xr} - 2c_{xr}^2 - 3\overline{c}_{yr}) \qquad a_{uy} = \tfrac{1}{2}\overline{c}_{yr}$$

where $c_{xr} = u_{i+1/2,j}^n \Delta t / \Delta x$, $\overline{c}_{yr} = \overline{v}_{i+1/2,j}^n \Delta t / \Delta y$ are Courant numbers, and the *y* velocity component at point (*i*+1/2,*j*) is computed as suitable average of the *v* values at surrounding points where the component is defined. Analogous expressions are derived for the *y* flux and for different signs of the velocity components, according to the upwind concept. The QUICKEST scheme (Leonard, 1979) is recovered for the case of 1D pure advection.

The domain boundary is assumed to be closed, i.e. it does not allow for the PDF to leave nor enter the domain. Zero fluxes are then prescribed. This is appropriate if the steady state solution is of interest, whatever the time evolution of the PDF, provided the support of the asymptotic PDF is located entirely away from the boundary. The same condition must be fulfilled at any time instant if the PDF evolution with time is also of interest. Zero normal derivatives are also imposed at the boundary for fluxes at inner cell sides where the drift normal velocity is directed inward.

Guidance for selecting the time step to achieve stability is given by the stability conditions of the schemes for pure advection and pure diffusion deduced from Fourier analysis. Details of the truncation error and stability analysis for the pure advection scheme are given in Balzano (1999). The sufficient conditions $\max\{c_x, c_y, \} \leq 1$ for pure advection and $\Delta t \leq 1/[2(D_{xx}/\Delta x^2 + D_{yy}/\Delta y^2)]$ for pure diffusion hold.

Although originally designed for 1D advection schemes, use of an adaptation of the ULTIMATE flux limiter (Leonard, 1991) has proven effective in avoiding negative PDF values of significant magnitude to affect the solution, which is a drawback of most schemes presented in the literature for solving the FP equation.

3.2. Split QUICKEST (SQ)

The Split QUICKEST scheme is based on both dimensional and operator splitting, it being given by a sequence of factorized fractional steps of the form:

$$p_i^{n+1} = L_{QUI_x}^{\Delta t/2} L_{QUI_y}^{\Delta t/2} L_D^{\Delta t} L_{QUI_y}^{\Delta t/2} L_{QUI_x}^{\Delta t/2} p_i^n \tag{5}$$

where pure 1D advection is computed in steps 1-2 and 4-5 with the QUICKEST scheme (Leonard, 1979) and 2D pure diffusion is computed in step 3, with discretization of second spatial derivatives as in Eqn. (4). Boundary conditions are as for MOSQUITO.

Compared to MOSQUITO or, in general, non-splitting schemes, Split QUICKEST is of easier implementation, which makes it more and more convenient with increasing number of dimensions of the problem.

Stability conditions mentioned above, which are only approximate for MOSQUITO, are exact sufficient conditions for Split QUICKEST.

Unlike the case of MOSQUITO, use of the ULTIMATE flux limiter (Leonard, 1991) can be proved to ensure positive PDF values to be computed in each of the fractional 1D advection steps. Diffusive terms do not give rise to spurious oscillations.

3.3. Split Centred- Matsuno (SCM)

The third method tested is a splitting scheme of the same form as SQ (Eqn. (5)), but making use of 1D advection steps based on the second central derivative and Matsuno time stepping (Mesinger and Arakawa, 1976), which for the first fractional step reads:

$$\begin{aligned} p_{i,j}^* &= p_{i,j}^n - \tau_x (F_{i+1/2,j}^n - F_{i-1/2,j}^n) & F_{i+1/2,j}^n &= 0.5(p_{i+1,j}^n + p_{i,j}^n) \\ p_{i,j}^{n+1/5} &= p_{i,j}^n - \tau_x (F_{i+1/2,j}^* - F_{i-1/2,j}^*) & F_{i+1/2,j}^* &= 0.5(p_{i+1,j}^* + p_{i,j}^*) \end{aligned} \tag{6}$$

where the time index $n+1/5$ is merely symbolic, relating to the result of the first of five fractional steps. Diffusion step, boundary conditions and use of the ULTIMATE flux limiter are as for Split QUICKEST.

4. Computational results

The integration schemes described above have been compared in terms of their accuracy and CPU requirements, considering two numerical examples proposed in the literature.

4.1. Damped Duffing oscillator

The first test case considered is a damped Duffing oscillator governed by the equation:

$$\ddot{x} + \eta\dot{x} + \alpha x + \beta x^2 = gR(t) \quad , \tag{7}$$

considered by Kumar et al. (2009), with $\alpha = -15$, $\beta = 30$, $\eta = 10$, $g = 1$, for which the bimodal PDF shown in Figure 2a can be derived as stationary solution, of equation:

$$p(x,\dot{x}) = \wp \exp\left[-2\frac{\eta}{g^2 Q}\left(\frac{\alpha x^2}{2} + \frac{\beta x^4}{4} + \frac{\dot{x}^2}{2} \right) \right] \quad , \tag{8}$$

where $\wp$ is a normalization constant. Root mean squared (RMS) errors and CPU times obtained with four grids of sizes in the range 25×25 to 200×200 are shown in Table 1 for 100 time units. Substantial convergence conditions are obtained with the largest grid. Errors are virtually equivalent for the three methods, with MOS being slightly the most accurate, probably due to its lacking errors related to dimensional and operator splitting. Moreover, in this case MOS is also by far the fastest method.

Accurate and efficient solution of distributed dynamical system models

Table 1. RMS errors and CPU times for test case of damped Duffing oscillator

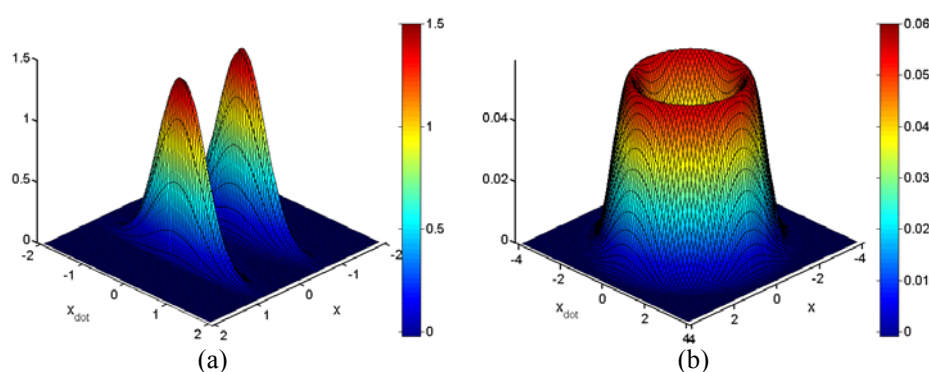| Grid size	SCM		MOS		SQ	
	RMS error	CPU time (s)	RMS error	CPU time (s)	RMS error	CPU time (s)
25×25	$3.80 \; 10^{-2}$	43.0	$3.78 \; 10^{-2}$	26.8	$3.79 \; 10^{-2}$	50.2
50×50	$1.88 \; 10^{-2}$	508.8	$1.88 \; 10^{-2}$	320.5	$1.88 \; 10^{-2}$	600.4
100×100	$3.38 \; 10^{-3}$	$11.2 \; 10^3$	$3.18 \; 10^{-3}$	$22.08 \; 10^2$	$3.33 \; 10^{-3}$	$66.76 \; 10^2$
200×200	$6.48 \; 10^{-4}$	$13.4 \; 10^4$	$6.31 \; 10^{-4}$	$23.19 \; 10^3$	$6.38 \; 10^{-4}$	$95.94 \; 10^4$

Figure 2. Stationary PDFs for (a) Duffing oscillator and (b) energy dependent damped oscillator.

4.2. Energy dependent damped oscillator

The second test case considered is an energy dependent damped oscillator governed by the equation (Muscolino et al., 1997; Kumar et al., 2009):

$$\ddot{x} + \eta\dot{x} + x + \beta(x^2 + \dot{x}^2) = gR(t) \; , \tag{9}$$

with parameters $\beta = 0.125$, $\eta = -0.5$, $g = 0.86$, for which the ring-like PDF shown in Figure 2b can be derived as stationary solution, of equation:

$$p(x,\dot{x}) = \wp \exp\left\{ -\frac{1}{2g^2}\left[\frac{\beta}{2}(x^2 + \dot{x}^2)^2 + \eta(x^2 + \dot{x}^2) \right] \right\} \; . \tag{10}$$

RMS errors and CPU times obtained for four grid sizes are shown in Table 2. Related error distributions obtained with MOS are shown in Figure 3. MOS is now by far the most accurate method; however, unlike test 1, SQ is the fastest scheme. This is due to the different relative importance of the advective and diffusive components, and their related influence on the stability condition, which is used to set the time step.

Table 2. RMS errors and CPU times for test case of energy dependent damped oscillator

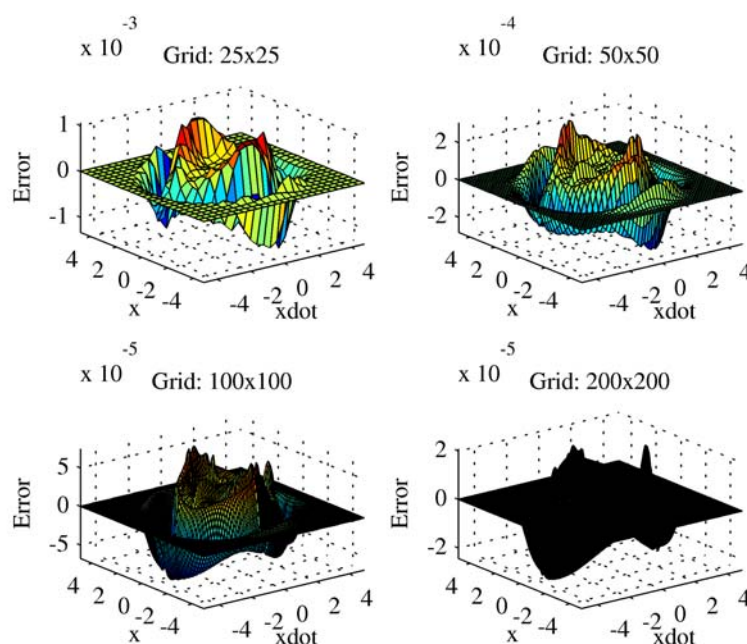| Grid size	SCM		MOS		SQ	
	RMS error	CPU time (s)	RMS error	CPU time (s)	RMS error	CPU time (s)
25×25	$5.46 \; 10^{-4}$	1.7	$3.44 \; 10^{-4}$	1.4	$3.52 \; 10^{-4}$	1.5
50×50	$1.94 \; 10^{-4}$	13.0	$7.66 \; 10^{-5}$	13.0	$8.85 \; 10^{-5}$	11.5
100×100	$8.31 \; 10^{-5}$	101.8	$1.98 \; 10^{-5}$	116.6	$2.92 \; 10^{-5}$	89.0
200×200	$3.97 \; 10^{-5}$	747.4	$6.42 \; 10^{-6}$	1203.3	$1.30 \; 10^{-5}$	682.6

Figure 3. Dependence of stationary PDF error distribution on mesh size for test case of energy dependent damped oscillator (MOSQUITO scheme).

5. Conclusions

In the present work different methods to solve numerically, efficiently and accurately, the time evolution of the state variables in distributed systems were presented and discussed. The presented methods were tested by solving the Fokker-Planck equation for two mechanical systems used as a benchmark in literature. It was found that the proposed methods fulfill the normalization and positivity conditions for the PDF, even for very small diffusivities values. Being this the case, it is evident their applicability also for chemical distributed process systems, also in cases characterized by very large Peclet numbers, as, e.g. high performance crhomatography and tubular reactor.

References

Balzano, A. (1999). MOSQUITO: An efficient finite difference scheme for numerical simulation of 2D advection. Intl. J. Numer. Meth. Fluids 31, 481-496.

Leonard, B.P. (1979). A stable and accurate convective modelling procedure based on quadratic upstream interpolation. Comp. Meth. App. Mech. Engng. 19, 59-98.

Leonard, B.P. (1991). The ULTIMATE conservative difference scheme applied to unsteady one-dimensional advection. Comp. Meth. Appl. Mech. Engng. 88, 17-79.

Tronci, S., Grosso, M., Alvarez, J., and Baratti, R. (2009), Stochastic dynamical nonlinear behavior analysis of a class of single-state CSTRs, ADCHEM 2009, Istanbul, Turkey.

Kumar, M. Chakravorty, S., Junkins, J. L. (2009). A homotopic approach to domain determination and solution refinement for the stationary Fokker-Planck equation, Probabilistic Engineering Mechanics, 24, 265-277.

Mesinger, F., and Arakawa, A. (1976). Numerical methods used in atmospheric models. Garp Publication Series No. 17. WMO/ICSU Joint Organizing Committee, Geneva.

Muscolino, G., Ricciardi, G., and Vasta, M. (1997). Stationary and non-stationary probability density function for non linear oscillators. Int. J. Non-Linear Mech., 32(6), pp. 1051-1064.

Discretize-Then-Relax Approach For State Relaxactions In Global Dynamic Optimization

Ali M. Sahlodin and Benoît Chachuat

Department of Chemical Engineering, McMaster University, 1280 Main Street West, Hamilton, ON L8S 4L7, Canada. E-mail: {sahlodam,benoit}@mcmaster.ca

Abstract

This paper presents a discretize-then-relax approach to construct convex/concave bounds for the solutions of a wide class of parametric nonlinear ODEs. The procedure builds upon interval-based techniques implemented in state-of-the-art validated ODE solvers and uses McCormick's relaxation technique to propagate the convex/concave bounds. At each time step, a two-phase procedure is applied: a priori convex/concave bounds that are valid over the entire time step are calculated in the first phase; then, pointwise-in-time convex/concave bounds at the end of the time step are obtained in the second phase. This approach is implemented in an object-oriented manner using templates and operator overloading. It is demonstrated by a case study of a Lotka-Volterra system.

Keywords: Interval analysis, Convex relaxations, McCormick relaxations, Ordinary differential equations, Global dynamic optimization

1. Introduction

Pivotal to deterministic global optimization methods for nonconvex dynamic optimization is the ability to construct tight convex and concave relaxations for the solutions of parametric ordinary differential equations (ODEs),

$$\dot{\mathbf{x}}(t) = \mathbf{f}(\mathbf{x}(t), \mathbf{p}), \quad t \in (t_0, t_f]; \qquad \mathbf{x}(t_0) = \mathbf{h}(\mathbf{p}), \tag{1}$$

where $t \in [t_0, t_f]$ refers to the independent variable (hereafter *time*), and $\mathbf{p} \in \mathbf{P}$ are the continuous time-invariant parameters, with $\mathbf{P} \triangleq [\mathbf{p}^L, \mathbf{p}^U] \subset \mathbb{R}^{n_p}$ an interval vector. By convex (resp. concave) relaxations, we mean functions that underestimate (resp. overestimate) the actual ODE solutions and are convex (resp. concave) on $\mathbf{P}$, pointwise in time. Using McCormick composition's technique [1], such state relaxations allow construction of convex relaxations for dynamic optimization problems, the solutions of which can then be used as lower bounds in a branch-and-bound algorithm [2, 3].

Most of the emphasis so far has been on constructing an auxiliary system of parametric ODEs that describes convex/concave bounds of the parametric solutions pointwise in time [4]; that is, a relax-then-discretize approach. Concurrently, simple discretize-then-relax procedures have been proposed recently that employ McCormick-based relaxations of algorithms [5]. However, neither of these procedures are rigorous in that they discard truncation errors, thereby potentially leading to invalid bounds.

In this paper, a new discretize-then-relax approach is presented to construct tight convex/concave bounds for a wide class of parametric nonlinear ODEs. The procedure builds upon interval-based techniques implemented in state-of-the-art validated ODE solvers, such as VNODE [6], and uses (a generalization of) McCormick's relaxation technique [1, 7] to propagate the convex/concave bounds (§2). Specifically, a two-phase procedure

is applied at each time step (§3): in the first phase, a priori convex/concave bounds are calculated, which enclose the ODE solutions over the entire time step; in the second phase, pointwise-in-time convex/concave bounds are obtained that enclose the ODE solutions at the end of the time step. This approach is implemented in an object-oriented manner using templates and operator overloading and is demonstrated by a case study of a Lotka-Volterra system (§4). Finally, a number of concluding remarks close the paper (§5).

2. Preliminaries

2.1. Validated Solution of Parametric ODEs

A number of methods have been developed over the years to construct interval bounds that enclose the solutions of parametric ODEs, whereby both round-off and truncation errors are accounted for. Of particular interest in this work is the high-order enclosure (HOE) method developed by [6]. This method assumes that the function $\mathbf{f}$ in (1) is $(k-1)$-times continuously differentiable in $\mathbf{x}$ and $\mathbf{p}$, with $k \geq 2$, and that its representation contains only a finite number of unary and binary operations and univariate intrinsic functions. Consider a grid $t_0 < t_1 < \cdots < t_m$, not necessarily evenly spaced, and denote the step-size from t_j to t_{j+1} by h_j. In order to compute interval bounds $\mathbf{X}_j \overset{\Delta}{=} [\mathbf{x}_j^L, \mathbf{x}_j^U] \supseteq \mathbf{x}(t_j; \mathbf{P})$ at each time step, the HOE method proceeds in two phases:

Phase I. Given interval bounds $\mathbf{X}_j$ at t_j, existence and uniqueness of the solutions on $[t_j, t_{j+1}]$ are established and an a priori enclosure $\widetilde{\mathbf{X}}_j \supseteq \mathbf{x}(t; \mathbf{P}, t_j, \mathbf{X}_j)$ is computed for all $t \in [t_j, t_{j+1}]$. Specifically, $\widetilde{\mathbf{X}}_j$ can be obtained as [8]:

$$\widetilde{\mathbf{X}}_j = \sum_{i=0}^{k-1} [0, h_j]^i \mathbf{f}^{[i]}(\mathbf{X}_j, \mathbf{P}) + [0, h_j]^k \mathbf{f}^{[k]}(\widetilde{\mathbf{X}}_j^0, \mathbf{P}) \subseteq \widetilde{\mathbf{X}}_j^0, \quad \text{for some } \widetilde{\mathbf{X}}_j^0 \supsetneq \mathbf{X}_j, \qquad (2)$$

with $\mathbf{f}^{[i]}$ denoting the ith Taylor coefficient of $\mathbf{x}$ expanded with respect to t,

$$\mathbf{f}^{[0]}(\mathbf{x}, \mathbf{p}) = \mathbf{x}; \qquad \mathbf{f}^{[i]}(\mathbf{x}, \mathbf{p}) = \frac{1}{i} \left(\frac{\partial \mathbf{f}^{[i-1]}}{\partial \mathbf{x}} \mathbf{f} \right)(\mathbf{x}, \mathbf{p}), \quad \text{for } i \geq 1.$$

Phase II. Given interval bounds $\mathbf{X}_j$ at t_j and a priori interval bounds $\widetilde{\mathbf{X}}_j$ on $[t_j, t_{j+1}]$, an enclosure $\mathbf{X}_{j+1} \supseteq \mathbf{x}(t_{j+1}; \mathbf{P})$ is computed at t_{j+1}. This is typically done by applying a high-order Taylor series expansion combined with the mean-value theorem,

$$\mathbf{X}_{j+1} = \sum_{i=0}^{k-1} h_j^i \mathbf{f}^{[i]}(\hat{\mathbf{x}}_j, \hat{\mathbf{p}}) + h_j^k \mathbf{f}^{[k]}(\widetilde{\mathbf{X}}_j, \mathbf{P}) + \sum_{i=0}^{k-1} h_j^i \frac{\partial \mathbf{f}^{[i]}}{\partial \mathbf{x}}(\mathbf{X}_j, \mathbf{P}) \ [\mathbf{X}_j - \hat{\mathbf{x}}_j]$$

$$+ \sum_{i=0}^{k-1} h_j^i \frac{\partial \mathbf{f}^{[i]}}{\partial \mathbf{p}}(\mathbf{X}_j, \mathbf{P}) \ [\mathbf{P} - \hat{\mathbf{p}}], \quad \text{for some } (\hat{\mathbf{x}}_j, \hat{\mathbf{p}}) \in \mathbf{X}_j \times \mathbf{P}. \qquad (3)$$

The QR-factorization technique [6] can be applied to compute the interval-matrix-vector product terms in (3), which are the critical terms contributing to the wrapping effect.

2.2. Convex and Concave Relaxations

In addition to computing interval bounds for the state variables, the proposed algorithm also propagates convex and concave relaxations at each time step. As such, it relies heavily on the ability to construct convex and concave relaxations of functions that are known in closed form. The emphasis in this paper is on McCormick's relaxation technique [1], which can be used recursively to construct the desired convex and concave relaxations for so-called *factorable* functions. The factorable functions are those defined as a finite recursive composition of binary sums, binary products and univariate functions, an extremely inclusive class of functions.

The algorithm described subsequently also requires convex and concave relaxations of composite functions of the form $\phi(\boldsymbol{\xi}(\mathbf{p}))$, where ϕ is factorable, but $\boldsymbol{\xi}$ is not. Such relaxations cannot be obtained by applying the standard McCormick technique. Provided that convex and concave relaxations (as well as interval bounds) are known for $\boldsymbol{\xi}$ on $\mathbf{P}$,

$$\boldsymbol{\xi}^{cv}(\mathbf{p}) \leq \boldsymbol{\xi}(\mathbf{p}) \leq \boldsymbol{\xi}^{cc}(\mathbf{p}), \text{ for each } \mathbf{p} \in \mathbf{P}; \quad \boldsymbol{\xi}^{cv} \text{ convex on } \mathbf{P}; \quad \boldsymbol{\xi}^{cc} \text{ concave on } \mathbf{P},$$

the *generalized McCormick relaxations* introduced recently by [7] provide a framework for the recursive computation of convex and concave relaxations of $\phi(\boldsymbol{\xi}(\cdot))$ on $\mathbf{P}$, in the following form (the compact notation $[\alpha, \beta](\cdot) \triangleq [\alpha(\cdot), \beta(\cdot)]$ is used subsequently):

$$\phi^{cv}\left([\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}](\mathbf{p})\right) \leq \phi(\boldsymbol{\xi}(\mathbf{p})) \leq \phi^{cc}\left([\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}](\mathbf{p})\right).$$

Note that the convex/concave bounds $[\boldsymbol{\xi}^{cv}, \boldsymbol{\xi}^{cc}](\mathbf{p})$ can be obtained with any method [3]. Note also that subgradients for the standard and generalized McCormick relaxations can be propagated along with the relaxations [5].

3. State Relaxation Algorithm

The proposed discretize-then-relax algorithm computes convex and concave bounds for the state variables at given time instants $t_0 < t_1 < \cdots < t_n$, in addition to (validated) interval bounds. More precisely, for a specified parameter value $\mathbf{p} \in \mathbf{P}$, the algorithm calculates relaxed state values $\mathbf{x}_j^{cv}(\mathbf{p})$ and $\mathbf{x}_j^{cc}(\mathbf{p})$ such that:

  i. $\mathbf{x}_j^{cv}(\mathbf{p}) \leq \mathbf{x}(t_j; \mathbf{p}) \leq \mathbf{x}_j^{cc}(\mathbf{p})$;　and

  ii. $\mathbf{x}_j^{cv}$ is convex on $\mathbf{P}$, $\mathbf{x}_j^{cc}$ is concave on $\mathbf{P}$.

The procedure starts by constructing the state relaxations $\mathbf{x}_0^{cv}(\mathbf{p}), \mathbf{x}_0^{cc}(\mathbf{p})$ at t_0. Provided that the function $\mathbf{h}$ in (1) is factorable, $\mathbf{x}_0^{cv}(\mathbf{p})$ and $\mathbf{x}_0^{cc}(\mathbf{p})$ can be computed easily, e.g. by applying the recursive (standard) McCormick relaxation technique. Next, the state relaxations $\mathbf{x}_j^{cv}(\mathbf{p})$ and $\mathbf{x}_j^{cc}(\mathbf{p})$ are propagated through each time step t_j, $j = 1, \ldots, n$, based on a two-phase procedure similar to the one described earlier in §2.1:

Phase I. Given convex/concave bounds $\mathbf{x}_j^{cv}(\mathbf{p}) \leq \mathbf{x}(t_j; \mathbf{p}) \leq \mathbf{x}_j^{cc}(\mathbf{p})$ at t_j, functions $\widetilde{\mathbf{x}}_j^{cv}, \widetilde{\mathbf{x}}_j^{cc} : \mathbf{P} \to \widetilde{\mathbf{X}}_j$ are constructed such that:

  i. $[\widetilde{\mathbf{x}}_j^{cv}, \widetilde{\mathbf{x}}_j^{cc}](\mathbf{p}) \supseteq \{\mathbf{x}(t; \mathbf{p}, \tau, \boldsymbol{\xi}) : t_j \leq \tau \leq t_{j+1}, \mathbf{x}_j^{cv}(\mathbf{p}) \leq \boldsymbol{\xi} \leq \mathbf{x}_j^{cc}(\mathbf{p})\}$;　and

  ii. $\widetilde{\mathbf{x}}_j^{cv}$ is convex on $\mathbf{P}$, $\widetilde{\mathbf{x}}_j^{cc}$ is concave on $\mathbf{P}$.

This is most easily achieved as follows:

$$[\widetilde{\mathbf{x}}_j^{cv}, \widetilde{\mathbf{x}}_j^{cc}](\mathbf{p}) = \sum_{i=0}^{k-1} [0, h_j]^i \otimes \left[\mathbf{f}^{[i],cv}, \mathbf{f}^{[i],cc}\right] \left([\mathbf{x}_j^{cv}, \mathbf{x}_j^{cc}](\mathbf{p}), \mathbf{p}\right) + [0, h_j]^k \mathbf{f}^{[k]}(\widetilde{\mathbf{X}}_j^0, \mathbf{P}),$$

where the interval bounds $\widetilde{\mathbf{X}}_j^0$ are obtained in the same way as in (2); the result of the binary operation $\otimes$ is meant to be convex/concave bounds for the product between two pairs of convex/concave bounds for the operands (e.g., McCormick product rule); and $\mathbf{f}^{[i],cv}, \mathbf{f}^{[i],cc}$ denote convex and concave relaxations on $\mathbf{P}$ of the Taylor coefficients $\mathbf{f}^{[i]}(\mathbf{x}(\cdot), \cdot)$, for each $i = 0, \ldots, k-1$. In particular, $\mathbf{f}^{[i],cv}$ and $\mathbf{f}^{[i],cc}$ can be obtained from the application of the generalized McCormick relaxation technique (see §2.2).

Phase II. Given convex/concave bounds $\mathbf{x}_j^{cv}(\mathbf{p}) \leq \mathbf{x}(t_j; \mathbf{p}) \leq \mathbf{x}_j^{cc}(\mathbf{p})$ at t_j and a priori convex/concave bounds $[\widetilde{\mathbf{x}}_j^{cv}, \widetilde{\mathbf{x}}_j^{cc}](\mathbf{p})$ on $[t_j, t_{j+1}]$, functions $\mathbf{x}_{j+1}^{cv}, \mathbf{x}_{j+1}^{cc} : \mathbf{P} \to \mathbf{X}_{j+1}$ are constructed such that:

   i. $\mathbf{x}_{j+1}^{cv}(\mathbf{p}) \leq \mathbf{x}(t_{j+1}; \mathbf{p}) \leq \mathbf{x}_{j+1}^{cc}(\mathbf{p});$ and

   ii. $\mathbf{x}_{j+1}^{cv}$ is convex on $\mathbf{P}$, $\mathbf{x}_{j+1}^{cc}$ is concave on $\mathbf{P}$.

Applying a high-order Taylor series expansion along with the mean-value theorem gives:

$$[\mathbf{x}_{j+1}^{cv}, \mathbf{x}_{j+1}^{cc}](\mathbf{p}) = \sum_{i=0}^{k-1} h_j^i \mathbf{f}^{[i]}(\hat{\mathbf{x}}_j, \hat{\mathbf{p}}) + h_j^k \left[\mathbf{f}^{[k],cv}, \mathbf{f}^{[k],cc}\right]\left([\widetilde{\mathbf{x}}_j^{cv}, \widetilde{\mathbf{x}}_j^{cc}](\mathbf{p}), \mathbf{p}\right)$$

$$+ \sum_{i=0}^{k-1} h_j^i \left[\frac{\partial \mathbf{f}^{[i],cv}}{\partial \mathbf{x}}, \frac{\partial \mathbf{f}^{[i],cc}}{\partial \mathbf{x}}\right]\left([\boldsymbol{\xi}_j^{cv}, \boldsymbol{\xi}_j^{cc}](\mathbf{p}), [\boldsymbol{\rho}^{cv}, \boldsymbol{\rho}^{cc}](\mathbf{p})\right) \otimes \left([\mathbf{x}_j^{cv}, \mathbf{x}_j^{cc}](\mathbf{p}) - \hat{\mathbf{x}}_j\right)$$

$$+ \sum_{i=0}^{k-1} h_j^i \left[\frac{\partial \mathbf{f}^{[i],cv}}{\partial \mathbf{p}}, \frac{\partial \mathbf{f}^{[i],cc}}{\partial \mathbf{p}}\right]\left([\boldsymbol{\xi}_j^{cv}, \boldsymbol{\xi}_j^{cc}](\mathbf{p}), [\boldsymbol{\rho}^{cv}, \boldsymbol{\rho}^{cc}](\mathbf{p})\right) \otimes (\mathbf{p} - \hat{\mathbf{p}}),$$

for some $(\hat{\mathbf{x}}_j, \hat{\mathbf{p}}) \in \mathbf{X}_j \times \mathbf{P}$. Here, $\frac{\partial \mathbf{f}^{[i],cv}}{\partial \mathbf{x}}, \frac{\partial \mathbf{f}^{[i],cc}}{\partial \mathbf{x}}$ and $\frac{\partial \mathbf{f}^{[i],cv}}{\partial \mathbf{p}}, \frac{\partial \mathbf{f}^{[i],cc}}{\partial \mathbf{p}}$ denote convex and concave relaxations on $\mathbf{P}$ of the derivatives $\frac{\partial \mathbf{f}^{[i]}}{\partial \mathbf{x}}$ and $\frac{\partial \mathbf{f}^{[i]}}{\partial \mathbf{p}}$ of the Taylor coefficients, respectively, for each $i = 0, \ldots, k-1$; and the functions $\boldsymbol{\xi}_j^{cv}, \boldsymbol{\xi}_j^{cc} : \mathbf{P} \to \mathbf{X}_j$ and $\boldsymbol{\rho}^{cv}, \boldsymbol{\rho}^{cc} : \mathbf{P} \to \mathbf{P}$ are defined as:

$$[\boldsymbol{\xi}_j^{cv}, \boldsymbol{\xi}_j^{cc}](\mathbf{p}) = \hat{\mathbf{x}}_j + [0, 1] \otimes \left([\mathbf{x}_j^{cv}, \mathbf{x}_j^{cc}](\mathbf{p}) - \hat{\mathbf{x}}_j\right), [\boldsymbol{\rho}^{cv}, \boldsymbol{\rho}^{cc}](\mathbf{p}) = \hat{\mathbf{p}} + [0, 1] \otimes (\mathbf{p} - \hat{\mathbf{p}}).$$

It should be stressed that these latter functions follow directly from the application of the mean-value theorem. Interestingly, their interval counter-parts can be shown to be $\mathbf{X}_j$ and $\mathbf{P}$, respectively, which explains that they do not appear in (3). As with validated ODE methods, the QR-factorization technique is used in computing relaxations of the matrix-vector product terms to mitigate the wrapping effect.

Numerical Implementation. The proposed discretize-then-relax algorithm is implemented in a C++ program that makes extensive use of templates and operator overloading. Several publicly available third-party programs are used in the current implementation:

- The successive Taylor coefficients $\mathbf{f}^{[i]}$, $i = 0, \ldots, k$, as well as their derivatives $\frac{\partial \mathbf{f}^{[i]}}{\partial \mathbf{x}}$ and $\frac{\partial \mathbf{f}^{[i]}}{\partial \mathbf{p}}$, are computed by using the automatic differentiation (AD) package FADBAD++ (http://www.fadbad.com/)

- All the computations involving the interval data type are performed with the C++ library PROFIL (http://www.ti3.tu-harburg.de/).

- The computation of convex and concave relaxations is automated in our in-house package MC++ (http://mcplusplus.mcmaster.ca/), which implements the generalized McCormick relaxation technique for factorable functions. It can be used in combination with FADBAD++ for AD and supports various interval arithmetic libraries (including the library PROFIL). MC++ can also propagate sub-gradients of the computed relaxations.

4. Case Study

The discretize-then-relax approach is illustrated for the Lotka-Volterra system

$$\dot{x}_1(t) = px_1(t)\left[1 - x_2(t)\right]; \quad x_1(t_0) = 1.2$$
$$\dot{x}_2(t) = px_2(t)\left[x_1(t) - 1\right]; \quad x_2(t_0) = 1.1$$

with $t \in [0, 2]$, and the unique parameter $p \in [2.95, 3.05]$. The Taylor expansion order is set to $k = 10$, and QR factorization is used to fight the wrapping effect.

The bound and relaxation trajectories corresponding to the variable x_1 are displayed in Fig. 1. On the left plot, the comparison between interval bounds obtained from differential inequalities [4] and bounds calculated by the validated two-phase procedure (§2.1) shows a better performance of the latter on this problem, which can be attributed to its ability to mitigate the wrapping effect. Note in particular that the bounds obtained with differential inequalities blow up before $t = 2$. The convex and concave bounds calculated by the proposed discretize-then-relax approach for the parameter value $p = 3$ are shown on the right plot. They are guaranteed to be at least as good as interval bounds, by construction.
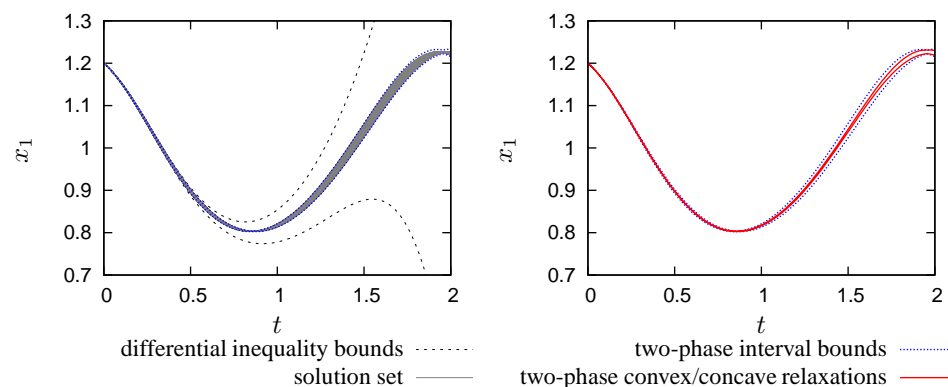

Figure 1. Interval bounds (left plot) and convex/concave bounds for $p = 3$ (right plot).

Pointwise-in-time convex and concave relaxations of x_1 and x_2 at final time $t = 2$ are displayed in Fig. 2. These plots are generated by applying the discretize-then-relax approach at a number of parameter values $p \in [2.95, 3.05]$. It is seen that the relaxations provide much tighter bounds than simple interval bounds, thereby making their utilization in a branch-and-bound procedure for global dynamic optimization very appealing.
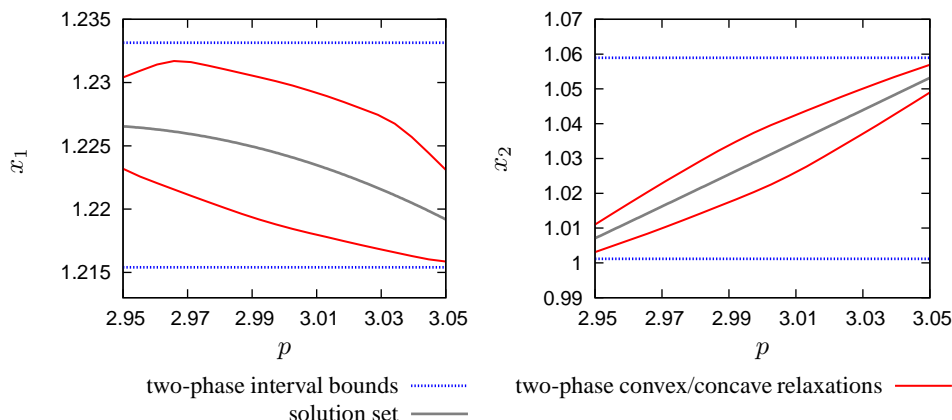
Figure 2. Interval and convex/concave bounds at $t = 2$ for x_1 (left plot) and x_2 (right plot).

5. Conclusions

An algorithm that constructs convex and concave relaxations for the solutions of nonlinear parametric ODEs has been presented in this paper. It builds upon validated ODE methods and McCormick's relaxation technique and, unlike other existing ODE relaxation methods, is rigorous in its accounting of truncation errors. The proposed algorithm also has built-in capabilities to efficiently mitigate the wrapping effect. In terms of computational effort, the time needed to compute state relaxations is a fixed multiple factor of that of interval bounds, typically of the order of 2-3. All these features make the discretize-then-relax algorithm a viable and promising technique for use in a branch-and-bound algorithm for global dynamic optimization.

References

[1] G. P. McCormick. Computability of global solutions to factorable nonconvex programs: Part I – Convex underestimating problems. *Math Program*, 10:147–175, 1976.

[2] B. Chachuat, A.B. Singer, and P.I. Barton. Global methods for dynamic optimization and mixed-integer dynamic optimization. *Ind Eng Chem Res*, 45(25):8373–8392, 2006.

[3] C. A. Floudas and C. E. Gounaris. A review of recent advances in global optimization. *J Global Optim*, 45(1):3–38, 2009.

[4] A. B. Singer and P. I. Barton. Bounding the solutions of parameter dependent nonlinear ordinary differential equations. *SIAM J Sci Comput*, 27(6):2167–2182, 2006.

[5] A. Mitsos, B. Chachuat, and P. L. Barton. McCormick-based relaxations of algorithms. *SIAM J Optim*, 20(2):573–601, 2009.

[6] N. S. Nedialkov. *Computing Rigorous Bounds on the Solution of an Initial Value Problem for an Ordinary Differential Equation*. PhD thesis, University of Toronto, Toronto, ON, 1999.

[7] J. K. Scott, M. D. Stuber, and P. I. Barton. Generalized McCormick relaxations for ODEs. Technical report, Department of Chemical Engineering, MIT, July 2009. (available at http://web.mit.edu/jkscott/www/Supplements/).

[8] Y. Lin and M. A. Stadtherr. Validated solutions of initial value problems for parametric ODEs. *Appl Numer Math*, 57(10):1145–1162, 2007.

Development and Validation of a Dynamic Model for Regeneration of Passivating Baths using Membrane Contactors

Eugenio Bringas, Rosa Mediavilla, Ana María Urtiaga, Inmaculada Ortiz

Department of Chemical Engineering and Inorganic Chemistry, University of Cantabria, Avda. Los Castros, 39005 Santander, Spain, urtiaga@unican.es

Abstract

Selective liquid membranes have been traditionally employed for liquid/liquid and gas/liquid mass transfer in a wide range of applications. In particular, the Emulsion Pertraction Technology (EPT) using hollow fiber membrane contactors is a promising alternative to carry out the selective separation of metals from complex mixtures. However, the application of a new technology requires of reliable mathematical models and parameters that serve for design and optimization purposes allowing to accurate scale-up processes. This work reports the methodology for the development of a dynamic model to describe the kinetics of the EPT separation-concentration process applied to the regeneration of spent trivalent chromium-based passivating baths. The regeneration stage aims at the selective removal of Zn^{2+} dragged from previous steps in the plating line, not affecting the level of Cr^{3+} concentration in the passivating bath. In the case study the mathematical model was initially developed in a rigorous way and in a further analysis, a systematic method for its simplification was established. Then, the system of partial differential and algebraic equations obtained was integrated using the commercial software package ASPEN CUSTOM MODELER (from ASPENTECH) making possible the analysis of the model sensibility under different values of the operation variables. Finally, the model was validated with kinetic data obtained at laboratory scale.

Keywords: Passivating Baths, Membrane Contactors, Zinc Separation, Emulsion Pertraction Technology, Facilitated Transport

2. Introduction

The use of selective liquid membrane systems represents a promising alternative which excels the limitations of most conventional membrane-based separation processes [1]. However, in spite of the known advantages and applications of liquid membrane separation processes in hollow fiber contactors, there are scarce examples of industrial application. The industrial application of a new technology requires a reliable mathematical model and parameters that serve for design, cost estimation and optimisation purposes allowing to accurate scale-up processes [2].

The rigorous description of a membrane contactor comprises the characterization of diffusive mass transfer phenomena and interfacial chemical reactions, but also the mathematical modeling of the fluid flow on either side of the membrane (tube and shell sides) by the development of the appropriate conservation and continuity equations [3,4].

This work reports the methodology for the development of a dynamic model to describe the kinetics of the Emulsion Pertraction Technology (EPT) that is a novel configuration of the selective membrane technology that allows the simultaneous selective removal and recovery of the target species using one hollow fiber membrane contactor [5]. Finally, the methodology is illustrated by its application to a selected case study that deals with the regeneration of spent trivalent chromium-based passivating baths which are complex solutions having as major components Cr^{3+} as the passivation agent, and Zn^{2+} and Fe^{3+} as the tramp ions dragged from previous steps in the plating line; iron is removed by the use of ion exchange resins [6], so that the target compound in this work is zinc that should be removed from the spent bath and concentrated for possible recovery.

3. Mathematical modeling of the EPT process

Figure 1 presents a flow diagram describing the EPT experimental set-up. Two tanks were used: one stirred tank for homogenization of the spent passivation bath, and one vessel for the emulsion phase prepared by dispersion of the stripping solution into the organic phase that contains the selective carrier for zinc extraction. Both fluid streams are contacted in a microporous hollow fiber membrane contactor (Liqui-Cel® Extra-Flow 2.5 x 8, Hoechst Celanese), with an effective area of 1.4 m^2 [7].
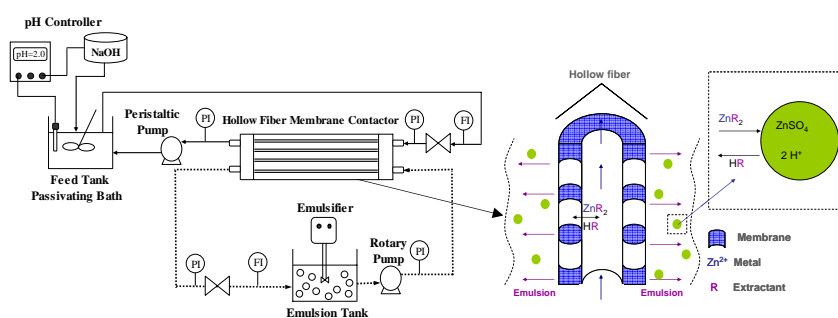

Figure 1. Experimental set-up of EPT system.

The mathematical modeling of the EPT process is divided into two steps: i) the mathematical modeling of the diffusive mass transport flux through the liquid membrane at steady state and, ii) the solute mass balances to the flowing phases in both the tanks and the membrane contactor (tube side and shell side) assuming suitable flow models.

3.1. Modeling of the diffusive mass transport flux

After a literature survey, Cyanex®272 and sulphuric acid were selected respectively as selective extractant and stripping agent to carry out the selective removal and recovery of zinc. As it is shown in Figure 2, several in-series steps are considered to describe the mass transfer of zinc ions from the spent bath to the stripping solution: i) the solute mass transport through both the stagnant layers and the supported liquid membrane, and ii) the interfacial extraction and back-extraction reactions [8].

Development and Validation of a Dynamic Model for Regeneration of Passivating Baths using Membrane Contactors
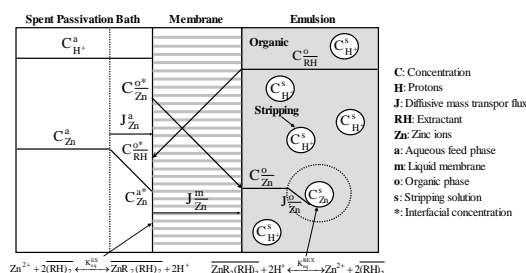

Figure 2. Enlarged view of the hollow fiber membrane.

From the analysis of the particular characteristics of the system under study, several simplifications are taken into account: i) due to the high concentration of zinc (2-7 g L^{-1}), the concentration polarization phenomena is considered negligible and thus, the mass transport resistances in both the aqueous and the organic phase stagnant layers are neglected and, ii) the interfacial reactions are considered fast enough to reach equilibrium instantaneously. Therefore, the global rate of the process depends on the membrane mass transport resistance and the overall mass transport flux (J) is given by:

$$J = k_m \cdot (C_{Zn}^{o*} - C_{Zn}^{o}) = k_m \cdot (C_{RH}^{o} - C_{RH}^{o*}) \tag{1}$$

where k_m is the membrane mass transport coefficient that can be calculated with correlations that depend on the membrane characteristics as well as on the diffusivity of the organic complexes through the liquid membrane.

The interfacial equilibrium reactions are described by a simple mass action law assuming ideal behavior of both phases (aqueous and organic). Therefore the interfacial condition can be written as follows:

$$K_{eq}^{EX} = \frac{C_{Zn}^{o*} \cdot (C_{H^+}^{a})^2}{C_{Zn}^{a} \cdot (C_{RH}^{o*})^2} \tag{2}$$

In spite of the different physicochemical characteristics of the spent bath and the stripping solution, ideal conditions are assumed for the sake of simplicity and thus:

$$K_{eq}^{BEX} = \frac{1}{K_{eq}^{EX}} \tag{3}$$

In conclusion, the solution of the proposed mass transport model requires the knowledge of the design parameters which are the membrane mass transport coefficient (k_m) and the equilibrium parameter of the extraction reaction (K_{eq}^{EX}).

3.2. Development of the solute mass balances

After describing the diffusive mass transport flux, the solute mass balances in the different fluid phases (feed, organic and stripping) are developed in both the membrane contactor and the stirred tanks. For this purpose, two main assumptions are undertaken:

- The residence time of the stripping solution in the emulsion tank is higher (up to five times) than the corresponding value in the membrane contactor. Therefore, the stripping step is assumed to take place in the emulsion tank and thus, the solute mass balances to the stripping solution in the membrane contactor are omitted.
- The mathematical description of the stripping solution in the emulsion tank is carried out by the replacement of the physical emulsion tank by three different

artificial units: two stirred tanks for the organic and stripping phases and an equilibration unit where the back-extraction reaction takes place. Figure 3 shows a diagram with the units to be modeled as well as, a qualitative description of the different model equations.
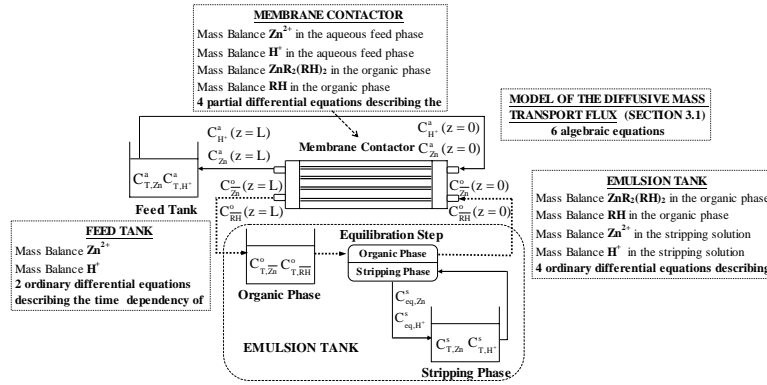

Figure 3. Diagram of the modeling approach.

The solute mass balances in the membrane contactor are developed under the following hypothesis: i) ideal axial plug flow and negligible axial diffusion, ii) each fiber is of identical specifications, iii) hollow fibers are hydrophobic thus obtaining:

$$\frac{V^j}{F^j \cdot L} \cdot \frac{\partial C_i^j}{\partial t} + \frac{\partial C_i^j}{\partial z} = \pm \frac{A}{F^j \cdot L} \cdot J \quad z \in [0,L]; \forall t; \forall i = Zn, H^+, \overline{ZnR_2(RH)_2}, \overline{RH}; \forall j = a,o$$

$$t = 0 \Rightarrow C_i^j = C_i^j(t=0); \tag{4}$$

$$z = 0 \Rightarrow C_i^j = C_i^{j,in}$$

The mass balances in the stirred tanks for both the feed and the organic solutions are developed under the assumption of ideal mixed flow as follows:

$$V_T^j \cdot \frac{dC_{T,i}^j}{dt} = \pm F^j \cdot [C_i^j(z=L) - C_{T,i}^j] \qquad \forall t; \forall i = Zn, H^+, \overline{ZnR_2(RH)_2}, \overline{RH}; \forall j = a,o \tag{5}$$

$$t = 0 \Rightarrow C_i^j = C_i^j(t=0)$$

The mathematical description of the back-extraction process is given by the following equations:

$$V_T^s \cdot \frac{dC_{T,i}^s}{dt} = F^s \cdot [C_{eq,i}^s - C_{T,i}^s] \qquad \forall t; \forall i = Zn, H^+ \tag{6}$$

$$t = 0 \Rightarrow C_i^s = C_i^s(t=0)$$

where the equilibrium concentrations are obtained by the combination of Eq.(3) and the following expression

$$K_{eq}^{BEX} = \frac{C_{eq,Zn}^s \cdot (C_{T,\overline{RH}}^o)^2}{C_{T,\overline{Zn}}^o \cdot (C_{eq,H^+}^s)^2} \tag{7}$$

4. Simulation and model validation

Once the model was developed, a simulation analysis was performed in order to check the model sensibility under different values of the main operation variables: (i) pH of

Development and Validation of a Dynamic Model for Regeneration of Passivating Baths using Membrane Contactors

the feed solution (pH), (ii) initial concentration of zinc in the aqueous feed solution ($C_{T,Zn}^a[t=0]$), (iii) feed flowrate (F^a) and (ii) concentration of sulphuric acid ($C_{T,H^+}^s[t=0]$). Table 1 shows the values of the operation variables used in the simulation analysis.

Table 1. Values of the main operation variables used in the simulation analysis.

Variable	Value	Variable	Value
pH	2-3	$F^{emulsion}$	40 L h^{-1}
$C_{T,H^+}^a[t=0]$	10^{-3}-10^{-2} mol L^{-1}	F^s/F^o	1/4
$C_{T,Zn}^a[t=0]$	0.05-0.15 mol L^{-1}	V_T^s / V_T^o	1/4
F^a	9-26 L h^{-1}	V_T^s / V_T^a	1/5
$C_{T,H^+}^s[t=0]$	1-2 mol L^{-1}	A	1.4 m^2
$C_{T,\overline{Zn}}^o[t=0]$	0	Time	3 h

The model was solve using ASPEN CUSTOM MODELER 2004.1 (ASPENTECH) being the values of the design parameters $K_{eq}^{EX} = 2.8\,10^{-4}$ (experimentally obtained) and k_m=2.9 10^{-4} m h^{-1} (calculated with correlations). Figure 4 shows the simulated evolution with time of the zinc concentrations in the feed tank and in the stripping solution at different values of the initial pH of the feed solution (Figure 4a) and different sulfuric acid concentrations (Figure 4b).
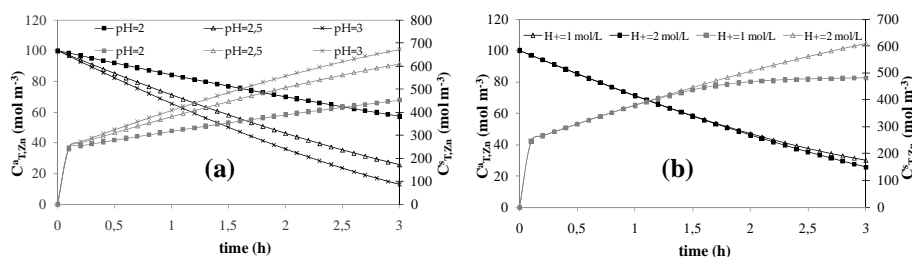

Figure 4. Evolution with time of the simulatd concentration of zinc in the feed and stripping phases. (a) Influence of the initial pH of the feed solution; (b) Influence of the sulphuric acid concentration.

From Figure 4 it can be concluded that the higher the initial pH is, the higher zinc extraction percentages are. These results are in agreement with the equilibrium data reported in the technical data sheets of the selective extractant Cyanex®272. Regarding the influence of the stripping acid concentration (proton concentration), it can be concluded that this operation variable does not exert a significant influence on the extraction and back-extraction kinetics. However, it can be observed that the kinetic curve corresponding to the simulation performed with a lower acid concentration slows down after 2 hours of experimental running due to the depletion of the stripping agent. Therefore, it can be concluded that the proposed mathematical model is able to predict the kinetic behavior of the EPT process under different process operation conditions.

Finally, some kinetic experiments were performed at laboratory scale in order to obtain data that allow the model validation. These experiments were performed under the same

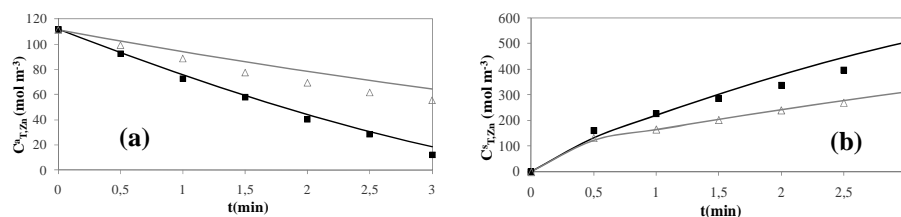operation conditions employed in the simulation analysis (Table 1).

Figure 5. Comparison between experimental (dots) and simulated (solid lines) data. (a) Extraction results; (b) back-extraction results.

From the results shown in Figure 5 it is thought that the model permits satisfactory description of the separation and concentration of zinc from spent passivation baths by the EPT process using Cyanex®272 and sulfuric acid as reagents.

5. Conclusions

This study addressed the methodology for the development of a dynamic model to describe the kinetics of the EPT process applied to the removal and concentration of zinc from spent trivalent chromium-based passivating baths. A rigorous mathematical model was developed consisting of two different sub-models: the description of the diffusive mass transfer flux and the solute mass balances in the different process units. Then, a simulation analysis was carried out and finally the model validation was performed using experimental data. In a further step the model will be adapted to make possible the integration of the separation step into the global surface treatment process.

6. Acknowledgements

7. Nomenclature

C: Concentration (mol m^{-3}); V: Volume (m^3); F: Flowrate (m^3 h^{-1}); t: time (h); A: Membrane area (m^{-2}); L: Fiber length (m); z: Axial distance (m); k_m: Membrane mass transport coefficient (m h^{-1}); K_{eq}: equilibrium parameter; J: mass transport flux (mol m^{-2} h^{-1})-*Subscripts*: i: solute; T: Tank-*Superscripts*: a: aqueous phase; o:organic phase; s:stripping solution; *: equilibrium; *in*: initial.

References

[1] X.J. Yang, A.G. Fane and K. Soldenhoff, Ind. Eng. Chem. Res., 42 (2003) 392.
[2] J. de Gyves and E. Rodríguez de San Miguel, Ind. Eng. Chem. Res., 38 (1999) 2182.
[3] J. Marriot and E. Sørensen, Chem. Eng. Sci., 58 (2003) 4975.
[4] A.M. Urtiaga and J.A. Irabien, AIChE J., 39 (1993) 521.
[5] S.B. Hu and J.M. Wiencek, AIChE J., 44 (1998) 570.
[6] I. Fernández-Olmo, A. Ortiz, A. Urtiaga and I.Ortiz, J. Chem. Technol. Biotechnol., 83 (2008) 1616.
[7] J.A. Carrera, E.Bringas, M.F. San Román and I. Ortiz, J. Membr. Sci., 326 (2009) 672.
[8] E. Bringas, M.F.San Román and I. Ortiz, Ind. Eng. Chem. Res., 45 (2006) 4295.

Cell-based dynamic heat exchanger models – direct determination of the cell number and size

Petar Varbanov, Jiří Klemeš, Ferenc Friedler

EC Marie Curie Chair (EXC), Faculty of Information Technology, University of Pannonia, Egyetem utca 10, Veszprém H-8200, Hungary, varbanov@cpi.uni-pannon.hu

Abstract

Large amounts of thermal energy are transferred to for heating or cooling in the industry as well as in the other sectors. Typical examples are crude oil preheating, ethylene plants, exothermic and endothermic reactions and many others. Heat exchangers frequently operate under varying conditions. Their appropriate use in flexible heat exchanger networks as well as maintenance/reliability related calculations require adequate models for estimating their dynamic behaviour. Cell-based dynamic models are very often used to represent heat exchangers with varying arrangements. The paper describes a direct method and a visualisation technique for determining the number of the modelling cells and their size.

Keywords: dynamic heat exchanger modelling, cell models

1. Introduction

Heat exchangers are usually parts of larger heat recovery networks and frequently operate under varying conditions. The variability as well as the uncertainty of operating conditions have been generally in the framework of the concepts of flexibility, controllability and operability (Oliveira *et al.* 2001; Skogestad and Postletwaite, 1996).

The appropriate use of heat exchangers under varying conditions requires adequate dynamic models. There are two general approaches to modelling the dynamics of a heat exchanger – distributed and lumped. The lumped cell-based models are more popular (Roetzel and Xuan, 1998; Mathisen *et al.*, 1994; Varga *et al.*, 1995). There have been noticeable advances in the field of dynamic simulation of heat exchangers. Recent examples include: Luo *et al.* (2003) models the dynamic behaviour of multi-stream heat exchangers; Konukman and Akman (2005) heat integrated plant; Ansari and Mortazavi (2006) presents a distributed heat exchanger model; and Díaz *et al.* (2001); Varshney and Panigrahi (2005), Peng and Ling (2009) featuring a neural network based model. All these models are quite complex and difficult to understand by process design engineers. The current paper is a stem in direction of alleviating this problem.

Cell models can result in a potentially large number of equations, but the equations are very simple and the approach offers a uniform framework and modelling flexibility to accommodate any type of surface heat exchanger with any flow arrangement. The model complexity can be controlled by the number of cells, allowing a trade-off between the accuracy and the ability of the model to tackle large and complex process systems such as heat exchanger networks. Usually dynamic heat exchanger models (Roetzel and Xuan, 1998) are based on certain assumptions:

    (1)   The heat transfer area is uniformly distributed throughout the apparatus.
    (2)   All thermal properties (film heat transfer coefficients, specific heat capacities) of the fluids and the exchanger wall are constant.

   (3)   The heat conduction along the axial direction (i.e. direction of the fluid flow) is negligible both within the fluid and within the wall.

   (4)   The wall thermal resistance to heat transfer is negligible. The imprecision resulting from this assumption can be compensated by an equivalent increase in the values of the film transfer coefficients.

   (5)   No heat is lost to the ambient through the exchanger casing.

The cell-based models are based on the fact that combining a sufficient number of perfectly mixed model tanks, called cells. The final modelling result is equivalent to applying simulation with a distributed model. Two mass and three energy balances around the elements of a heat exchange cell are performed.

2. Heat exchange cell

2.1. Definition

A simple heat exchange cell is defined as two perfectly stirred tanks, exchanging heat only with each other through a dividing wall. This type of arrangement is illustrated in Fig 1.
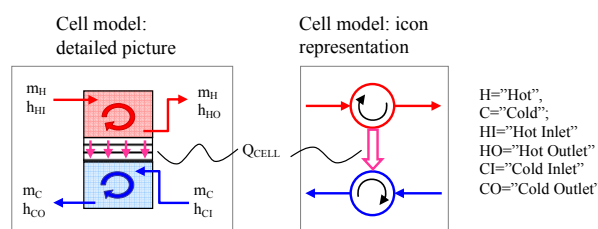

Fig 1. Representation of a modelling cell

2.2. Assumptions for the modelling cells

The following modelling assumptions are employed to derive the dynamic cell model:

   (i)   Both tanks cell feature perfect mixing.

   (ii)   The fluid densities are constant. This holds completely for liquids. For gases, this would be true as long as the pressures are kept approximately constant.

   (iii)   The tanks are completely full with the corresponding fluid.

   (iv)   As this model aims mainly at controlling the fluid temperatures, the streams are assumed to have finite constant specific heat capacities, effectively excluding process streams with pure vaporization or condensation.

   (v)   The wall resistance to heat transfer is neglected; its temperature is uniform within the cell volume. The main reason for adopting this assumption is the complexity of the dynamics of the heat transfer through the wall. The wall heat capacity (kJ/^{o}C), *i.e.* the metal heating and heat buffering, could cause significant delay in the heat flow and– influence the temperature distribution in time. For this reason the wall heat capacity is taken into account.

2.3. Equations

Regarding the material balances, assumptions (ii) and (iii) above make them trivial, eliminating any change in the amount of fluid holdup. The heat transfer rate for the hot and the cold tanks are:

$$Q_{CELL,H} = \alpha_{H,CELL} \cdot A_{CELL} \cdot (T_{HO} - T_W) \tag{1}$$

Cell-based dynamic heat exchanger models – direct determination of the cell number and size

$$Q_{CELL,C} = \alpha_{C,CELL} \cdot A_{CELL} \cdot (T_W - T_{CO}) \tag{2}$$

According to the adopted assumptions the overall heat transfer coefficient U_{CELL} would be calculated using only the film transfer coefficients of the fluids in the two tanks:

$$U_{CELL} = \left(\frac{1}{\alpha_{H,CELL}} + \frac{1}{\alpha_{C,CELL}} \right)^{-1} \tag{3}$$

As a result, the following related energy balances are obtained:

$$m_{H,CELL} \cdot C_{P,H} \cdot \frac{dT_{HO}}{dt} = m_H \cdot C_{P,H} \cdot T_{HI} - m_H \cdot C_{P,H} \cdot T_{HO} - \alpha_{H,CELL} \cdot A_{CELL} \cdot (T_{HO} - T_W) \tag{4}$$

$$m_{C,CELL} \cdot C_{P,C} \cdot \frac{dT_{CO}}{dt} = m_C \cdot C_{P,C} \cdot T_{CI} - m_C \cdot C_{P,C} \cdot T_{CO} + \alpha_{C,CELL} \cdot A_{CELL} \cdot (T_W - T_{CO}) \tag{5}$$

$$m_W \cdot C_{P,W} \cdot \frac{dT_W}{dt} = \alpha_{H,CELL} \cdot A_{CELL} \cdot (T_{HO} - T_W) - \alpha_{C,CELL} \cdot A_{CELL} \cdot (T_W - T_{CO}) \tag{6}$$

For modelling complete heat exchangers, a number of cells are combined following the flow arrangement. It is more convenient to refer to the volumes of the cell tanks rather than to the mass holdups. The following equations are given by Varga *et al.* (1995):

$$\frac{dT_{HO}}{dt} = \frac{v_H}{V_{H,CELL}} \cdot (T_{HI} - T_{HO}) - \frac{\alpha_{H,CELL} \cdot A_{CELL}}{V_{H,CELL} \cdot \rho_H \cdot C_{P,H}} \cdot (T_{HO} - T_W) \tag{7}$$

$$\frac{dT_{CO}}{dt} = \frac{v_C}{V_{C,CELL}} \cdot (T_{CI} - T_{CO}) + \frac{\alpha_{C,CELL} \cdot A_{CELL}}{V_{C,CELL} \cdot \rho_C \cdot C_{P,C}} \cdot (T_W - T_{CO}) \tag{8}$$

$$\frac{dT_W}{dt} = \frac{\alpha_{H,CELL} \cdot A_{CELL}}{V_W \cdot \rho_W \cdot C_{P,W}} \cdot (T_{HO} - T_W) - \frac{\alpha_{C,CELL} \cdot A_{CELL}}{V_W \cdot \rho_W \cdot C_{P,W}} \cdot (T_W - T_{CO}) \tag{9}$$

They assume insignificance of the wall capacitance and work with the overall heat transfer coefficient directly. While this simplifies the computations, it also has a relatively high probability to distort the dynamics estimates. A comparison made by Mathisen *et al.* (1994) shows that neglecting the capacitance of the wall leads to estimates of the dynamic response of a heat exchanger that are much less inert than those with accounting for the wall capacitance. The cell outlet temperatures are the state variables for the model. They constitute the vector of controlled variables for the cell. This leaves the stream flowrates and the inlet temperatures as variables that influence the outlets. One usual case is: (i) The inlet temperatures are disturbances – they are beyond the operator control; (ii) The stream flow-rates are manipulated variables.
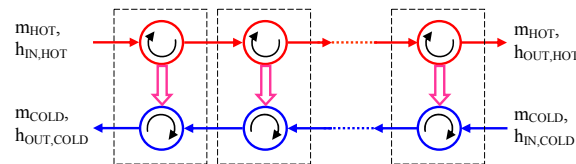
2.4. Cell model of a heat exchanger

Fig 2. Cell arrangement for a single-pass shell-and-tube heat exchanger

A complete heat exchanger can be represented by a combination of heat exchange cells, arranged in a way to most accurately reflect the flow patterns in the actual device Mathisen et al., 1994; Varga et al., 1995). An example for a single-pass heat exchanger is shown in Fig 2. More complex cell configurations are also possible. The cell numbering is assumed to start at the inlet of the exchanger tube-side stream and to follow its path. Usually this is the hot stream.

3. Derivation of the cell parameters

3.1. Driving force effects of cell modelling

The cell model uses the temperature differences between the modelling tanks and the wall as estimates of the driving forces. They are smaller than the actual temperature differences (Fig 3). The effect is stronger for smaller number of cells and weaker for larger number of cells. To compensate for it, the cell heat transfer coefficients must be larger than those for the exchanger as a whole.
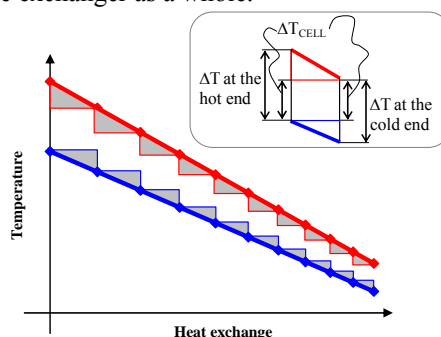

Fig 3. Driving forces decrease in the cell model

3.2. How many cells are needed to model a heat exchanger adequately?

Conceptually, the thermodynamically possible minimum number of modelling cells is given by the number of heat transfer units for the exchanger (Mathisen *et al.*, 1994). As it has been shown, the number of modelling cells is closely linked to the values of the kinetic heat transfer coefficients. This ensures that the overall efficiency of the heat exchanger at steady state will be accurately estimated.

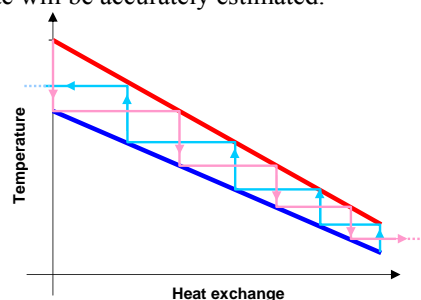Fig 4. Lower bound on the number of cells based on temperature differences

Consider the heat exchanger driving forces at steady state. Fewer cells produce lower estimates of the temperature differences, and vice versa (Fig 3). Below a certain number of cells the cell temperature differences would become negative, making the model thermodynamically incorrect. This can be used for estimating the lower bound on the

Cell-based dynamic heat exchanger models – direct determination of the cell number and size

number of cells needed. If the temperature differences in all heat exchange cells are assumed exactly zero, a sequence of steps can be projected starting from one of the exchanger ends, as is shown in Fig 4.

The procedure can start from any exchanger end. For instance, from the exchanger cold end, a vertical line is drawn from the cold stream to the hot stream. From that point, a horizontal line in the left direction is drawn to the cold stream, finishing the construction of the first cell. This geometrical operation is equivalent to assuming zero temperature difference in a heat exchanger cell at the cold end of the device. Minimum temperature difference leads to maximum possible cell size in terms of heat transfer flow and in turn – minimum number of cells. Further cells are drawn on the diagram in the same way, until the total heat flow reaches or exceeds the one for the heat exchanger at the hot end. For the particular case in Fig 4 the minimum number of cells is four. The described procedure can be performed in the opposite direction, resulting in the same number of cells (Fig 4). The procedure is completely analogous to the classical method for determination of the theoretical number of stages for a binary distillation column (McCabe and Thiele, 1925). For obtaining a feasible cell model, the number of cells must be larger than the identified minimum.

It is also necessary account for the estimated responsiveness of the heat exchanger, which can be expressed through the apparent dead time in reaction of the outlet temperatures to inlet temperature variations. The general trend is in favour of increasing the number of cells. Usually, the best trade-off for a given heat exchanger can be found by varying the number of cells and registering the resulting apparent dead time (Fig 5) (Mathisen *et al.*, 1994). The latter features a typical pattern of asymptotic approach of the dead time estimate to the real one. Therefore, one could start with a number of cells one or two above the lower bound – equal number of cells per heat exchanger pass. The cell number can be gradually increased with one cell per tube pass at each step. The increase should stop when the estimates of the apparent dead time approach the actual ones closely enough.
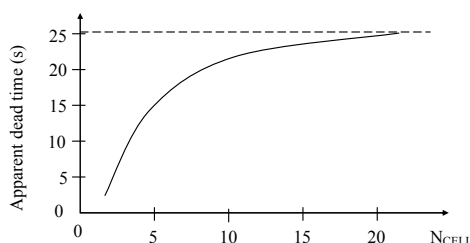
Fig 5. Trend of the apparent dead time prediction of the hot stream outet response to itlet temperature changes (after Mathisen et al. 1994)

3.3. Cell-based film heat transfer coefficients

As it has become apparent, the heat transfer coefficients used in the cell model must be larger than those used for calculations of the exchanger as a whole. One option for obtaining their values is solving the heat transfer equations for the cells in steady state as part of an optimization formulation. The procedure description follows.

Initialization. A number of parameters has to be known, including the stream heat capacity flow-rates, their film transfer coefficients, the tube wall parameters: wall thickness and heat conductivity, the number of cells, the inlet and outlet temperatures of one of the streams, say the cold one, and the inlet or outlet temperature of the other stream, say the inlet temperature of the hot stream, the exchanger heat transfer area.

Step 1. The following heat exchanger properties are calculated directly (only once): duty, the outlet temperature of the second stream, the temperature differences at the hot and the cold ends, the average driving force, the average area per cell, the overall heat transfer coefficient.

Step 2. The steady-state energy balances of the cell tanks plus the equations for heat transfer across the cell wall segments are formulated.

Step 3. The cell temperature differences and the overall cell heat transfer coefficients are used as optimization variables. Further, the calculation of the heat transfer flows in each cell and of the cell tank temperatures are also added to the problem. Finally, an objective function involving the squared error for the total heat exchanger load to be minimized is set. The formulation uses a set of additional inequalities on the cell overall heat transfer coefficient that help in coping with the non-linearity.

4. Conclusions

A method for direct identification of the number of cells in the model, mostly applicable to shell-and-tube heat exchangers has been developed. A useful visualisation of the cell number identification procedure is provided. The method can be further extended to the other kinds of heat exchangers. The case studies shown promising results.

5. Acknowledgements

References

Ansari M. R., MortazaviV., 2006. Simulation of dynamical response of a countercurrent heat exchanger to inlet temperature or mass flow rate change. Applied Thermal Engineering 26(17-18), 2401-2408.

Konukman A.E.S., Akman U., 2005 Flexibility and operability analysis of a HEN-integrated natural gas expander plant, Chemical Engineering Science 60 (24) 7057-7074

Luo X., Guan X., Li M., Roetzel W., 2003. Dynamic behaviour of one-dimensional flow multistream heat exchangers and their networks. In J of Heat and Mass Transf. 46(4), 705-715.

Mathisen K.W., Morari M., Skogestad S., 1994. Dynamic Models for Heat Exchangers and Heat Exchanger Networks. Computers and Chemical Engineering 18, S459-S463.

McCabe W. L., Thiele E.W., 1925. Graphical design of distillation columns, Ind. Eng. Chem. 17(16), 605-611.

Oliveira S. G., Liporace F. S., Araújo,O. Q. F., Queiroz E .M., 2001. The Importance of Control Considerations for Heat Exchanger Network Synthesis: A Case Study., Braz. J. Chem. Eng. 18, 195.

RoetzelW., XuanY., 1998. Dynamic Behaviour of Heat Exchanges, Computational Mechanics; Volume 3, ISBN: 1853125067.

Skogestad S., Postlethwaite I., 1996. Multivariable Feedback Control - Analysis and Design, John Wiley & Sons Inc., New York.

Peng H., Ling X. 2009, Neural networks analysis of thermal characteristics on plate-fin heat exchangers with limited experimental data**,** Applied Thermal Engng 29 (11-12,) 2251-*2256*

Varshney K., Panigrahi P.K., 2005, Artificial neural network control of a heat exchanger in a closed flow air circuit, Applied Soft Computing, 5 (4) 441-465

Varga E. I., Hangos K. M., Szigeti F., 1995. Controllability and Observability of Heat Exchanger Networks in the Time-Varying Parameter Case. Control Eng Practice 3(10), 1409-1419.

Modelling and control of a three-phase catalytic slurry intensified continuous reactor

S. Bahroun[a], S. Li[a], C. Jallut[a], C. Valentin[a], F. De Panthou[b]

[a] Université de Lyon, F-69622, Lyon, France ; Université Lyon 1, Villeurbanne ;
LAGEP, UMR 5007, CNRS, CPE, 43, Bd du 11 Novembre 1918, 69100 Villeurbanne
cedex, France ;(e-mail: bahroun@lagep.univ-lyon1.fr, Tel: (+33)4 72 43 18 81)

[b] AETGROUP SAS, 6 montée du Coteau, 06800 Cagnes-sur-mer, France,

(e-mail: fabrice.depanthou@aetgroup.com, Tel: (+33)6 24 05 25 83).

Abstract

In this paper we propose a hierarchical two-layer approach for the control and optimization of a three-phase continuous chemical reactors. An optimization layer calculates the set points of an advanced nonlinear controller that is based on the concavity of the entropy function and the use of the thermodynamic availability as Lyapunov function. A knowledge based dynamic model of the process in the case of the catalytic o-cresol hydrogenation reactor is briefly described and closed-loop simulations are provided to illustrate the approach.

Keywords: Entropy based control; Dynamic modeling; Intensification; Three-phase continuous reactor; thermodynamic.

1. Introduction

The use of intensified tubular mini/micro reactors instead of batch or fed-batch agitated reactors may be an inherently safe solution to overcome the problem of the heat generated by highly exothermic chemical reactions. However, the question of their control is an issue since these reactors are smaller and more sensitive to perturbations. This paper deals with the modeling and control of the RAPTOR®, an intensified stirred continuous reactor designed by AETGOUP SAS Company. This reactor is well suited for chemical synthesis requiring extreme conditions of pressure and temperature (up to 300°C and 250 bar). For catalytic reactions, this reactor can handle dispersed catalyst. For confidentiality reasons, a detailed description of the RAPTOR® cannot be given. The chemical reaction that is used to illustrate our approach is the hydrogenation of o-cresol. Such a chemical system has already been studied from the control point of view in the case of classical slurry reactors [1,2] and kinetic and thermodynamic data are available in literature [3].

This paper is organized as follow: in section II, the dynamic modeling of a three phase catalytic continuous intensified reactor is briefly described. In section III, the way to build a Lyapunov function for thermodynamic systems is presented. In Section IV is presented the two layers controller that we propose. Finally, the efficiency of the proposed controller is highlighted in section V by some simulation results by comparing its performances with those of a classical PID controller.

2. Dynamic model of the continuous mini reactor

2.1. Principle of the model

Since the equations of the model are available elsewhere [9], there are not given here. Only the principle of the model is described. This model is used to describe the dynamic behavior of the reactor during the hydrogenation of o-cresol on the heterogeneous Ni/SiO2 catalyst. The reaction under consideration occurs without any solvent between the dissolved hydrogen and liquid o-cresol in presence of the catalyst. The reaction rate is available in [3]. Due to the presence of the stirrer, the tubular continuous intensified reactor is treated as an association of J perfectly stirred tank reactors in series with back mixing effect represented by the parameter α. This flow model is assumed to hold for the three phases of the slurry: the liquid phase, the gaseous phase and the catalyst. Similarly, the jacket content is treated as the association of J perfectly stirred tank reactors in series, as shown in fig. 1 [9]. In order to take into account the influence of the reactor metallic body, J pieces of the metal body are also included in the description of the system, the temperature of each being assumed to be uniform. Such an influence of the reactor body has been emphasized in the case of intensified heat-exchanger reactors for example [10].

The model consists in material and energy balance equations for the gaseous phase, the liquid-bulk phase and the catalyst particles, and also energy balance for the metal body and the cooling medium in the jacket. These balances are written for each stirred tank reactor used for the representation of the bulk flow, the jacket liquid flow and the corresponding piece of metal body. The net volumetric flow rates of the liquid and catalyst phases are assumed to be constant. Due to the hydrogen consumption, the gaseous phase volumetric flow rate varies with the axial position. This variation is accounted by the diminution of gaseous phase volumetric hold-up leading to a decreasing of the gas-liquid interface.
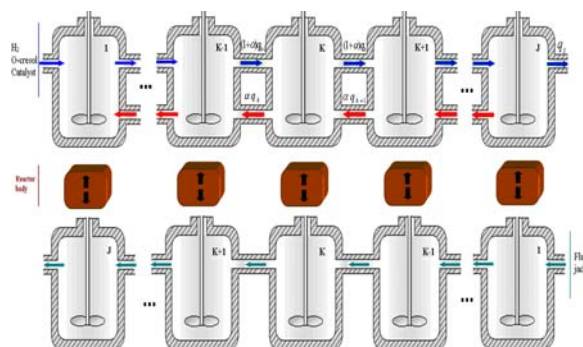
Fig.1. Flow model of the mini-reactor

2.2. Open loop simulation and the choice of controller variable

Open loop dynamic simulations have been published elsewhere [9]. The main conclusion of these simulations is the great effect of the metal body of the reactor on the thermal behavior of the reactor, thanks to its mass. This characteristic has been pointed out to be an inherently safer characteristic of intensified reactors [10]. These simulations have also shown the inlet jacket fluid temperature can be chosen as an efficient manipulated variable for the control purpose.

Modelling and control of a three-phase catalytic slurry intensified continuous reactor

3. Construction of a Lyapunov function

We briefly summarize the way the availability function can be used as a Lyapunov function for thermodynamic systems. This approach has been extensively developed by Ydstie and co-workers [5-8]. The availability function A is a non-linear extension of the entropy curvature $\delta^2 S$ as defined by the Brussel School of Thermodynamic for stability studies [11]. It is positive for thermodynamic systems that are assumed to remain homogeneous. For such systems, the entropy function $S = S(U,V,N_i)$ (U and V are the internal energy and the volume of the system, T, P its temperature and pressure, μ_i the chemical potential and N_i the number of moles of component i) is necessarily concave [12]. According to the local equilibrium principle and in the case of great perturbations from a given steady state of the system, a finite distance between the entropy function and its tangent plane can be considered as follows:

$$S(U,V,N_i) = S\left(\underline{U},\underline{V},\underline{N_i}\right)$$

$$+ \frac{1}{\underline{T}}\left(U-\underline{U}\right) + \frac{P}{\underline{T}}\left(V-\underline{V}\right) - \sum_i^I \frac{\mu_i}{\underline{T}}\left(N_i-\underline{N_i}\right) \tag{1}$$

$$+ U\left(\frac{1}{T}-\frac{1}{\underline{T}}\right) + V\left(\frac{P}{T}-\frac{P}{\underline{T}}\right) - \sum_i^I N_i\left(\frac{\mu_i}{T}-\frac{\mu_i}{\underline{T}}\right)$$

where the overbar stands for the steady-state point. The quantity

$S(\overline{U},\overline{V},\overline{N_i}) + \frac{1}{\overline{T}}\left(U-\overline{U}\right) + \frac{\overline{P}}{\overline{T}}\left(V-\overline{V}\right) - \sum_i^I \frac{\overline{\mu_i}}{\overline{T}}\left(N_i-\overline{N_i}\right)$ is the equation of the tangent plane

to the entropy surface at the point $\left(\overline{U},\overline{V},\overline{N_i}\right)$. The following condition must be satisfied for the homogeneous phase to be stable:

$$A = U\left(\frac{1}{\overline{T}}-\frac{1}{T}\right) + V\left(\frac{\overline{P}}{\overline{T}}-\frac{P}{T}\right) - \sum_i^I N_i\left(\frac{\overline{\mu_i}}{\overline{T}}-\frac{\mu_i}{T}\right) > 0 \tag{2}$$

In order to use the availability concept according to the approach proposed by Ydstie and co-workers, its time derivative $\frac{dA}{dt}$ has to be calculated and the way the condition $A = 0$ is satisfied has to be analyzed. In order the entropy to be strictly concave and the condition $A = 0$ to be satisfied only at the steady-state point under consideration, a constraint has to be imposed to the system, specifically on at least one extensive variable.

4. The two layer control structure

We propose to control the outlet concentration of o-cresol in the liquid phase through the control of the metal body temperature at the outlet axial position of the intensified continuous reactor (see Fig. 1). Indeed, the metal body of the reactor has an essential effect on the thermal behavior of the reactor, thanks to its high mass, as highlighted in [9,10].

4.1. Optimization layer synthesis

The objective of the optimization layer is to calculate an optimal set point for the metal body temperature of the reactor at the outlet position $\overline{T}_m^J$ to guaranty a fast convergence of the o-cresol desired outlet concentration despite the presence of the disturbances. The optimization layer is based on the steady-state version of the dynamic model and an

objective function. Due to the reactor construction, its temperature should not be greater then a given threshold $T_{m_{\max}}$ so that a constrained optimization problem can be defined:

$$\begin{cases} \min\limits_{T_j^0, D_w}(\overline{C}_{Bl} - C_{Bl}^J)^2 \\ T_m^k \underset{k \in \{1..J\}}{\leq} T_{m_{\max}} \end{cases} \tag{3}$$

The constrained optimization problem is solved by a sequential quadratic programming (SQP) algorithm.

4.2. Control layer synthesis

The proposed control law is based on the measurement of the outlet fluid temperature, T_f^J, and of the axial outlet reactor body temperature, T_m^J. The control variable is classically the inlet jacket temperature, T_j^0.

The optimization layer calculates, for each control horizon k, a fixed reference set point $\overline{T}_{m,k}^J$ for the reactor metal body temperature at the outlet axial position (see Fig.1). The feedback law acting during this control horizon allows calculating the control variable T_j^0.

The synthesis of this inner loop control law is based on the availability of the metal body of the Jth CSTR with respect to $\overline{T}_m^J$ as given by:

$$A_m^J = \left(\frac{1}{\overline{T}_m^J} - \frac{1}{T_m^J}\right)U_m^J + \left(\frac{\overline{P}}{\overline{T}_m^J} - \frac{P}{T_m^J}\right)V_m^J - N_m^J\left(\frac{\overline{\mu}_m^J}{\overline{T}_m^J} - \frac{\mu_m^J}{T_m^J}\right) > 0 \tag{4}$$

A state feedback control is designed to calculate the manipulated variable T_j^0 such that $\dfrac{dA_m^J}{dt} < 0$. Since V_m^J, N_m^J are constant, the time derivative of A_m^J is given by:

$$\frac{dA_m^J}{dt} = \left(\frac{1}{\overline{T}_m^J} - \frac{1}{T_m^J}\right)\frac{dU_m^J}{dt} \tag{5}$$

The energy balance for the reactor body of the CSTR number J can be written as follows:

$$\frac{dU_m^J}{dt} = \lambda(T_f^J - T_m^J) + \lambda_j(T_j^0 - T_m^J) \tag{6}$$

where λ and λ_j are the global heat transfer coefficients on the bulk and jacket sides respectively. From the equation (6), the time derivate of A_m^J can be written as follows:

$$\frac{dA_m^J}{dt} = \left(\frac{1}{\overline{T}_m^J} - \frac{1}{T_m^J}\right)\left(\lambda T_f^J - (\lambda + \lambda_j)T_m^J + \lambda_j T_j^0\right) \tag{7}$$

Let us choose:

$$T_j^0 = \frac{1}{\lambda_j}\left(-\lambda T_f^J + (\lambda + \lambda_j)T_m^J - K_c\left(\frac{1}{\overline{T}_m^J} - \frac{1}{T_m^J}\right)\right) \tag{8}$$

with K_c strictly positive. We obtain for $\dfrac{dA_m^J}{dt}$:

$$\frac{dA_m^J}{dt} = -K_c\left(\frac{1}{\overline{T}_m^J} - \frac{1}{T_m^J}\right)^2 \leq 0 \tag{9}$$

Modelling and control of a three-phase catalytic slurry intensified continuous reactor

5. Simulation results and discussion

On the Fig. 2(a) are represented the open-loop (OL) and closed-loop (CL) behaviors of the mini-reactor when a ramp disturbance on the inlet bulk fluid flow is applied showing the good performance of the control law to fully compensate for this disturbance.
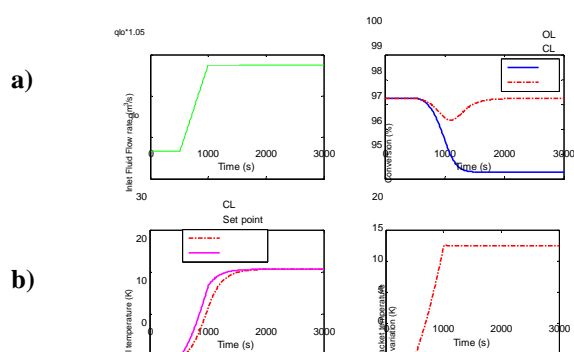

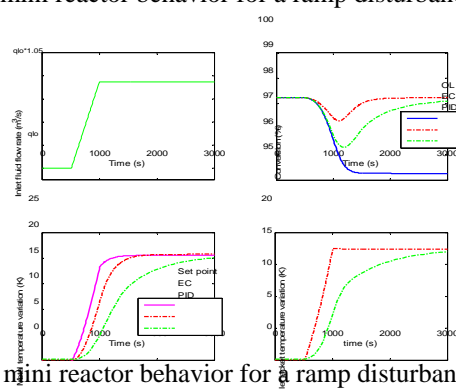Fig. 2. The mini reactor behavior for a ramp disturbance on q_{l0}.

Fig. 3. The mini reactor behavior for a ramp disturbance on q_{l0}.

On the Fig. 2(b) are represented the set point evolution $T_{\underline{m}}^{J}$ as it is calculated by the optimization layer and the metal body temperature as it is assumed to be measured at the outlet position of the mini-reactor. The metal body temperature follows the set point with a satisfying evolution that guaranty an industrially applicable evolution of the manipulated variable T_{j}^{0}.

A conventional PID regulation is compared with the advanced controller for the same situation. The parameters of PID controller are designed by a first-order transfer function obtained by model simulation. The set point of PID controller is the optimal conversion (98.23%). The set point of entropy based controller is calculated by optimization. On the Fig. 3 are presented the results obtained with the PI controller and the Entropy based Controller (EC), in the case of the ramp perturbation on the inlet bulk fluid flow. This figure highlights the efficiency of the new entropy controller with respect to a traditional industrial controller to reject inlet disturbances.

6. Conclusion

In this paper we proposed a novel controller based on the concavity of the entropy and the use of the thermodynamic availability function as Lyapunov function.

The two main objectives of the new two-layer entropy-based control method presented in this paper are to reject disturbances in inputs variables that affect the outlet product quality and to operate safely. The optimization layer, using SQP algorithms, calculates a set point for the control layer. The obtained results show the efficiency of the controller to reject input disturbances with a safe behavior. Indeed the proposed control scheme is simple and based on measurements of the fluid temperature and of the metal temperature at the outlet of the mini-reactor. Simulated scenarios show an industrially applicable evolution of the inlet jacket temperature, which is the manipulated variable.

References

[1] M. C. Rezende, A. C. Costa and M. Filho, Control and optimization of a Three Phase Industrial Hydrogenation Reactor, International Journal of Chemical Engineering. 2 (2004) A21.

[2] D. N.C. Melo, C. B. Borba Costa, E. C. V. De Toledo, M. M. Santos , M. R. Wolf Maciel, and R. M. Filho, Evaluation of control algorithms for three-phase hydrogenation catalytic reactor, Chemical Engineering Journal. 141 (2008) 250–263.

[3] H. Hichri, A. Armand and J. Andrieu, Kinetics and slurry-type reactor modelling during catalytic hydrogenation of o-cresol on Ni/SiO2, Chemical Engineering Process (1991) 30, 133–140.

[4] A. A. Alonso, B. E. Ydstie, Process systems, passivity and the second law of thermodynamics, Computers and Chemical Engineering. 20(2) (1996) 1119-1124.

[5] B. E. Ydstie, A. A. Alonso, Process systems and passivity via the Clausius-Planck inequality, Systems Control Letters. 30(5) (1997) 253-264.

[6] A. A. Alonso, B. E. Ydstie, Stabilization of distributed systems using irreversible thermodynamics, Automatica. 37 (2001) 1739-1755.

[7] M. Ruszkowski, V. Garcia-Osorio, B. E. Ydstie, Passivity based control of transport reaction systems, AIChE Journal. (2005) 51(12) 3147-3166.

[8] H. Hoang, F. Couenne, C. Jallut, Y. Le Gorrec, Lyapunov based control for non isothermal continuous stirred tank reactor, in: Proceedings of the 17th World Congress of the IFAC, July 6-11, 2008, Seoul, Korea.

[9] S. Bahroun, C. Jallut, C. Valentin, F. De Panthou, Dynamic modelling of a three-phase catalytic slurry intensified chemical reactor, Proceedings of the IFAC ADCHEM 2009, International Symposium on Advanced Control of Chemical Processes, Istanbul, Turkey, July 12-15 (2009) 862-867.

[10] W. Benaissa, N. Gabas, M. Cabassud, D. Carson, S. Elgue, M. Demissy, Dynamic behaviour of a continuous heat exchanger/reactor after flow failure, Journal of Loss Prevention in the Process Industries. 21 (2008) 528-536.

[11] D. Kondepudi, I. Prigogine, Modern thermodynamics. From heat engines to dissipative structure, Wiley and Sons (1998).

[12] H. B. Callen, Thermodynamics and an introduction to thermostatistics, 2nd edition, Wiley and Sons (1985).

Practical Aspects of Dynamic Simulation in Plant Engineering

Gabriele Engl, Andreas Kröner, Martin Pottmann

Linde AG, Linde Engineering Division, Dr.-Carl-von-Linde-Str.6-14, Pullach, Germany
E-mail: gabriele.engl@linde-le.com, andreas.kroener@linde-le.com,
martin.pottmann@linde-le.com

Abstract

This paper illustrates the increasing trend to integrate dynamic simulation into the workflow of a plant engineering and contracting company. While methods and tools for dynamic simulation are well-established in academia, the industrial world still faces a lot of challenges when it comes to their practical application. These challenges are discussed by example of a carbon capture and storage (CCS) technology: an oxyfuel power plant. Dynamic simulation is first applied to the individual plant components such as the Gas Processing Unit (GPU) and the Air Separation Unit (ASU), where the early integration into the plant engineering workflow plays a major role. In order to study the dynamic behavior of the complete oxyfuel power plant in a second step, the models which are implemented in different simulation tools are reduced for integration into a common simulation platform.

Keywords: Dynamic simulation, carbon capture and storage, oxyfuel, model reduction

1. Introduction

Computer Aided Process Engineering (CAPE) plays a key role in the industrial practice of the Linde AG, Engineering Division. As a leading international engineering and contracting company, Linde Engineering designs and builds turnkey process plants for a wide variety of industrial users and applications: The chemical industries, air separation, hydrogen and synthesis gas production, natural gas processing and more. Being able to call on its own extensive process engineering know-how in the planning, project development and construction of turnkey plants, Linde Engineering is also pursuing a CAPE strategy based on a long tradition of internal know-how: High quality methods and tools available from universities or commercial suppliers are combined with internal developments to achieve optimal solutions. For process simulation and optimization, Linde Engineering uses commercial process simulation tools as well as its in-house process simulation program OPTISIM[®] [1,2].

While steady-state simulation and optimization are well established in the engineering workflow, the use of dynamic simulation is still restricted to a small number of experts and selected projects – even though the methods and tools have been available for a long time. However, there is an increasing trend to apply dynamic simulation, in particular for new process developments such as in the area of CCS. This technology is briefly described in Section 2. Section 3 discusses challenges of integrating dynamic simulation into the engineering workflow, illustrated by the GPU which is modelled with the commercial tool UniSim[®]. Section 4 presents approaches for the reduction and integration of models implemented in different simulation tools, illustrated by the ASU which is modelled with the in-house program OPTISIM[®].

2. Example: Carbon Capture and Storage (CCS)

CCS technologies allow separating carbon dioxide (CO_2) when generating energy, particularly in coal-fired power plants, and thus support the target of zero-emissions. Linde Engineering is working closely with industry partners to develop technical solutions for CCS [3], among them the oxyfuel process where coal is combusted in an atmosphere of oxygen, produced by an Air Separation Unit (ASU), and CO_2. This means that the nitrogen content in the resultant flue gas is significantly reduced, and the flue gas primarily consists of CO_2 and water. This CO_2–rich gas is further concentrated and compressed in a Gas Processing Unit (GPU) and then transferred to storage. The ASU and GPU are part of Linde Engineering's product portfolio [4,5].

3. Dynamic Simulation within the Plant Engineering Workflow

The complexity of new process steps and novel plant configurations calls for intensive dynamic simulations right from project start. The basic task is to identify the most appropriate process concepts for the ASU and GPU out of numerous process variants. Both technical requirements such as guaranteed product purities at specified load change rates, and business requirements such as lowest investment costs aligned with highest energy efficiency and flexibility have to be satisfied. Originating from steady-state design process models, dynamic simulation models are derived for verification of static equipment design and development of control strategies. Some challenges arising from integrating this approach into the plant engineering workflow are discussed using the GPU design process as an example.

3.1. Communication between subject matter experts

Dynamic simulation can play a key role in a variety of engineering tasks - process design, control and operational strategy design including startup and shutdown procedures, safety and abnormal situation handling, etc. However, its use still faces barriers as dynamic simulation is often considered an "expert tool" and is not easily accessible to engineers of various backgrounds and design responsibilities. In order to remove these barriers and facilitate efficient communication between subject matter experts, Linde Engineering uses even at an early project stage Human Machine Interface (HMI) visualization to develop graphical user interfaces for dynamic models. This approach provides an intuitive and common environment in which process configurations, control and operational strategies can be demonstrated and discussed. It lets the user better focus on the relevant (e.g. operational) information than by using the simulation software interface directly, which typically provides a detailed process design view. For the GPU example, such a process visualization consisting of a process overview (see Fig. 1) and a series of more detailed process area views has been implemented in UniSim® Operations, resulting in significantly improved acceptance of the dynamic model.

In this particular GPU design [5] CO_2 is separated from the oxyfuel flue gas stream using a cryogenic approach - raw gas compression, CO_2 liquefaction via a main heat exchanger and two separators close to the triple point of CO_2, vaporization and consequent product compression. Before being partially used for adsorber regeneration, vent gas is routed through the main exchanger and two expansion turbines, which drive booster compressors in the raw gas stream and in one of the product streams. The process objectives are to maintain a required CO_2 recovery rate while satisfying the product purity specifications for a given range of feed conditions.

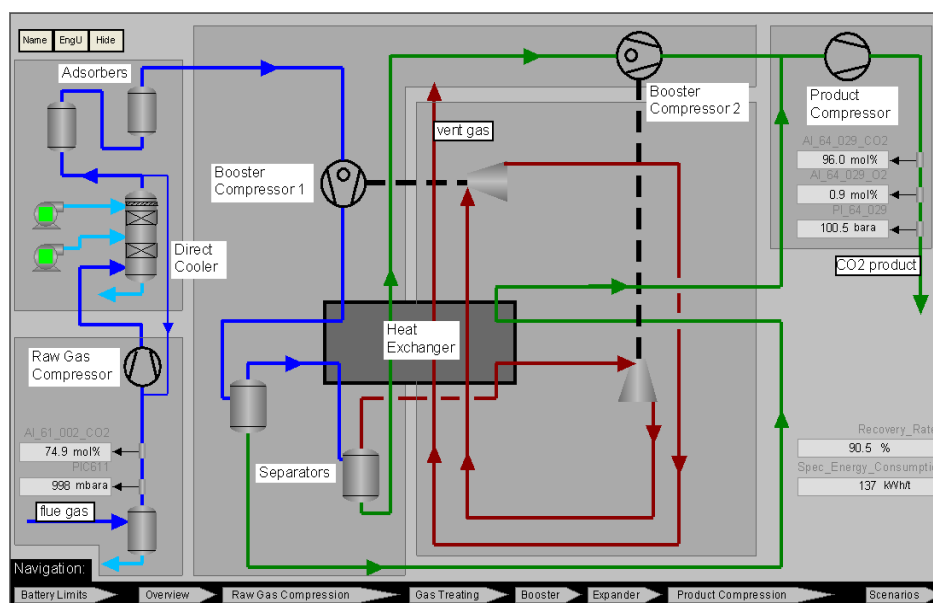*Practical Aspects of Dynamic Simulation in Plant Engineering*

Figure 1. GPU Overview

3.2. Iterative nature of design procedure

It is generally accepted that the benefits of dynamic simulation tend to be largest if the findings become available early enough in a project so that design changes can still be incorporated. Yet, a high-fidelity dynamic model can only be developed if detailed design information is available. To reconcile these opposing trends it is imperative to integrate dynamic simulations tightly into the process design workflow. Initially, steady-state design has not advanced sufficiently for specifying detailed simulation models for equipment such as turbo machines, heat exchangers, or piping. Hence, the fidelity of the initial dynamic models is low as e.g. constant efficiencies for turbo machines are assumed. At this stage the focus of dynamic simulation is on evaluating general operability and control concepts, and understanding the dominant dynamic modes of the process. As the design advances, more equipment details get specified (e.g. performance maps for turbo machines) which have to be transferred to the dynamic model. Therefore, the simulation software needs to be flexible to allow for different modeling levels within individual process units. Findings from dynamic analysis impact the static process concepts, resulting in an iterative design procedure.

Typical applications of high-fidelity dynamic models are discussed below:

 (i) *Design evaluation*: It is determined whether the design objectives are met during transitions between steady-state design points, and how robust the design is with respect to disturbances (such as varying feed and utility supply conditions) and equipment failures (such as turbine trips or loss of cooling water). As an example, the GPU model is subjected to a 30% rate reduction introduced over 7.5 min and a concurrent 2% reduction in feed CO_2 concentration. The simulation results in Figure 2 confirm that the product purity specifications (>95% CO_2, <1% O_2) are maintained throughout the transition. However, the specific energy consumption increases due to the lower efficiency of the turbo machines which at around t=7min start to move into recycle mode.
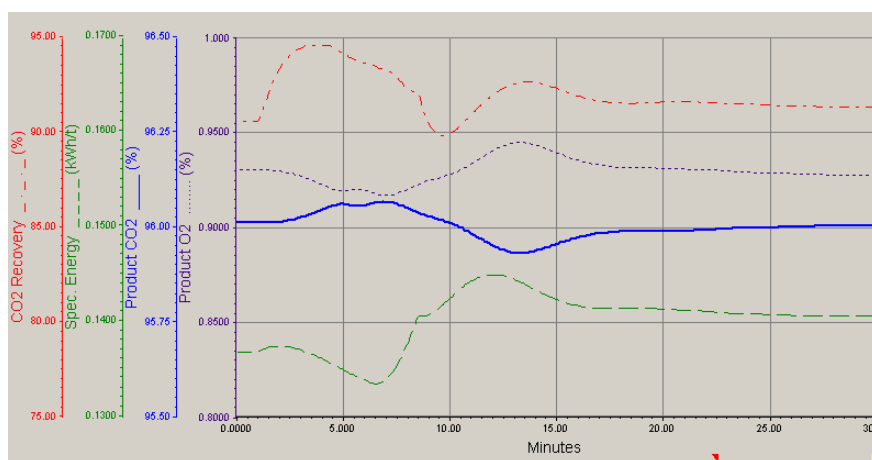
Figure 2. GPU model predictions during a feed transition

(ii) *Design improvements:* Knowledge of the dynamic behavior can be used to improve the design of the overall process as well as the individual units. As an example, the temperature peak resulting from a switchover between adsorber beds impacts the operation of the main heat exchanger. While the exchanger was initially designed based on steady-state (worst-case) inlet temperature predictions, resulting in a certain level of overdesign, the dynamic temperature predictions are used to further improve the exchanger design.

(iii) *Control strategy development:* While the basic control strategies can be developed and evaluated using a low-fidelity model, a thorough evaluation of control and operational strategies requires both accurate process unit models and specification of all control-relevant equipment and logic such as compressor anti-surge and expander speed limiting controllers. For the GPU process, simulations covering a wide range of operating conditions revealed the need for additional temperature control loops in the vent gas path.

4. Model Reduction

Once dynamic simulation models have been developed for each of the oxyfuel process units ASU, GPU and others, the next step is to combine theses models to a plant wide dynamic model for operability analysis and control system design. The fact that each industry partner contributing process units to the oxyfuel process usually applies its own design and dynamic simulation software, calls for a strategy to efficiently combine dynamic models from different dynamic simulation programs. Options to set up a plant wide dynamic model composed of process unit models originating from commercial simulation tools, e.g. UniSim® and APROS, generic modeling software, e.g. gPROMS, MATLAB® and Dymola, or in-house developments, e.g. OPTISIM®, are discussed.

Without identifying a specific target platform for integration in the current example its selection must consider various aspects such as the effort of transferring dynamic models to the target platform, reasonable computing times, acceptable level of model fidelity of the transferred model, protection of each contributor's model know-how, etc. The latter may be important in case the integrated model shall be available to all

Practical Aspects of Dynamic Simulation in Plant Engineering

contributing partners. What clearly needs to be avoided is the attempt to re-implement an existing dynamic model in a new software tool.

Several options have been evaluated for transferring models from the original dynamic simulation tool to a target platform for integration:

(i) Applying an *operating point linearization* of a nonlinear dynamic process unit model in an equation based simulation tool, e.g. OPTISIM® [2], results in a linear dynamic state space model. The transfer to the target platform then includes a linear model reduction step [6] to gain a reduced dimensional linear state space model.

(ii) For *identification of a linear or nonlinear dynamic system,* responses of the original system to step inputs, PRBS (pseudo-random binary sequence), or other suitable signals are generated [7]. This information is input to commonly available system identification software for linear systems or e.g. artificial neural networks. Input signals and system responses can be generated with models in any simulation tool.

(iii) A *direct coupling of the original heterogeneous dynamic simulation tools* in an overall integration platform has been reported in [8].

For the oxyfuel example, the first two options have been pursued due to aspects of know-how protection, reducing computing time and maintaining flexibility for integrating the models into different simulation target platforms. Numerous reduction and identification algorithms can be found in free and commercially available software tools. Either approach to model transfer results in an encoded model of reduced order. Prior to integration into the target platform, the range of validity of the reduced model must be evaluated via comparison runs with the original models.
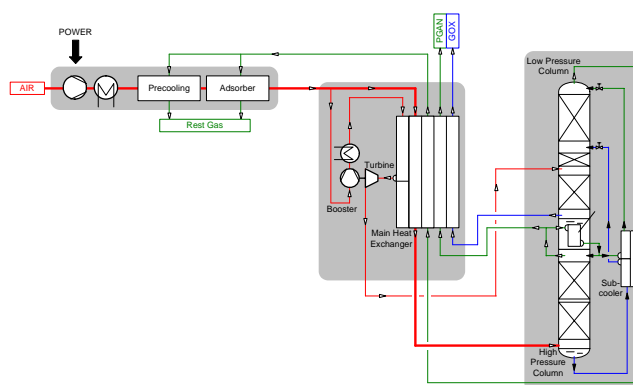
Figure 3. View of an ASU for oxyfuel

As an example, the transfer of the ASU model from OPTISIM® to a linear reduced state space model is explained. Figure 3 shows a typical ASU configuration for an oxyfuel process [4]. The main AIR stream is compressed and cooled in the main heat exchanger and fed to the bottom of the high pressure column. The product stream GOX with 95 % oxygen is withdrawn from the bottom of the low pressure column and reheated in the main heat exchanger. Nitrogen is removed as high pressure product PGAN or as low pressure rest gas.

The nonlinear model set up in OPTISIM® consists of a set of 7482 differential and algebraic equations (DAE). This model is linearized at the steady state operating point of 100% load to yield a linear DAE system. After eliminating the algebraic subsystem, a

balanced residualization algorithm for model reduction is applied to obtain a linear ODE system of size 70. The original nonlinear model is compared to the linear reduced model in various test runs. Figure 4 shows a typical system response in selected oxygen concentrations of both models to changes in AIR and GOX flow rates.
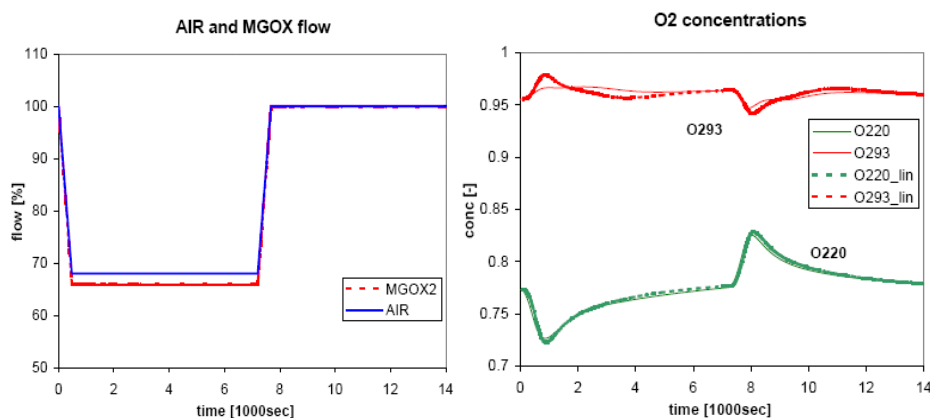

Figure 4. Load change simulation with nonlinear model and linear reduced model

5. Conclusion

The paper presented practical aspects of dynamic simulation from the perspective of a plant engineering and contracting company. Typical challenges and selected solutions were discussed and illustrated using the oxyfuel process as an example. It was shown how dynamic simulation can improve the process and control design. Additional benefits were realized by an early workflow integration and HMI visualization. Furthermore, results of a successful model reduction were presented, which provides the basis for integrating models from different simulation tools.

References

[1] G. Engl, A. Kröner, 2006, Success Factors for CAPE in the Engineering Practice of a Process Plant Contractor, In: W. Marquardt, C. Pantelides (eds..): Computer-Aided Chemical Engineering 21A, Proc. of ESCAPE-16 and PSE'2006, Elsevier, Amsterdam, 763-768.
[2] A. Kröner, 2006, An Engineering Company's Approach to Filling "CAPE Gaps" in Process Simulation, In: W. Marquardt, C. Pantelides (eds.): Computer-Aided Chemical Engineering 21A, Proc. of ESCAPE-16 and PSE'2006, Elsevier, Amsterdam, 781-786.
[3] Linde AG, 2008, Carbon Capture and Storage, The Linde Corporate Responsibility Report 2008, 59-63.
[4] G. Beysel, 2009, Enhanced Cryogenic Air Separation: A Proven Process Applied to Oxyfuel, In: S. Santos (ed.): Proc. of 1st Oxyfuel Combustion Conference, Cottbus, Germany.
[5] R. Ritter, A. Kutzschbach, T. Stoffregen, 2009, Energetic Evaluation of a CO2 Purification and Compression Plant for the Oxyfuel Process, In: S. Santos (ed.): Proc. of 1st Oxyfuel Combustion Conference, Cottbus, Germany.
[6] A. C. Antoulas, D. C. Sorensen, and S. Gugercin, 2001, A survey of model reduction methods for large scale systems, Contemporary Mathematics, AMS Publication, 280, 193-219.
[7] M. Pottmann, D. E. Seborg, 1992, Identification of nonlinear processes using reciprocal multiquadric functions, Journal of Process Control, 2, 189-203.
[8] V. Kulikov, H. Briesen, R. Grosch, A Yang, L. v. Wedel, W. Marquardt, Modular dynamic simulation of integrated process flowsheets by means of tool integration, Chemical Engineering Science, 2005, 60(7), 2069-2083.

Multicriteria dynamic optimization of an emulsion copolymerization reactor

B. Benyahia, M. A. Latifi, C. Fonteix, F. Pla

Laboratoire des Sciences du Génie Chimique, CNRS-ENSIC
1 rue Grandville, BP 20451, 54001 Nancy Cedex, France,
E-mail addresses : brahim.benyahia,latifi,fonteix,pla}@ensic.inpl-nancy.fr

Abstract

A multicriteria optimization procedure based on an evolutionary algorithm has been developed to determine the optimal control policies for a fed-batch emulsion copolymerization reactor, particularly for styrene and butyl acrylate in the presence of n-C12 mercaptan as chain transfer agent (CTA). The process model was elaborated and validated experimentally in order to predict the global monomer conversion, the number and weight average molecular weights, the particle size distribution and the residual monomers mass fraction. The process objectives were to produce core-shell particles (hard core and smooth shell) with specific end use properties and high productivity. This has been achieved by the maximization of the conversion at the end of the process and the minimization of the error between the glass transition temperature and a designed profile subject to a set of operational constraints. The nondominated Pareto solutions obtained were ranked according to a decision aid strategy based on a decision maker preferences and experience using multiple attribute utility theory (MAUT). Finally, the best solution was implemented experimentally.

Keywords: Fed-batch emulsion copolymerization, multicriteria optimization, modeling, core-shell particles, experimental implementation.

1. Introduction

Multiobjective optimization problems are encountered in most real-world applications and more recently in chemical processes (Fonteix et al. (2004), Garg and Gupta (1999), Mitra et al. (2004), Sakar et al. (2007)). Since such problems involve several objective functions with conflicting nature, the final optimum is not unique but a set of non dominated solutions (the Pareto front) which show a trade-off between the whole objectives. Genetic algorithms (GAs) are well adapted tools to solve multiobjective problems. This kind of technique stands for a class of stochastic optimization methods that simulate the process of natural evolution (mainly genetic algorithms, evolutionary programming, and evolution strategies). These algorithms have proven themselves as a general, robust and powerful search mechanism. Moreover, evolutionary algorithms (EAs) seem to be especially suited for multiobjective optimization because they are able to find multiple Pareto-optimal solutions in a single simulation run.

Emulsion polymerization is an important industrial process used to produce a great variety of polymers of multiple uses (e.g. paints, adhesives, coatings, varnishes). The end-use properties of the products obtained by this process are governed by the molecular weight distribution (MWD), polymer microstructure, glass transition temperature (T_g), particle size distribution (PSD) and particles morphology. These

parameters must be involved in the process design, optimization and control in order to produce latex particles with specific and controlled properties.

The present paper deals with a multiobjective dynamic optimization of an emulsion copolymerization fed-batch reactor. The aim is to produce core-shell particles with specific mechanical and film-forming properties with high productivity. These characteristics are achieved by using two objective functions subject to a set of tight operational constraints and the mathematical model of the system. The first objective function is related to the glass transition temperature of both core and shell while the second deals with the final conversion rate.

The nondominated solutions (Pareto's front) are obtained by using an evolutionary algorithm (EA). This set of optimal solutions is ranked according to the decision maker preferences by using multiattribute utility theory (MAUT) which leads to the selection of the unique solution to be implemented.

2. Formulation of the problem

The process model of the emulsion copolymerization of styrene and butyl acrylate in the presence of n-C12 mercaptan as chain transfer agent (CTA) was developed and validated experimentally for a batch reactor and extended to the fed-batch case (Benyahia et al.(2009)). The objective of the model is to predict different variables including overall monomers conversion, number and weight average molecular weights, particles average diameter and residual monomer fractions.
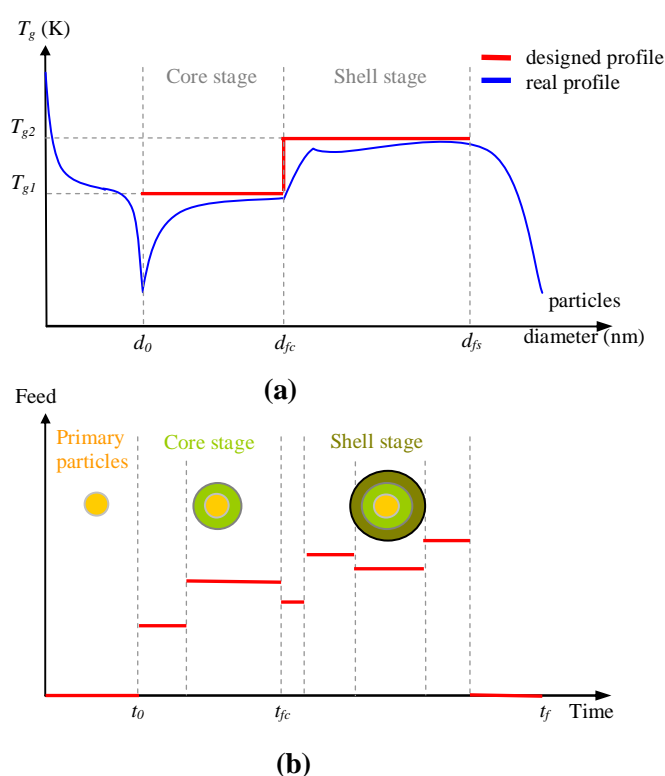

Figure 1: **a**- feed rate profile, **b**- instantaneous glass transition temperature (T_{gi}) profile.

Multicriteria dynamic optimization of an emulsion copolymerization reactor

The objective of the process is to produce core-shell particles with a specific end-use properties depending on the glass transition temperature profile (Fig.1a). The monomers used (styrene and butyl acrylate) in the copolymerization process have different reactivity ratios and their polymers have very different glass transition temperature (-54 °C for butyl acrylate and 100 °C for styrene).

The key feature of the optimization problem is to determine optimal feed rate and time periods profiles necessary to control polymerization reactions in order to produce particles with a designed morphology and glass transition temperature (Fig. 1b).

Two objective functions have been selected for the optimization of this process. The first one is to minimize the error between the glass transition temperature and the desired profile. The second objective is to maximize the conversion rate at the end of the process which leads to higher productivity (equations 1).

$$Min\ \boldsymbol{f} = \left[f_1, f_2 \right]^T$$

$$f_1 = \frac{1}{t_{fc} - t_0} \int_{t_0}^{t_{fc}} \left| T_g - T_{g1} \right| dt + \frac{1}{t_{fs} - t_{fc}} \int_{t_{fc}}^{t_{fs}} \left| T_g - T_{g2} \right| dt$$

$$f_2 = -X(t_f)$$

s.t. $\quad \dot{\boldsymbol{x}} = \boldsymbol{f}(\boldsymbol{x}(t), \boldsymbol{u}(t), \boldsymbol{p}, t) \ ; \quad x(t=0) = x_0$

$$\frac{1}{t_{fc} - t_0} \int_{t_0}^{t_{fc}} \left(0.9 - X(t) \right)^2 dt \le \varepsilon^2 \tag{1}$$

$$X(t_0) = 0.90$$

$$M_n(t_f) \le 4 \times 10^4$$

$$M_w(t_f) \le 1.5 \times 10^5$$

$$\boldsymbol{u}_{inf} \le \boldsymbol{u}(t) \le \boldsymbol{u}_{sup}$$

$$\boldsymbol{x}^T = \left[V_R, M_1, M_2, M_{T1}, M_{T2}, I, Z, TA, S, N_p, \right.$$
$$\left. R_1, R_2, \chi_1, \chi_2, \tilde{n}, \lambda_1, \lambda_2, N_m, L_1, L_2 \right]$$

$$\boldsymbol{u}^T = \left[\Delta t_1, \Delta t_2, \dots, \Delta t_n, Q_1, Q_2, \dots, Q_n \right]$$

where T_g is the time dependent glass transition temperature, T_{g1}, T_{g2} the desired glass transition temperature for the core (5°C) and the shell (10°C) respectively, t_{fc} and t_{fs} the times necessary to produce the core and the shell respectively, $X(t_f)$ is the overall conversion at the end of the process and $\boldsymbol{u}$ the control vector (feed rates and time periods).

At the first stage of the process, the primary particles are produced under batch conditions. This stage ends when the overall conversion reaches the value of 0.9. The reactor is then fed with pre-emulsioned monomers and chain transfer agent (CTA). Core stage is designed to be under starving conditions (no droplets are produced and the feed rate is equal to the polymerization rate). Styrene is consumed faster than butyle acrylate due to the difference between there reactivity ratios. As a result, the instantaneous glass temperature will grow to reach the desired value. This stage is operating under a

constraint on the overall conversion rate. The shell stage is conducted without required conditions or constraints. Only the objective to reach the second step of the designed glass temperature profile is kept. Feed rates are more important at this stage to allow the growth of the glass temperature by adding more quantities of styrene. The final stage is operating under batch conditions. Since no styrene is added the residual butyl acrylate is consumed leading to lower glass temperatures. The objective at this stage is to maximize the overall conversion which means maximizing productivity and minimizing residual volatile organic compounds (VOC's).

3. Results and discussion

The first result of the multiobjective optimization is the Pareto-optimal set of solutions depicting tradeoffs between the competing objectives. This set was generated by using an evolutionary algorithm (EA) with different initial populations. The best results obtained are presented in figure 2a. The best value of the objective functions taken individually are 4.4 (the error between the designed and the resulting profiles) and -0.948 for the criterion related to the final conversion rate.

The Pareto front is divided into 10 subsets or classes (Fig.2a) according to the decision aid method developed in this work (MAUT). Moreover, the feed profiles (feed rates and time periods) of the best subset (class 1) are presented in figure 2b. It is noteworthy that the feed profiles show clearly the limits of the second and the third stages. The second one (core stage) is characterized by low feed rates which corresponds to starving condition where styrene is more consumed leading to the designed glass temperature. Similarly the third stage shows higher feed rate necessary to increase once more the glass temperature.

The best profile presented in figure 3a was experimentally implemented and the results are given in figure 3. First, the glass temperature profile obtained (Fig. 3b) corresponds to the designed profile ($T_{g1} = 5$ °C and $T_{g2} = 10$ °C). The first stage (the primary particles formation or seeding) ends with a fall in the glass temperature value. This is quite realistic since butyle acrylate is more consumed when no styrene is added. This phenomenon is also noticed at the end the process.

| (a) | (b) |
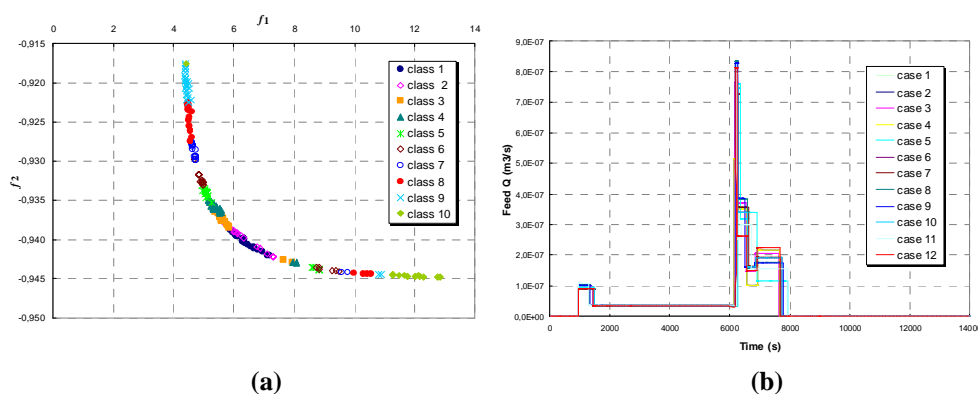
Figure 2: **a**- the Pareto front, **b**- feed profiles of the class 1

Multicriteria dynamic optimization of an emulsion copolymerization reactor

The overall conversion at the stage of the core formation lies within the limit of the constraint (Fig. 3c). The conversion rate falls during the shell stage as a result of higher feed rates (no constraint on the conversion rate is applied). On the other hand, the last stage (batch process) shows that the overall conversion rate grows to reach the final conversion which is high enough compared to the best solution obtained for the second objective function.

The time evolution of the number and weight average weights presented in figure 3d show that the constraints on the final values ($M_n(t_f)$ and $M_w(t_f)$) are respected. More over good agreement between experimental measurements and simulations is achieved.

More results of styrene residual mass fraction, average particles diameter and the number of particles are not presented here for the sake of brevity.

(a)

(b)

(c)

(d)

Figure 3: **a**-feed profile of the best solution, **b**- instantaneous glass transition temperature profile, **c-** overall conversion, **d**- number and weight average molecular weights

4. Conclusions

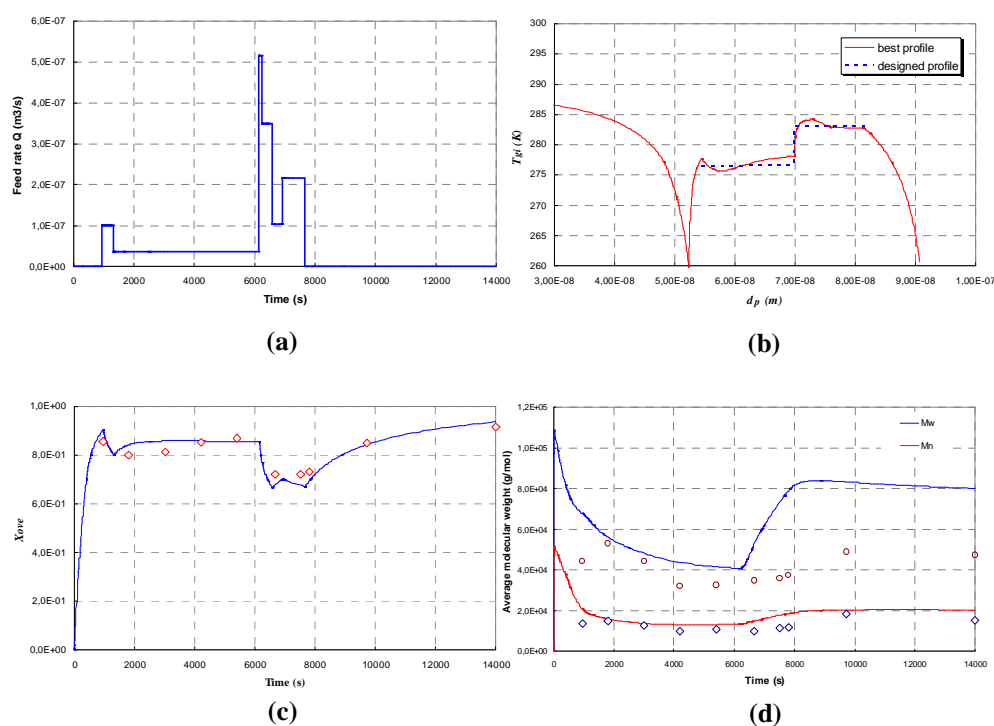In this work, multiobjective optimization problem has been addressed to determine optimal feed profiles necessary to produce core-shell latex particles with specific end-use properties depending on the considered application (e.g. paints or adhesives). This has been achieved with a designed glass temperature profile and maximum final conversion necessary to maximize production and minimize residual volatile organic compounds.

The non-dominated solutions (Pareto set) were obtained by an evolutionary algorithm developed and tested for this purpose. This set of solution give a wide range of operational options necessary for the improvement of the process. Pareto solutions were ranked by using MAUT strategy. This approach based on the decision maker experience and preferences, leads to the unique solution that was experimentally implemented. Good agreement was achieved between experimental and simulated data. The results showed the efficiency of the multicriteria strategy to produce polymer products with specific end-use properties.

References

B. Benyahia, M. A. Latifi, C. Fonteix, F. Pla, S. Nacef, 2009, Emulsion copolymerization of styrene and butyl acrylate in the presence of a chain transfer agent. Part 1: Modeling and experimentation of batch and fed-batch processes, Chemical Engineering Science, doi/10:1016/j. ces. 2009. 09.036.

C. Fonteix, S. Massebeuf, F. Pla, 2004, Nandor Kiss L, Multicriteria optimization of an emulsion polymerization process, Europeen Journal of Operational Research, 153, 350-359.

S. Garg, S. K. Gupta, 1999, Multiobjective optimization of a free radical bulk polymerization reactor using gentic algorithm, Macromolecular Theory Simulation, 8, 46-53.

R. R. Gupta, S. K. Gupta, 1999, Multiobjective optimization of an industrial Nylon-6 semi-batch reactor system using genetic algorithm, Journal of Applied Polymer Science, 73, 729-739.

S. K. Kim, O. Song, 2009, A MAUT approach for selecting a dismantling scenario for the thermal column in KRR-1, Annals of Nuclear Energy, 36, 145-150.

K. Mitra, S. Majundar, S. Raha, 2004, Multiobjective optimization of a semibatch epoxy polymerization process using the eliptic genetic algorithm, Industrial Engineering and Chemical Research, 43, 6055-6063.

D. Sakar, S. Rohani, A; Jutan, 2007, Multiobjective optimization of semibatch reactive crystallization processes, AIChE J., 53, 5, 1164-1174.

C. M. Silva, E.C. Biscaia, 2003 , Genetic algorithm development for multi-objective optimization of batch free-radical polymerization reactors, Computer Chemical Engineering, 27, 1329-1344.

State Variable Reconstruction From A Reduced Set Of Measurements

Erik C. Herrera-Hernández[a*] and Ramiro Rico-Martínez[a]

[a]*Instituto Tecnológico de Celaya, Departamento de Ingeniería Química.*
Avenida Tecnológico S/N, C.P. 38010, Celaya, Gto. México, cesarerik@gmail.com

Abstract

A methodology for reconstructing low-dimensional dynamical systems using a combination of artificial neural networks (ANNs) and the shooting method (SM), for solving differential equations with mixed boundary conditions, is presented. Previously, the reconstruction of dynamical systems based in experimental results has been addressed using several approaches that rely in the availability of all the state variables, the use of high-order polynomial equations to relate variables in order to obtain a set of Ordinary Differential Equations (ODEs), and/or the use of numerical derivatives as surrogate variables. These approaches may exhibit undesirable numerical sensitivity, and a difficult physical interpretation of the reconstructed variables. The methodology presented here bypasses the use of polynomial approximations and does not require of time-series of all state variables.

Keywords: Dynamical Systems Reconstruction, Artificial Neural Networks, Shooting Method

1. Introduction

Understanding the global behavior of a dynamical system often relies on the reconstruction of its states from a reduced set of time-series. Such analysis is assisted by a procedure that involves finding a closed model, in the form or a set of Ordinary Differential Equations (ODEs) [1-7]. These procedures address simultaneously the approximation of the complex behavior and the phase-space reconstruction of unmeasured states. The approximation capabilities are based in classical fitting strategies and statistical analysis [8,9]. Reconstruction of states based on Takens theorem is achieved through the use of derivative, lagged or proper orthogonal decomposition coordinates.

All of these methodologies usually ignore the parametric dependence of the system on operating parameters, and fail to address the physical meaning of reconstructed variables. A notable exception comes from the efforts involving ANNs (e.g [10]). Addressing the physical meaning is a nontrivial aspect of phase-space reconstruction, unless one relies in coordinates with a physical meaning in themselves such as derivative coordinates. However, even the use of derivative coordinates places a significant challenge related to the consistency [11,12] in the presence of noisy measurements, and the dynamical system evolution as the operating parameters are varied. In this paper, we address these challenges by introducing, during the training of an ANN approximator, based on derivative coordinates, the use of the shooting method as the means to achieve consistency with the original observed time-series. The method is illustrated using a second-order ODE with periodic behavior over a window of an operating parameter, and relying in measurements from the non-derivative state only.

2. Background

From measurements of a single state variable, trajectories of a conjugate set of complementary variables are constructed as initial target, for example, successive derivatives of the measured variables. An ANN approximates the underlying set of ODEs that describes the time evolution contained in the trajectories, while the SM is applied in order to retain consistency among the complementary states. The procedure is stopped when the ANN predictions have the same qualitative behavior as the experimental trajectories for a set of reference parameter values.

The proposed methodology is illustrated using the van der Pol system [13], a second order ODE with periodic behavior.

$$\frac{d^2x}{dt^2} + \gamma\left(x^2 - \delta\right)\frac{dx}{dt} + \omega\,x = 0 \tag{1}$$

Which can be easily transformed in a set of two first order ODEs as follows:

$$\frac{dx_1}{dt} = x_2 \tag{2}$$

$$\frac{dx_2}{dt} = \gamma\left(\delta - x_1^2\right)x_2 + \omega\,x_1 \tag{3}$$

Where the parameters δ and ω were fixed at 4 and 1, respectively. γ is used as operating parameter and different values were assigned in order to capture the parametric dynamical response of the system.

It is assumed that only measurements of one of the state variables (x_1) are available. The second state, nominally the time-derivative of the first state (x_2), is reconstructed using the algorithm. Several trajectories were obtained for different values of γ in order to have a representative training set. The ANN training is started by approximating the unmeasured derivative using finite differences; forward for the first point, backward for the last point, and central for all points in between.

The ANN was firstly trained for 15 conjugated gradient cycles, to ensure a reasonable first approximation, followed by application of the SM correction every 10 conjugated gradient cycles until convergence. Training was performed using least squares by minimizing the usual energy function:

$$E(\Theta, W) = 0.5 \sum_{j=1}^{2} \sum_{i=1}^{P} \left(x_j^i - x_j^{*i}\right)^2 \tag{4}$$

Here the vector x_j^i involves both the measured variable and the reconstructed variable, and x_j^{*i} represents the ANN prediction, for all trajectories.

The role of the SM consists in seeking the trajectory of the unmeasured states that is consistent with the ANN prediction and the measured time state, this means that for each segment of the measured variable one solves a mixed boundary value problem "guessing" the value of the unmeasured state that is consistent with the ANN predictions and the observed trajectory of the measured state, as shown in the next equation.

$$F = x_1(t_f) - x_1^*\left(t_f, x_1(t_i,), x_2(t_i), \Theta, W\right) \tag{5}$$

where F is the target function used by the SM which must be satisfied within a prefixed tolerance, ε, by using the measurement $x_1(t_i)$, the guessed value $x_2(t_i)$, the time step between consecutive points in a trajectory, and the ANN parameters Θ and W. Once F $F \approx \varepsilon$, the x_2 values are used as states for the ANN training.

Convergence of the procedure is declared by comparing the phase-space behavior for parameter values not included in the training set. Posing the reconstruction of the unmeasured variables as a mixed boundary problem may prove useful in assigning a meaning to the reconstructed trajectories beyond the purely statistical significance.

2.1. Noisy data

Noise is ubiquitous in experimental dynamical systems. Its presence complicates the reconstruction process, affecting numerical derivatives estimation and adding uncertainty to "embedded" dimension estimation.

In order to test the proposed algorithm, white noise was added to the time-series from the van der Pol system, to mimic the presence of a noisy measurement of the variable x_1, according to the follwing equation:

$$x_1 = x_1 + \eta; \ \eta = k(2\alpha - 1) \cdots \alpha \in [0,1] \tag{6}$$

Where α is a random number, η is the white Gaussian noise (with mean zero and standard deviation one) and k is the magnitude of the noise (approximately 1.5 percent of the active range for x_1).

3. Results and discussion

The reconstruction process involves three different stages: the first is training of the ANN with a numerical derivative (ND) approximation of x_2, without the application of the SM. The second stage is the application of the SM in order to enforce consistency of the trajectories of the unknown reconstructed variable. Finally, the third involves further training of the ANN to refine the prediction of the x_2 trajectories. The goal in the first and in the third stages is to take advantage of ANN capabilities to learn the dynamic behavior of any system, while the second stage has the task of preserving consistency of the reconstructed trajectory by solving a nonlinear algebraic equation at each point along its way.

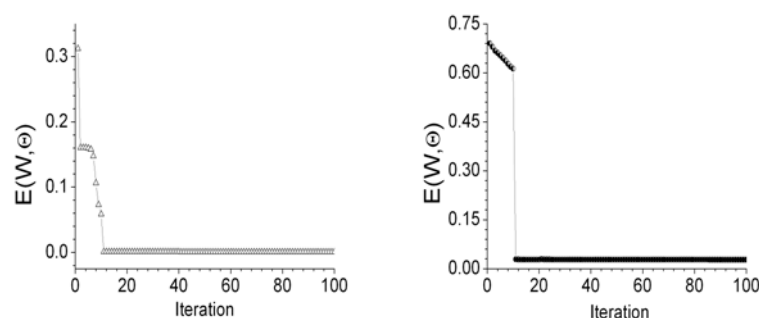The energy function change during training is plotted in Figure1.

Figure 1.Energy function evolution for noise-free data (left) and for noisy data (right)

The energy function evolution for the case of noise-free and noisy data has almost the same behavior: an initial abrupt descent is observed after a few iterations. The apparent discontinuity is due to the application of the SM.
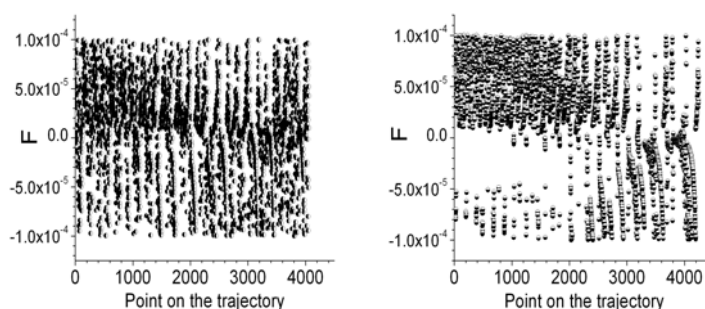

Figure 2. Convergence error distribution for the training data using the SM. Noise-free data (left) and noisy data (right).

The error distribution for the SM convergence is shown in Figure 2. For data without noise there is a seemingly random error distribution; while for noisy data the convergence has a very discernible pattern. This pattern is related to the number of points used in each trajectory and to the number of trajectories. The same number of data points was used for each trajectory seeking to avoid overweighting a specific behavior during the reconstruction process; However, at larger γ values, the stiffness of the periodic trajectory leads to uneven ANN prediction accuracy along a portion of the trajectory, and thus to a certain level of inconsistency in the predictions, that the SM method seeks to compensate. This fact is further appreciated by looking at the predicted trajectories, Figures 3 and 4. The combined SM-ANN methodology improved prediction is particularly relevant in the prediction of the sharp edges of the attractor.

Figure 3. Comparison between reconstructed (ANN-SM), original, and approximated trajectories using numerical derivative for the ANN training (ND-ANN) for γ = 1.75 and noise-free data.

For noisy data, the contribution of the SM to improve the ANN predictions is even more dramatic. While for the SM-ANN the predictions are qualitatively correct, for the case in which the unknown variable is approximated by numerical derivatives without enforcing consistency (ND-ANN), the ANN fails to predict the presence of a periodic trajectory.

Furthermore, the SM methodology also improves, in an indirect way, the prediction of the period of the oscillations.

The incorporation of the SM method to the reconstruction process retains, in general, the approximation capabilities of the ANN methodology while improving the consistency of the inferred trajectory with the original data.
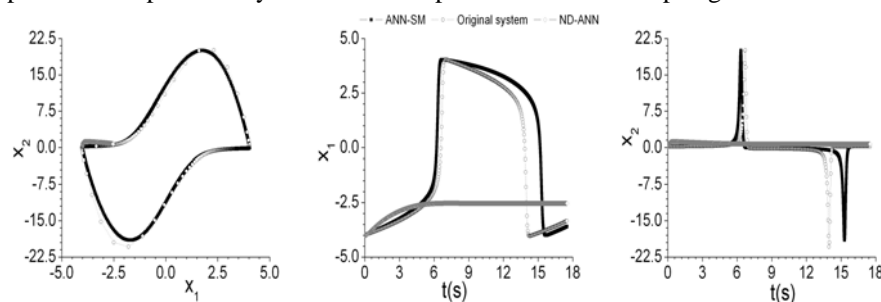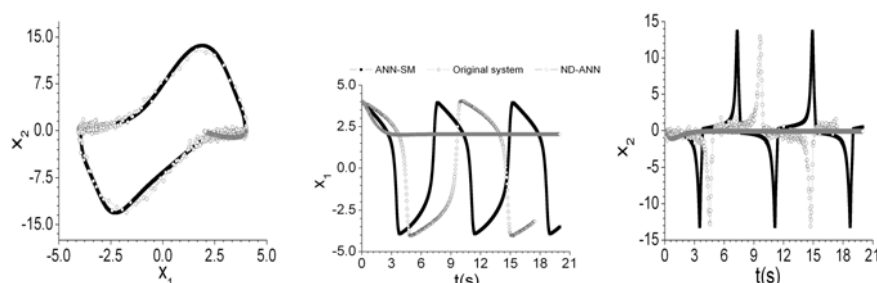

Figure 4. Comparison between reconstructed (ANN-SM), original, and approximated trajectories using numerical derivative for the ANN training (ND-ANN) for $\gamma = 1.0$ and noisy data.

4. Summary and conclusions

The proposed reconstruction/approximation methodology for nonlinear dynamical systems can be applied for the reconstruction of unmeasured states. The methodology uses derivative coordinates as the basis to obtain a set of conjugate states, potentially allowing for a physical interpretation of the reconstructed states to be sought.

The application of the SM under the ANN training framework leads to improved consistency of the predicted trajectories, and may affect indirectly the prediction of other quantities, as the period of oscillation in the illustration presented here. The coupling of the SM with the ANN training may be manipulated to enforce higher accuracy during the resolution of the equivalent mixed boundary value problem, thus providing a mechanism to control the degree of consistency sought, at the cost of a higher computational expense.

Additionally, the ANN capability to include an operating parameter as input during training, leads to a description that includes the dependence of the system with respect to such a parameter. For the van der Pol system, the constructed approximator is capable to correctly infer the presence of a Hopf bifurcation.

We are currently extending these results to higher dimensional systems, where the boundary value problem to be solved involves more than one equation.

References

[1] M. J. Bunner and M. Popp and Th. Meyer and A. Kittel and J. Parisi, 1997, Recovery of the time-evolution equation of time-delay systems from time series, Physical Review E, 56(5), 5083-5089

[2] J. Cremers and A. Hubler, 1987, Construction of differential equations from experimental data, Chemical Science, 42, 797-802

[3] B. P. Bezruchko and A. S. Karavaev and V. I. Ponomarenko and M. D. Prokhorov, 2001, Reconstruction of time-delay systems from chaotic time series, Physical Review E, 64, 056216-056221

[4] F. Lu, D. Xu, and G. Wen, 2004, Estimation of initial conditions and parameters of a chaotic evolution process from a short time series, Chaos, 14, 1050-1055

[5] T. Eisenhammer and A. Hubler and N. Packard and J. A. S. Kelso, 1991, Modeling experimental time series with ordinary differential equations, Biological Cybernetics, 65, 107-112

[6] P. Perona and A. Porporato and L. Ridolfi, 2000, On the trajectory method for the reconstruction of differential equations from time series, Nonlinear Dynamics, vol. 23, 13-33

[7] S. Tronci and M. Giona and R. Baratti, 2003, Reconstruction of chaotic time series by neural models: a case study, Neurocomputing, 55, 581-591

[8] H. Voss and J. Kurths, 1999, Reconstruction of nonlinear time-delayed feedback models from optical data, Chaos, Solitons and Fractals, vol. 10, 805-809

[9] A. Sitz and U. Schwarz and J. Kurths and H. U. Voss, Estimation of parameters and unobserved components for nonlinear systems from noisy time series, Physical Review E, 2002, 66, 01621.1-0162210.9

[10] R. Rico-Martínez, J. S. Anderson and I. G. Kevrekidis, 1994, Continuous-time nonlinear signal processing: a neural network based approach for gray box identification, IEEE Workshop on Neural Networks Signal Processing

[11] R. Rico-Martinez, J. S. Anderson, and I. G. Kevrekidis, 1996, Self-consistency in neural network-based NLPC analysis with applications to time-series analysis, Computers and Chemical Engineering, 20, Suppl, S751-S756

[12] I. Steinwart and M. Anghel, 2009, Consistency of support vector machines for forecasting the evolution of an unknown ergodic dynamical system from observations with unknown noise, The Annals of Statistics, vol. 27, 841-856

[13] B. Appelb, 2006, Existence of multiple cycles in a van der Pol system with hysteresis in the inductance, Journal of Physics: Conference Series, 55, 1-11

Dynamic optimization of an Intensive Energetically Integrated Large-Scale Process

Mariela A. Rodriguez, J. Ignacio Laiglecia, Patricia M. Hoch, M. Soledad Diaz

Planta Piloto de Ingeniería Química, PLAPIQUI, Universidad Nacional del Sur, CONICET. Camino La Carrindanga km 7, 8000 Bahía Blanca, ARGENTINA
sdiaz@plapiqui.edu.ar

Abstract

In this work we propose a first principles dynamic optimization model for an intensive energetically integrated process, a natural gas processing plant, within a simultaneous dynamic optimization framework. We have developed rigorous models, including thermodynamics with a cubic equation of state, for separation tanks, distillation columns, turboexpanders and cryogenic countercurrent heat exchangers with partial condensation. The resulting partial differential algebraic equation system is transformed into an ordinary differential algebraic equation one (DAE) by applying the method of Lines for the spatial coordinate. The high integration between process units as well as path constraints have been efficiently handled by a simultaneous dynamic optimization approach in which the DAE is transformed into a large nonlinear programming problem through orthogonal collocation over finite elements in time and solved with an interior point algorithm. In the case study, we maximize ethane recovery under a ramp change in natural gas feed flowrate. The model provides temporal and spatial profiles of controlled and manipulated variables that are in good agreement with plant data.

Keywords: Dynamic Optimization, Simultaneous Approach, Heat exchanger with phase change

1. Introduction

Dynamic optimization of entire plants has proven to be of great importance for properly studying process operation and control. Recent advances in dynamic optimization algorithms have paved the way to entire plant optimization, being the simultaneous approach (Biegler and Zavala, 2008) the most appropriate one to address highly integrated processes with numerous path constraints. Natural gas processing plants are examples of highly integrated cryogenic processes. They provide ethane as raw material for olefin plants. Ethane yield must be high, while minimizing energy consumption and complying with environmental regulations related to carbon dioxide emissions and maximum carbon dioxide content in the residual gas injected to the pipeline. Mandler (2000) presents Air Products dynamic modeling efforts since 1990 for analysis and control of cryogenic liquified gas plants (LNG). Dynamic optimization models have been proposed for cryogenic demethanizing columns (Diaz et al., 2003; Raghunathan et al., 2004) and cryogenic heat exchangers (Rodriguez and Diaz, 2007). Vinson (2006) has presented recent advances in air cryogenic separation.

In this work, the dynamic optimization of the entire cryogenic sector of a natural gas plant, which includes separation tanks, turboexpanders, distillation columns and countercurrent shell and tube heat exchangers with partial phase change, has been addressed. The distributed parameter model has been spatially discretized by the Method of Lines (Schiesser, 1991) and the optimization problem subject to the DAE

system has been solved through a simultaneous dynamic optimization approach (Biegler and Zavala, 2008) in which the resulting Nonlinear Programming (NLP) problem is solved with an Interior Point method (Waechter and Biegler, 2006). Optimal profiles have been obtained for main operating variables to achieve an enhanced product recovery under a ramp change. Numerical results have been compared to available plant data.

2. Case study

The cryogenic sector is the most important part in a natural gas processing plant, being the core of the turboexpansion process. As it is shown in Fig. 1, part of the feed gas is used for heat exchanging with the demethanizer top product in cryogenic heat exchangers (HE1 and HE2), while the rest of it is heat integrated with the distillation column reboilers.
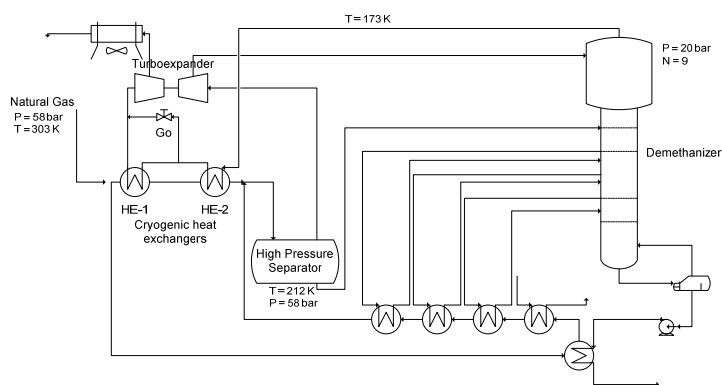
Figure 1 Basic turboexpansion process

After heat integration, both streams are mixed and sent to a high pressure separator (HPS). The vapor stream is expanded through a turboexpander in order to achieve the low temperatures required for demethanization. The liquid stream from the HPS enters the demethanizer in its lowest feed point. The top product has mainly methane and nitrogen, while higher hydrocarbons are obtained in the bottom product. Carbon dioxide distributes between top and bottom streams. It is necessary to verify that the operating conditions of the columns are such that carbon dioxide does not solidify in the top stages. Top product, referred to as residual gas, is used to cool the feed gas and it is then recompressed to the pipeline pressure, and distributed for sale. The bottom product from the demethanizer can be further processed to obtain ethane, propane, butane and gasolines.

3. Mathematical modeling

In previous work, we have developed first principles dynamic models for countercurrent heat exchangers with partial phase change (Rodriguez and Diaz, 2006, 2007) and cryogenic distillation columns (Diaz et al., 2003; Raghunathan et al., 2004). In this work, we have formulated horizontal vessel dynamic models and static models for turboexpanders and we have integrated process unit models within a simultaneous dynamic optimization framework. A brief description of main features for process unit models is given below.

Dynamic Optimization of an Intensive Energetically Integrated Large-Scale Process

3.1 Cryogenic heat exchangers with phase change

Main simplifying assumptions for countercurrent heat exchangers with partial phase change (HE2, in Fig. 1) are that we assume thermodynamic equilibrium between the vapor and liquid phases, but they can have different velocities, and one dimensional flux. Thermodynamic predictions are carried out with the Soave, Redlich, Kwong cubic equation of state (Soave, 1972) and analytical derivatives of thermodynamic functions (compressibility factor, fugacity coefficient, residual enthalpy) with respect to temperature, pressure and compositions are included in the model. To transform the resulting partial differential algebraic equation system describing this distributed parameters problem into a DAE, we have applied the Method of Lines (Schiesser, 1991), using backward finite differences. In this system, we have selected the residual gas flowrate through the bypass valve between HE1 and HE2 as an optimization variable (Go in Fig. 1), as it is used to achieve a desired outlet natural gas temperature to the high pressure separator. The heat exchanger where partial natural gas condensation takes place (HE2) has been modeled with 6 cells, while the heat exchanger where only sensible heat is exchanged (HE1) is modeled with 10 cells. A detailed description for the cryogenic system model is given in Rodriguez and Diaz (2007) and Rodriguez (2009).

3.2 High pressure separator

The model includes an overall dynamic mass balance and geometric equations relating liquid content in the tank to liquid height and liquid flowrate as function of pressure drop over the liquid stream valve. Detailed equations are presented in Rodriguez and Diaz (2007) and Rodriguez (2009).

3.3 Turboexpander

The turboexpander is the main process unit in cryogenic natural gas processing plants, as it provides the low temperatures required for methane/ethane separation. Due to its fast dynamics, it can be represented with a static model. A pressure-temperature diagram is shown in Fig. 2.
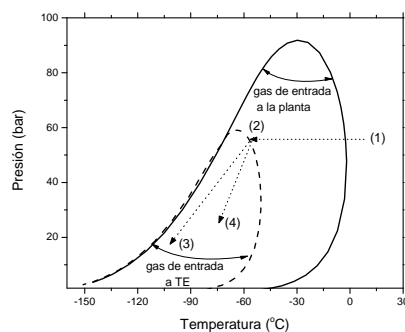
Figure 2. P-T diagram for natural gas in turboexpansion process

Natural gas enters the turboexpander at point 2 and, as it undergoes expansion, it moves from 2 to 3, which represents the turboexpander outlet conditions. It has been modeled as an isentropic expansion and corrected by the expander efficiency. Residual entropy calculations with SRK equation of state have been included. The procedure proposed in GPSA Engineering Data Book (2004) for turboexpander calculation has been implemented and the equation oriented approach efficiently avoids the iterative routine. Details can be found in Rodriguez (2009).

3.4 Demethanizing column

The demethanizer model includes dynamic energy and component mass balances at each stage, equilibrium calculations with SRK equation of state and hydraulic calculations, rendering an index one model. It also includes solubility calculations for carbon dioxide with SRK at each stage, as the low temperatures in the upper section of the column can produce carbon dioxide precipitation. Path constraints on CO_2 fugacities have been included to avoid precipitation as:

$$\overline{f_{i,CO2}^{V}} \leq f_{i,CO2}^{S}, \tag{1}$$

which can be calculated with low computational effort in a simultaneous approach as:

$$\overline{f_{i,CO2}^{V}} = y_{i,CO2} P_i \overline{\phi_{i,CO2}^{V}} \tag{2}$$

Further details can be found in Diaz et al (2003), Raghunathan et al. (2004).

4. Optimization model and algorithm

We have formulated the following dynamic optimization problem:

$$\min \int_0^{tf} \left(\eta_{ethane} - \eta_{SP} \right)^2 dt$$

st DAE system + Solubility constraints (1)-(2) $\tag{3}$

$$15 \leq P_{top}(t) \leq 22 (bar) \; ; \; 0 \leq X_{bypass}(t) \leq 1 \; ; \; 0 \leq x_{methane-bottoms}(t) \leq 0.005$$

$$y_L \leq y \leq y_U \; ; \; z_L \leq z \leq z_U \; ; \; z(t=0) = z_0$$

where the objective function is the minimization of the offset between ethane recovery and a target value; optimization variables are demethanizer top pressure (Ptop) and flowrate fraction derived through the bypass valve in cryogenic heat exchangers (Xbypass). The dynamic optimization problem has been formulated in Fortran 90 environment within a simultaneous approach in which both control and state variables are approximated by piecewise polynomials and the DAE system is discretized by orthogonal collocation over finite elements (Cervantes abd Biegler, 1998; Biegler and Zavala, 2008). The large scale nonlinear programming model is solved with an Interior Point method with reduced SQP techniques within program IPOPT (Waechter and Biegler, 2006).

5. Numerical results

To validate the optimization model, we have compared model results for a ramp perturbation with available plant data from an actual turboexpansion plant in operation (Cia. Mega SA). The objective is to analyze model performance to represent the type and speed of response of main plant variables under a 3.5% increase of natural gas feed flowrate ramp perturbation, corresponding to a current plant situation, as shown in Figs. 3 and 4. Due to confidentiality, it is not possible to report actual values for data, but for analysis purposes the pressure scale was divided into 1 bar ticks and flowrates in 20 kmol/h ticks. Figure 3 shows profiles for feed gas flowrate and demethanizer top pressure; it can be seen that at time=128 min a ramp type perturbation is introduced, reaching a final value 3.5% higher than the initial one. Figure 4 shows the response of a controlled variable, the outlet temperature of the cryogenic heat exchangers, which is controlled through opening the bypass valve between HE1 and HE2. The top temperature of the demethanizer column is also shown.

To compare plant data with model results, the dynamic optimization problem posed as Eqn. (1) has been solved for a natural feed flow rate perturbation represented as a ramp that approximates very well the real plant perturbation (Figure 5, solid line).

Dynamic Optimization of an Intensive Energetically Integrated Large-Scale Process
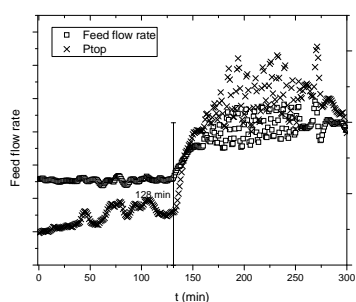
Figure 3 Natural gas feed flowrate and top pressure of demethanizer (plant data)
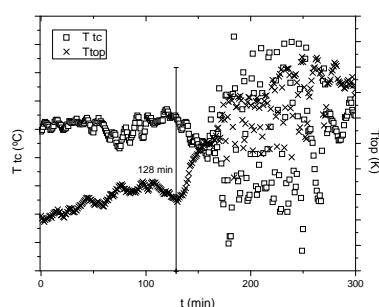
Figure 4 Cryogenic HE temperature and demethanizer top temperature (plant data)
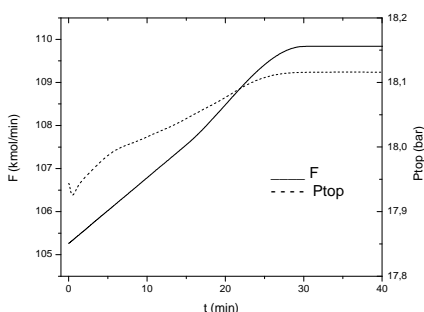
Figure 5. Feed flowrate to each train of cryogenic heat exchangers (solid line) and top pressure profile of demethanizer
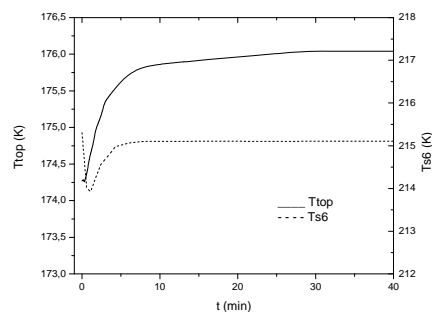
Figure 6. Demethanizer top temperature (solid line) and outlet temperature of cryogenic heat exchangers
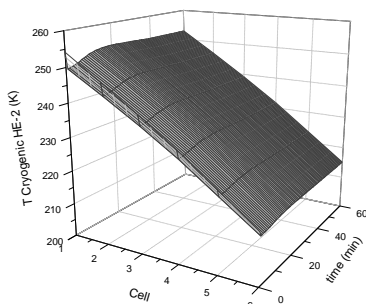
Figure 7: Condensate flowrate profile in HE2
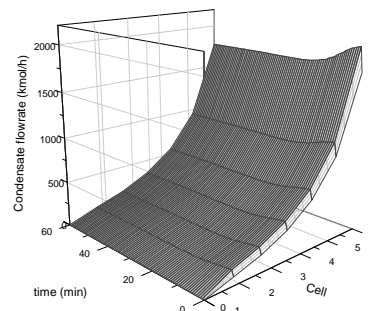
Figure 8: Temperature profile in HE2

The discretization has been carried out with 20 finite elements with two collocation points, rendering an NLP with 39550 variables and 39510 equality constraints. Numerical results are shown in Figures 5 to 10. The main variables show a behavior comparable to plant data. Figures 5 and 6 show the responses of top pressure and temperature of the demethanizer column, as well as the temperature of the partially condensed natural gas entering the HPS (Ts6). They do not show delay and follow the shape of the feed gas profile, reaching the new steady state at approximately 30 minutes. Figure 6 shows that the temperature of the partially condensed vapor from the

cryogenic heat exchangers (Ts6) shows an initial decrease due to the instantaneous shut-down of the bypass control valve (in the initial steady state 10% of the stream is deviated to the first cryogenic heat exchanger). The model determines the immediate shut-down of the valve, emulating the behavior of the control system regulating the cold tank temperature. Then, the increasing demethanizer top flow rate to a higher thermal level gives rise to a new steady state, where the temperature is 2K higher than the initial one(215.10K). The behavior is analogous to plant behavior reported in Fig. 4, starting from T=128 min. Figures 7 and 8 show temporal and spatial profiles of condensate flow rate and temperature for the cryogenic heat exchanger with phase change, respectively. Total CPU time has been 864 s. The objective function, ethane recovery has changed from an initial value of 73.64% to 75.76%.

Conclusions

We have addressed dynamic optimization of a highly energy integrated large-scale process through the formulation of first principles models for process units within a simultaneous dynamic optimization approach. Even though the resulting NLP problem has a large number of equations, the Interior point method with reduced SQP techniques, as well as appropriate handling of the Jacobian structure within program IPOPT (Waechter and Biegler, 2006; https://projects.coin-or.org/Ipopt) allows an efficient resolution. Numerical results have been favorably compared to plant data of an actual plant in operation.

6. Acknowledgements

References

Biegler,L.T., V.M. Zavala (2008), Large-scale nonlinear programming using IPOPT: An integrating framework for enterprise-wide dynamic optimization. Computers and Chemical Engineering 33 (2009) 575–582

Cervantes A. M., L. T. Biegler (1998). Large-scale DAE optimization using a simultaneous NLP formulation. AIChE Journal,44,1038-1050.

Diaz, S., S. Tonelli, A. Bandoni, L.T. Biegler (2003), "Dynamic optimization for switching between steady states in cryogenic plants", Found Comp Aided Process Oper 4, 601-604.

Mandler J.A. (2000). "Modelling for control analysis and design in complex industrial separation and liquefaction processes", J. Process Control, 10, 2, 167-175.

Raghunathan, A., M.S. Diaz, L.T. Biegler (2004). An MPEC Formulation for Dynamic Optimization of Distillation Operations", Computers & Chemical Engineering, 28, 2037-2052.

Rodriguez, M., (2009) "Dynamic Modeling And Optimization of Cryogenic Separation Processes", PhD Dissertation, Universidad Nacional del Sur, Bahia Blanca, Argentina.

Rodríguez, M., M. S. Diaz, (2007), "Dynamic Modelling And Optimisation of Cryogenic Systems", Applied Thermal Engineering (ISSN 1359-4311), 27, 1182-1190.

Schiesser, W.E., (1991), "The Numerical Method of Lines", San Diego, CA: Academic Press

Soave G. Equil. Constants for a Modified Redlich-Kwong Eq. of State. Chem. Eng. Sci. 1972; 27: 1197-1203.

Vinson, D.R., (2006), Air separation control technology, Computers & Chemical Engineering, 30, 10-12, 1436-1446.

Waechter, A., Biegler, L.T. (2006), On the implementation of an interior-point filter-line search algorithm for large-scale nonlinear programming Research Report RC 23149, IBM T.J. Watson Research CenterYorktown, New-York.

Effective Decomposition Algorithm for Multistage Batch Plant Scheduling

Pedro M. Castro,[a] Iiro Harjunkoski,[b] Ignacio E. Grossmann[c]

[a]*Energy Systems Modeling and Optimization Unit, LNEG, 1649-038 Lisboa, Portugal*
[b]*ABB Corporate Research Center, Wallstadter Str. 59, 68526 Ladenburg, Germany*
[c]*Dept. Chemical Engineering, Carnegie Mellon University, Pittsburgh 15213, USA*

Abstract

This paper presents a new algorithm for the scheduling of batch plants with a large number of orders and sequence-dependent changeovers. Such problems are either intractable or yield poor solutions with full-space approaches. We use decomposition on the entire set of orders and derive the complete schedule in several iterations. The key idea is to allow for partial rescheduling without altering the main decisions in terms of unit assignments and sequencing, so that the complexity is kept at a manageable level. It has been implemented with a unit-specific continuous-time model and tested for different decomposition settings. The results show that a real-life 50-order, 17-unit, 6-stage problem can effectively be solved in roughly 6 minutes of computational time.

Keywords: Optimization, Continuous-time, Sequence-dependent changeovers.

1. Introduction

The vast literature in the scheduling area highlights the successful application of optimization approaches to an extensive variety of challenging problems. This important achievement comes from the remarkable advances in modeling techniques, algorithmic methods and computer hardware that have been made in the last two decades. However, there is still a significant gap between theory and practice. New academic developments are mostly tested on relatively small problems whereas real-world applications consist of hundreds of batches, dozens of equipments and long scheduling horizons. In order to make exact methods more attractive for real-world applications, efforts should be oriented towards the development of systematic techniques that allow maintaining the number of decisions at a reasonable level, even for large-scale problems. Although these techniques can no longer guarantee optimality, this may not be critical in practice due to: (i) very short time available to generate a solution; (ii) theoretical optimality can easily get lost due to the dynamic nature of industrial environments; (iii) implementing the schedule as such is often limited by the real process; (iv) only a subset of the actual scheduling goals are taken into account.

In this paper, we address the short-term scheduling of multistage batch plants with sequence-dependent changeovers. A complex algorithm is proposed that can be parameterized for the fast and efficient solution of problems of varying size. The key idea is essentially the one proposed by Röslof et al. (2001) and further explored by Méndez and Cerdá (2003). The novel aspect is the use of a multiple time grid continuous-time model, where the solution process itself overcomes the difficulty in specifying the number of event points that will be different from grid to grid, instead of a model based on sequencing variables. The algorithm is validated through the solution of example problems of moderate size for which the optimal solution is known, in order to measure the optimality gap and difference in total computational effort.

2. Problem Definition

Given a multistage, multiproduct plant with processing stages $k \in K$, product orders $i \in I$ and units $m \in M$, the goal is to determine the assignment of orders to units and the sequence of orders in each unit so as to minimize the makespan. Orders may be subject to release (r_i) and due dates (d_i) that are enforced as hard constraints. The processing times are unit dependent $(p_{i,m})$ and the duration of sequence-dependent changeovers are given by $cl_{i,i',m}$. A particular unit can handle all orders belonging to set I_m and is allocated to a single stage, with set M_k including those belonging to stage k. Set I_k gives the orders that are handled in stage k. Unlimited intermediate storage and wait policies (UIS/UW) are assumed, while transfer times between units are negligible. The process representation is given in Fig. 1 in the form of a Resource-Task Network.
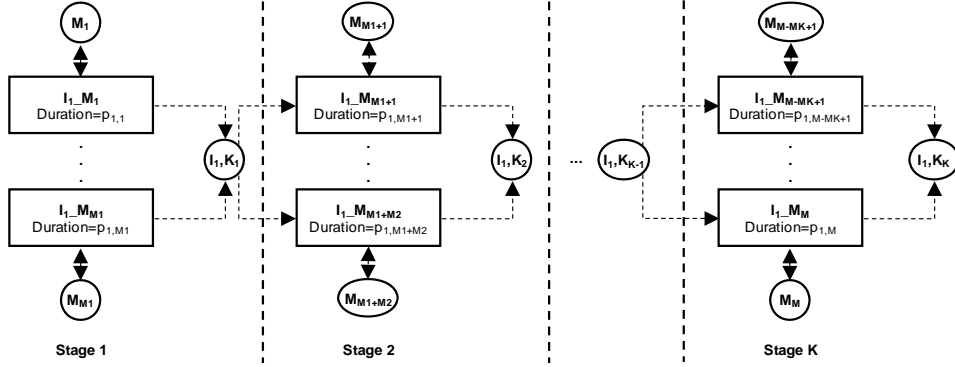
Figure 1. Schematic of a multistage multiproduct plant

3. Key Idea of Decomposition Approach

Finding a schedule for a multistage plant with parallel units involves two decision levels: (i) assigning orders to units; (ii) sequencing orders on every unit. We follow this hierarchy to set different degrees of freedom for the orders. Those being considered for the first time can be assigned to all possible units and take any position in the sequence. In contrast, previously scheduled orders have significantly less freedom. While the timing of events is allowed to change, orders cannot be reassigned to other units. Furthermore, their relative position in the sequence remains unchanged.

Let *NOS* be the number of orders to schedule per iteration. The higher the value, the fewer the iterations (set *J*) and the larger the feasible region up to that of the full-space model (*NOS=|I|*) so better solutions are likely to result. However, the resulting mathematical problems also become more complex and may become intractable, so there is an obvious tradeoff.

The second decision concerns order-iteration assignments. We will use the increasing slack times heuristic (MST) that prioritizes orders with a smaller time span (Eq. 1).

$$span_i^{MST} = d_i - r_i - \sum_{\substack{k \in K}} \min_{\substack{m \in M_i \\ m \in M_k}} p_{i,m} \quad \forall i \in I \tag{1}$$

In the following we heavily rely on the concepts of multiple time grids and combined process and changeover tasks (Castro et al., 2006). More specifically, the execution of combined task (i,i',m) comprises the processing time of order i plus the required changeover from i to i' so that unit m is ready for order i' to immediately follow. Also, bear in mind that postulating a single time slot per task is enough to ensure generality.

Effective Decomposition Algorithm for Multistage Batch Plant Scheduling
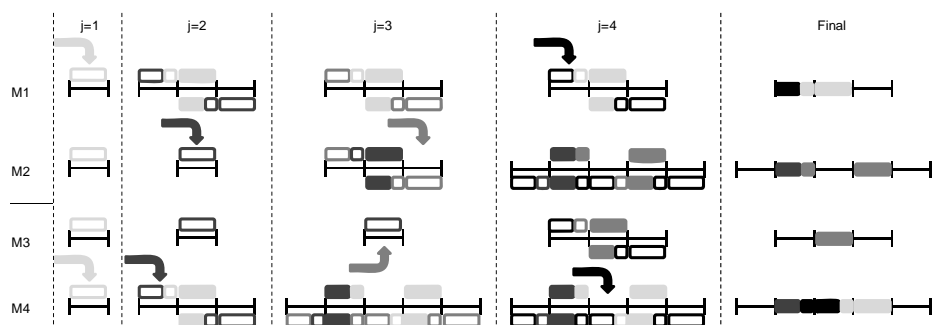
Figure 2. Illustration of decomposition algorithm for *NOS*=1.

The scheduling algorithm is illustrated in Fig.2 for *NOS*=1. In the first iteration (*j*=1), all units get a single time slot to allow for the execution of tasks of type (*i,i,m*). From *j*=1 to 2, idle units remain with a single time slot, whereas the remaining (M1 and M4) receive two more. This makes it possible to produce the new dark-grey order before or after the already assigned light-grey order. More specifically, the latter can only be executed in slot #2 with either combined task (light-grey, light-grey) or (light-grey, dark-grey). In general, we need to postulate *NOS*+1 additional slots for each previously assigned order. Depending on their relative positioning form the previous iteration ($pos_{i,m}$), such orders will be assigned to slot number $pos_{i,m} \times (NOS+1)$.

Things become increasingly more complicated as we go through iterations. For *j*=3 and M4, we need (dark-grey, light-grey)-A, (dark-grey, medium-grey); (light-grey, medium-grey) and (light-grey, light-grey)-B and, for the new order, (medium-grey, dark-grey)-C, (medium-grey, light-grey) and (medium-grey, medium-grey). It can be easily checked that all possible sequences with dark-grey before light-grey are accounted for. As an example (medium-grey)-(dark-grey)-(light-grey) is achieved with C in slot #1, A in slot #2 and B in slot #4. The procedure continues until the final schedule is obtained. It is important to highlight that there may be idle slots between orders or empty last slots and that no task extends past the duration of a slot. Although it looks like it, in M1 and M4 the first light-grey box accounts for the black-(light-grey) changeover.

4. Scheduling Algorithm

The proposed algorithm comprises two parts. In the first, constructive scheduling, the goal is to find a good initial solution, which is improved afterwards by performing a local search. In this rescheduling part, a couple of orders are released from the schedule to try to find better unit assignments or sequencing. It can be viewed as repeating the last iteration of the constructive step several times, for different order candidates.

Every iteration *j* starts with the selection of orders that are under consideration, I^{act}. As explained in section 3, the number of active time slots for unit *m* (T_m^{act}) is a function on the number of orders previously assigned to it as well as *NOS*. As illustrated in Fig. 2, any previous order can be assigned to a single time slot, determined by its position in the sequence. New orders can be assigned to any other slot, either before previous orders or after the last one. Such information is then used to generate sets $I_{m,t}$ (orders that in unit *m* can be executed in slot *t*) and $I_{i',m,t}$ (orders that can be assigned to slot *t* of unit *m* and be followed by order *i'*). Finally, the last slot is the sole element of T_m^{last}.

5. Mathematical Formulation

The underlying unit-specific continuous-time formulation is essentially CT4I proposed by Castro et al. (2006). The novelty is the introduction of slack variables $S^1_{t,m}$ to allow for due dates violation, an absolute necessity when decomposition algorithms are involved. These are penalized in the objective function (Eq. 2) through weight $\alpha=10$, which essentially minimizes the makespan, *MS*. 4-index binary variables $\overline{N}_{i,i',m,t}$ identify the execution of the combined tasks at time slot t, while continuous variables $C_{i,m,t}$ keep track of unit states ($C^0_{i,m}$ for the initial state). Timing variable $T_{t,m}$ gives the time of event point t in time grid m, while the transfer time of order i in stage k is $TD_{i,k}$.

$$\min MS + \sum_{m \in M_{|K|}} \sum_{t \in T^{active}_m} \alpha \cdot S^1_{t,m} \tag{2}$$

$$C_{i,m,t} = C^0_{i,m}\Big|_{t=1} + C_{i,m,t-1}\Big|_{t\neq1} + \sum_{\substack{i' \in I_{i,m,t-1} \\ i' \neq i}} \overline{N}_{i',i,m,t-1} - \sum_{\substack{i' \in I_m \\ i \in I_{i',m,t}}} \overline{N}_{i,i',m,t} \quad \forall i \in I^{act}, m \in M_i, t \in T^{act}_m \tag{3}$$

$$\sum_{i \in I_m} C^0_{i,m} \leq 1 \; \forall m \in M \tag{4}$$

$$T_{t+1,m}\Big|_{t \notin T^{last}_m} + MS\Big|_{t \in T^{last}_m} - T_{t,m} \geq \sum_{i' \in I_m} \sum_{i \in I_{i',m,t}} \overline{N}_{i,i',m,t} \times (p_{i,m} + cl_{i,i',m}) \; \forall m \in M, t \in T^{act}_m \tag{5}$$

$$T_{t,m} \geq \sum_{i' \in I_m} \sum_{i \in I_{i',m,t}} \overline{N}_{i,i',m,t} \times (r_i + \sum_{k' \in K, k'<k} \min_{m' \in M_{k'} \cap M_i} p_{i,m'}) \; \forall k \in K, m \in M_k, t \in T^{act}_m \tag{6}$$

$$T_{t,m} \leq \sum_{i' \in I_m} \sum_{i \in I_{i',m,t}} \overline{N}_{i,i',m,t} ub_{i,i',m} + S^1_{t,m} + H(1 - \sum_{i' \in I_m} \sum_{i \in I_{i',m,t}} \overline{N}_{i,i'm,t}) \; \forall m \in M, t \in T^{act}_m \tag{7}$$

$$TD_{i,k} \geq T_{t,m} + \sum_{i' \in I_m, i \in I_{i',m,t}} \overline{N}_{i,i',m,t} p_{i,m} - H(1 - \sum_{t' \in T^{active}_m, t' \geq t} \sum_{i' \in I_m, i \in I_{i',m,t'}} \overline{N}_{i,i',m,t'}) \tag{8}$$
$$\forall k \neq |K|, m \in M_k, t \in T^{act}_m, i \in I_{m,t}$$

$$TD_{i,k-1} \leq T_{t,m} + H(1 - \sum_{t' \in T^{active}_m, t' \leq t} \sum_{i' \in I_m, i \in I_{i'm,t'}} \overline{N}_{i,i',m,t'}) \; \forall k \neq 1, m \in M_k, t \in T^{act}_m, i \in I_{m,t} \tag{9}$$

$$TD_{i,k} = TD_{i,k-1} \; \forall k \in K, i \in I^{act}, i \notin I_k \tag{10}$$

$$\sum_{m \in M_k} \sum_{t \in T^{act}_m} \sum_{i' \in I_m, i \in I_{i',m,t}} \overline{N}_{i,i',m,t} = 1 \; \forall k \in K, i \in I^{act} \cap I_k \tag{11}$$

Equation 3 gives the resource balances over the equipment states. Notice in the third term on the right-hand side that combined tasks with a single order index (i,i,m) do not need to regenerate the equipment state since no order follows. Eq. 4 limits the initial states of units to a single order. The central timing constraint is given by Eq. 5. It ensures that the difference in time between two consecutive event points must be greater than the processing time of the order being executed plus the required changeover time for the following order. Eqs. 6-7 are the release and due date constraints, where slacks $S^1_{t,m}$ allow for violation of the due dates (note that $ub_{i,i',m}=f(d_i)$, see Castro and Novais, 2009). These are needed for all pairs (slot, unit) and not just for those penalized in the objective function (Eq. 2). The transfer time of order i in stage k (e.g. in hours) must be greater than its ending time in stage k and lower than its starting time in $k+1$ (Eqs. 8-9).

The transfer times of orders not involved in stage k are equal to those in the previous stage (Eq. 10). Finally, all orders need to be executed once on every stage (Eq. 11).

6. Computational Results

The performance of the scheduling algorithm is illustrated through the solution of ten example problems. P7-P13 are taken from Castro and co-workers (2006, 2009) and can be solved to global optimality by the full-space continuous-time formulation. Their purpose is to evaluate the quality of the solution returned by the algorithm. P16 is the challenging industrial problem, provided by a pharmaceutical batch plant, with P14-P15 considering the first 30/40 orders to measure the effect of problem size on computational effort. The algorithm and underlying models were implemented in GAMS 23.2 with CPLEX 12.1 as the MILP solver. The termination criteria were a relative optimality tolerance=10^{-6} and maximum computational time per iteration=3600 CPUs on the constructive part; 60 CPUs on the rescheduling part. The hardware was a HP laptop with an Intel Core2 Duo T9300 2.5 GHz processor running Windows Vista.

The solutions obtained are listed in Table 1 for a total of 30 trials resulting from different choices of parameter *NOS*. All generated problems from the smaller instances were solved to optimality, so failure to find the global optimal solution is entirely due to the decomposition strategy. Interestingly, there was not a single successful (resulting in an optimal solution) run and solutions with due date violations were observed in 19% of the cases. Increasing the value of *NOS* leads to an average increase in solution quality, which comes as no surprise since one is working closer to full-space mode (*NOS*=|*I*|).

The focus is really on difficult large-scale problems and on obtaining near optimal solutions fast. From the last columns, one can see that reasonably good solutions can be found in roughly 6 minutes of computational time. For *NOS*=2, P14 and P15 can still be solved in less than 1 hour but for P16 it is no longer possible to solve the problems generated in the last few iterations to optimality. As a consequence, solution quality degrades from *NOS*=2 to 3, contrary to what happens for most of the smaller instances.

The major strength of the proposed algorithm is the ability to address problems systematically for a wide variety of settings. In particular, one can switch to a continuous-time model with general precedence sequencing variables. While a similar performance was observed, we did found a better solution for the most challenging problem (*NOS*=3) featuring a makespan=47.722 h, see Fig. 3. It is also important to highlight that the full-space version of the model could only find a makespan=36.121 h for P14 before running out of memory at 51,900 CPUs, a value 16% higher than the best solution found by the decomposition algorithm in considerably less time.

7. Conclusions

This paper has addressed the scheduling of large-scale multistage batch plants with sequence-dependent changeovers. A new decomposition algorithm was proposed to construct the full schedule in several iterations. The complete set of orders is handled sequentially according to the increasing slack times heuristic. Newly considered orders are given total freedom in terms of unit assignments and sequencing, while those handled in previous iterations are allowed to change their starting times provided that their unit assignments and relative positions in the sequence are kept constant. By changing the number of orders tackled per iteration, the algorithm can effectively handle problems of different size in an efficient way and, hence, take the most of available computational resources. In particular, it successfully tacked an industrial scale problem in a few minutes of computational time.

Table 1. Computational Statistics (**Best Solution**, *Solution with Due Dates Violation*)

Problem	(I,M,K)	Optimum	Solution for NOS=			CPUs for NOS=		
			1	2	3	1	2	3
P7	(8,6,2)	542	591	595	**558**	3.89	2.26	4.10
P8	(8,6,2)	584	664	611	**587**	3.78	2.38	2.27
P9	(8,6,3)	915	*1355*	**981**	*1166*	3.93	2.34	13.0
P10	(8,6,3)	914	970	**938**	938	3.79	3.22	4.91
P11	(12,6,2)	233	261	**245**	261	5.86	4.01	12.9
P12	(8,8,4)	265	*374*	*507*	**275**	4.12	7.42	27.2
P13	(15,4,2)	273	314	328	**309**	7.54	5.50	21.1
P14	(30,17,6)	?	33.424	**31.018**	31.934	40.9	1607	20540
P15	(40,17,6)	?	42.056	**40.766**	42.151	99.7	3251	30409
P16	(50,17,6)	?	52.849	**48.929**	51.444	371	12690	29413
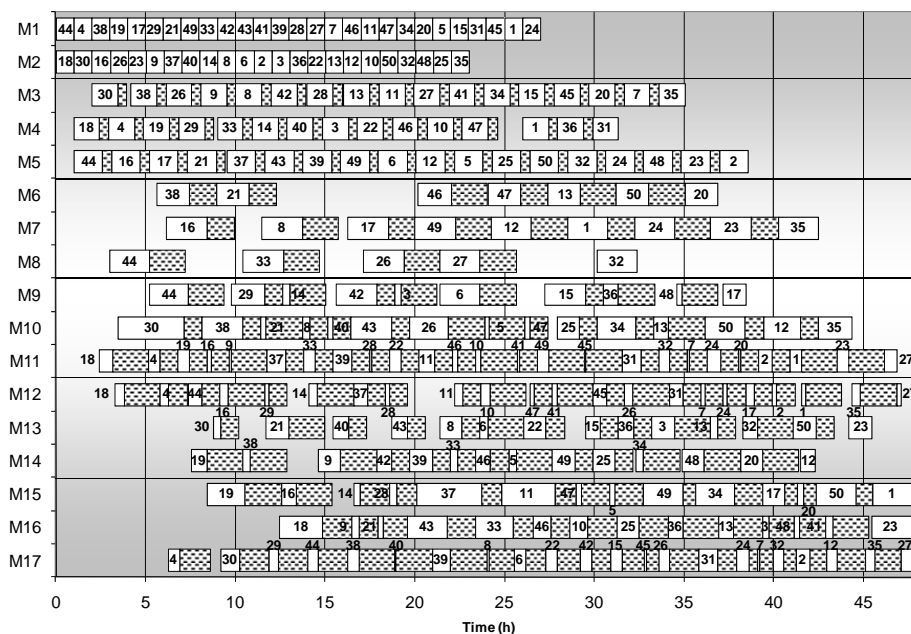
Figure 3. Best found solution for P16.

References

P. Castro , I. Grossmann, A. Novais, 2006, Two New Continuous-Time Models for the Scheduling of Multistage Batch Plants with Sequence-Dependent Changeovers. Ind. Eng. Chem. Res. 45, 6210.

P. Castro, A. Novais, 2009, Scheduling Multistage Batch Plants with Sequence-Dependent Changeovers. AIChE J. 55, 2122.

C. Méndez, J. Cerdá, 2003, Dynamic Scheduling in Multiproduct Batch Plants, Comput. Chem. Eng. 27, 1247.

J. Roslöf, I. Harjunkoski, J. Björkqvist, S. Karlsson, T. Westerlund, 2001, An MILP-based Reordering Algorithm for Complex Industrial Scheduling and Rescheduling. Comput. Chem. Eng. 25, 821.

A toolkit for efficiently generating Pareto sets in (bio)chemical multi-objective optimal control problems

Filip Logist,[a] Boris Houska,[b] Moritz Diehl,[b] Jan F. Van Impe[a]

^aBioTeC & OPTEC, Chemical Engineering Department, Katholieke Universiteit Leuven,
W. de Croylaan 46, 3001 Leuven, Belgium, {filip.logist,jan.vanimpe}@cit.kuleuven.be
^bSCD & OPTEC, Electrical Engineering Department, Katholieke Universiteit Leuven,
Kasteelpark Arenberg 10, 3001 Leuven, Belgium, {boris.houska,moritz.diehl}@esat.kuleuven.be

Abstract

Many practical (bio)chemical engineering problems involve the determination of optimal trajectories given multiple and conflicting objectives. These conflicting objectives typically give rise to a set of Pareto optimal solutions. To enhance real-time decision making efficient approaches are required for determining the Pareto set in a fast and accurate way. Hereto, the current paper integrates efficient multiple objective scalarisation strategies (e.g., Normal Boundary Intersection and Normalised Normal Constraint) with fast deterministic simultaneous approaches for dynamic optimisation (e.g., Multiple Shooting and Collocation). All techniques have been implemented as an add-on to the freely available automatic control and dynamic optimisation toolkit ACADO (www.acadotoolkit.org). ACADO Multi-Objective's features are discussed and its use is illustrated on different (bio)chemical examples.

Keywords: multi-objective optimisation, dynamic optimisation, open source, Pareto set

1. Introduction

In practical (bio)chemical optimal control problems multiple and conflicting objectives are often present, giving rise to a set of Pareto optimal solutions instead of one single solution [1]. The most often exploited approaches to generate this Pareto set are *(i)* Weighted Sum (WS) approaches in which for a grid of different weights optimal control problems are solved with deterministic optimisation routines, or *(ii)* stochastic genetic algorithms [2] in which a population of solutions is updated based on repeated cost computations in order to evolve gradually to the Pareto frontier. Unfortunately, both approaches exhibit certain restrictions. For the weighted sum it is known that an equal distribution of weights does not necessarily lead to an even spread along the Pareto front, and that points in a non-convex part of the Pareto front cannot be obtained [3]. Stochastic approaches, although quite successful over the years [4], *(i)* may become time consuming due to the repeated model simulations required, *(ii)* are less suited to incorporate constraints exactly, and *(iii)* are limited to rather low dimensional search spaces, which restricts the control discretisations to coarse approximations. Recent scalar multiple objective optimisation techniques as Normal Boundary Intersection (NBI) [5] and Normalised Normal Constraint (NNC) [6] have been found to mitigate the disadvantages of the Weighted Sum, while still allowing the exploitation of fast deterministic solvers. Therefore, the aim is to use NBI and NNC in deterministic direct optimal control approaches in order to efficiently solve multiple objective optimal control problems in the (bio)chemical industry. In addition, these approaches are implemented in the software toolkit ACADO Multi-Objective, which is freely available at www.acadotoolkit.org in order to facilitate real-time decision making.

2. Multi-objective optimal control problems

A general *multi-objective optimal control problem* (MOOCP) is defined as follows:

$$\min_{\mathbf{x}(\cdot),\mathbf{u}(\cdot),\mathbf{p},\xi_f} \{J_1,...,J_m\} \tag{1}$$

subject to:
$$\frac{d\mathbf{x}}{d\xi} = \mathbf{f}(\mathbf{x}(\xi),\mathbf{u}(\xi),\mathbf{p},\xi) \qquad \xi \in [0,\xi_f] \tag{2}$$

$$\mathbf{0} = \mathbf{b}_c(\mathbf{x}(0),\mathbf{x}(\xi_f),\mathbf{p}) \tag{3}$$

$$\mathbf{0} \geq \mathbf{c}_p(\mathbf{x}(\xi),\mathbf{u}(\xi),\mathbf{p},\xi) \tag{4}$$

$$\mathbf{0} \geq \mathbf{c}_t(\mathbf{x}(\xi_f),\mathbf{u}(\xi_f),\mathbf{p},\xi) \tag{5}$$

Here, $\mathbf{x}$ are the state variables, while $\mathbf{u}$ and $\mathbf{p}$ denote the time-varying and time-constant control variables, respectively. The vector $\mathbf{f}$ represents the dynamic system equations (on the interval $\xi \in [0,\xi_f]$) with initial and terminal boundary conditions given by the vector $\mathbf{b}_c$. The vectors $\mathbf{c}_p$ and $\mathbf{c}_t$ indicate respectively path and terminal inequality constraints on the states, the controls and/or the independent variable ξ. Each individual cost function can consist of Mayer and Lagrange terms.

$$J_i = h_i(\mathbf{x}(\xi_f),\mathbf{p},\xi) + \int_0^{\xi_t} g_i(\mathbf{x}(\xi),\mathbf{u}(\xi),\mathbf{p},\xi)d\xi \tag{6}$$

In multi-objective (MO) optimisation, typically no single optimal solution exists, but a set of *Pareto optimal* solutions. Broadly speaking, a solution is called Pareto optimal if there exists no other feasible solution that improves at least one objective function without worsening another. (For a formal definition, see, e.g., [1].)

3. ACADO Multi-Objective

ACADO Multi-Objective extends the ACADO toolkit for automatic control and dynamic optimisation [7] with several MO approaches. Due to the self-contained, object-oriented, C++ implementation, the toolkit is easy-to-use, it does not require third-party software, and allows a flexible control over algorithmic settings.

The idea behind ACADO Multi-Objective is the integration of efficient *multi-objective scalarisation* techniques with fast *deterministic direct optimal control* approaches. Scalarisation methods convert the original multi-objective optimisation problem (MOOP) into a (series of) parametric single objective optimisation problem (SOOPs) whose solution each time yields one point of the Pareto set. By consistently varying the method's parameter(s) (often referred to as *weights*) an approximation of the Pareto set is obtained. Despite its intrinsic drawbacks, combining the different objectives into a convex Weighted Sum (WS) is still one of the most popular scalarisation methods. However, alternatives as NBI and NNC, that mitigate these drawbacks, exist. Direct optimal control approaches transform the original infinite dimensional optimal control problem via discretisation into a finite dimensional NLP. Sequential strategies (e.g., Single Shooting (SiS)) discretise only the controls leading to small but dense NLPs, while simultaneous approaches (e.g., Multiple Shooting (MuS) and orthogonal collocation (OCol)) discretise both the controls and states, resulting in large but structured NLPs. The NLPs can be solved fast by deterministic optimisation routines.

It should be noted that although a number of commercial (e.g., gPROMS, PROPT) and non-commercial (e.g., DynoPC, MUSCOD-II, DYOS, DOTcvpSB) optimal control packages exist, none of them offers systematic and advanced multi-objective features. Fig. 1 shows the structure of ACADO Multi-Objective. Its features are the following.

A toolkit for efficiently generating Pareto sets in (bio)chemical multi-objective optimal control problems

- **Multiple objective optimisation methods.** Three scalarisation methods have been implemented: WS, NBI and NNC. The implementation is generic such that problems with any number of objectives can be tackled.
- **Weight generation.** The weights w_i are generated automatically based on a given step size. Currently, only convex combinations can be generated, i.e., satisfying $w_1+w_2+...+w_m = 1$ and $w_i \geq 0$. However, more advanced generation schemes, which leave out the positivity constraints [8], can be incorporated as well.
- **Single objective optimisation problem initialisation.** Different initialisation strategies for the series of SOOPs can be selected. First, all SOOPs can be initialised using the same fixed values provided by the user. Second, a hot-start

Figure 1. Schematic structure of ACADO Multi-Objective.

strategy, which exploits the result of the previous SOOP to initialise the next one, allows a significant decrease in computation time. Finally, the weights can also be employed to use as an initial guess a linear combination of the minimisers of the individual objectives.
- **Direct optimal control methods.** In the current version of ACADO, only SiS and MuS are available, OCol approaches are currently being implemented.
- **Integration and collocation methods.** Different integration routines are available for ordinary differential equation (ODE) systems, e.g., explicit integrators as Runge-Kutta45 or Runge-Kutta78 and implicit integrators as BDF, which can also be exploited to solve systems of differential and algebraic equations (DAEs). All integrators have been equipped with forward and backward automatic differentiation options such that exact first and second order derivatives are available for the optimiser. Settings involve, e.g., the absolute and relative integration tolerances. In the near future several collocation schemes will be made available.
- **Optimisation routines.** As optimisation routines SQP and interior point methods are available. Settings to be specified relate to, e.g., the Karush-Kuhn-Tucker (KKT) optimisation tolerance.
- **Pareto filter.** As NBI and NNC may produce non-Pareto optimal points, the solution set can be filtered using a *Pareto filter algorithm* (see [6] for details).
- **Visualisation.** The resulting Pareto sets can be directly plotted for cases with up to three objectives. The optimal states and controls and their corresponding Pareto optimal cost values can be exported in different formats.

4. Case studies

To test the approaches and the toolkit, three (bio)chemical cases are studied.

4.1. Case I: catalyst mixing problem in a tubular reactor

This problem considers a steady-state plug flow reactor of fixed length, packed with two catalysts, involving a series of a reversible and an irreversible reaction ($S_1 \leftrightarrow S_2 \rightarrow S_3$):

$$\begin{aligned} \frac{dx_1}{dz} &= -u(x_1 - 10x_2) \\ \frac{dx_2}{dz} &= +u(x_1 - 10x_2) - (1-u)x_2 \end{aligned} \tag{7}$$

with x_1 and x_2 the concentrations of S_1 and S_2, u the fraction of catalyst 1 and z the spatial coordinate. The original problem [9] considered the optimal mixing policy of the two catalysts in order to maximise the production of species S_3: $J_1 = -(1-x_1-x_2)$. The fraction of catalyst 1 is bounded: $0 \leq u \leq 1$. The reactor has a length z_f equal to 1 and at the inlet only S_1 is fed: $\mathbf{x}(0) = [1,0]^T$. To introduce a MO nature, the minimisation of the amount of the most expensive catalyst, i.e., catalyst 1, is added as an objective: $J_2 = \int u \, dz$.

4.2. Case II: fed-batch bioreactor problem

This problem relates to a fed-batch fermenter for the production of lysine [10]:

$$\frac{dx_1}{dt} = \mu x_1 \qquad\qquad \frac{dx_3}{dt} = \pi x_1 \tag{8}$$
$$\frac{dx_2}{dt} = -\sigma x_1 + u \, C_{s,F} \qquad\qquad \frac{dx_4}{dt} = u$$

with x_1 the biomass, x_2 the substrate, x_3 the product (lysine), and x_4 the fermenter volume. u is the volumetric rate of the feed, which contains a limiting substrate concentration $C_{s,F}$ of 2.8. The specific rates are given as follows for growth $\mu = 0.125 C_s$, substrate consumption $\sigma = \mu/0.135$, and production $\pi = -384\mu^2 + 134\mu$, with $C_s = x_2/x_4$ the substrate concentration. The initial conditions are specified as $\mathbf{x}(0) = [0.1, 14, 0, 5]^T$. The aim is to derive optimal feeding profiles and batch times that maximise productivity: $J_1 = x_3(t_f)/t_f$ and yield: $J_2 = x_3(t_f)/(\int u C_{s,F} \, dt)$. For practical reasons constraints are imposed on the volume: $5 \leq x_4 \leq 20$, the feed rate: $0 \leq u \leq 2$, the total amount fed: $20 \leq \int u C_{s,F} \, dt$, and the operation time: $20 \leq t_f \leq 40$.

4.3. Case III: temperature control of a jacketed tubular reactor

Case III concerns a jacketed reactor with an exothermic, irreversible first order reaction:

$$\frac{dx_1}{dz} = \frac{\alpha}{v}(1 - x_1)\exp(\gamma/(1 + x_2)) \tag{9}$$
$$\frac{dx_2}{dz} = \frac{\alpha\delta}{v}(1 - x_1)\exp(\gamma/(1 + x_2)) + \beta(u - x_2)$$

with x_1, x_2 and u the dimensionless reactant concentration, reactor temperature and jacket temperature. (For parameters values, see [11]). The original aim was to derive an optimal jacket fluid temperature that maximises conversion: $J_1 = C_F(1-x_1(z_f))$ and heat recovery via the jacket: $J_2 = \beta/z_f \int (u-x_2)dz$. For practical reasons bounds are imposed on all variables: $x_{1,\min} \leq x_1 \leq x_{1,\max}$, $x_{2,\min} \leq x_2 \leq x_{2,\max}$, and $u_{\min} \leq u \leq u_{\max}$. The initial conditions are given as: $\mathbf{x}(0) = [0,0]^T$. To make this case more challenging the reactor length z_f is allowed to vary is between 0.5 and 1, and is added as a third objective: $J_3 = z_f$.

5. Results

In the current study, a Multiple Shooting approach with an SQP optimiser has been employed. Table 1 provides an overview of the computational results for the different cases and multi-objective methods: WS, NBI and NNC. Figs 2, 3 and 4, depict the Pareto sets and a selection of the optimal controls. Despite the fine control discretisations (i.e., 25 and 50), the toolkit is able to solve the MOOCPs fast and accurately for both the bi- and tri-objective cases. For integration and optimisation tolerances of at most 1E-6, CPU times on a 1.86 GHz machine with 2GB RAM are between 0.1 and 2 s per Pareto point and below 2 minutes for the entire Pareto set. With respect to the different multi-objective approaches, NBI and NNC have been found to yield identical results for all three cases. Although equidistant weights have been generated and used for all three methods, NBI and NNC clearly produce Pareto points with a more uniform spread than WS (see Fig. 2 and 3). Moreover, these methods are also able to find points in non-convex regions of the Pareto set (see Fig. 3).

A toolkit for efficiently generating Pareto sets in (bio)chemical multi-objective optimal control problems

Table 1.	Case I			Case II			Case III		
	WS	NBI	NNC	WS	NBI	NNC	WS	NBI	NNC
# control pieces	25	25	25	50	50	50	50	50	50
SQP iterations	91	93	91	633	519	521	835	518	528
CPU time [s]	1.41	1.45	1.38	69.90	61.00	54.20	52.25	34.22	34.97
# Pareto points	11	11	11	41	41	41	66	66	66
CPU/Pareto point [s]	0.13	0.13	0.13	1.70	1.49	1.32	0.79	0.52	0.53

For Case I, there is a clear and continuous trade-off (see the convex Pareto set in Fig. 2). When conversion is focused on, the optimal control consists of a *max-singular-min* structure, similar to the results in [9]. However, when more emphasis is put on limiting the use of catalyst 1, the *max* and *singular* arc shrink and eventually vanish. In Case II (see Fig. 3), a *max-min-singular-min* sequence and a short batch time are observed, when productivity is emphasised. Here, the fast feeding stimulates the fast biomass growth and lysine production. When more attention is paid to the yield, the initial *max* arc vanishes, the batch times increase and the height of the *singular* arc decreases. These actions limit the substrate consumption, while maintaining the production. Note that the non-convex region in the Pareto set is induced by the requirement of adding at least 20 g of substrate. In Case III, the 3D Pareto front is clearly continuous and convex (see Fig. 4). The arc structure is typically *max-min-constrained-min*. In the first reactor part the reactor temperature has to be as high as possible (without violating the upper bound) to stimulate conversion and heat production, while in the last part it has to be decreased to recover heat. When solely conversion or heat recovery are aimed at, the entire reactor is used (i.e., $z_f = 1$). However, in the former case the control is used to maintain the upper reactor temperature until the outlet (i.e., no second *min* arc), whereas in the latter the upper temperature is only maintained in a small part, in favour of a large
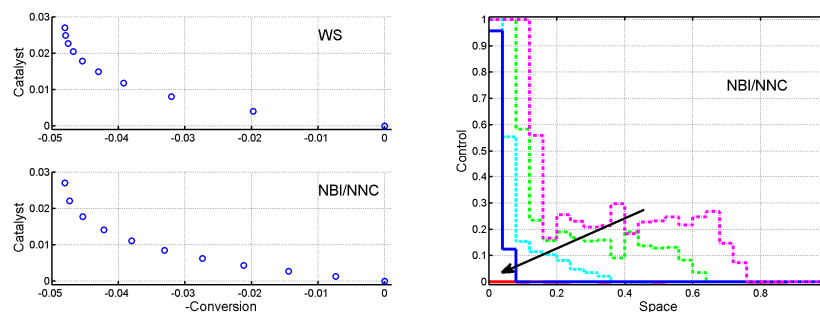

Figure 2. Case I: Pareto set obtained with WS, NBI and NNC (left); control profiles (right).
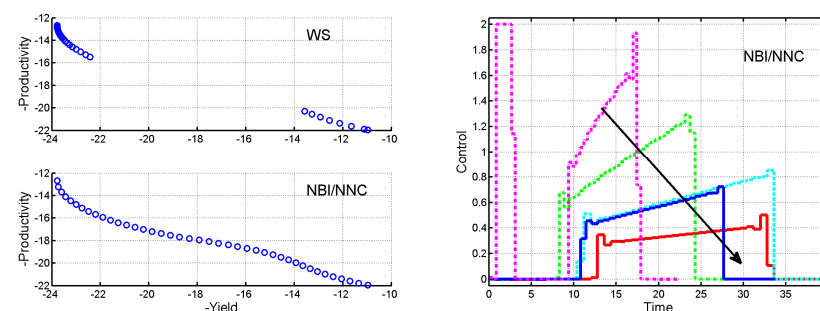

Figure 3. Case II: Pareto set obtained with WS, NBI and NNC (left); control profiles (right).
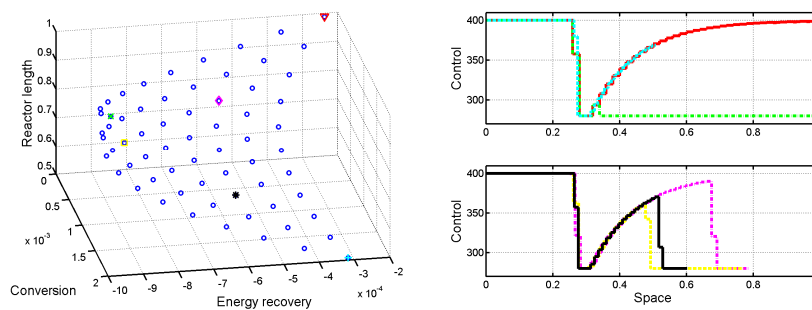
Figure 4. Case III: Pareto set obtained with NBI ($\times$,+,$\bigtriangledown$: individual optima; *,□,◊: intermediate points) (left); controls for individual minima (top right) and intermediate points (bottom right).

heat recovery section at the end (i.e., a large second *min* arc). Alternatively, the reactor length is reduced to its minimum value, when this objective is solely concentrated on. The intermediate Pareto optimal points exhibit also intermediate behaviour, as expected.

In summary, the ACADO Multi-Objective has been shown to be able to solve different types of MOOCPs (e.g., with/without singular arcs, with fixed/free end times, with control/state constraints) for different numbers of objectives.

6. Conclusions

This paper deals with the fast and efficient solution of (bio)chemical optimal control problems with multiple objectives. Hereto, several scalarisation techniques for multi-objective optimisation, e.g., WS, NBI and NNC have been integrated with fast deterministic direct optimal control approaches (e.g., SiS and MuS). All techniques have been implemented in the ACADO Multi-Objective toolkit (www.acadotoolkit.org).

7. Acknowledgements

References

[1]   K. Miettinen (1999) Nonlinear multiobjective optimization. Kluwer Academic Publishers, Boston.
[2]   K. Deb (2001) Multi-objective optimization using evolutionary algorithms. John Wiley, London.
[3]   I. Das and J. Dennis (1997) A closer look at drawbacks of minimizing weighted sums of objectives for Pareto set generation in multicriteria optimization problems. Structural Optimization 14:63–69.
[4]   V. Bhaskar, S. Gupta, A. Ray (2000) Applications of multiobjective optimization in chemical engineering. Reviews in Chemical Engineering 16:1–54.
[5]   I. Das and J. Dennis J (1998) Normal-Boundary Intersection: A new method for generating the Pareto surface in nonlinear multicriteria optimization problems. SIAM Journal on Optimization 8:631–657.
[6]   A. Messac, A. Ismail-Yahaya and C. Mattson (2003) The normalised normal constraint method for generating the Pareto frontier. Structural and Multidisciplinary Optimization 25:86–98.
[7]   B. Houska, H. Ferreau and M. Diehl (2009) ACADO Toolkit homepage. http://www.acadotoolkit.org
[8]   A. Messac and C. Mattson (2004) Normal constraint method with guarantee of even representation of complete Pareto frontier. AIAA Journal 42:2101–2111.
[9]   V. Vassiliadis, E. Balsa-Canto and J. Banga (1999) Second-order sensitivities of general dynamic systems with application to optimal control problems. Chemical Engineering Science 54:3851-3860.
[10] F. Logist, P. Van Erdeghem and J. Van Impe (2009) Efficient deterministic multiple objective optimal control of (bio)chemical processes. Chemical EngineeringScience 64:2527-2538
[11] F. Logist, P. Van Erdeghem, I. Smets and J. Van Impe (2009) Optimal design of dispersive tubular reactors at steady-state using optimal control theory. Journal of Process Control 19:1191–1198.

Model reduction as a tool for robust predictive control: application to OPR

Estefanía López-Quiroga, Luís T. Antelo, Antonio A. Alonso[*]

Process Engineering Group IIM-CSIC, Eduardo Cabello 6, Vigo 36208, Galiza, Spain,

[*] *Corresponding Author:* antonio@iim.csic.es

Abstract

The production of some chemicals or pharmaceuticals has been improved over the years by the development of a new concept of reactors, the reactor-heat exchangers, that overcome the classical constrains affecting many reaction units related to dissipation of heat and dilution/separation of products. The aim of this work is to propose a methodology for robust predictive control which aims at capturing the slow and most relevant, dynamics of the system. This model reduction constitutes the preliminary step to apply a real time optimization (RTO) framework for the robust control of reaction systems. The Open Plate Reactor (OPR) developed by the Swedish company Alfa Laval is used as a benchmark to validate the proposed methodology.

Keywords: OPR, Process Dynamics, Time Scales, Reduced Order Models, Robust Control.

1. Introduction

The synthesis of some chemicals or pharmaceuticals products constitutes a highly constrained process due to restrictions related to the dissipation of heat generated by the reactions [1]. One of the most common ways of solving these dissipation problems is to dilute the reactants, so that the reaction rate and therefore the heat released are reduced, which facilitates the process. Once the reaction ends, a separation of the solvent is carried out in order to increase the product concentrations. This separation process implies energy and time expenses which should be avoided. In order to overcome these problems, a new concept of heat-exchange reactors which combine both the advantages of plate exchangers for heat transferring as well as the mixing efficiency and reaction control of micro-reactors was developed [2,3]. As a consequence, full process performance is improved and time and energy costs are reduced. In this new framework, the Open Plate Reactor (OPR) developed by Alfa Laval presents such features. The dynamic behavior of the OPR is described by a large nonlinear system of partial differential equations involving temperature and reactant concentrations [4]. Due to the high dimensionality of the full model, reduced order representations of partial differential equations are developed. These follow the strategies proposed in [5], based on Laplacian Spectral Decomposition (LSD) and Proper Orthogonal Decomposition (POD). The aim is to capture the most relevant (slow) dynamics of the system without significant loss of accuracy in the representation with respect to those corresponding to the full model. These reduced models constitute the core of a real time optimization (RTO) methodology for controlling the OPR. The objective will be the minimization of process, so time to attain a desired yield while keeping the maximum temperature inside the unit within the established security margins defined for the process.

In this paper, the full dynamic model for OPR is presented in Section 2. Model reduction is described in Section 3. Finally, in Section 4, the formal statement of the RTO control problem for the Open Plate Reactor is proposed.

2. The Open Plate Reactor dynamic model.

The OPR unit is based on plate heat exchangers, which separate the hot and cold fluids in alternate plates with a large surface area, so to attain a higher heat transfer. The mixing of the reactants is performed in reactor plates, being each of them located between two cooling plates. This configuration allows flexible operation conditions such as several reactant injection points all along the reactor, changes in residence time, etc.

Since the cross section of the horizontal channels in the OPR is too small when compared with its length [4], an equivalent design consisting in a continuous plug reactor with cooling jacket can be used. In this work, the following exothermic reaction is considered:

$$A + 2B \rightarrow \frac{1}{2}C + \frac{1}{2}D \qquad \text{Eq.(1)}$$

The system of nonlinear partial differential equations (PDE) describing the dynamic behavior of the resulting plug flow reactor [6] is obtained from the energy balance inside the OPR and the reactants mass balances when a unique reactant injection point is considered. The equations are as follows:

$$\rho\, c_p \frac{\partial T_r}{\partial t} = \mathrm{k}\, \frac{\partial^2 T_r}{\partial z^2} - \rho\, c_p\, \mathrm{v_r} \frac{\partial T_r}{\partial z} + \frac{4h}{D_r}\left(T_c - T_r\right) + \Delta H\ \mathrm{r}\!\left(C_A, C_B, T_r\right) \qquad \text{Eq.(2)}$$

$$\frac{\partial C_A}{\partial t} = D \frac{\partial^2 C_A}{\partial z^2} - \mathrm{v_r} \frac{\partial C_A}{\partial z} - r\ \left(C_A, C_B, T_r\right) \qquad \text{Eq.(3)}$$

$$\frac{\partial C_B}{\partial t} = D \frac{\partial^2 C_B}{\partial z^2} - \mathrm{v_r} \frac{\partial C_B}{\partial z} - 2r\ \left(C_A, C_B, T_r\right) \qquad \text{Eq.(4)}$$

where: z is the position in the reactor, t is the time variable, D_r is the diameter of the tube and h is the heat transfer coefficient. The reaction heat is represented by ΔH; T_r and T_c are the temperatures in the reactor and of the cooling medium, respectively. The reacting mixture defined by its density ρ, thermal conductivity k, and heat capacity c_p flows through the reactor with velocity v_r. Finally, D is the mass diffusion coefficient, and C_A, C_B are the concentrations of reactants A and B, respectively. It must be noted that the reaction rate is given by the Arrhenius law: $r(C_A, C_B, T_r) = k_0 C_A C_B(-E_a / RT_r)$; and that the initial conditions are: $C_A(z,0) = 0$, $C_B(z,0) = 0$, $T_r(z,0) = 273\,\mathrm{K}$, $\forall\, 0 \le z \le L$, being L the reactor length.

In order to both facilitate the analysis and control of the OPR dynamic model, and to reduce the amount of parameters while achieving a well-conditionated system for numerical simulation, the model described by Eqs.(2-4) is transformed into a dimensionless one, by taking as independent spatial variable $\xi = z / L$, and as independent temporary variable $\tau = t\mathrm{v_r} / L$. This leads to:

Model reduction as a tool for robust predictive control: application to OPR.

$$\frac{\partial y}{\partial \tau} = \frac{1}{Pe_1}\frac{\partial^2 y}{\partial \xi^2} - \frac{\partial y}{\partial \xi} + W(u-y) + Da_3 \cdot f(x_A, x_B, y)$$ Eq.(5)

$$\frac{\partial x_A}{\partial \tau} = \frac{1}{Pe_2}\frac{\partial^2 x_A}{\partial \xi^2} - \frac{\partial x_A}{\partial \xi} - Da_{1a} \cdot f(x_A, x_B, y)$$ Eq.(6)

$$\frac{\partial x_B}{\partial \tau} = \frac{1}{Pe_2}\frac{\partial^2 x_B}{\partial \xi^2} - \frac{\partial x_B}{\partial \xi} - 2Da_{1b} \cdot f(x_A, x_B, y)$$ Eq.(7)

where $f(x_A, x_B, y) = (1 + x_A)(1 + x_B)\exp\left(\frac{E_a}{RT_{ref}}\left(\frac{y}{1+y}\right)\right)$; x_A, x_B, y are the dimensionless

variables for reactants A, B, and reactor temperature, respectively, with:
$Pe_1 = \rho L v_r c_p / k$; $Pe_2 = L v_r / D$; $Da_{1a} = L r_{ref} / v_r C_{Ae}$; $Da_{1b} = L r_{ref} / v_r C_{Be}$;
$Da_3 = r_{ref}L\Delta H / \rho v_r c_p T_{ref}$; $W = 4hL / c_p \rho D_r v_r$ and $r_{ref} = k_0 C_{Ae}C_{Be}\exp(-E_a / RT)$.

Finally, the dimensionless model is completed with the following boundary conditions:

- At $\xi = 0$
- At $\xi = 1$

$$-\frac{\partial y}{\partial \xi} + Pe_1 y = 0 \quad \text{Eq.(8)}$$

$$\frac{\partial y}{\partial \xi} = 0 \quad \text{Eq.(11)}$$

$$-\frac{\partial x_A}{\partial \xi} + Pe_2 x_A = 0 \quad \text{Eq.(9)}$$

$$\frac{\partial x_A}{\partial \xi} = 0 \quad \text{Eq.(12)}$$

$$-\frac{\partial x_B}{\partial \xi} + Pe_2 x_B = 0 \quad \text{Eq.(10)}$$

$$\frac{\partial x_B}{\partial \xi} = 0 \quad \text{Eq.(13)}$$

and initial conditions:

$$x_A(\xi,0) = -1 \qquad x_B(\xi,0) = -1 \qquad y(\xi,0) = -0,099 \ , \qquad 0 \le \xi \le 1 \qquad \text{Eq.(14)}$$

For its characteristics, the Matlab toolbox MatMOL (www.matmol.org) has been selected to solve the EDP system by using Finite Element Method (FEM). In this case, a non homogeneous spatial discretization, with 121 Lagrange P1 elements (half in the first third of the OPR and half in the rest) has been chosen. The results obtained for this model will be presented in comparison with those corresponding to the reduced model in the next Section.

3. Construction of reduced order models.

Solving PDE systems by using classical methods, for instance FEM, Finite Difference or Finite Volumes, implies the handling of a high number of ODEs. This fact difficults the application of these methods in many fields as real time optimization or predictive control, where the computing times are very important. In order to overcome this issue, new techniques capable of both simplifying the analysis and control of a given dynamic system as well as to reduce the computational effort have been developed. In this aim, Reduced Order Models will be employed to approximate the system by its slow and possibly unstable dynamics. Such approach will, in consequence, reduce the number of equations involved and thus the computation times [5,7]. In this work, the Laplacian Spectral Decomposition (LSD) and the Proper Orthogonal Decomposition (POD) are

applied to the OPR model. Both methods consist of projecting the system of partial differential equations on a lower dimensional subspace characterized by a set of globally defined basis functions satisfying the boundary conditions [8]. These basis functions are also required to form an orthogonal set in L^2 on the spatial domain considered.

Laplacian Spectral Decomposition (LSD). The reduced model is obtained by using the `matlsd` function from the MatMOL toolbox. The procedure consists on solving the following eigenvalue problem:

$$-\frac{1}{Pe}\Delta\phi_i = \lambda_i\phi_i \text{ , } en \text{ } (0,1), \text{ i} = 1,..., \text{ neig ; } -\frac{1}{Pe}\phi'_i(0) + \phi_i(0) = 0 \text{ , } \phi'_i(1) = 0 \qquad \text{Eq.(15)}$$

where λ_i are the eigenvalues (increasing ordered), ϕ_i are the eigenfunctions and *neig* is the number of basis functions. Finally, the EDP system with boundary conditions is projected over the selected set of eigenfunctions ϕ_i .

One of the dimensionless parameters present in the dynamic model is the Péclet number, *Pe*. Each reactor has a characteristic *Pe* value which plays an important role on system reduction: the eigenvalues and eigenfunctions are determined by the diffusion coefficient (1/*Pe*), according to Eq.(15). In this section, two different OPR, with its own *Pe* values, are analyzed. For an OPR with a low *Pe* number (*Pe*=7), a spectral gap of characteristic time scales that separates slow and fast components of the system dynamics exists [9], as shown in Figure 1. For an OPR with a high *Pe* number (*Pe*=59), the identification of different time scales is not as clear since the gap is smaller (all eigenvalues are in the same

Figure 1. Comparison between λ_i obtained from Eq.(15) for Pe =7 and Pe=59.

order of magnitude), being the dynamics more homogeneous. According to this, the LSD method will be more efficient for obtaining a model reduction when dealing with low Péclet values (i.e. reaction systems in which diffusion is important).

Proper Orthogonal Decomposition (POD). In this method, the basis functions are calculated through sets of data, called *snapshots*, by solving the following eigenvalue problem:

$$\phi_i(\xi) = \lambda_i \int_\Omega K(\xi,\xi')\phi_i(\xi')d\xi' \text{ , } \qquad \text{Eq.(16)}$$

where $K(\xi,\xi') = \frac{1}{l}\sum_{i=1}^{l} z_i(\xi)z_i(\xi')$ is the kernel defined by the *snapshots* $z_i(\xi)$.

The *snapshots* are chosen so to reproduce the leading dynamics of the system. In this work, the *snapshots* are obtained from the numerical simulation of the OPR model and the function `matpod` is employed to perform the reduction. As shown in Table 1, the POD method achieves a reduction in the degrees of freedom (DOF) of the model of around a 50% for each variable, while maintaining the accuracy with respect to the full model. The temperature evolution in the OPR for both full and reduced model is presented in Figure 2, confirming the good agreement between them.

Model reduction as a tool for robust predictive control: application to OPR.
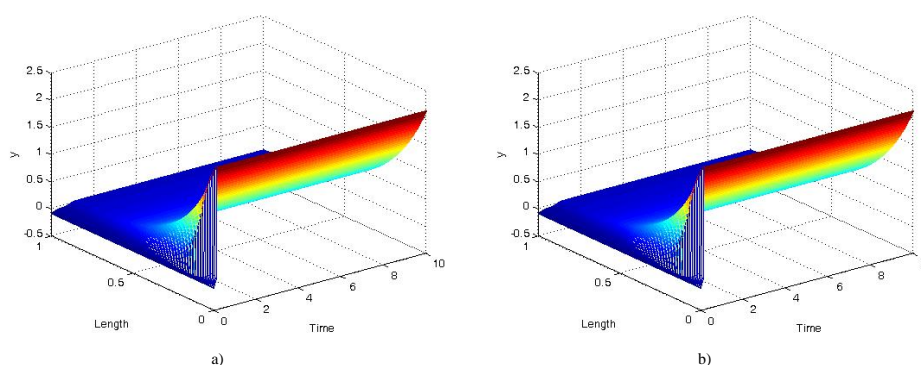
Figure 2. Temperature evolution in the OPR reactor: a) Full model; b) POD reduced model.

Table 1. POD results for the OPR model

Variable	DOF full model	DOF reduced model	Maximum absolute error
$x_A \in (-1, 0)$	121	56	0.0011
$x_B \in (-1, 0)$	121	49	0.0003
$y \in (-0.099, 2.15)$	121	54	0.0025

4. The OPR real time optimization problem.

The high dimension and nonlinearity inherent to the mass and energy balance equations of the OPR result into sets of control problems computational involved which justify a hierarchy two level approach:

1) The *upper layer* is responsible of obtaining different operation scenarios (i.e. different inlet configurations for reactants A and B or several reactant injection points distributed all along the reactor's length -Figure 3), which can be evaluated as a function of the reaction yield. Such assessment will be formally stated as a dynamic optimization problem to provide the required profile for the considered control variable u (cooling water temperature, T_c).

Figure 3. OPR scheme for the case of multiple reactant injection points. Each block can be described by the reduce model defined for the OPR with a single injection point.

The aim in this case is to minimize the process time t_f so to achieve a given yield γ in the OPR while satisfying both process dynamics f and operation constraints related to the reactor temperature T_r. Mathematically, the problem can be stated as follows:

$$\min_u J = t_f$$
$$s.t. \ f(\dot{x}, x, u, z, t) = 0; \ \gamma(L, t_f) = 96\%; \ T_r(z,t) \leq 155^{\circ}C; \ 20^{\circ}C \leq T_c \leq 80^{\circ}C \qquad \text{Eq.(17)}$$

As discussed above, the high dimensionality of the full model for the OPR could make the solution of problem in Eq.(17) computational involved. This calls for methodologies as the one proposed for model reduction to facilitate its analysis and solution [10].

2) The *lower layer* consists on the implementation of the pre-computed policies on a real time optimization framework designed to react to possible disturbances and to

minimize the adverse effect of final product variability. It is at this level where once again model reduction techniques together with on-line observers and estimators could be employed to cope with parameter and structural uncertainty as well as to minimize the computational burden associated with on–line dynamic optimization. As mentioned in Section 3, such identification methods will be based on first principle models covering the relevant operation time scales of the process (slow dynamics).
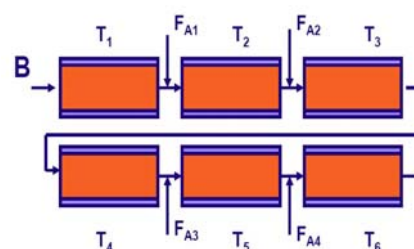
5. Conclusions and future work

Reduced order models for the Open Plate Reactor were obtained by applying two different techniques, the Laplacian Spectral Decomposition (LSD) and the Proper Orthogonal Decomposition (POD). It was found that model reduction through LSD is efficient for low Péclets numbers. In addition, the POD has been revealed as an effective method to obtain the desired reduced model for the OPR, no matter the importance of the diffusive phenomenon (i.e. for all the Péclet values). The latter reduced model is the core of the RTO problem proposed in this paper. As a future work, the resulting NLP control problems defined on the *upper layer* will be solved by employing the NDOT MATLAB toolbox [11] developed in the Process Engineering Group of the IIM-CSIC (http://www.iim.csic.es/~gingproc). Moreover, different strategies such as the distribution of reactant injection points all along the reactor's length together with some modifications on the OPR model design (to consider extensive variables –inventories- instead the current intensive ones) and on the operation conditions will be considered. These extensions are believed to provide a considerable improvement of the operation.

References

[1] L.T. Antelo, J.R. Banga, A.A. Alonso, Hierarchical design of decentralized control structures for the Tennessee Eastman Process, Comput. Chem. Eng. 32(9):1995-2015, (2008).

[2] C.H. Phillips, G. Lauschke, H. Peerhossaini, Intensification of batch chemical processes by using integrated chemical reactor-heat exchangers, Appl. Therm. Eng. 17, 809-824, (1997)

[3] J. Nilsson, F. Sveider, Characterising mixing in a hex reactor using a model chemical reaction, http://www.chemeng.lth.se/exjobb/002.pdf, (2000).

[4] S. Haugwitz, Modeling, control and optiomization of a Plate Reactor, Ph.D. thesis, Department of Automatic Control, Lund Institute of Technology, (2007).

[5] C. Vilas, M.R. García, J.R. Banga, A.A. Alonso, Robust feed-back control of distributed chemical reaction systems, Chem. Eng. Sci. 62(11):2941-2957, (2007).

[6] P.D. Christofides, Nonlinear and Robust Control of PDE Systems: Methods and Applications to Transport-Reaction Processes. Birkhäauser. Boston, (2001)

[7] R. Courant, D. Hilbert, Methods of Mathematical Physics, Wiley & Sons, New York, (1989).

[8] A.A. Alonso, C.E. Frouzakis, I.G Kevrekidis, Optimal sensor placement for state reconstruction of distributed process systems. AIChE Journal, 50(7):1438-1452, (2004).

[9] A. Armaou, I.G. Kevrekidis, C. Theodoropoulos, Equation-free gaptooth-based controller design for distributed complex/multiscale processes, Comput. Chem. Eng. 29:731-740, (2008).

[10] M.R. García, C. Vilas, E. Balsa-Canto, A.A. Alonso, Real time optimization for thermal processes, Proceedings of the European Control Conference - ECC'09. Budapest (Hungary), August 23-26, 2009.

[11] M.G. García, E. Balsa-Canto, A.A. Alonso, J.R. Banga, A software toolbox for the dynamic optimization of nonlinear processes, Computer Aided Chemical Engineering, Elsevier, 121-126, (2005).

Stable Model Predictive Control tuning considering asymmetric bounded signals

Mario Francisco[a], Hernán Álvarez[b], Pastora Vega[a], Silvana Revollar[c]

[a]Dpto. Informática y Automática, E.T.S.I.I. de Béjar. Universidad de Salamanca, Av. Fernando Ballesteros. 37700 Béjar, Salamanca (Spain); mfs@usal.es; pvega@usal.es
[b]Universidad Nacional de Colombia, sede Medellín (Colombia). Escuela de Procesos y Energía. Facultad de Minas; hdalvare@unal.edu.co
[c]Universidad Simón Bolívar. Departamento de Procesos y Sistemas. Caracas 1081A, Venezuela; srevolla@usb.ve

Abstract

In this paper the consideration of asymmetric bounded signals (disturbances, inputs) in a general procedure for tuning infinite horizon model predictive controllers (MPC) with constraints is presented. These measures are stated by means of an asymmetric objective functional, instead of the standard l_1 norm. The MPC is implemented with a terminal penalty to guarantee stability for all tuning parameters. Moreover, multiple models have been considered to ensure robust performance of the closed loop process, in the face of high non linear dynamics and load disturbances. The methodology has been applied to tune an MPC for the activated sludge process in a simulated wastewater treatment plant (WWTP). The problem is stated as a non linear programming problem with constraints and solved by Sequential Quadratic Programming (SQP).

Keywords: model predictive control, stability, l_1 norm, asymmetric signals, activated sludge process.

1. Introduction

Frequently, real plants are subject to constrained inputs (e. g. control signal saturations), and it is typical to have asymmetric input signals (control signals or disturbances), for example due to technological and safety reasons in the actuators or to a large variability of load disturbances. For those reasons, in this work an MPC tuning methodology developed by the authors [1, 2] is improved with the consideration of asymmetric bounded signals using a specific objective functional [3]. Moreover, the MPC implemented here operates over infinite horizon, and it is implemented as a finite horizon with terminal penalty, in order to solve finite dimensional QP problems. The terminal weight is obtained either as the solution of a Lyapunov equation or a Riccati equation, providing that the system is stable [4, 5].

The advantages of the development of a stable MPC tuning method are clear. MPC controllers have been tuned traditionally through a number of different parameters including the number of control moves, and input/output weights in the objective function. The tuning task is usually difficult because many of these parameters have overlapping effects on the closed-loop performance and robustness. Disturbance rejection capability is also a key issue in the controllability of a process, and therefore it is crucial to find a good controller that reduces the effect of disturbances.

The controllability indices used in this work are based on the H_∞ and l_1 norms of different weighted closed loop transfer functions matrices of the system, representing the process disturbances rejection capability and control efforts [2, 6], together with an asymmetric objective functional considered in some constraints.

The tuning approach developed has been validated on a simulated activated sludge process based on a real WWTP [7]. The paper is organized as follows. First, the MPC formulation and the controllability indices are presented, including the specific asymmetric functional and the optimization problem for tuning. Then, the activated sludge process is explained and some results are discussed, to end up with some conclusions.

2. MPC formulation

The MPC formulation consists of the on-line calculation of the future control moves by solving the following constrained optimization problem subject to constraints on inputs, predicted outputs and changes in inputs. The prediction model is a linear discrete state space model of the plant obtained by linearizing the first-principles nonlinear equations of the process. The infinite horizon MPC is implemented with finite horizon and terminal penalty, guaranteeing closed loop stability. The objective function is the following:

$$\min_{\Delta \mathbf{u}} V(k) = \min_{\Delta \mathbf{u}} \left(\left\| \mathbf{x}\left(k+Hc\right) \right\|_P^2 + \sum_{i=0}^{Hc-1} \left(\left\| \mathbf{x}\left(k+i\right) \right\|_Q^2 + \left\| \Delta \mathbf{u}\left(k+i\right) \right\|_R^2 \right) \right) \tag{1}$$

where k denotes the current sampling point, $\mathbf{x}(k+i)$ is the predicted state vector at time $k+i$, depending of measurements up to time k, $\Delta \mathbf{u}$ are the changes in the manipulated variables, Hc is the control horizon, R and Q are positive definite diagonal matrices representing the weights of the change of control variables and the weights of the set-point tracking errors respectively. In this work the reference is fixed to zero (steady state) and the outputs are equal to the states. Matrix P is the terminal penalty calculated solving the following Lyapunov equation, where A is the process state matrix:

$$P - A'PA = Q \tag{2}$$

This formulation is based on [4], where an infinite horizon MPC is developed with constraints both on states and outputs. The feasibility of the constraints guarantees nominal stability of the closed loop system for any choice of the tuning parameters, because the objective function is a Lyapunov function. The implementation of this controller only requires the solution of finite QP (Quadratic Programming) problems to obtain the control increments in a receding horizon strategy.

The use of objective function with terminal penalty (1) guarantees MPC stability, but in some cases the performance might get worse. This behaviour improves considerably if from sampling time Hc, an optimal state feedback controller (LQR) is implemented [5]. For this reason, in the example of this work the terminal penalty comes from the solution of the Riccati equation:

$$P = A'PA - A'PB\left(B'PB + R\right)^{-1} B'PA + Q \tag{3}$$

The MPC can be expressed as a combined feedforward-feedback control system, with the following output ($S_0(s) \cdot R_{d0}(s)$) and control (M_0) sensitivity functions to disturbances:

$$S_0(s) \cdot R_{d0}(s) = \frac{R_{d0}}{1 + G_0 K_1} \; ; \; M_0(s) = \frac{K_2 - K_1 G d_0}{1 + G_0 K_1} \; ; \; R_{d0}(s) = G d_0(s) - K_2 G_0(s) \tag{4}$$

where K_i are the transfer functions between the control signal and the different inputs ($r(s)$-$y(s)$ error signal, $d(s)$ disturbances) which depend on the control system tuning parameters (Q, R, Hc), and the nominal transfer functions are denoted by G_0 and Gd_0.

3. Controllability indices and optimization for automatic MPC tuning

In this work norm based indices are used in the tuning procedure to assess process controllability. Although those functions are only defined for linear control systems, it can be shown that it is also valid when the set of active constraints of the MPC is fixed [8, 2]. This assumption is sometimes a bit strong, and for that reason the constraints imposed are also used to keep the variables within the feasibility region. The tuning is stated as a mixed sensitivity optimization problem that takes into account disturbance rejection and control effort objectives in the same tuning function:

$$\min_R \|N_0\|_\infty = \min_R \left(\max_\omega |N_0(j\omega)| \right) \text{ where } N_0 = \begin{pmatrix} Wp \cdot S_0 \cdot R_{d0} \\ Wesf \cdot s \cdot M_0 \end{pmatrix} \tag{5}$$

The dependence on s of the transfer functions has been omitted for brevity. *Wp(s)* and *Wesf (s)* are suitable weights to achieve closed loop performance specifications and to reduce the control efforts respectively.

In order to ensure disturbance rejection (considering normalized disturbances) the following constraint must be added to (5) [2, 9].

$$\|Wp \cdot S_0 \cdot R_{d0}\|_\infty < 1 \tag{6}$$

The new functional proposed in [3] is also included as constraint in the MPC tuning methodology, giving information on the signal amplitudes, including the asymmetricity. For a generic discrete signal $z(k)$ it is defined by the following expression, and the specific constraints will be defined in the paragraph of results:

$$\|z(k)\|_a = \begin{bmatrix} \max\left\{0, \max_k (z(k))\right\} \\ \max\left\{0, -\min_k (z(k))\right\} \end{bmatrix} \equiv \begin{bmatrix} z_{max} \\ z_{min} \end{bmatrix} \tag{7}$$

4. Activated sludge process and control problem

The plant layout is represented in Fig. 1, consisting of one aerobic tank and one secondary settler. The basis of the process lies in maintaining a microbial population (biomass) into the bioreactor that transforms the biodegradable pollution (substrate) when dissolved oxygen is supplied through aeration turbines. Water coming out of the reactor goes to the settler, where the activated sludge is separated from the clean water and recycled to the bioreactor to maintain there an adequate concentration of microorganisms.
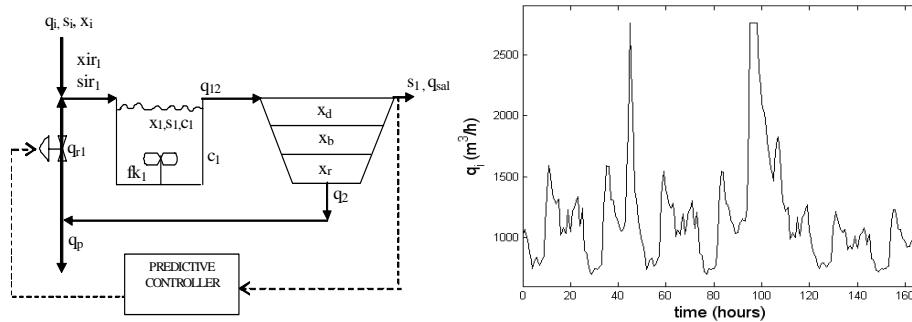
Fig. 1: Plant layout and storm weather disturbances at the influent (q_i)

The whole set of variables is presented also in Fig. 1. Generically, "*x*" is used for the biomass concentrations (mg/l), "*s*" for the organic substrate concentrations (mg/l) and "*q*" for flow rates (m^3/h). The complete set of non linear differential equations (the process order is 5) and model parameters are given in [7].

The control of this process aims to keep the substrate at the output (s_1) below a legal value despite the large variations of the flow rate and the substrate concentration in the incoming water (q_i and s_i), which are the input disturbances. The recycling flow (qr_1) is the manipulated variable and the substrate (s_1) is the controlled variable. The biomass (x_1) is only a constrained variable for a good performance of the process. The different sets of disturbances used in dynamic simulations (Fig. 1) have been determined by the COST 624 European research program and its benchmark [10].

5. Results

The tuning methodology begins with the selection of a fixed plant with V_1=7668 m^3 (reactor volume) and A=2970.88 m^2 (settler cross-sectional area), together with a steady state working point. Then the MPC is automatically tuned using linearized state space models of the system, and the MPC obtained is tested on the linearized model of the plant, i.e., the prediction model and plant model are the same (in this case the variables are deviations from the steady state, and they are represented with upper bar notation). Finally, the MPC is tested on the nonlinear plant, using the differential equations of the activated sludge process.

In the WWTP, the influent disturbances variations are asymmetric. For example, for storm weather q_{imin}=923.59 m^3/h and q_{imax}=1956.7 m^3/h, whose normalized value (deviations from the steady state) is $\bar{d}_{max} = 1; \bar{d}_{min} = 0.2866$ .Then, the q_i domain is determined by $D^t = \left[\bar{d}_{max}, \bar{d}_{min} \right] = \left[1, 0.2866\right]$. On the other hand the zero saturation of $u=qr_1$ puts a stricter bound $\overline{qr}_{1min}$ while $\overline{qr}_{1max}$ can be much larger depending on the pump characteristics. Then, the functional (7) is used, and the following constraint is included in the tuning procedure:

$$\left\| \overline{qr_1} \right\|_a \leq U \text{ , where } U^t = \left[\overline{qr}_{1max}, \overline{qr}_{1min} \right] \tag{8}$$

In order to solve the optimization problem, constraint (8) has been substituted by the approach in [3], which gives a simple condition (based on the impulse response of the system) that guarantees (8) for any disturbance defined in the asymmetric domain:

$$\left\| d(k) \right\|_a \leq D \text{ , where } D^t = \left[\bar{d}_{\max}, \bar{d}_{\min} \right] \tag{9}$$

In this work, only results for R tuning are shown, using the SQP method, because the influence of R in performance is more relevant. The value of Hc is empirically fixed to 10, large enough to provide a good response. However, the tuning of Hc could also been performed using a two iterative steps algorithm that combines a directed random search for tuning Hc (integer), and the SQP for tuning the weight R (real variable) [2, 6]. The sampling period is T=0.5 hours, and disturbances s_i and q_i are assumed to be measured. Multiple linearized models changing the substrate concentration in the reactor have been considered for robust performance tuning, imposing constraint (6) for every local model [2].

Table 1: Results for different bounds U in qr_1 (dry weather disturbances)

	Unconstrained qr_1	$U^t = \begin{bmatrix} 2200, 2200 \end{bmatrix}$	$U^t = \begin{bmatrix} 2200, 1450 \end{bmatrix}$
R	0.00259	0.00418	0.00444
max(qr_1)-min(qr_1)	735.09	712.06	707.56
max(s_1)-min(s_1)	2.71	3.36	3.46
$\left\| \overline{qr_1} \right\|_a$	[534.3; 275.3]	[503.9; 260.9]	[506.4; 258.3]
$\left\| Wp \cdot S_0 \cdot R_{d0} \right\|_\infty$	0.590	0.798	0.831

Table 2: Results for different bounds U in qr_1 without H_∞ performance constraint (6) (storm weather disturbances)

	$U^t = \begin{bmatrix} 5500, 5500 \end{bmatrix}$	$U^t = \begin{bmatrix} 2200, 2200 \end{bmatrix}$	$U^t = \begin{bmatrix} 2200, 1000 \end{bmatrix}$
R	0.00997	0.03614	0.18222
max(qr_1)-min(qr_1)	1741.1	1472.9	685.24
max(s_1)-min(s_1)	23.18	30.19	42.89
$\left\| \overline{qr_1} \right\|_a$	[1273.3; 627.0]	[1147.9; 384.3]	[428.6; 240.7]
$\left\| Wp \cdot S_0 \cdot R_{d0} \right\|_\infty$	1.391	4.857	7.067

First, the automatic tuning has been performed including the H_∞ constraint (6) for proper disturbance rejection, considering scaled disturbances for dry weather and different asymmetric bounds in qr_1. In Table 1 the numerical results are presented, including the optimal MPC. Then, in Table 2 and Figure 5 some results are presented changing the conditions over qr_1, for scaled storm weather disturbances. Here the H_∞ constraint (6) has not been considered to clarify the effect of constraint (8). In all results it can be seen that disturbance rejection is better when U is increased ($\overline{qr}_{1\min}$ is relaxed), because the MPC obtained allows for larger control variations. The SQP convergence tolerances are 10^{-8} for all results, and it is reached in less than 20 iterations.

6. Conclusions

In this work a general method for tuning MPC considering asymmetric bounded signals for disturbance rejection has been developed. This method has been tested in MPC

applied to a simulated activated sludge process, and the closed loop responses show that obtained controllers are properly tuned, taking into account the large magnitude of influent disturbances. The use of asymmetric bounded signals allows for a more realistic selection of the tuning controllability criteria, providing better control performance. The methodology proposed here is general, so any other chemical processes or performance criteria could be considered. Finally it is important to show that the developed method is particularly suitable for its inclusion in the resolution of the Integrated Design optimization problem, which determines the optimum controller and the optimum plant at the same time.
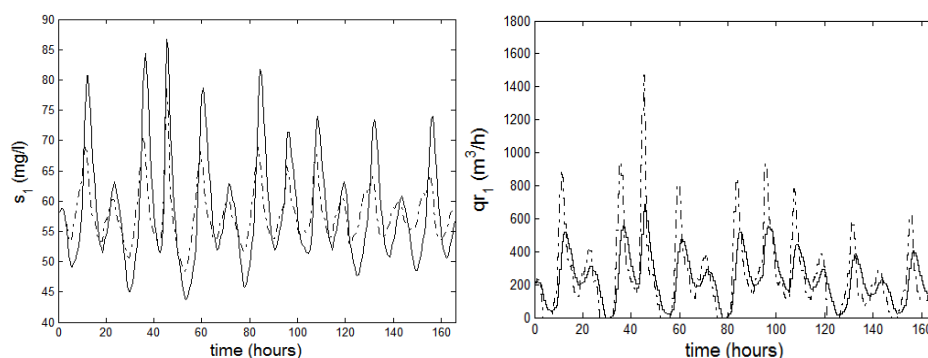
Fig. 4: Output substrate concentration (s_1) and recycling flow (qr_1) for the process with MPC tuned with $U^t = \begin{bmatrix} 2200, 2200 \end{bmatrix}$ (dashed dotted line) and $U^t = \begin{bmatrix} 2200, 1000 \end{bmatrix}$ (solid line).

7. Acknowledgements

References

[1]   M. Francisco, P. Vega. Diseño Integrado de procesos de depuración de aguas utilizando Control Predictivo Basado en Modelos. Rev. Iberoamericana de Automática e Informática Industrial, Vol. 3, nº 4 (2006) 88-98.

[2]   M. Francisco, P. Vega. Automatic tuning of Model Predictive Controllers based on Multiobjective Optimization. Latin American Applied Research (2009, in press)

[3]   M. Naib, A. Benzaouia, F. Tadeo, l_1- optimal control with asymmetric bounds. European Control Conference (Cambridge, UK) (2003)

[4]   J. B. Rawlings, K. R. Muske, The stability of constrained receding horizon control. IEEE transactions on automatic control, 38, nº 10 (1993) 1512-1516.

[5]   P. O. M. Scokaert, J. B. Rawlings. Constrained Linear Quadratic Regulation. IEEE Trans. on Automatic Control, vol. 43, nº 8 (1998) 1163-1169.

[6]   P. Vega, M. Francisco, E. Sanz. Norm based approach for automatic tuning of Model Predicitive Controllers. Proceedings of ECCE-6 (Copenhague) (2007)

[7]   R. Moreno, C. De Prada, J. Lafuente, M. Poch, G. Montague. Non-linear predictive control of dissolved oxygen in the activated sludge process. IFAC BIO 2, Ed. Pergamon Press. (1992) 289-298.

[8]   J. M. Maciejowsky. Predictive Control with Constraints. Prentice Hall, (2002).

[9]   S. Skogestad, I. Postlethwaite. Multivariable Feedback Control, Analysis and Design. John Wiley & Sons, (1997).

[10]   J. B. Copp. The COST Simulation Benchmark: Description and Simulator Manual. Office for Official Publications of the European Community. ISBN 92-894-1658-0. (2002).

Plantwide Control for Economic Operation of a Recycle Process

Rahul Jagtap[a], Nitin Kaistha[b] and Sigurd Skogestad[c]

[a] Chemical Engineering, IIT Kanpur 208016, India, Email: jagtap@iitk.ac.in
[b] Chemical Engineering, IIT Kanpur 208016, India, Email: nkaistha@iitk.ac.in
[c] Chemical Engineering, NTNU, Trondheim 7491, Norway, Email: skoge@ntnu.no

Abstract

Plant-wide control system design for economically optimum operation of a recycle process with side reaction is studied. The process consists of a liquid phase CSTR followed by two simple distillation columns. The exothermic irreversible reactions A + B → C (main reaction) and C + B → D (side reaction) occur in the CSTR. The reactor effluent is distilled in the recycle column to recycle the light reactants (A and B) back to the CSTR. The column bottoms is further distilled in the product column to produce nearly pure C as the overhead product with side-product D leaving from the bottoms. For a base-case design, the steady-state operating degrees of freedom are optimized to maximize operating profit for two modes of operation – Mode I: Given fresh A feed rate and Mode II: Maximum through-put. The set of active constraints at the economic optimum significantly simplifies the plant-wide control design problem by forcing structural decisions for process operation close to and where possible, at the active constraints. The economic performance of the control structure so synthesized is compared with other reasonable regulatory structures with and without a supervisory optimizing constraint controller. Quantitative process operation back-off results show that the incorporation of economic considerations in plantwide control system design can significantly improve profitability.

Keywords: Plantwide control, control system design, economic operation

1. Introduction

The plantwide control system for chemical processes typically consists of a regulatory layer that ensures safe and stable operation and an economic optimization layer that adjusts key setpoints in the regulatory layer for optimizing an economic criterion such as operating profit or energy consumption. The design of the regulatory plantwide control system has been extensively studied in the literature. The combinatorial complexity of the plantwide control structure design problem results in several reasonable structures that provide safe and stable process operation.

To systematize the choice of the loop pairings in the regulatory layer, Luyben et. al. proposed a nine-step bottom-up heuristic design procedure for "smooth" process operation [1]. An inherent disadvantage of this bottom-up approach is that economic considerations are inadvertently ignored. Given that the optimum economic operating point typically lies at the intersection of process constraints, the implemented regulatory control system affects the transients in these "active" constraint variables and hence a "back-off" is necessary to avoid transient hard constraint violation. Structures minimizing the transients in the active constraints would require smaller back-offs with consequently better economic performance

while ensuring safe and stable operation. Skogestad [2] termed such regulatory structures with an acceptable economic loss as self-optimizing. Based on the concept, a systematic design procedure that uses *apriori* knowledge of active constraints at the economic optimum to synthesize the regulatory control system has been proposed [3,4].

Not withstanding the simplicity of self-optimizing structures, what constitutes "acceptable" economic loss is quite subjective. In particular, even a small relative increase in production (say 1%) can translate into millions of dollars of additional revenue for the volume driven process industry. Quantification of the benefit of a supervisory economic optimizing controller on top of the regulatory layer is thus highly desirable. Further, even as self-optimizing control of complex chemical processes has been demonstrated in the literature, studies that quantify the back-off due to dynamic transients with or without a supervisory controller are lacking. This work is intended to fill this void through a case-study on a recycle process with side reaction.

In the following, a brief process description is provided followed by optimized operating conditions for two modes of operation corresponding to a given fresh feed processing rate (Mode I) and throughput maximization (Mode II). The active constraints at the optimum for each Mode are used to synthesize regulatory control structures using the systematic procedure of Skogestad. The economic performance of the synthesized control structures is quantitatively compared with other reasonable regulatory structures with and without an explicit supervisory optimizing controller. The article ends with the conclusions.

2. Process Description, Design and Optimum Operation

The process consists of a liquid phase CSTR followed by two distillation columns. The exothermic reactions $A + B \rightarrow C$ (main reaction) and $C + B \rightarrow D$ (side reaction) occur in the CSTR. The reactor effluent is distilled in the recycle column to recycle the light reactants (A and B) back to the CSTR. The column bottoms is further distilled in the product column to produce nearly pure C as the overhead product with side-product D leaving from the bottoms. The reaction chemistry necessitates reactor operation in excess A environment to suppress the side reaction. Figure 1 shows a schematic of the process along with salient design and base-case operating parameters for processing 100 kmol/h of fresh A to produce 99 mol% pure C. Table 1 reports the reaction kinetics and hypothetical component properties for modeling in Hysys.

Table 1: Modeling details of recycle process

Kinetics	$A+B \rightarrow C$	$r_1 = k_1 x_A x_B$	$k_1 = 2 \times 10^8 \exp(-60000/RT)$
	$B+C \rightarrow D$	$r_2 = k_2 x_B x_C$	$k_2 = 1 \times 10^9 \exp(-80000/RT)$
Hypotheticals	MW	NBP (°C)	Hydrocarbon estimation procedure used to estimate parameters for thermodynamic property calculations
A	50	80	
B	80	100	
C	130	130	
D	210	180	
VLE	Soave-Redlich-Kwong		

Reaction rate units: $kmol.m^{-3}.s^{-1}$

There are a total of eight steady state operational degrees of freedom for this process; two for the feeds (two feed rates), two for the reactor (temperature and holdup) and two each for the two columns. We choose the following variables as steady-state degrees of freedom (any independent set may be chosen): The fresh A feed rate, the reactor feed A to B excess ratio, the reactor level and temperature, the recycle column distillate C mol fraction (or reflux

rate) and bottoms B to C mol ratio and the product column distillate C mol fraction and bottoms C component flow rate.
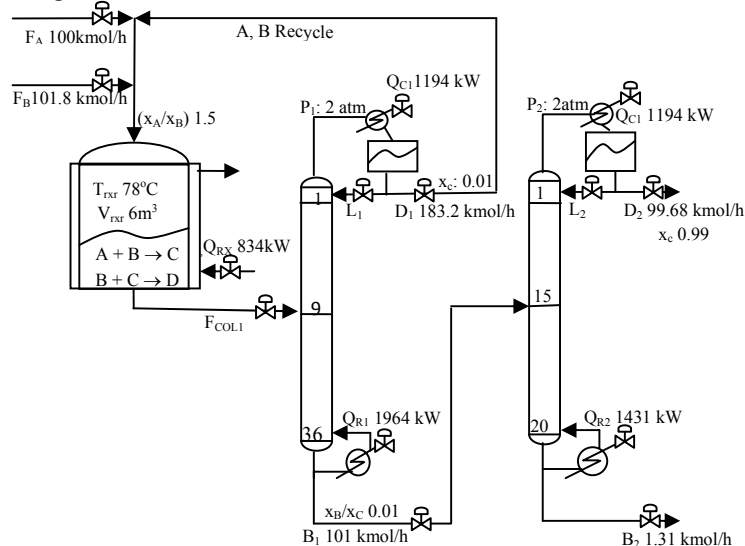
Figure 1: Schematic of Simple Recycle Process

Table 2: Optimization Summary

Objective	Max. (Product price – raw material price – energy costs)	
Constraints	$60\,°C < T_{rxr} < 100\,°C$ $0 <$ Material flows < 2(base-case) $0 <$ Energy flows < 2(base-case) $0 <$ Column boilups < 1.5(base-case) $0 <$ Rxr Holdup $< 6\ m^3$	

Optimized Operating Conditions		
Variable	*Mode I*	*Mode II*
F_A	100 kmol/h	184.6 kmol/h
V_{rxr}	6 m^3 (max)	6 m^3 (max)
T_{rxr}	70.47 °C	100 °C (max)
$(x_A/x_B)^{RxrFeed}$	2.3270	1.6698
L_1	~0 kmol/h	~0 kmol/h
$(x_B/x_C)_{Col1}^{Bottoms}$	0.01	0.01
$(x_C)_{Col2}^{Distillate}$	0.99	0.99
$(b_C)_{Col2}$	0.5 kmol/h	0.5 kmol/h
Profit per yr	\$2.793x10^6	\$4.615x10^6
Active Constraints	Col 1 Maximum Boilup Maximum reactor holdup	

Optimally, the values of the eight steady state operating degrees of freedom should be chosen to optimize an economic criterion such as maximizing plant operating profit (product price – raw material price – energy costs) subject to operating constraints (maximum/minimum flows, pressures and temperatures). Optimization for two modes of operation is studied. In Mode I, the fresh A feed to be processed is specified (e.g. dictated by product demand). In Mode II, the fresh A feed to be processed is also an optimization variable for maximizing operating profit. This mode typically corresponds to maximizing the process through-put for the highest possible production rate of the value-added product.

The optimization problem and results for Mode I and Mode II operation are summarized in Table 2. In both modes of operation, the recycle column maximum boil-up constraint and the maximum reactor level constraint are

active. In Mode I, the fresh A feed rate is fixed while in Mode II, the maximum reactor temperature constraint is active. Also, the recycle column reflux rate is small at the optimum for both modes. This corresponds to the recycle column essentially operating as a stripper (no enriching). Setting the reflux rate to zero and reoptimizing gives a maximum profit very close to the actual optimum (difference in hundreds of dollars). Accordingly, column operation at zero reflux ie as a stripper is considered optimal. The remaining three specifications for both modes correspond to the recycle column bottoms B to C mol ratio of 1%, a product purity of 99% and holding the C loss in the byproduct stream at a small value (0.5 kmol/h in the study). In both modes of operation, all the operating degrees of freedom are thus exhausted.

The optimization results may be interpreted as follows. For Mode I operation, the active maximum reactor volume constraint allows for a lower operating temperature with increased yield to the desired product C. The active maximum boil-up constraint for the recycle column corresponds to increasing the recycle of A so that the reactor excess A is as large as possible for higher product yield. For Model II operation (maximum through-put), operating the reactor at maximum level and temperature maximizes the conversion and hence the production rate. Operating the column at maximum boil-up causes higher amounts of fresh A to be processed without compromising selectivity (maximum allowable recycle).

3. Plantwide Control Structures

The active constraints for the different modes of operation dictate control structure decisions. In both operating modes, the recycle column vapor boil-up constraint is active and the column should be operated as a stripper. Accordingly, the reflux rate is fixed at zero and the reboiler duty is set for maximum permissible boil-up. A tray temperature in the stripping section is maintained by adjusting the column feed for maintaining B impurity in the bottoms. The reactor level is maintained by adjusting the F_B with F_A being maintained in ratio. The ratio set-point is adjusted to maintain the A/B composition ratio of the reactor feed. In Mode I operation, since the desired fresh A to be processed is specified, a discrepancy from this specification is used to adjust the A/B composition ratio set-point. Alternatively, the reactor temperature set-point may be adjusted. We found the former to entail slightly lower steady state economic loss due to disturbances. In Mode II operation, the A/B composition ratio set-point is kept fixed. In both modes, the plantwide control structure is the same except for the adjustment in A/B

Table 3: Plantwide control structures evaluated

Control Task	Adjusted Variable			
	CS1	CS2	CS3	CS4
Regulatory Control System				
TPM[*]	Q_{R1}	F_{COL1}	F_{CSTR}	F_A
Fresh Feed Ratio[#]	F_A/F_B	F_A/F_B	F_A/F_B	F_B/F_A
$x_{A(rxr\ feed)}$	$(F_A/F_B)^{SP}$	$(F_A/F_B)^{SP}$	$(F_A/F_B)^{SP}$	$(F_B/F_A)^{SP}$
T_{RX}	Q_{RX}	Q_{RX}	Q_{RX}	Q_{RX}
Reactor level	F_B	F_B	F_{COL1}	F_{COL1}
Col 1 T_{12}	F_{COL}	Q_{R1}	Q_{R1}	Q_{R1}
Col 1 Top Level	D_1	D_1	D_1	D_1
Col 1 Sump Level	B_1	B_1	B_1	B_1
Col 2 Top Level	D_2	D_2	D_2	D_2
Col 2 Sump Level	Q_{R2}	Q_{R2}	Q_{R2}	Q_{R2}
Col 2 (B_2*x_c)	B_2	B_2	B_2	B_2
Supervisory Control Loops				
F_A (Mode I)	$x_{A(rxrfeed)}^{SP}$	$x_{A(rxrfeed)}^{SP}$	$x_{A(rxrfeed)}^{SP}$	TPM
Boilup (Mode I)	TPM	TPM	TPM	$x_{A(rxrfeed)}^{SP}$
Boilup (Mode II)	TPM	TPM	TPM	TPM

*: Throughput manipulator
#: Stream in denominator is the wild stream

excess ratio set-point for Mode I. The remainder of the control structure is standard and is referred to as CS1 (see Table 3 for loop pairings).

CS1 utilizes *a priori* knowledge of the active constraints to locate the throughput manipulator at the principal bottleneck, the recycle column boilup. The remainder of the inventory control system is then "radiating" around it (Price and Georgakis [5]). For comparison, we also consider other reasonable regulatory control structures with alternative throughput manipulator locations. These structures (CS2-CS4) are summarized in Table 3.The feed to the column, total feed to the reactor and the fresh A feed are respectively the throughput manipulators in CS2, CS3 and CS4. For mode I operation, an excess ratio controller similar to CS1 is required for CS2 and CS3 but not CS4 since the latter directly fixes the fresh A feed. For CS2-CS4, a supervisory optimizing controller that adjusts the throughput manipulator to control the boilup near maximum can also be implemented for tighter boilup control to reduce the back-off in the boilup due to disturbances in both operating modes.

4. Results and Discussion

Of the active constraints, the maximum level and maximum boil up constraints are considered as hard. Rigorous dynamic simulations are performed to quantify the back-off necessary to avoid violating these constraints during transients due to disturbances. A 5% step increase in the heavy impurity in the fresh B feed stream is considered as the worst case disturbance. To quantify the impact of the supervisory boilup optimizing controller, the back-off is performed for operation at constant throughput manipulator setpoint (optimizing controller is off) and with the optimizing controller on.

The back-off in the level for both modes and CSTR temperature for mode II is about the same in all structures. The boilup back-off however varies significantly. Table 4 reports the salient operating parameters at the final steady state for derated process operation with and without the boilup optimizing controller. Notice that the boilup back-off increases in order CS1 < CS2 < CS3 < CS4. Also note that in both modes, an optimizing boilup controller reduces the back-off. In mode I, as the back-off necessary in the boilup increases, the A/B reactor feed excess ratio must decrease to process the same amount of A feed. In mode II operation, back-off in the boilup is directly related to the amount of fresh feed processed and hence the product rate. The lower the back-off, the higher the amount of feed processed with consequently better economic performance.

The plant operating profit results show that while the operating profit is relatively insensitive to back-off in mode I operation (because the profit curve is very flat near the optimum), the implemented control structure can significantly affect profitability in mode II. For example, the difference in profit between CS1 and CS4 with and without a boilup optimizing controller is respectively about $200,000 and $350,000, a relative change of more than 4% and 7.5%, respectively. Also, the application of an optimizing controller improves operating profit by more than 1% in all the structures where a boilup optimizing controller can be implemented (CS2-CS4). Notice that as the throughput manipulator location moves away from the bottleneck, the relative increase in profit using an optimizing controller improves over process operation at a constant derated throughput manipulator setpoint. In cases where the regulatory control system is already implemented and the

throughput manipulator location is away from the principal bottleneck, there exists significant incentive for implementing a supervisory controller for improving plant profitability. However, when possible, the plantwide regulatory control system should be designed so that the throughput manipulator is at (or close) to the principal bottleneck.

Table 4: Salient parameters for derated process operation for CS1-CS4(dynamics)

	F_A kmol/hr		F_C kmol/hr		$(x_A/x_B)_{rxr\ feed}$		Col 1 Boilup kmol/hr		Profit x10^6 \$/year	
	a	b	a	b	a	b	a	b	a	b
Mode I Operation										
Base	100		97.57		2.327		321.1		2.793	
CS1	100	100	98.06	98.06	2.318	2.318	321.1	321.1	2.793	2.793
CS2	100	100	98.11	98.07	2.234	2.289	311.2	317.7	2.803	2.813
CS3	100	100	98.17	98.13	2.154	2.22	301.5	309.2	2.815	2.81
CS4	100	100	98.21	98.21	2.143	2.143	299.3	299.3	2.817	2.817
Mode II Operation										
Base	184.6		180.62		1.635		321.1		4.615	
CS1	179.1	179.1	176.7	176.7	1.627	1.627	321.1	321.1	4.595	4.595
CS2	174	175.7	171.6	173.4	1.629	1.627	309.4	313.6	4.444	4.502
CS3	170.3	174.1	168	171.8	1.627	1.628	299.2	308	4.314	4.424
CS4	167.8	173.4	165.5	171.1	1.627	1.628	294.2	307	4.248	4.416

a: Without boilup optimizing controller.   b: With boilup optimizing controller

5. Conclusions

In conclusion, this case study demonstrates that the regulatory plantwide control system can significantly affect process profitability. *A priori* knowledge of the active constraints at the economic optimum operating point should be used to synthesize a control structure that mitigates the transient variability in the principal hard bottleneck constraint for reduced back-off from the optimum and hence better economic performance. Quantitative results show that locating the throughput manipulator at the principal bottleneck constraint mitigates the back-off.

References

1. M.L. Luyben, B.D. Tyreus and W.L. Luyben, 1997, Plant-wide Control Design Procedure, AIChE J. 43, 3161-3174.
2. S. Skogestad, 2000, Plantwide Control: The Search for the Self Optimizing Control Structure, J. Proc. Cont. 10(5), 487-507.
3. T. Larsson and S. Skogestad, 2000, Plantwide control- A review and a new design procedure, Mod. Ident. Cont. 21, 209-240.
4. S. Skogestad, 2004, Control Structure Design for Complete Chemical Plants, Comp. Chem. Eng. 28, 219-234.
5. R.M. Price and C. Georgakis, 1993, Plantwide regulatory control design procedure using a tiered framework, Ind. Eng. Chem. Res. 32, 2693-2705.

Optimal Process Control and Operation of an Industrial Heat Integrated Fluid Catalytic Cracking Plant Using Model Predictive Control

Mihaela Iancu[a], Paul Serban Agachi[a]

[a]*Babes-Bolyai University, Faculty of Chemistry and Chemical Engineering, Chemical Engineering Department, Arany Janos St., No. 11, 400028, Cluj-Napoca, Romania,*
mmorar@chem.ubbcluj.ro

Abstract

Nowadays, the optimal control of a heat integrated industrial plant becomes one of the most important research areas in the chemical industry. There are at least two main reasons why this topic is interesting: first, the reduction of production costs applying the heat integration techniques and second, process optimization through advanced control alternatives, when taking into account the improvement of the plant safety in operation and the increasing of the products quality. It is known that the heat integration destabilizes the whole plant, advanced control being needed to make the plant operational.

Due to its complexity, the fluid catalytic cracking (FCC) process is a good candidate to apply heat integration and advanced control techniques.

It is well known that the investigation of an entire FCC plant taking into account the complex dynamic behavior in conditions of heat integration has not been studied yet.

In this study a real FCC plant from a Romanian refinery was used for simulation and at the same time for the implementation of a model predictive control (MPC) strategy in conditions of a previous retrofitted heat integration plant configuration.

The aim of this research is to study the complex dynamic behavior of the heat integrated plant under the effect of the main disturbances and to develop an optimal advanced control scheme for the same heat integrated FCC industrial plant. The implemented MPC strategy focused on the response of the heat integrated process in terms of operation, product quality and cost reduction of the heat integrated plant.

To simulate the FCC heat integrated process Aspen HySys software was used. In the simulation, the reactor-regenerator section, the main fractionator and the retrofitted heat exchanger network (used for preheating the feedstock before entering the riser) are included.

Keywords: fluid catalytic cracking, heat integration, dynamic behavior, model predictive control

1. Introduction

The modern process plants are continuously improved for a flexible production and for maximization of the energy and material savings. These plants are becoming more complex with a strong interaction between the process units. Consequently, the failure of one unit might have a negative effect on the overall productivity. This situation reveals important control problems. Another problem is that the techniques developed by now can't solve all the control problems that appear in modern plants because different plants have different requirements.

However, the appearance and the continuously development of the advanced control techniques provide better solutions for plants control at any level of complexity of the process.

The benefits of the advanced control implementation can be observed first in the operating costs of the plant. The operating costs can decrease with 2% - 6% [1]. The second benefit is the reduction of the process variability. As a consequence, the plants can be operated to their designed capacity.

One of few advanced control methods used successfully in industrial control applications is the model predictive control (MPC). This technique represents an advanced method of optimal process control since 1980s when it was developed to meet the specialized control needs of power plants and petroleum refineries.

Therefore, the aim of our research is to develop a MPC control scheme capable to control an industrial fluid catalytic cracking (FCC) plant with the new HEN design obtained in a previous work [2], [3]. The implementation of a MPC control scheme can provide a high stability of the plant knowing that the heat integration induces more instability in the process.

Due to its complexity the interest of solving problems related to the FCC process is worldwide dissipated. There has been a continuous effort to improve the efficiency and yield of the FCC unit during the time. There are many articles that present the problem of the FCC process modeling, simulation and control the most significant being [4] - [12] and just a few articles in which the study is related to the problem of energy integration [13].

For reaching the purpose of this research, a FCC plant dynamic simulator was build in Aspen HYSYS using real industrial data related to material fluxes, temperatures, pressures, equipments size and geometry, etc. The data have been provided by a Romanian refinery.

2. FCC dynamic simulator presentation

The FCC process description is very difficult due to several reasons: the complexity of the chemical reactions mechanism, complex hydrodynamics, strong interaction between the operation of the main reactor and the regenerator and due to the operation constrains imposed by the new HEN.

Taking advantage of the Aspen HYSYS capabilities the model of the FCC heat integrated plant has been built.

In the HYSYS interface the steady state FCC model appears as a main flowsheet and two sub-flowsheets. The main flowsheet contains the FCC reaction block with the riser and the regenerator, a simplified scheme of the FCC column and the preheating train for the raw material – the HEN. One sub-flowsheet consists in the reactions block and the other one represents the FCC column.

In the dynamic environment the FCC simulator contains the main flowsheet only with the simplified scheme of the FCC column and the preheating train and the sub-flowsheet representing the FCC column.

The FCC dynamic model has been fitted to simulate the real behavior of the industrial plant.

Because the case study of this paper is represented by a real industrial plant witch already has a PID control scheme implemented, the development of the MPC control scheme for the heat integrated plant was started from the real PID control scheme.

Therefore, the dynamic simulations were performed discovering that only 5 controllers are necessary from the entire PID control scheme in order to obtain a proper stability of

Optimal Process Control and Operation of an Industrial Heat Integrated Fluid Catalytic Cracking Plant Using Model Predictive Control

the FCC column knowing that the stability of the column is reflected in the quality of heat transfer through the HEN.

The next step will be the development of the MPC controller for controlling the FCC column based on the data provided by those 5 PID controllers selected to be able to set the FCC column behavior and, consequently, to stabilize the HEN at normal functioning conditions.

3. The development of the MPC control scheme

The MPC is differentiated from the other advanced control techniques by three key elements: the predictive model, the optimization in range of a temporal window, and the feedback correction. These three steps are usually carried on continuously by computer programs on-line.

The need of the MPC strategy appear due to the fact that in order to improve the operation of the plant it is needed to have implemented a control strategy capable to provide appropriate results in a very short time. It was observed that the PID control scheme is not capable to manage the operation in a short time.

The development of the MPC controller of the FCC column is based on the data of the selected PID controllers.

Therefore 5 inputs (controlled variable – CV) and 5 outputs (manipulated variable – MV) MPC is implemented for the FCC column.

The controlled and the manipulated variables are summarized in Table 1.

Table 1. MPC control scheme selected variables

Controlled Variable		**Manipulated Variable**	
CV1	To Condenser stream temperature	**MV1**	Reflux flow rate
CV2	Condenser Liquid % Level	**MV2**	Gasoline flow rate
CV3	FV-13 bottom HCN flow rate	**MV3**	FV-13 HCN feed flow rate
CV4	FV-10 bottom LCO flow rate	**MV4**	FV-10 LCO feed flow rate
CV5	Slurry temperature	**MV5**	Column recycle slurry flow rate

"To Condenser" stream represents the top column output stream that passes two heat exchangers and enters in the condenser. The FV-10 and FV-13 are the side strippers of the FCC column. The bottom product stream of FV-13 (HCN stripper) is named HCN-FP-19. The bottom product stream of FV-10 (LCO stripper) is named LCO-FP-6. The Slurry stream is the bottom product of the FCC column.

Having all these specified the way of the MPC strategy implementation is presented as follows.

From the Aspen HYSYS options to specify the MPC controller model was selected the First order model. The first order model implementation it is possible if the process gain (K_p), the process time constant (T_p) and the delay (τ) are known. With these process parameters it is possible to obtain the step response matrix necessary to build the internal MPC model.

The identification of the process parameters is realized by developing step response tests for each manipulated variable using the implemented PID control scheme from the dynamic state model. Consequently, a +10% step was used for each manipulated variable and the effect on the controlled variable was registered.

The step response tests provided the process gains, the time constants and the delays necessary for building the MPC controller. These process parameters, presented in Table 2, were used to determine the 5x5 MPC step response matrix necessary for implementation of the MPC controller internal model.

Table 2. First order model process parameters

	MV1	MV2	MV3	MV4	MV5
CV1	Kp = -2.35 T = 2.56 τ = 0	Kp = 0.567 T = 15.66 τ = 0	Kp = -1.5182 T = 45.83 τ = 6.66	Kp = -1.5526 T = 77.083 τ = 0	Kp = -0.9611 T = 16.183 τ = 0
CV2	Kp = -0.15685 T = 3.033 τ = 0	Kp = -0.312 T = 8.33 τ = 0	Kp = -0.068 T = 58.75 τ = 0	Kp = -0.0775 T = 58.583 τ = 0	Kp = -0.0738 T = 20.016 τ = 0
CV3	Kp = -0.2043 T = 29.03 τ = 0	Kp = 0.1147 T = 32.16 τ = 0	Kp = 1.41 T = 0.01 τ = 0	Kp = -0.061 T = 113.5 τ = 7.33	Kp = -0.00456 T = 42.316 τ = 0
CV4	Kp = -0.66 T = 34.783 τ = 0	Kp = 0.39 T = 24.08 τ = 0	Kp = 0.437 T = 21.75 τ = 1.583	Kp = 7.436 T = 0.01 τ = 0	Kp = -0.1452 T = 26.85 τ = 0
CV5	Kp = -0.3741 T = 14.06 τ = 0	Kp = 0.32 T = 12.33 τ = 0	Kp = 0.28 T = 8.916 τ = 0	Kp = 1.156 T = 6.5 τ = 0	Kp = -1.151 T = 4.28 τ = 0

After establishing the MPC controlled and manipulated variables and after determining the internal MPC model, the MPC controller was implemented in the FCC heat integrated plant dynamic model.

The MPC control performances were analyzed in order to verify if the MPC controller was properly developed.

The molar flow of the column feed stream was selected to perform a test concerning the performance of the MPC controller when handling the disturbances that can occur in the FCC plant. Consequently, in Figure 1 are represented the results obtained after a +5% disturbance applied on the column feed molar flow. In this figure the red line is representing the setpoint, the green line is the controlled variable and the blue line is the manipulated variable. One division on the diagrams corresponds to 10 minutes.

The MPC controller presented a small delay in identifying the disturbance and taking action on it. The delay was less than a minute.

The top column product stream temperature (To Condenser stream) was brought back to the setpoint in 30 minutes (Figure 1, b). During this time the temperature variation values were 0.5 ^{0}C above and below the set point (108^{0}C). The temperature control obtained by the MPC controller is very good.

Also, the condenser liquid level control is stabilized in 30 minutes with a level variation of approx. 3 % (Figure 1, c).

Regarding the flow control of the bottom product stream, of the FCC column side-strippers, the MPC controller behavior proofed to be very efficient. The flows were stabilized in 10 minutes (Figures 1, d,e).

As it can be seen in Figure 1 f, 15 minutes were necessary for the MPC controller to set the Slurry temperature back to the initial value from the moment of disturbance appearing.

*Optimal Process Control and Operation of an Industrial Heat Integrated Fluid
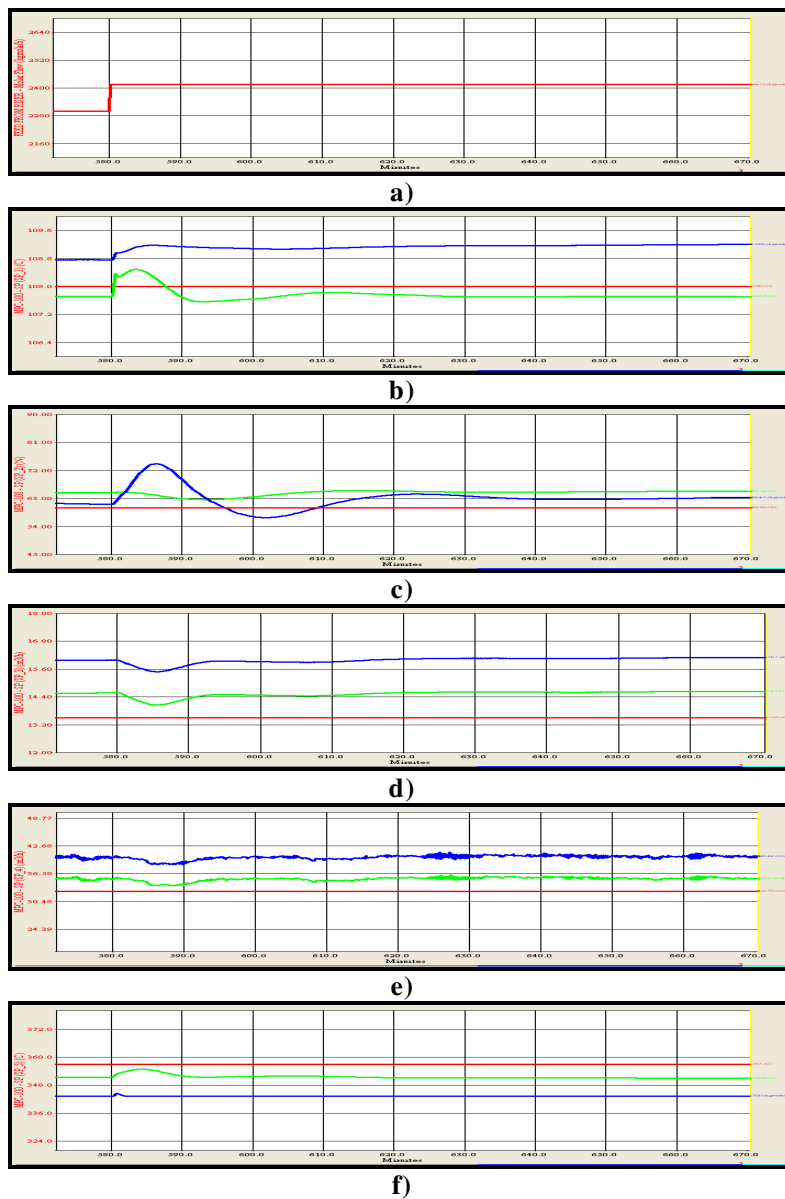Catalytic Cracking Plant Using Model Predictive Control*

a)

b)

c)

d)

e)

f)

Figure 1: MPC controller disturbance test
a) +5% column feed molar flow step disturbance;
b) Top column product stream temperature control;
c) Condenser liquid level control;
d) HCN stripper bottom product stream flow control;
e) LCO stripper bottom product stream flow control;
f) Slurry temperature control.

4. Conclusions

The control analysis and MPC development was based on a real industrial case. The data for modeling and design have been obtained from a FCC plant that is in function at this time in a Romanian refinery.

It was observed that a good stability of the FCC column provides an appropriate heat transfer through the new HEN in the imposed conditions of the real plant. For good stability of the column a MPC controller was developed and the simulations demonstrated that the new HEN design can successfully be implemented in the real plant with significantly plant costs reduction.

Regarding the MPC controller development, an easy method was used for the internal MPC model development. This method is fast and very useful in implementing an advanced control scheme at industrial scale. The step response tests in the real plant are insecure and costly especially in a continuous process. Any kind of these step control tests are able to compromise the quality of the products and to destabilize the industrial plant followed by undesirable incidents and costs.

The study results revealed an improvement of process operation and more importantly that the MPC scheme was able to assure higher products quality, increased energy recovery and reduced operating costs.

References

[1] J.S. Anderson, Process control opportunities and benefits analysis, Proc. Advanced Control for the Process Industries, Cambridge, 9-11th Sept. 1992.

[2] E. Jara-Morante, M. Morar, P. Ş. Agachi, Studia Universitatis Babes-Bolyai Chemia, LIV(1) (2009) 69.

[3] M. Morar, P. S. Agachi, Computer Aided Chemical Engineering, 26 (2009) 465.

[4] A. Arbel, I. H. Rinard, R. Shinnar, Industrial & Engineering Chemistry Research, 35 (1996) 2215.

[5] P. Grosdidier, A. Mason, A. Aitolahti, P. Heinonen, V. Vanhamäki, Computer and Chemical Engineering, 17 (1993) 165.

[6] I. S. Han, C. B. Chung, Chemical Engineering Science, 56 (2001) 1951.

[7] S. M. Jacob, B. Gross, S. E. Voltz, V. M. Weekman, AIChE Journal, 22 (1976) 701.

[8] H. Kurihara, Optimal Control of Fluid Catalytic Cracking Processes, PhD. Thesis, MIT, 1967.

[9] W. Lee, A. M. Kugelman, Industrial and Engineering Chemistry Process Design and Development, 12 (1973) 197.

[10] M. Hovd, S. Skogestad, AIChE Journal, 39 (1993) 1938.

[11] R. C. McFarlane, R. C. Reineman, Multivariable optimizing control of a model IV fluid catalytic cracking unit, AIChE Spring National Meeting, Orlando, FL., 1990.

[12] V. W. Weekman, D. M. Nace, AIChE Journal, 16 (1970) 397.

[13] A. B. Al-Riyami, J. Klemes, S. Perry, Applied Thermal Engineering, 2001, 21, 1449.

[14] P. S. Agachi, Z. K. Nagy, M. V. Cristea, A. Imre-Lucaci, "Model Based Control. Case Studies in Process Engineering", Wiley – VCH, 2006.

[15] N.,Alsop, J. M. Ferrer, "Step-test free APC implementation using dynamic simulation", Process Control Spring National Meeting, Orlando, FL, April 24-27, 2006.

Advanced Control Strategies for Dividing-Wall Columns

Anton A. Kiss, Ruben C. van Diggelen

AkzoNobel – Research, Development and Innovation, Velperweg 76, 6824 BM Arnhem, The Netherlands, tony.kiss@akzonobel.com

Abstract

This study explores the controllability of dividing-wall columns (DWC) and makes a comparison of various control strategies based on PID loops, within a multi-loop framework (DB/LSV, DV/LSB, LB/DSV, LV/DSB) versus more advanced controllers such as LQG/LQR and high order controllers obtained by H$_\infty$-controller synthesis and μ-synthesis. The controllers are applied to a DWC used in an industrial case study – the ternary separation of benzene-toluene-xylene. The performance of these control strategies and the dynamic response of the DWC is investigated in terms of products composition and flow rates, for various persistent disturbances in the feed flow rate and composition. Significantly shorter settling times can be achieved using the advanced controllers based on LQG/LQR, H$_\infty$ and μ-synthesis.

Keywords: DWC control, multi-loop PID controller, LQG/LQR, H$_\infty$ and μ-synthesis

1. Introduction

Distillation remains among the most important separation technologies in the chemical industry. In the last decades, ternary separations progressed via thermally coupled columns such as Petlyuk configuration to a novel design that integrates two columns into one shell – a setup known today as dividing-wall column. The DWC concept is a major breakthrough in distillation technology, as it brings significant reduction in the capital invested as well as savings in the operating costs, up to 25–40% (Kiss, 2009).

This study explores various DWC control strategies based on PID loops, within a multi-loop framework versus more advanced controllers. The controllers are applied to an industrial DWC used for the ternary separation of benzene-toluene-xylene. The performance of these control strategies is investigated in terms of products composition and flow rates, for various persistent disturbances in the feed flow rate and composition.
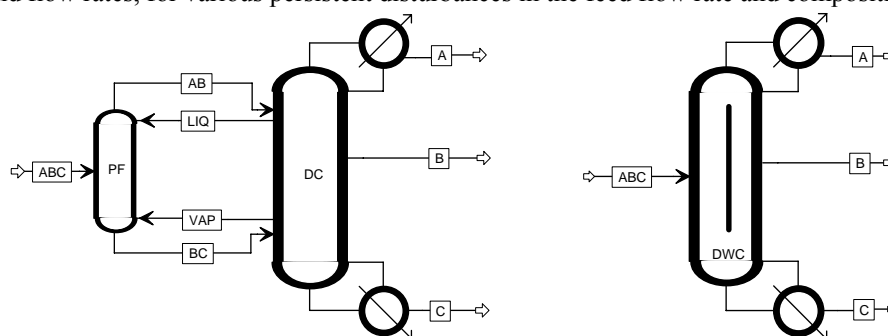

Figure 1. Petlyuk configuration (left). Dividing-wall column (right).

2. Problem statement

The integration of two columns into one shell leads also to changes in the operating mode and ultimately in the controllability of the system. Although much of the literature focuses on the control of binary distillation columns, there are just a few studies on the controllability of DWC (Serra, 2000; Adrian, 2004; Ling, 2009). The problem is that different DWC separation systems were used hence no fair comparison of controllers is possible. To solve this problem, we explore the DWC control issues on one system and compare various multi-loop PID control strategies versus more advanced controllers such as LQG/LQR, GMC, and high order controllers based on the H_∞ norm μ-synthesis.

3. Dynamic model

Several reasonable simplifying assumptions were made: 1. constant pressure, 2. no vapor flow dynamics, 3. linearized liquid dynamics and 4. neglected energy balances and enthalpy changes. The dynamic model of the DWC is implemented in Mathworks Matlab combined with Simulink and it is based on the Petlyuk model previously reported in literature by Halvorsen and Skogestad (1997). A rigorous steady-state simulation was also developed in AspenPlus to validate the assumptions of the model.
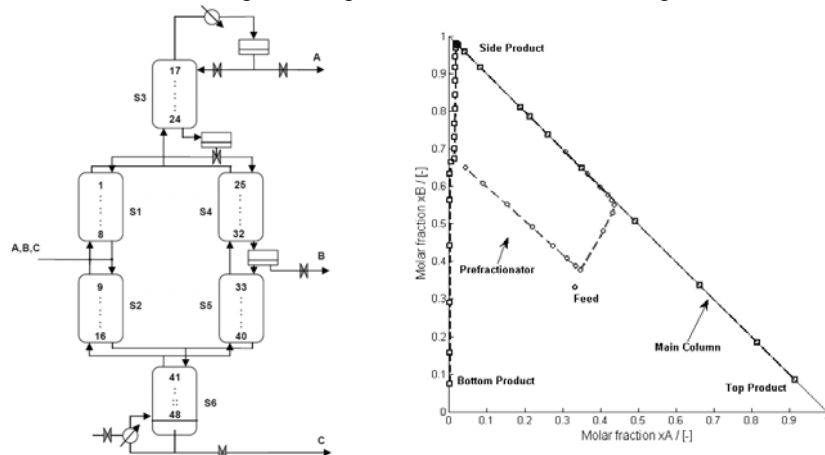

Figure 2. Schematics of the simulated DWC (left). Composition profiles inside the DWC (right).

4. Control strategies

PID loops within a multi-loop framework. The most used controllers in industry are the PID controllers. In case of a DWC, two multi loops are needed to stabilize the column and another three to maintain the set points specifying the product purities. As there are six actuators (D S B L_0 V_0 R_L) using PID loops within a multi-loop framework, many combinations are possible. However, there are only a few configurations that make sense from a practical viewpoint. The level of the reflux tank and the reboiler can be controlled by the variables L_0, D, V_0 or B respectively. Hence, there are four inventory control options to stabilize the column and to control the level in the reflux tank and the level in the reboiler, the combinations: D/B, L/V, L/B and V/D (Diggelen et al., 2010).
Linear Quadratic Gaussian control (LQG) is a combination of an optimal controller LQR and optimal state estimator (Kalman filter) based on a linear state-space model with measurement and process noise. LQG is an extension of the optimal state feedback that is a solution of the *Linear Quadratic Regulation* (LQR) which assumes no process

noise and availability of the full state for control. An additional feed-forward controller can be added or the LQG controller can be extended with an integral action.

Multivariable controller synthesis. Two advanced controller synthesis methods were used in order to obtain a robust controller: *loop shaping design procedure* (LSDP) and the *μ-synthesis* procedure. In contrast to previous studies, the inventory control and regulatory control problems are solved simultaneously in this work.
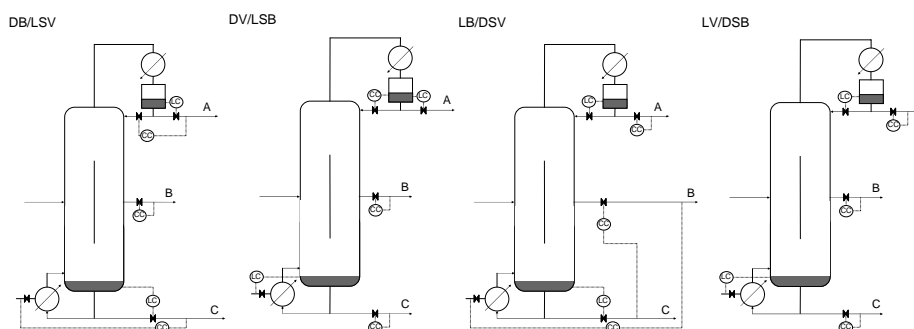

Figure 3. Control structures based on PID loops within a multi-loop framework.

5. Results and discussion

In this work, we use the ternary mixture benzene-toluene-xylene (equivalent to A, B, C) and the purity setpoints [0.97 0.97 0.97] for the product specifications. For the dynamic simulations performed in this study, disturbances of +10% in the feed flow rate (F) and +10% in the feed composition (xA) were used, as these are among the most significant ones at industrial scale. Note that persistent disturbances give a better insight of the quality of the controller than zero mean disturbances, as typically after a temporary disturbance the product compositions return to their given setpoints. Moreover, the effect of measurement noise on the control performance was also investigated.

As shown next by the results of the dynamic simulations, all PI control structures cope well with persistent disturbances. However, the control structure DV/LSB and LV/DSB make the DWC return to steady state only after a long settling time (>1000 min).

The LQG controller with feed forward control has only good results for disturbances in the feed flowrate. For other disturbances the tuning of the feed forward terms is less straightforward. The controller has no feedback on the error term that is the difference of the setpoints and the measured values. As a result offset in the product purities appears. This problem is solved by combining the LQG controller with an integral term.

A stop criterion is used for all test cases in order to have a fair comparison of the controllers – the simulation is stopped if the condition $\|(x_A, x_B, x_C) - (0.97, 0.97, 0.97)\|^2 < 1e^{-10}$ holds at time t_1 and also holds at time $t_2 = t_1 + 40$ min, where $t_1 < t_2$. The dynamic responses of the DWC at persistent disturbances – smaller settling times meaning better control – are shown in the next figures.
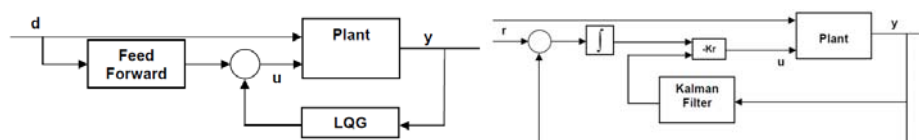
Figure 4. LQG controller with feed-forward (left), or extended with integral action (right).

Plotting the RGA number vs frequency clearly distinguishes between the LV/DSB and DB/LSV control structures, where the DB/LSV option is preferable to LV/DSB. The pairing x_A-V and x_B-V is predicted and proved to be more effective than x_B-L and x_C-L.

Figure 5. Dynamic response of the DB/LSV control structure, at a persistent disturbance of +10% in the feed flow rate (left) and +10% x_A in the feed composition (right).
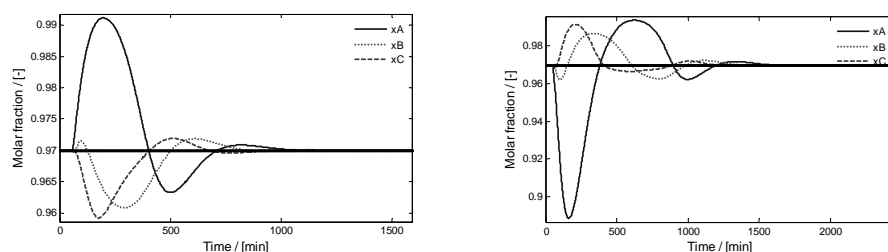
Figure 6. Dynamic response of the DV/LSB control structure, at a persistent disturbance of +10% in the feed flow rate (left) and +10% x_A in the feed composition (right).
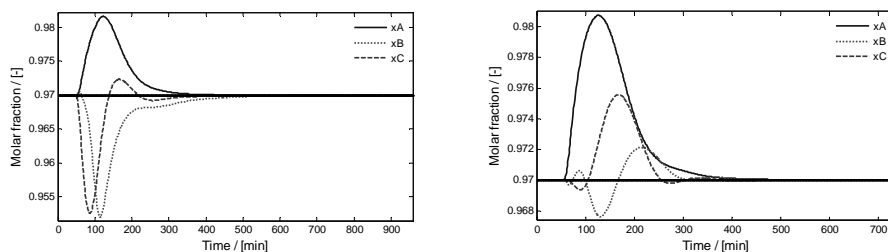
Figure 7. Dynamic response of the LB/DSV control structure, at a persistent disturbance of +10% in the feed flow rate (left) and +10% x_A in the feed composition (right).
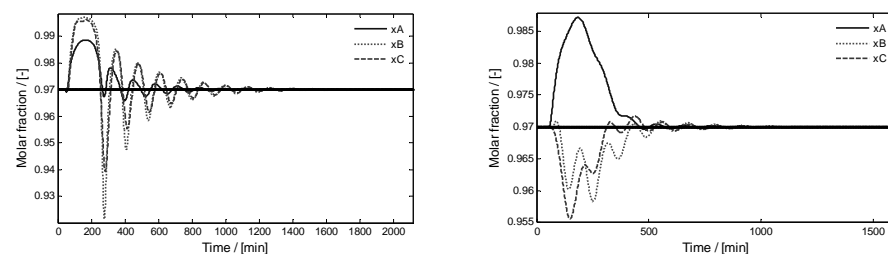
Figure 8. Dynamic response of the LV/DSB control structure, at a persistent disturbance of +10% in the feed flow rate (left) and +10% x_A in the feed composition (right).

Advanced Control Strategies for Dividing-Wall Columns
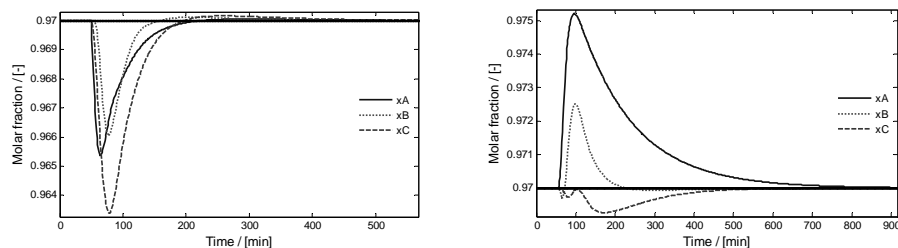
Figure 9. Dynamic response of the LQG controller combined with integral action, at a persistent disturbance of +10% in the feed flow rate (left) and +10% xA in the feed composition (right).
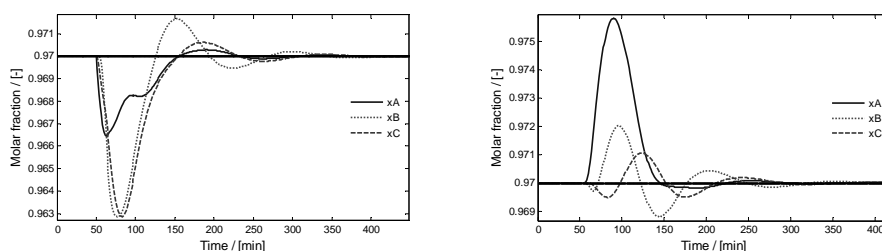
Figure 10. Dynamic response of the LSDP-controller, at a persistent disturbance of +10% in the feed flow rate (left) and +10% xA in the feed composition (right).
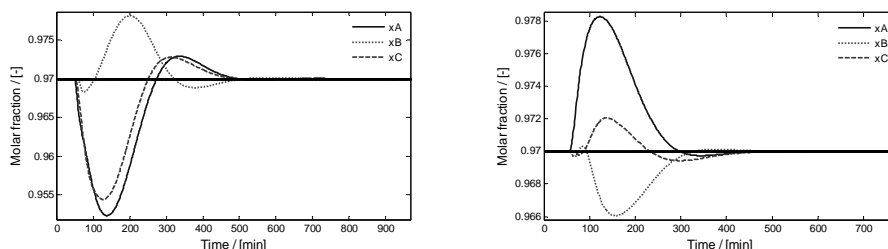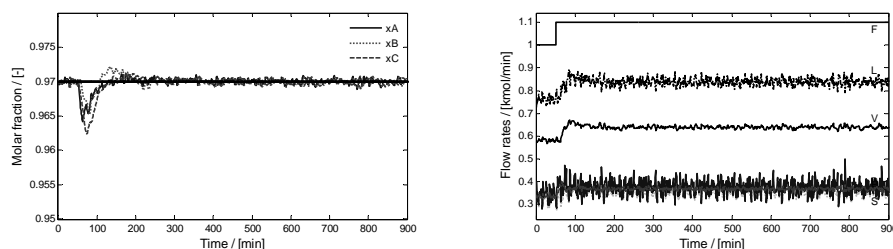

Figure 11. Dynamic response of the μ-controller, at a persistent disturbance of +10% at t=50min in the feed flow rate while there is white measurement noise and a time delay.

The dynamic simulations show no control or stability problems of the closed loop system. Furthermore, there is a trade off between a short settling time in the case of no measurement delay and noise, and a very smooth control action in case of measurement noise. A short settling time results in a more chaotic control if noise is present.
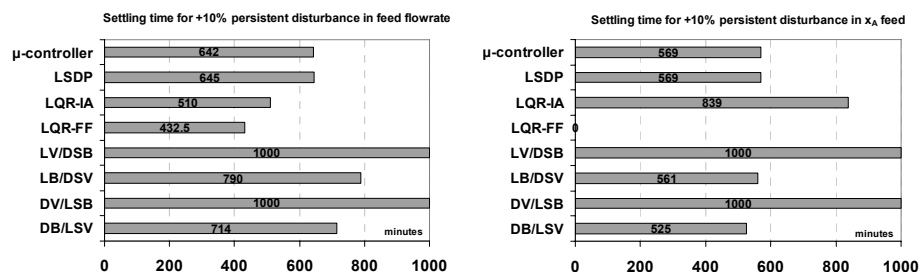
Figure 12. Settling time for +10% disturbance in the feed flow rate and X_A in the feed.

6. Conclusions

The advanced control strategies presented in this work were applied to a DWC separating the ternary mixture benzene-toluene-xylene. The results provide significant insight into the controllability of DWC, and gives important guidelines for selecting the appropriate control structure. The dynamic model of the DWC used in this study is not a reduced one, but a full-size non-linear model that is representative for industrial separations. Due to practical considerations based on the physical flows, there are basically four control strategies possible based on PID loops within a multi-loop framework: DB/LSV, LB/DSV, DV/LBS, LV/DSB. The results of the dynamic simulations show that the first two are the best among the decentralized multivariable PI structured controllers, being able to handle persistent disturbances in short times.

The DWC model is not only non-linear but also a true multi-input multi-output (MIMO) system, hence the applicability of a MIMO control structure starting with a LQG controller was also investigated. Two options were explored: feed forward control and addition of an integral action. The LQG-FF controller has good results for a persistent disturbance in the feed flowrate. However, for changes in the feed composition and condition it is difficult to find a good tuning. Moreover, persistent disturbances other than the ones used for tunning cannot be controlled with LQG. Nevertheless, combining LQG with an integral action and reference input solves the problem. Moreover, robustness against measurement noise results in a more conservative tuning.

The loop-shaping design procedure (LSDP) used in this work leads to a feasible μ-controller that has some additional benefits, while specific model uncertainties can be incorporated in the control structure. However, reduction of the LSDP controller is not possible since the reduced controller is unable to control the column. In contrast, the μ-controller can be reduced while still having a good control performance. In the DWC case described here, the obtained μ-controller is able to minimize the settling time when handling persistent disturbances. While PI control structures are also able to control the DWC, significantly shorter settling times can be achieved using MIMO controllers. Moreover, persistent disturbance are efficiently controlled using a MIMO controller.

7. Acknowledgements

References

R. Adrian, H. Schoenmakers, M. Boll, 2004, MPC of integrated unit operations: Control of a DWC, Chemical Engineering & Processing, 43, 347-355.

R. C. van Diggelen, A. A. Kiss, A. W. Heemink, 2010, Comparison of Control Strategies for Dividing-Wall Columns, Industrial & Engineering Chemistry Research, 49, 288-307.

I. J. Halvorsen, S. Skogestad, 1997, Optimizing control of Petlyuk distillation: understanding the steady-state behaviour, Computers & Chemical Engineering, 21, 249-254.

A. A. Kiss, J. J. Pragt, C. J. G. van Strien, 2009, Reactive Dividing-Wall Columns - How to get more with less resources?, Chemical Engineering Communications, 196, 1366-1374.

H. Ling, W. L. Luyben, 2009, New control structure for divided-wall columns, Industrial & Engineering Chemistry Research, 48, 6034-6049.

M. Serra, A. Espuña, L. Puigjaner, 2000, Study of the divided wall column controllability: influence of design and operation, Computers & Chemical Engineering, 24, 901-907.

Modelling and Explicit MPC of PEM Fuel Cell Systems

Christos Panos[a], Konstantinos I. Kouramas[a],Michael C. Georgiadis[b], Nigel Brandon[a], Efstratios N. Pistikopoulos[a]

[a]*Centre for Process Systems Engineering, Imperial College London, London SW7 2AZ, UK, E-mail: [christos.panos08,k.kouramas, e.pistikopoulos]@imperial.ac.uk*
[b]*Department of Engineering Informatics and Telecommunications, University of Western Macedonia, 50100 Kozani, Greece E-mail: mgeorg@uowm.gr*

Abstract

We present an analytical dynamic model and a general framework for the optimal design and control of a PEM fuel cell system. The mathematical model includes detailed model for the PEM fuel cell stack and simplified models for the compressor, humidifier and cooling system. The framework features (i) a detailed dynamic process model, and (ii) an explicit/multi-parametric model predictive controller design step. For the model based controller design, a reduced order approximate model is obtained from the dynamic simulation of the system.

Keywords: PEM fuel cell, Explicit/Multi-Parametric Model Predictive Control, Multi-parametric programming,

1. Introduction

Fuel cells are a promising technology for the electrical power generation, widely regarded as a potential alternative to internal combustion engines for the mobile applications. The electrical efficiency of the fuel cell is greater than the most conventional processes for electricity generation. The primary type of fuel cells under consideration for the automotive industry is Proton Exchange Membrane (PEM) fuel cells. They have the most suitable properties for vehicle applications such as fast start-up, low sensitivity to orientation and favorable power-to-weight ratio [2, 8]. However to compete with internal combustion engines, fuel cells must reach similar level of performance and life time. The main technical issue is that ground vehicle propulsion is a high challenging problem due to variety of transient behavior within the fuel cell system. The control system has to ensure that critical parameters are in their optimal values in order to optimize the operation of the system and avoid degradation that could damage the fuel cell stack.

There are many published mathematical models of PEM fuel cell in the literature with various level of complexity and could be classified into three categories, (i) the steady state mathematical models based on empirical equations, (ii) one dimensional dynamical models and (iii) two and three dimensional models based on navier-stokes equations. Amphlett et al. (1996) proposed a simple dynamic model which given a set of feed and operating condition could predict fuel cell voltage and stack temperature. Del Real et al.(2007) [1] presented a semi-empirical one dimension mathematical model with novel algorithm to calculate empirical polarization curve. Pathapati et al.[4] concentrated in the electrochemical behavior trying to best model the start-up sequence. Some authors developed more detailed models considering a gradient in stack temperature, two phase effect or implemented a two or three dimensional model [3].
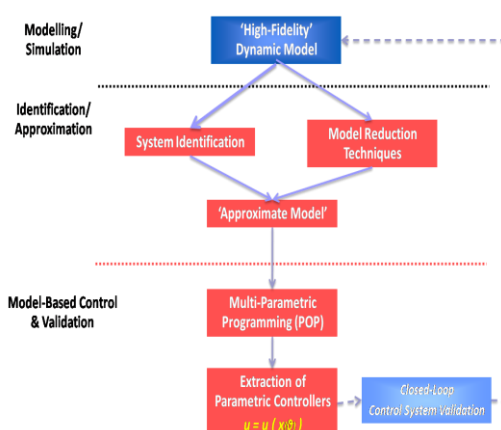
Figure 1. Framework for design explicit/Multi-Parametric MPC (Pistikopoulos, 2009)
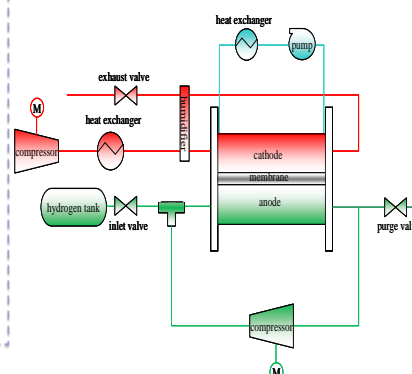
Figure 2. Overall scheme of the fuel cell system

In this paper we present a systematic framework for the optimal design and advanced control of the PEM fusl cell system. A detailed mathematical model is first presented, based on the work of [3] with the addition of heat balances. Dynamic simulation are performed based on which a reduced order state space (SS) model, suitable for the design of advanced model based controllers, is derived. Finally, the controller introduced in the actual process and its performance is validated.

2. A framework for optimal design, optimization and advanced control of PEM fuel cell systems

The proposed framework, presented in Pistikopoulos (2009) [5], is illustrated in figure (1) and consists of four key steps. Step 1 involves the development of a high-fidelity mathematical model, for performing dynamic simulation and design/operational optimization studies. In this work, we present a 1D semi-empirical mathematical model for PEM fuel cell system. In step 2, a reduced order approximated model is derived by performing system identification on the simulation data. Step 3 corresponds to the design of the multi-parametric/explicit Model Predictive Controllers (mp-MPC), by applying the available theory and tools of multi-parametric programming and control [6,7]. Finally step 4 involves the off-line validation of the derived explicit /multi-parametric controllers.

3. Mathematical Model of PEM Fuel Cell Systems

The system under consideration is a PEM fuel cell stack, a compressor, humidifier and a cooling system to maintain the temperature of the stack. Hydrogen is channeled in the anode side of the fuel cell while air in the cathode side. The compressor and the electric drive motor are used to achieve the desired air massflow and pressure, while the humidifier has been used to achieve proper humidity of the air in order to minimize the danger of dehydration of the membrane. In addition, a recycling system for hydrogen is applied to minimize the hydrogen consumption.

The 1D mathematical model, developed in this work, includes mass balances for the anode and cathode side and recirculation, semi-empirical equations for the membrane, electrochemical equations, heat balances for the fuel cell and mass and energy equation for the humidifier, compressor and the cooling system. In this section we present the

most important equations of the fuel cell mathematical model. The complete model consists of 82 dynamical algebraic equations (DAE).

The mass continuity is used to balance the mass of the elements inside the cathode and anode side,

$$\frac{dm_{O_2,ca}}{dt} = \dot{m}_{O_2,ca,in} - \dot{m}_{O_2,ca,out} - \dot{m}_{O_2,reacted}, \frac{dm_{N_2,ca}}{dt} = \dot{m}_{N_2,ca,in} - \dot{m}_{N_2,ca,out} \tag{1}$$

$$\frac{dm_{v,ca}}{dt} = \dot{m}_{v,ca,in} - \dot{m}_{v,ca,out} + \dot{m}_{v,membr} + \dot{m}_{evap,ca} \qquad \frac{dm_{H_2,an}}{dt} = \dot{m}_{H_2,an,in} - \dot{m}_{H_2,an,out} - \dot{m}_{H_2,reacted}$$

$$\frac{dm_{v,an}}{dt} = \dot{m}_{v,an,in} - \dot{m}_{v,an,out} - \dot{m}_{v,membr} + \dot{m}_{evap,an}, \frac{dm_{l,an}}{dt} = \dot{m}_{l,an,in} - \dot{m}_{l,an,out} - \dot{m}_{evap,an}$$

where $m_{i,ca}$ is the mass flowrate in cathode side (i is O_2, N_2, vapour) and $m_{j,an}$ is the mass flowrate in anode side(j is H_2, vapour, liquid), $m_{O2,react}$ is the reacted oxygen, $m_{v,memb}$ is the water mass flowrate across the membrane, m_{evap} is the evaporation mass and $m_{H2,react}$ is reacted mass of hydrogen.

The water transport through the membrane is achieved through two distinct phenomena. Water modules are dragged in the membrane from anode to cathode (electro-osmotic drag) while the humidity difference in anode and cathode flow is generating the back-diffusion of the water.

$$N_{v,memb} = nd\frac{I}{A_{fc}F} - D_w\frac{c_{v,ca} - c_{v,an}}{\delta_{memb}} \tag{2}$$

where I is the stack current, A_{fc} is the active area, F is the Faraday number, D_w is the diffusion coefficient, $c_{v,ca}$ and $c_{v,an}$ are the water concentration in cathode and anode channel, respectively, and δ_{memb} is the membrane thickness.

The total stack mass flowrate across the membrane can be calculated from

$$\dot{m}_{v,memb} = N_{v,membr}M_vA_{fc}N_{fc} \tag{3}$$

The steady state electrochemical model has been used to predict stack voltage output. The cell voltage was defined as follows,

$$V_{st} = N_{fc}\left(E_{Nernst} - V_{act} - V_{ohm} - V_{conc}\right) \tag{4}$$

where N_{fc} is the number of stacks, E_{Nerst} is the Nerst voltage, V_{act} the activation loss, V_{ohm} the ohmic loss and V_{conc} is the concentration voltage.

Heat is generated during the operation of the fuel cell, thus the cooling fluid is necessary in order to avoid damaging the fuel cell. The overall energy balance is expressed as follows

$$m_{st}Cp_{st}\frac{dT_{st}}{dt} = \dot{Q}_{in} - \dot{Q}_{out} - \dot{Q}_{chem} - \dot{Q}_{elec} - \dot{Q}_{cool} - \dot{Q}_{rad} - \dot{Q}_{latent} \tag{5}$$

where Q_{chem} is the heat generated by the chemical reaction, Q_{elec} is the heat flow in the form of electricity, the Q_{rad} is the heat exchange by radiation, Q_{latent} is the heat flowrate due to the change in water phase , Q_{cool} is the coolant heat flow and Q_{in}, Q_{out} are the inlet and outlet heat flow ,respectively. In the following section the model was used to perform dynamic simulation in order to design the reduced order linear multi-parametric MPC controller.

4. Model Identification

A reduced order SS model is designed with model identification from the data simulations of the PEM fuel cell process. The input/output data are obtained from simulation with Simulink of the nonlinear system along the given set points, while the SS model parameters are obtained with Matlab Identification Toolbox. The mathematical representation of the SS model without disturbances and with sampling time equal to 1 sec is as follows

$$x(t+1) = Ax(t) + Bu(t), \quad y(t) = Cx(t) \tag{6}$$

where y are the temperature and the voltage of the stack and u is the mass flowrate and temperature of the coolant, and the voltage of the compressor. The system matrices are given as follows,

$$A = \begin{vmatrix} 0.993 & 0.004 & 0.0009 & -0.001 \\ 0.013 & 0.992 & 0.022 & 0.007 \\ -0.015 & -0.003 & 0.707 & -0.086 \\ -0.001 & -0.005 & 0.0069 & 0.986 \end{vmatrix}, \quad B = \begin{vmatrix} -2.73e\text{-}006 & -0.00070165 & 2.090e\text{-}005 \\ -0.003405 & 0.0050739 & -2.53e\text{-}005 \\ 0.0016905 & -0.042051 & -0.0001598 \\ -2.5e\text{-}005 & 0.0022222 & -4.27e\text{-}006 \end{vmatrix},$$

$$C = \begin{vmatrix} 394.3 & -8.715 & 0.0438 & -0.369 \\ 45.10 & -19.16 & -0.112 & -0.104 \end{vmatrix}$$

A comparison between the reduced order SS model (green line) and the high-fidelity dynamic model (black line) is shown in Figure 3. The SS model closely approximates the behavior of the process with a small approximation error (difference between the actual process output and the SS output) of 7%.

5. Explicit/Multi-parametric Model Predictive Control (mp-MPC) Design

The next step in the general framework (Figure 1) involves the design of multi-parametric model predictive controller of the PEM fuel cell system. In model predictive control, the constrained optimal control problem is solved repetitively at each sampling instant given the current process measurements and references points, to obtain the current and future control actions up to a certain time horizon, based on the future predictions of the states and outputs by using a mathematical linear or nonlinear mathematical representation of the system. The benefits of MPC have long been recognized for optimal control design. Nevertheless, its application may be restricted due to increased online computational requirements related to the constraint optimization. In order to overcome this drawback, explicit/multi-parametric MPC was developed [6,7] which avoid the need for repetitive online optimization. In mp-MPC the optimization problem of the MPC is solved off-line by parametric optimization to obtain the optimal solution as an optimal mapping of the current state, output measurements and reference trajectory instead of demanding online optimization.
The following MPC formulation is considered for the PEM fuel cell control system

$$\min_{u_{t+1}, \ldots, u_{t+N_u}} J = \sum_{i=1}^{N_y} Q\left(y_i - y_{ref,i}\right)^2 + \sum_{j=0}^{N_u} R\left(u_j\right)^2$$

$$\text{s.t.} \quad x(t+1) = Ax(t) + Bu(t)$$

Modelling and Explicit MPC of PEM Fuel Cell Systems

$$y(t) = Cx(t) + d$$

$$\begin{bmatrix} 30 \\ 0.01 \\ 300 \end{bmatrix} \leq U_i \leq \begin{bmatrix} 150 \\ 0.8 \\ 340 \end{bmatrix}, \begin{bmatrix} 25 \\ 300 \end{bmatrix} \leq Y_i \leq \begin{bmatrix} 60 \\ 360 \end{bmatrix} \tag{7}$$

$$i = 1, \ldots N_y, \, j = 0, \ldots, N_u - 1$$

$$Q = 100, \, \rho = 0.01, \, R = 1$$

where y is the controlled variables, u are the manipulated variables, N_u is the control horizon (N_u=2) and N_y the prediction horizon (N_y=10). The MPC takes into account the operational limitation of the manipulated variables (input constraints) and the controlled variables (output constraints). The problem involves six optimization variables($m_{cool(t+1)}, T_{cool(t+1)}, V_{comp(t+1)}, m_{cool(t+2)}, T_{cool(t+2)}, V_{comp(t+2)}$) and eight parameters($\Theta = [x_t, T_{st}, V_{st}, T_{st,sp}, V_{st,sp}]^T$) which represent the states at time zero, the measurements of stack temperature and voltage and the reference trajectory of the stack temperature and the voltage. The objective function is set to minimize the quadratic norm of the error between the output variables and the reference points.
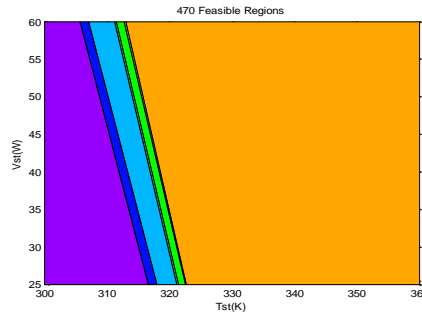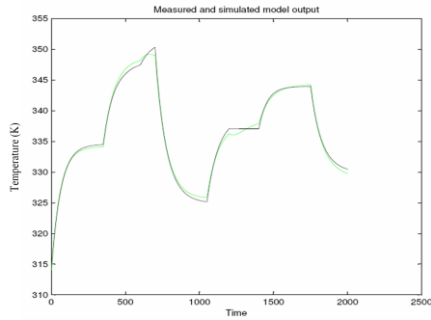
Figure 3. Process and SS model comparison   Figure 4. Critical regions of mp-MPC

The optimization problem (8) is a multi-parametric Quadratic Programming problem and can be solved with standard multi-parametric techniques (mp-QP) [6,7]. In our study the explicit parametric controller was derived with the Parametric Optimization (POP) software. The optimal map (Figure 4) of the control variables consists of 470 feasible regions and the corresponding control laws.

The derived controller introduced in the dynamical process model applying at each sampling time the following explicit control law

If x in $A_i x \leq b_i$ **Then** $U_f = K_i x + c_i$ $\tag{9}$
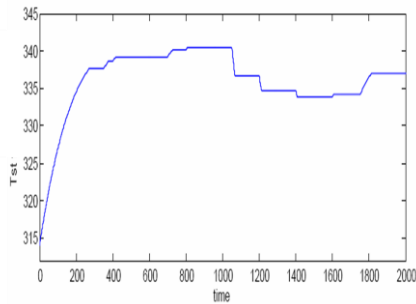
Figure 5. Stack temperature   Figure 6. Stack voltage
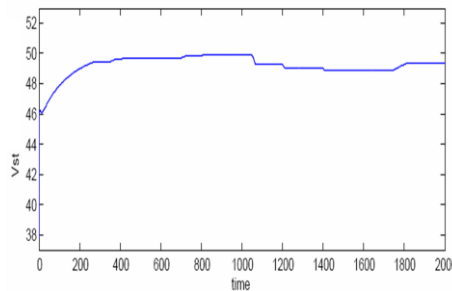
Figures 4 and 5 depict the simulation results of the mp-MPC implementation for different operating conditions (set points). The controller manages to maintain the variables at the desired reference values for different values of current while satisfied the constraints. The controller showed fast response to temperature and voltage set point changes for different values of current while managed to keep the output constraints in the feasible range.

6. Conclusions

In this work a general framework for multi-parametric programming and explicit/multi-parametric MPC control was applied in PEM fuel cell system. This framework allows the design and off-line implementation of multi-parametric controllers based on the simulation results of the high-fidelity mathematical model of the process. Ongoing work focuses on the implementation and evaluation of the explicit/multiparametric MPC for the case of real PEM fuel cell system.

7. Acknowledgements

References

[1] Del Real, A.J., Arce, A., Bordons, C. (2007) Development and experimental validation of a PEM fuel cell dynamic model. Journal of power sources, 173 (1). 310-324.

[2] Golbert, J. & Lewin, D.R. (2007) Model-based control of fuel cells: (2) optimal efficiency. Journal of power sources, 173 (1). 298-309.

[3] Muller, E. & Stefanopoulou , A.G. (2006) Analysis, modeling, and validation for the thermal dynamics of a polymer electrolyte membrane fuel cell system. Journal of fuel cell science and technology, 3 (2). 99-110.

[4] Pathapati, P.R., Xue, X. & Tang, J. (2005) A new dynamic model for predicting transient phenomena in a PEM fuel cell system. Renewable energy, 30, (1). 1-22.

[5] Pistikopoulos, E.N (2009), Perspectives in multiparametric programming and explicit model predictive control, Journal of American Institute of Chemical Engineering, 55, 1918

[6] Pistikopoulos, E.N., Georgiadis, M.C. & Dua, V. (2007a), Multi-parametric Programming: Theory, Algorithms and Applications, ISBN-978-3-527-31692-2, Wiley-VCH.

[7] Pistikopoulos, E.N., Georgiadis, M.C. & Dua, V. (2007b) Multiparametric model-based control, ISBN-978-3-527-31692-2, Wiley-VCH.

[8] Pukrushpan, J.T., Stephanopoulou, A.G. & Peng, H. (2005) Control of Fuel Cell Power Systems. Principles, Modelling, Analysis and Feedback Design. Grimble, M.J. et al (ed.), London Springer.

"Mega"-variate statistical process control in electronic devices assembling

Marco S. Reis,[a] Pedro Delgado [b]

[a]*CIEPQPF, Department of Chemical Engineering, University of Coimbra, Rua Sílvio Lima, 3030-790, Coimbra, Portugal, marco@eq.uc.pt*
[b]*Bosch Car Multimedia Portugal, Lda., Pedro.Delgado@pt.bosch.com*

Abstract

We analyze the assembly process of electronic devices, in particular the initial deposition of Solder Paste Deposits (SPD) over Printed Circuits Boards (PCB), that will later on provide the necessary fixation for all the electronic components as well as functionalize their operation. In this stage, thousands of SPD's, differing in shape and volume, are quickly and accurately placed in different positions of the PCB's. Monitoring the status of this operation raises very important problems, particularly during the initial production runs, as the number of quality features under monitoring is very large (order of thousands) and the number of samples available quite low (order of dozens). In this work, we propose an efficient approach for addressing the on-line and at-line monitoring of this process, addressing two hierarchically related problems: i) detection of faulty units (PCB's); ii) given that a faulty unit was detected, find a candidate set of SPD's responsible for the anomaly. Our methodology is based on a latent variable framework for effectively extracting the normal behavior of the process from the few reference samples available, and using it to classify the following samples as normal or abnormal and, in this case, analyze why it happens to be so. We have tested the proposed approach with real industrial data, and the results achieved illustrate its good discrimination ability, rendering it very promising for implementation in this class of scenarios.

Keywords: Printed Circuits Boards, Principal Components Analysis, Multivariate Statistical Process Control, Receiver Operating Characteristic

1. Introduction

The assembly of electronic devices usually consists of a sequence of stages performed with highly automated machinery under well controlled environmental conditions in order to produce functional units at high rates. Even though the high production rates achieved in these processes, the cost associated with each unit under processing may be economically significant and increases steeply as the product moves forward in the assembly line. Therefore, any malfunction in a unit under processing must be promptly signaled, analyzed for its relevancy according to well defined criteria after which a decision must be made concerning its acceptance or rejection. Furthermore, if a faulty pattern emerges or can be anticipated, it must be quickly corrected, in order to drive the process back to the desirable operation conditions. However, several important problems emerge in this context.
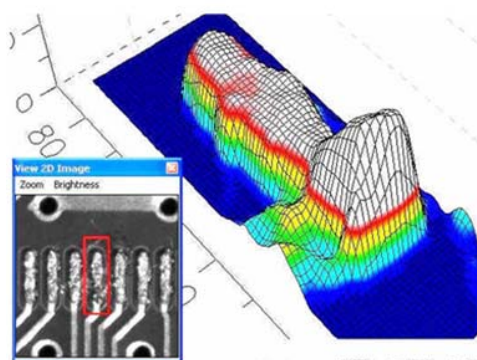
Figure 1. Several Solder Paste Deposits (SPD's) in a Printed Circuit Board (lower left), along with a three-dimensional profile (larger image in the background), obtained through Moiré interferometry, regarding an SPD with shape problems.

Firstly, in the assembly process, several thousands of Solder Paste Deposits (SPD) are placed at specific positions in the Printed Circuits Boards (PCB) (Figure 1). They will hold and connect all the electronic components that will be added at a later stage which, in the current situation, will be fixed through a "reflow" soldering process. The shape of these deposits varies significantly and is dependent upon the type and size of the components to be assembled, playing a central role in the proper constitution of the solder joints for each individual component and, ultimately, on the reliability of the whole electronic device. Therefore, one must properly address the problem of simultaneously monitor the shape and position of thousands of SPD's.

The second problem as to do with the low number of samples available for deriving a Statistical Process Control (SPC) approach for this problem, in order to assist operators in identifying abnormal samples. In fact, the most problematic period occurs in the beginning of each new production run, where the process parameters are not all optimized and fine tuned, which is also the one where normal operation data is still scarce. In our case, the set up process is speed up by applying a general set of conservative rules regarding the acceptable geometrical dimensions for the solder deposits depending on the type of components they are relative. This is good enough for initiating the production, but is not very sensitive in the detection of significant changes in the solder deposits shape. In this context, one must define the "normal operation conditions" with the few "good" samples initially available, in order to begin implementing SPC as soon as possible.

Finally, the third challenge is that everything must be implemented very efficiently, as the operator as less than one minute to decide what to do with the unit signaled as faulty. In this work, we propose an approach for hierarchically detect if a given unit is normal or abnormal taking into consideration information regarding all the SPD's applied to each printed circuit (in our case this amounts to 3084 different deposits) and, if this happens to be so, point out which deposits are potentially causing such abnormal behavior.

This article is organized as follows. In the next section, we describe the measurements collected from the process and present the proposed methodology for addressing the two hierarchically related problems. Then, in the following section, we present some results regarding the implementation of this methodology, in order to illustrate its validity and

"Mega"-variate statistical process control in electronic devices assembling

potential usefulness in real world scenarios. Finally, we summarize our contributions in the final conclusions section.

2. Materials and Methods

2.1. Dataset

A total of 31 records were collected for analysis, each one of them concerning a different unit under processing (PCB). Each record contains all the information acquired in a routine way, in particular, the height (h), area (a) and volume (v) for each of the 3084 SPD's (using Moiré interferometry), as well as their positioning offset in the X and Y directions, relative to the target points in the reference coordinate system, i.e. a total of 5x3084=15420 measurements for each unit. From the set of 31 records, 15 were classified as "good", whereas the remaining 16 present different kinds of problems that originate a final classification of "fail" for all of them. In order to develop a Statistical Process Control approach aimed at supporting process operators in their demanding task of quickly deciding about the status of each unit under processing (normal/abnormal), using the scarce amount of information available at the beginning of each new production run, we have selected, from the original 31 records, 10 classified as "good" for establishing the "normal operating conditions" (reference set), while the others, classified both as "fail" (16) and "good" (5), were used to assess the sensitivity (those labeled as "fail") and specificity (samples classified as "good") of the proposed approach (test set). The inclusion of "good" samples in the test set is important, because otherwise the results could be overlying optimistic. In particular, the detection of faulty units (sensitivity) could be inflated at the expense of an unacceptably high false alarms rate (specificity).

2.2. Methods

In order to effectively capture the structure of the "normal operation conditions" variability (NOC) provided by data, a latent variable approach based on Principal Components Analysis (PCA) was applied, based on which a Statistical Process Control approach was developed: Multivariate Statistical Process Control based on PCA (PCA-MSPC) [1, 2]. This approach essentially consists in extracting the structural relationships among the variables through PCA, leading to a lower dimensional subspace containing most of the NOC data variability. Normal operation data will then fall in a limited region of this space when orthogonally projected, and will cluster around this subspace with residuals of similar low magnitude. Therefore, by using two monitoring statistics, one relative to the quantification of the within PCA variability, T^2, and the other to the variability around the PCA subspace, Q, which are defined by equations (1) and (2), respectively, it is possible to conduct a Statistical Process Control activity that encompasses the simultaneous consideration of many variables (in our case much more than the number of observations or samples available), something that would be impossible to achieve with the traditional multivariate procedure based on the Hotelling's T^2 statistic (as the sample covariance matrix would not even be invertible) [3].

$$T_i^2 = t_i \Lambda^{-1} t_i^T \tag{1}$$

$$Q_i = e_i e_i^T \tag{2}$$

(t_i stands for the PCA scores of the ith observation, whose NOC covariance matrix is given by Λ; e_i is the ith residual vector). The PCA-MSPC approach is implemented in a stage-wise fashion, in order to address sequentially the two hierarchically related problems: *1st Level of Detection* - detect faulty units (PCB's); *2nd Level of Detection* - given that a faulty unit was signaled, find a candidate set of SPD's responsible for such abnormality. There are different possibilities for implementing these two conditionally related activities. A rational approach would be, for instance, to implement separate PCA-MSPC approach for each type of measurement available (h, a, v, off-set X and off-set Y), and signal a unit as faulty in the 1st Level of Detection if any of the T^2 or Q statistics were significant. As one must consider simultaneously 3084 quality features for each such model, this approach is highly justifiable, because combining all the measurements in a single large measurement vector would worsen the already highly unbalanced situation between the number of variables considered and the number of samples available and, furthermore, will not take advantage of the different nature of the measurements used. Alternatively, we have also tested an approach that consisted in combining all the T^2 or Q statistics made available from such a scheme (a total of 10), and again develop a second level of PCA-MSPC based on them, in order to capture some expected shared variability among these statistics (combined approach).

As to the 2nd Level of Detection, it was always based on the analysis of residuals for each new observation, regarding each variable (SPD). An SPD with a very high residual (positive or negative) would identify a potential faulty solder deposit. All data was previously autoscaled (i.e., transformed to zero mean and unit variance).

In the next section we present some of the results obtained for the 1st and 2nd levels of detection, using real data, by computing the associated empirical receiver operating characteristic (ROC) curves, that provide a clear and rigorous mean for assessing the power of the proposed methodology concerning the two (usually conflicting) goals to be observed: to achieve good detection ability, while avoiding high false alarm rates.

3. Results

3.1. *1st Level of Detection*

We have analyzed the performance in detecting faulty units (PCB's) using PCA-MSPC based on measurements of a given type (h, a, v, off-set X and off-set Y) as well as using the combined approach. In all situations, the PCA-MSPC parameters were first estimated using the reference data (10 samples classified as "good"), and then the ROC curves computed for the test set. The ROC curve is a representation of the rate of true detections (True Positive Rate, TPR) as a function of the rate of false detections (False Positive Rate, FPR), as the threshold parameter of the monitoring methods vary from very low values (where the TPR is high, but the FPR is also high) to very high values (where the TPR is low and the FPR is also low). Ideally, in a situation where "good" and "fail" samples were perfectly discriminated, the ROC curve will immediately take the value of TPR=100% as soon as one allows for some positive FPR, i.e., as soon as one lower the limits to such an extent that a false alarm occurs, then all true positives (samples "fail") will be detected. Figure 2 shows the ROC curves obtained when our approach is based on the area and for the combined approach, where it is possible to see that, notwithstanding the very low number of samples used in the reference set, the discrimination power achieved is still quite remarkable. This figure also contains the relative area under the ROC curve, which is a monotone increasing index of the discriminatory capability of the method, which is upwardly limited by 100%.

"Mega"-variate statistical process control in electronic devices assembling
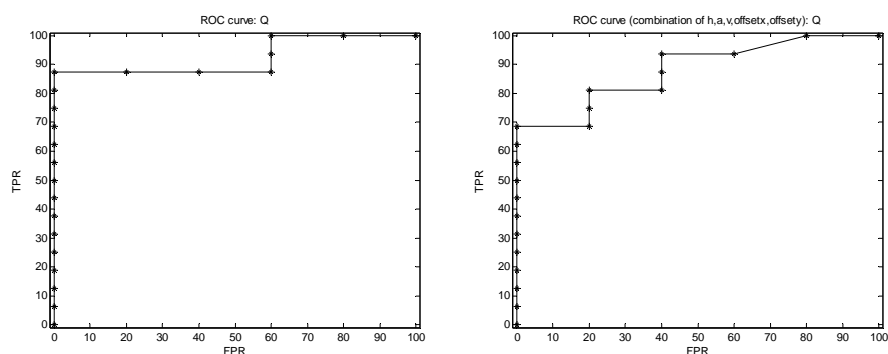
a)*Relative ROC area = 93.75%*   b) *Relative ROC area = 90.63%*

Figure 2. ROC curves for the 1st level detection using the Q statistic from PCA-MSPC based on (a) measurements of area and (b) the combined approach.

The results obtained for the remaining alternatives concerning this parameter are presented in Table 1, where one can verify that the discriminatory power attained with the Q statistic is always superior to that obtained with the T^2 statistic. The approaches based on height, area and the combined approach, present the best discrimination scores.

Table 1. Relative ROC area (%) for all the approaches tested for the 1st level of detection.

Detection statistics	Measurements used to compute the relative area under the ROC curve					
	Height (h)	*Area (a)*	*Volume (v)*	*Offset X*	*Offset Y*	*Combined approach*
T^2	70.00	62.50	85.63	76.88	70.63	90.00
Q	93.13	93.75	91.88	85.00	83.13	90.63

3.2. 2nd Level of Detection

Having presented the results regarding the capability for detecting faulty units, we turn now our attention to the second problem (2nd level of detection): given that a unit was signaled as faulty, which SPD's are responsible for such an anomaly? As already stated, the procedure followed here is entirely based on the analysis of residuals for each SPD. More specifically, after computing the projection of the vector of measurements for each new unit onto the PCA subspace estimated from reference data, one computes the vectorial difference from the original vector and the one obtained after the projection operation, leading to the residuals for each variable (SPD). Then, using reference data, it is possible to compute the standard deviation of such residuals of each SPD and to establish k-sigma statistical limits to use later on ($k \in R$), in order to decide which SPD's are faulty (i.e., falls outside the region bounded by the k-sigma control limits). By changing k, it is possible to compute the ROC curve for each unit or PCB, as the SPD's with problems were annotated, even though using quite conservative rules. The 2nd level of detection is a quite demanding problem, as is consists of figuring out which subset, among the 3084 SPD's, are causing the abnormality detected in the 1st level. Figure 3 illustrates the analysis for a faulty PCB (#4), where the SPD's with problems are signaled with circles in Figure 3.a) and the corresponding ROC is shown in Figure 3.b).
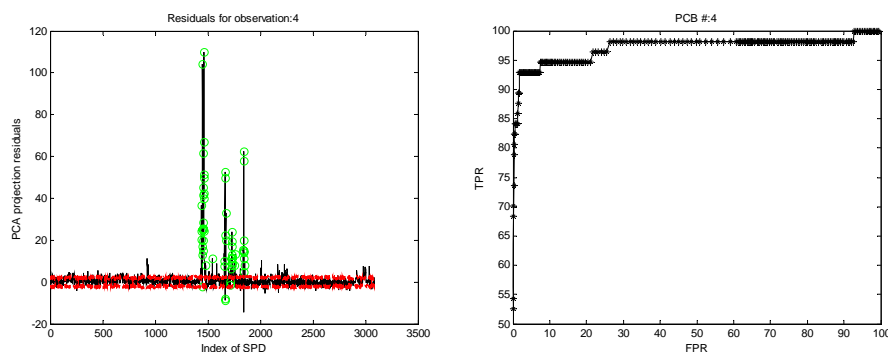
a) Residuals for sample (PCB) #4.   *b) Relative ROC area = 96.70%*

Figure 3. *a)* Residuals obtained for the test sample #4 (approach based on height; the circles signal SPD's annotated as abnormal, whereas the light/red line indicates 3-sigma control limits), and b) the ROC curve concerning the detection of abnormal SPD's in this sample.

The analysis for all the faulty units appears summarized in Table 2, once again for all PCA-MSPC approaches considered, where those with higher detection power are now the one based on height measurements and the combined approach.

Table 2. Mean and standard deviation obtained for the area under the ROC curves obtained for all the faulty test samples analyzed (2nd level of detection).

Summary statistics	Measurements used to compute the relative area under the ROC curve					
	Height (h)	*Area (a)*	*Volume (v)*	*Offset X*	*Offset Y*	*Combined approach*
Mean	80.33	65.82	76.04	60.23	54.38	80.33
Std. Dev.	20.07	29.97	21.02	21.23	29.93	20.07

4. Conclusions

In this work we address an important problem in the assembly of microelectronic components, namely the monitoring of its early stage when solder paste deposits are placed is specific positions to hold and functionalize electronic components. The current approach, based on components-specific conservative rules, is prone to overlook finer but still significant problems in solder deposits, as well as disregards the correlation structure among the thousands of SPD's deposited in each unit. Our procedure circumvents these difficulties, being able to detect quite well the existent abnormalities without incurring in large false alarms rates. Furthermore, all the computations can be conducted with a limited amount of samples, supporting operators in the initial production stages, where specific process knowledge and data availability is still scarce. Future work will optimize the combination of the several measurements available, namely through "OR" logical operations, by exploring their complementariness.

References

[1]   J.E. Jackson, Technometrics, 1:4 (1959) 359-377.
[2]   B.W. Wise, N.B. Gallagher, J. Process Control, 6:6 (1996) 329-348.
[3]   D.C. Montgomery, Introduction to Statistical Quality Control, Wiley, New York, 2001.

Influence of the moving window parameters on a posteriori standard deviations of estimated variables in dynamic data reconciliation

Christophe Ullrich[a], Carine Gerkens[a], Georges Heyen[a]

[a]Laboratoire d'Analyse et de Synthèse des Systèmes Chimiques, Université de Liège, *Sart-Tilman B6A, B-4000 Liège (Belgium), E-mail :* cullrich@ulg.ac.be

Abstract

The method previously proposed to estimate the uncertainty of validated variables in steady state data reconciliation has been extended to dynamic data reconciliation.

The approach used in this article to estimate a posteriori variances in the case of dynamic date validation is based on the one described in [4]. Orthogonal collocations are used to discretize ODE. The influence of the window parameters on posteriori variances is studied. Results are presented for a perfectly mixed reactor with heat exchange and first order kinetic.

Keywords: dynamic data reconciliation, moving horizon estimation, a posteriori variances, orthogonal collocations

1. Introduction

Efficient process monitoring is a key issue in plant operation, since measurement errors are always present. To address this issue, data validation is nowadays routinely performed for steady state processes, but dynamic systems still present some challenges. Data validation uses measurement redundancy and model constraints to reduce measurement uncertainty and to calculate non measured state variables of the system. A posteriori variance for validated variables compared to raw measurements can be calculated for linear or linearized steady state systems.

Several methods enable to solve the dynamic data reconciliation problem. In a previous paper [5] we used NLP technique and orthogonal collocation [3] to discretize the ODE systems, as described in [1]. The evaluation of variance of estimates in dynamic data reconciliation as been developed in [4], this method is similar to one developed by Heyen et al [2] for steady state case.

In this publication we study the influence of two window parameters on the a posteriori variance:

- The size of the interpolation interval of the input variables
- The size of the moving window

Results are shown for a perfectly mixed reactor with a heat exchange system and first order kinetic.

2. Estimation of a posteriori variances

The dynamic data reconciliation method from which we developed the estimation of a posteriori variances is a moving horizon method. Orthogonal collocations have been chosen to discretize differential equations so that they are transformed into algebraic equations. Inputs are represented either by linear interpolations or by Lagrange interpolation polynomials. Optimization of the data reconciliation objective function

and solution of the collocation variables is carried out simultaneously by an SQP algorithm.

The validation window is defined by five parameters which are represented on figure 1:

- h_1 : measurement frequency
- h_2 : size of the interpolation interval of the input variables
- h_3 : size of the discretization interval of the differential state variables
- h_4 : size of the moving window (reconciliation horizon)
- h_5 : the move of the window after optimization
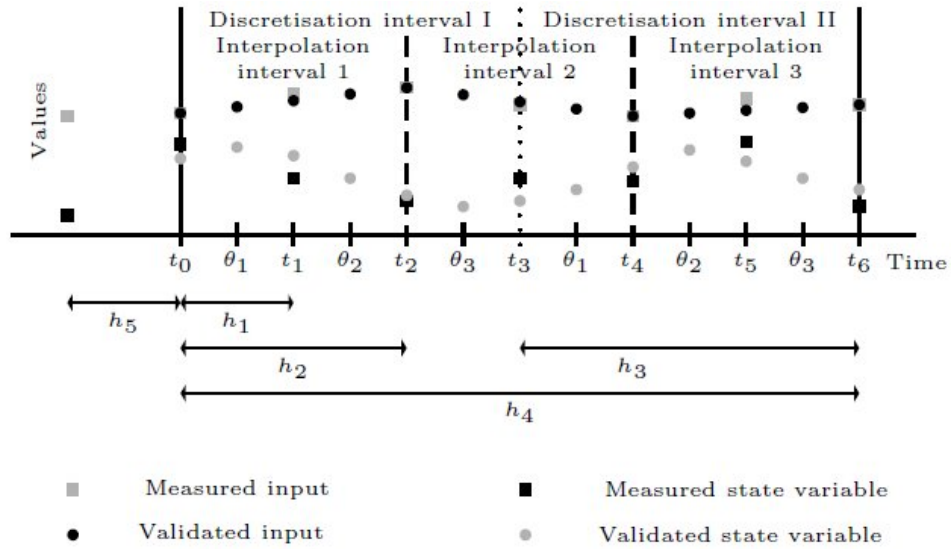
Figure 1. Moving window description

The objective function takes the following form in the case of our DDR method:

$$\min_{\mathbf{x}^c,\mathbf{u}} F = \left(\mathbf{x}-\mathbf{x}^m\right)^T \mathbf{P_x}\left(\mathbf{x}-\mathbf{x}^m\right) + \left(\mathbf{u}-\mathbf{u}^m\right)^T \mathbf{P_u}\left(\mathbf{u}-\mathbf{u}^m\right) + \left(\mathbf{z}-\mathbf{z}^m\right)^T \mathbf{P_z}\left(\mathbf{z}-\mathbf{z}^m\right) \tag{1}$$

Were **x**, **z** and **u** are respectively the differential, algebraic and input variables. This objective function is subjected to five kinds of constraints:

- A: the link equations (algebraic relationships between state variables and measured variables)
- B: relations between the differential state variables and the Lagrange interpolation polynomials
- C: linear interpolations of the values of input variables
- D: residuals of the differential state equations
- E: continuity constraints of the differential state variables between two discretization intervals

Optimization variables are the values of state and input variables at all collocation points. This constrained problem can be transformed into an unconstrained problem using Lagrange formulation. The necessary condition for optimality is expressed by setting to 0 the gradient of the Lagrangian. One obtains the sensitivity matrix M which is the Jacobian matrix of the equation system:

Influences of the moving window parameters on a posteriori standard deviations of estimated variables in dynamic data reconciliation

$$\mathbf{M} = \begin{pmatrix} P & E \\ E^T & 0 \end{pmatrix} \tag{2}$$

This matrix is composed of two submatrices :

- The P matrix whose elements are the inverses W of the a priori variances. In the case of input and differential state variables a relaxation term R is added for the first time of the window;
- The E matrix which is the Jacobian matrix of the constraint equations.

$$P = \begin{pmatrix} \mathbf{P_x} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{P_z} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{P_u} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \end{pmatrix}$$

$$\mathbf{P_x} = \mathbf{W_x} + \mathbf{R_x}$$
$$\mathbf{P_z} = \mathbf{W_z}$$
$$\mathbf{P_u} = \mathbf{W_u} + \mathbf{R_u}$$

$$E = \begin{pmatrix} \dfrac{\partial \mathbf{A}}{\partial \mathbf{x}} & \dfrac{\partial \mathbf{A}}{\partial \mathbf{z}} & \dfrac{\partial \mathbf{A}}{\partial \mathbf{u}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\[2mm] \mathbf{0} & \mathbf{0} & \mathbf{0} & \dfrac{\partial \mathbf{A}^c}{\partial \mathbf{x}^c} & \dfrac{\partial \mathbf{A}^c}{\partial \mathbf{z}^c} & \dfrac{\partial \mathbf{A}^c}{\partial \mathbf{u}^c} \\[2mm] \dfrac{\partial \mathbf{B}}{\partial \mathbf{x}} & \mathbf{0} & \mathbf{0} & \dfrac{\partial \mathbf{B}}{\partial \mathbf{x}^c} & \mathbf{0} & \mathbf{0} \\[2mm] \mathbf{0} & \mathbf{0} & \dfrac{\partial \mathbf{C}}{\partial \mathbf{u}} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\[2mm] \mathbf{0} & \mathbf{0} & \dfrac{\partial \mathbf{C}^c}{\partial \mathbf{u}} & \mathbf{0} & \mathbf{0} & \dfrac{\partial \mathbf{C}^c}{\partial \mathbf{u}^c} \\[2mm] \mathbf{0} & \mathbf{0} & \mathbf{0} & \dfrac{\partial \mathbf{D}}{\partial \mathbf{x}^c} & \dfrac{\partial \mathbf{D}}{\partial \mathbf{z}^c} & \dfrac{\partial \mathbf{D}}{\partial \mathbf{u}^c} \\[2mm] \mathbf{0} & \mathbf{0} & \mathbf{0} & \dfrac{\partial \mathbf{E}}{\partial \mathbf{x}^c} & \mathbf{0} & \mathbf{0} \end{pmatrix} \tag{3-4}$$

As for the stationary estimation [2], a posteriori variances can be deduced from this sensitivity matrix:

$$\mathrm{var}(\mathbf{x}_k) = \sum_{i=1}^{N_x} \left[\mathbf{M}_{k,i}^{-1} \mathbf{P}_{\mathbf{x}_i} \right]^2 \mathrm{var}(\mathbf{x}_i^m) + \sum_{i=1}^{N_z} \left[\mathbf{M}_{k,i+N_x}^{-1} \mathbf{P}_{\mathbf{z}_i} \right]^2 \mathrm{var}(\mathbf{z}_i^m)$$
$$+ \sum_{i=1}^{N_u} \left[\mathbf{M}_{k,i+N_x+N_z}^{-1} \mathbf{P}_{\mathbf{u}_i} \right]^2 \mathrm{var}(\mathbf{u}_i^m) \tag{5}$$

3. Example and results

This example comes from reference [1]. It consists of a perfectly mixed reactor with a heat exchange system, where a first-order exothermic reaction takes place. This system is described by two differential equations:

$$\frac{dC_A}{dt} = \frac{q}{V}(C_{A,0} - C_A) - \alpha_d k C_A$$
$$\frac{dT}{dt} = \frac{q}{V}(T_0 - T) - \alpha_d \frac{-\Delta H_r C_{A,r}}{\rho C_p T_r} k C_A - \frac{U A_R}{\rho C_p V}(T - T_c) \tag{6}$$
$$k = k_0 \exp\left(\frac{-E_A}{T T_r}\right)$$

This example is described by six variables:

- four input variables: the feed flowrate q, the feed concentration C_{A0}, the feed temperature T_0 and the cooling temperature T_C;

- two differential state variables: the concentration C_A and the temperature T.

The parameters of the reactor are listed in table 1.

Table 1. Reactor parameters

Parameters	Values	Units
V	1000	cm^3
ΔH_r	-27000	$cal\ mol^{-1}$
ρ	0.001	$g\ cm^{3\ -1}$
C_p	1	$cal\ mol^{-1}\ K^{-1}$
U	$5e^{-4}$	$cal\ cm^{2\ -1}s^{-1}K^{-1}$
A_R	10	cm^2
α_d	1	-
k_0	$7.86e^{12}$	s^{-1}
E_A	14090	K
$C_{A,r}$	$1e^{-6}$	$mol\ cm^{3\ -1}$
T_r	100	K

3.1. General influence of the window parameters on a posteriori variances

Input variables are represented by linear polynomials. State variables are discretized by second order Lagrange interpolation polynomials. The general influence of the window parameter studied on a posteriori variances are shown for two variables: an input variable, the feed concentration C_{A0} (figure 2) and a state variable, the concentration in the reactor C_A (figure 3). Reduction factors are the ratio of the variance of the original measurement to the variance of reconciled variables. We can see that increasing the size of the interpolation interval (from $h_2=4$ to $h_2=16$) of the input variable increases, allows to reduce the uncertainty of the estimates. In the same way if the length of the window increases, the reduction factor increases but soon reaches an asymptotic value (for instance, after a size of 16 for $h_2=4$)
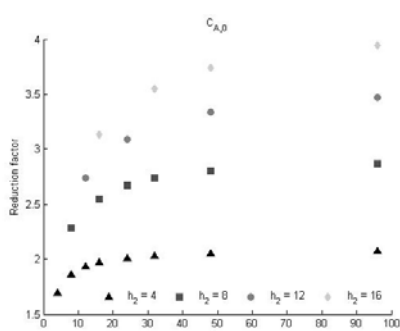
Figure 2. Reduction factor of a posteriori variance: C_{A0}
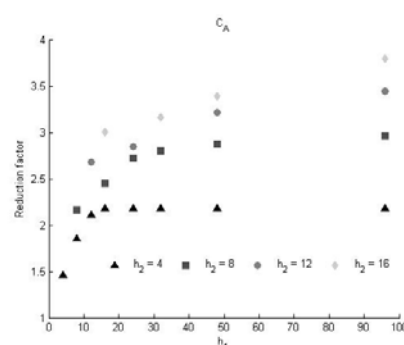
Figure 3. Reduction factor of a posteriori variance: C_A

3.2. Influence of the number of interpolation intervals for a fixed size of the window

In tables 2 and 3, we focus on the influence of the number of interpolation interval for a fixed size of the window. Results are described for a window of 48 measurements.

A reduction of the number of the interpolation interval (h_4/h_2) involves an increase of the size of interpolation interval (h_2) to conserve the window size fixed. The reduction

Influences of the moving window parameters on a posteriori standard deviations of estimated variables in dynamic data reconciliation

factor increases with the reduction of the number of interval, as shown in the column reduction factor. In the last column the percent of increase of the reduction factor is compared row by row. If we use $h_2=8$ instead of $h_2=4$ the reduction factor increase by a factor of 32 percent for C_A and 36 percent for C_{A0}. Similar conclusions can be obtained if we use other sizes of the window

Table 2. Windows size (h_4) 48: C_A

h_2	h_4/h_2	Constraint	Reduction factor	% gain
4	12	408	2,17	/
8	6	348	2,87	32,2%
12	4	328	3,21	11,8%
16	3	318	3,39	5,6%

Table 3. Windows size (h_4) 48: C_{A0}

h_2	h_4/h_2	Constraint	Reduction factor	% gain
4	12	408	2,05	/
8	6	348	2,8	36,5%
12	4	328	3,33	18,9%
16	3	318	3,74	12,3%

3.3. Influence of the size of the window for a fixed interpolation intervals

Table 4. $h_2=4$: C_A

h_4/h_2	h_4	Variable	Constraint	Reduction Factor	% gain
1	4	30	34	1,46	/
2	8	54	68	1,85	26,4%
3	12	78	102	2,11	14,1%
4	16	102	136	2,17	3%
6	24	150	204	2,17	0%
8	32	198	272	2,17	0%
12	48	294	408	2,17	0%
24	96	582	816	2,18	0,2%

Table 5. $h_2=4$: C_{A0}

h_4/h_2	h_4	Variable	Constraint	Reduction Factor	% gain
1	4	30	34	1,7	/
2	8	54	68	1,86	9,7%
3	12	78	102	1,93	3,9%
4	16	102	136	1,97	1,8%
6	24	150	204	2,01	2,1%
8	32	198	272	2,03	1%
12	48	294	408	2,05	0,9%
24	96	582	816	2,07	1%

We analyzed the influence of the window size for fixed interpolation intervals. Results are presented in tables 4 and 5 for interpolation intervals h_2 of size 4. We can note that if the size of the window increases, the reduction factor starts to increases, but reaches soon an asymptotic value, as can be seen with the percent gain between two rows equal

to 0. Similar conclusions can be obtained if we use other sizes of the interpolation interval as shown in the figures 2 and 3.

3.4. Other representation of the input variable

In the previous section the input variables are represented by a linear approximation. The same study of the influence of the window parameters has been carried out when input variables are represented by Lagrange interpolation polynomial. As can be seen on figures 4 and 5, similar conclusions can be obtained.
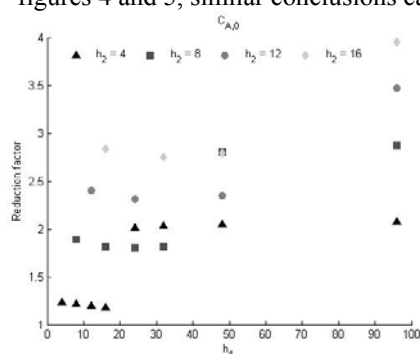
Figure 4. Reduction factor of a posteriori variance: C_{A0}
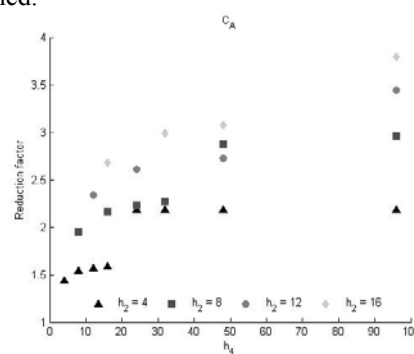
Figure 5. Reduction factor of a posteriori variance: C_A

4. Conclusions

In this paper, the influence of the size of the validation window and the size of the discretization intervals has been studied for a perfectly mixed reactor with a heat exchange and first-order exothermic reaction. Input has been represented by linear interpolations and Lagrange interpolation polynomials. Test show that an optimal validation horizon can be identified, allowing to maximize the reduction of variance of reconciled variables.

The study has been carried out for other examples such as networks of tanks and adiabatic reactor with different order of kinetic for which similar conclusions can be made. Nevertheless, optimal values of window parameters are problem specific.

5. Acknowledgements

References

[1] M.J. Liebman, T.F. Edgar, L.S. Lasdon, Computers & Chemical Engineering, No. 16, (1992), 963.

[2] Heyen, G., Maréchal, E., Kalitventzeff, B, Computers and Chemical Engineering, No. 20S, (1996), 539.

[3] J. Villadsen, M.L. Michelsen, (ed(s).), Solution of differential equation models by polynomial approximation, Prentice-Hall, Englewood Cliffs, New Jersey, 1978.

[4] Ullrich, C., Gerkens, C. Heyen, G., Variance of Estimates in Dynamic Data Reconciliation, Proceedings of ESCAPE-19, Cracow, Poland, June 14-17 2009

[5] Gerkens, C., Ullrich, C., Mateus, M., Heyen, G., Comparaison de techniques de validation dynamique de données, Proceedings of SIMO-06, Toulouse, France, October 11-12 2006

Plantwide Control of a Biodiesel Process by Reactive Absorption

Costin S. Bildea[1], Anton A. Kiss[2]

[1]*University "Politehnica" of Bucharest, Department of Chemical Engineering, Bucharest, Polizu 1-7, 011061 Bucharest, Romania. Email: s_bildea@upb.ro*
[2]*AkzoNobel – Research, Development and Innovation, Velperweg 76, 6824 BM, Arnhem, The Netherlands. Email: tony.kiss@akzonobel.com*

Abstract

Integrated biodiesel processes based on reactive separations powered by solid acid/base catalysts are available nowadays, offering significant advantages such as minimal capital investment and operating costs, as well as no catalyst-related waste streams and no soap formation. However, the controllability of the process is just as important as the capital and operating savings. In such processes the small number of degrees of freedom is a drawback which makes it difficult to set the reactants feed ratio correctly and consequently to avoid impurities in the products. This work considers the process control of biodiesel production by reactive absorption, the main result being an efficient control structure that ensures the stoichiometric ratio of reactants. Moreover, the excess of methanol operating constraint that is necessary for the total conversion of the fatty acids and for prevention of the difficult separations is fulfilled. Rigorous simulations were performed using Aspen Plus and Aspen Dynamics as efficient computer aided process engineering tools.

Keywords: process control, biofuels, reactive separation, solid acid catalysts, FAME

1. Introduction

Biodiesel is an alternative fuel produced from green sources such as waste vegetable oils, animal fat or even frying-oils from the food industry (Kulkarni et al., 2006). Nowadays, modern plants replaced the homogeneous catalysts with solid bases or acids, thus eliminating the salt waste streams and simplifying the downstream processing steps. Moreover, integrated processes based on reactive distillation (Kiss et al., 2006, 2008, 2009; Dimian et al., 2009) or reactive absorption (Kiss, 2009) are now available, offering significant advantages such as minimal capital investment and operating costs, as well as no catalyst-related waste streams and no soap formation.

This work considers the process control of biodiesel production by reactive absorption. The results are given for a plant producing 10 ktpy biodiesel from waste vegetable oil with high free fatty acids content (up to 100%), using solid acids as green catalysts. Compared to reactive distillation, the absence of a reboiler and a condenser makes reactive absorption a simpler process. However, the drawback is the small number of degrees of freedom that makes it difficult to set the reactants feed ratio correctly and consequently to avoid impurities in the products. Aspen Plus and Aspen Dynamics were used as efficient computer aided process engineering tool to perform rigorous simulations for testing various control structures. The main result of this study is an efficient control structure that ensures the stoichiometric ratio of reactants and fulfills the excess of methanol operating constraint that is sufficient for the total conversion of the fatty acids and for prevention of the difficult separations.
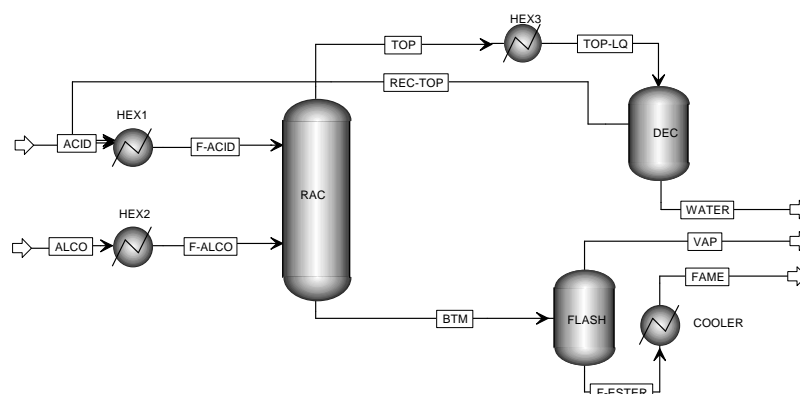
Figure 1. Flowsheet of a biodiesel production process by reactive absorption.

2. Process description

Figure 1 presents the flowsheet of a biodiesel production process based on a reactive absorption column. The production rate is 10 ktpy fatty acid methyl esters (FAME) manufactured from 100% fatty acids and methanol. The fatty components were lumped into one fatty acid and its fatty ester – according to the reaction: R-COOH + CH$_3$OH ↔ R-COO-CH$_3$ + H$_2$O. Lauric acid/ester was selected as lumped component due to the availability of experimental results, kinetics and VLLE parameters for this system (Kiss, 2009). The column has a diameter of 0.4 m. There are 15 theoretical stages with a liquid holdup of 18 L. Stages 3 to 12 are reactive, the catalyst loading being 6.5 kg. The fatty acid is pre-heated then fed as hot liquid in the top of the reactive column while a stoichiometric amount of alcohol is injected as vapor into the bottom of the column, thus creating a counter-current flow regime over the reactive zone. Water by-product is removed as top vapor, then condensed and separated in a decanter from which the fatty acids are recycled back to the column while water by-product is recovered at high purity. The fatty esters are delivered as high purity bottom product of the reactive column. The hot product is flashed first to remove the remaining methanol, and then it is cooled down and stored. Table 1 presents a summary of the feed and product streams.

Table 1. Summary of main process streams

	ACID	ALCO	F-ACID	F-ALCO	WATER	FAME
Mole flow rate / [kmol/h]	5.824	5.876	5.824	5.876	5.827	5.873
Mass flow rate / [kg/h]	1166.73	188.30	1166.73	188.30	105.05	1249.98
Temperature / [ºC]	20	20	160	65.44	51.77	30
Purity (%wt)	100	100	100	100	99.9	99.9

3. Problem statement

Reactive absorption offers indeed significant advantages such as minimal capital investment and operating costs, as well as no catalyst-related waste streams and no soap formation. However, the controllability of the process is just as important as the capital and operating savings. In processes based on reactive distillation or absorption, feeding the reactants according to their stoichiometric ratio is essential to achieve high products purity. Thus, the fatty acid is completely converted to fatty esters when there is an excess of methanol, but the excess of methanol becomes an impurity in the top stream

and thereafter in the water by-product. On the contrary, when there is an excess of fatty acids, the purity of water by-product remains high, but the conversion of fatty acids is incomplete. In the later case the bottom product contains unreacted fatty acids that can not be removed from the final product by simple flashing. Since the separation of fatty acids from fatty esters is more difficult than the separation of fatty acids from water, this situation should be avoided. It is important to remark that this constraint must be fulfilled not only during the normal operation, but also during the transitory regimes arising due to planned production rate changes or unplanned disturbances.

4. Plantwide control results

4.1. Fixing plant-inlet flow rates
Figure 2 shows the first control structure tested (CS-1a). The flow rate of fresh acid is fixed and determines the production rate. The flow rate controller on alcohol stream receives its setpoint as the measured column-inlet acid flow multiplied by the desired alcohol / acid ratio.
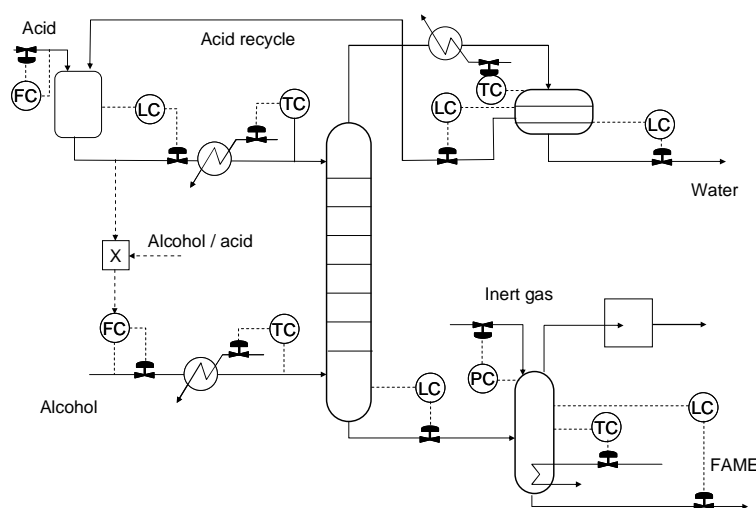
Figure 2. Control structure CS-1a.

The sensitivity analysis (Figure 3) shows that the purity depends on the ratio alcohol / acid and that there is an optimum value for which both purities are high (Figure 3, left).
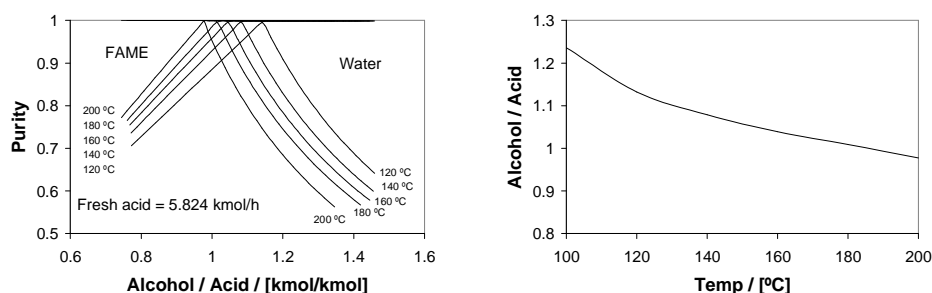
Figure 3. CS-1a – purity versus the alcohol / acid ratio at various temperatures of the column-inlet acid feed (left); Optimum ratio versus temperature of the column-inlet acid feed (right)
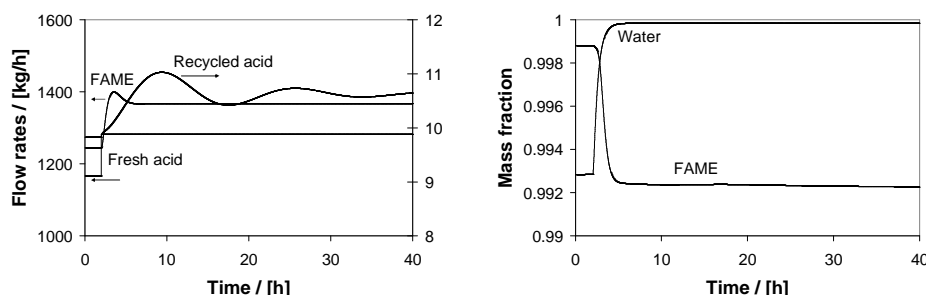
Figure 4. Dynamic simulation results for control structure CS-1a. (At t=2h, the fresh acid rate is increased by 10%, from 5.824 kmol/h to 6.4 kmol/h)

The optimal ratio versus temperature is shown in Figure 3, right. Remarkably, this optimal value appears to be independent of the reaction rate. Nevertheless, the challenge remains to find the variables to be measured (and controlled) such that the alcohol/acid and temperature are set at the best values.

Figure 4 proves that modifying the fresh acid flow rate is a direct way for changing the production rate. The dynamics of the acid recycle is slow. However, the purity of FAME decreases and the main impurity is the fatty acid, which is unacceptable. It should be stressed that in control structure CS-1a any inaccuracy of measuring the acid flow rate will be reflected by deviation of the alcohol / acid ratio from the required value and therefore by impure products. For this reason and considering FAME as the main product, the control structure CS-1b (Figure 5) adds a concentration controller that measures the concentration of acid in the bottom and adjust the ratio alcohol / acid.

Figure 6 presents the dynamic simulation results for control structure CS-1b. At t = 2h, the fresh acid rate is increased to 110% of the nominal value (from 5.824 kmol/h to 6.4 kmol/h). At t=20 h, is decreased to 90% of the nominal value (from 6.4 kmol/h to 5.2 kmol/h). The production rate changes properly, but the water purity is too low during large production times.
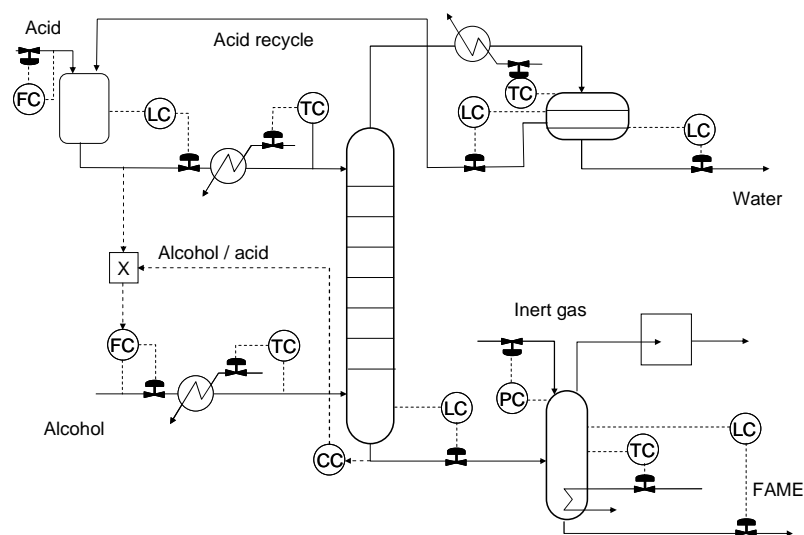
Figure 5. Control structure CS-1b – concentration of acid in bottom stream is controlled by manipulating the alcohol / acid ratio

Plantwide Control of a Biodiesel Process by Reactive Absorption
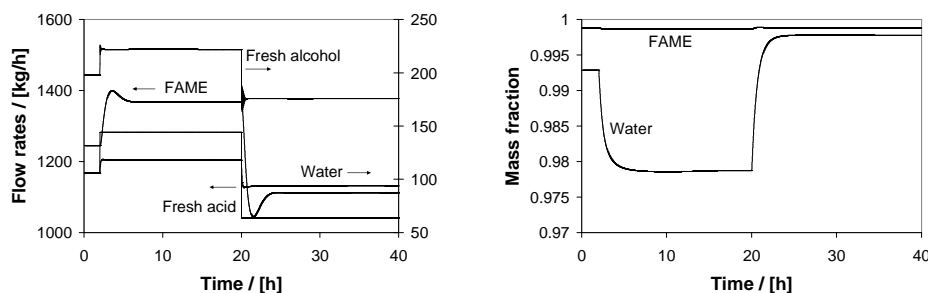
Figure 6. Dynamic simulation results for control structure CS-1b

The control structure CS-1c (Figure 7) adds another concentration controller. The setpoint of the column-inlet acid temperature is manipulated, in a cascade fashion, by a concentration controller which prevents that an excessive amount of methanol arrives in the top of the column by increasing the temperature and thus the reaction rate in the upper part of the column. Figure 8 shows the dynamic simulation results for the control structure CS-1c. The same scenario was tested as for CS-1b. The production rates are also properly changed and the purity is significantly better.
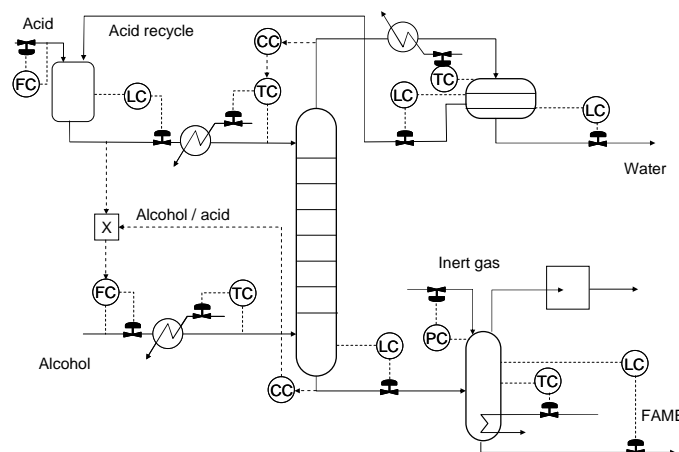
Figure 7. Control structure CS-1c – concentration of acid in bottom stream is controlled by manipulating the alcohol / acid ratio. The temperature of the column-inlet acid stream is manipulated by a methanol concentration controller.
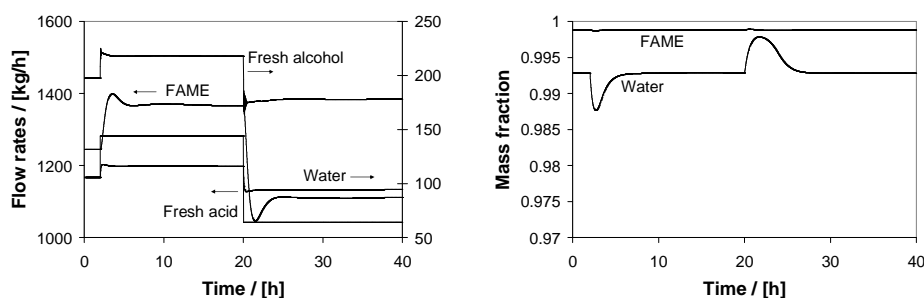
Figure 8. Dynamic simulation results for control structure CS-1c. Production rate changes are easily achieved and the products purity is maintained at high values

4.2. Dynamics of quality control

When an increase of the production rate is desired, the alcohol feed rate is firstly increased, followed by the acid. When a decrease of the production rate is desired, the acid rate is the first one to be reduced. The quality control structure is shown below.
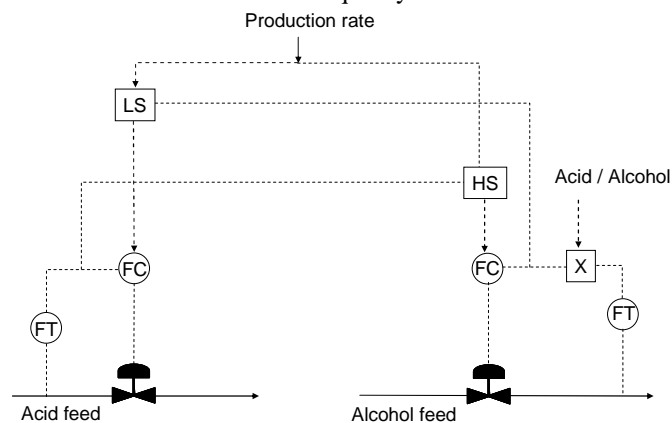
Figure 9. Control scheme ensuring that during the transient period the acid is never in excess.

5. Conclusions

Integrated biodiesel processes based on reactive absorption have fewer degrees of freedom compared to reactive distillation. This makes it difficult to set the reactants feed ratio correctly and consequently avoiding impurities in the products. Aspen Plus and Aspen Dynamics can be successfully used as efficient computer aided process engineering tools to perform rigorous simulations for testing various control structures. The main result of this study is an efficient control structure that can ensure the stoichiometric ratio of reactants and fulfills the excess of methanol operating constraint that is sufficient for the total conversion of the fatty acids and for prevention of the difficult separations (e.g. fatty acid – fatty ester).

References

A. C. Dimian, C. S. Bildea, F. Omota, A. A. Kiss, 2009, Innovative Process for Fatty Acid Esters by Dual Reactive Distillation, Computers & Chemical Engineering, 33, 743-750.

A. A. Kiss, A. C. Dimian, G. Rothenberg, 2006, Solid acid catalysts for biodiesel production - towards sustainable energy, Advanced Synthesis & Catalysis, 348, 75-81.

A. A. Kiss, G. Rothenberg, A. C. Dimian, F. Omota, 2006, The heterogeneous advantage: biodiesel by catalytic reactive distillation, Topics in Catalysis, 40, 141-150.

A. A. Kiss, A. C. Dimian, G. Rothenberg, 2008, Biodiesel by Reactive Distillation Powered by Metal Oxides, Energy & Fuels, 22, 598-604.

A. A. Kiss, 2009, Novel Process for Biodiesel by Reactive Absorption, Separation & Purification Technology, 69, 280-287.

M. G. Kulkarni, A. K. Dalai, 2006, Waste cooking oil-an economical source for biodiesel: A review, Industrial & Engineering Chemistry Research, 45, 2901-2913.

Benchmark for Hierarchical Plantwide Control of Hybrid Chemical Processes: Control of Coupled Batch and Continuous Reactors

Harvey Arellano-Garcia [a], Stephanie Geist [b], Günter Wozny [a], Jörg Raisch [b, c]

[a]*Chair of Process Dynamics and Operation, Technische Universität Berlin, Str. des 17. Juni 135, D-10623 Berlin, Germany*
[b]*Chair of Control Systems, Technische Universität Berlin, Str. des 17. Juni 135, 10623 Berlin, Germany*
[c]*Systems and Control Theory Group, Max-Planck-Institut für Dynamik komplexer technischer Systeme, Magdeburg, Germany*

Abstract

Chemical processes often exhibit a heterogeneous character and often have a highly complex behavior. In this work, a benchmark problem for hierarchical hybrid plantwide control is presented. The considered chemical process consists of coupled batch and continuous reactors. For this process, a hierarchical solution will be motivated. In order to compensate most disturbances, time-variant parameters, changes of the set point, and changes of the operating points, a control structure for each plant component is proposed. The benchmark high-level problem is to coordinate these component's control systems in a safe and optimal plantwide way.

Keywords: Hybrid Process, Plantwide Control, Hybrid Control, Hierarchical Control

1. Introduction

Industrial chemical processes are often operated in a number of plant components, which are connected by material or energy flows. Although most fundamental phenomena in chemical processes are continuous, chemical processes often feature a number of discrete event mechanisms which results in a hybrid control problem. In this contribution, we present a benchmark chemical process which consists of interacting continuous sub-processes and sub-processes performed in batch mode which imposes process discontinuities regarding the coordination of the entire chemical plant. The plant is controlled by both discrete and continuous control inputs. Moreover, some sensor information is available in form of discrete event signals only. The benchmark control problem is to design a plantwide optimal and safe control system which is robust to disturbances.

In hybrid control problems, which involve a large number of variables and highly nonlinear continuous dynamics, structured abstraction based approaches in combination with hierarchical methods seem a tractable way to overcome complexity, e.g. Raisch and Moor (2005); Skogestad (2004). In chemical industry, controllers for sub-processes often already exist. To increase productivity and safety, however, more effort in the control of the interaction of connected plant components is needed while "simplicity" should be retained. In Skogestad (2004), the advantages of a hierarchical structuring for plantwide control systems are described. A hierarchical approach can reduce control system complexity and, thus, the required effort in modeling and design. To follow this approach, we propose a low-level control system for the considered benchmark process.

The low level control system consists of controllers designed independently for each plant component. These controllers compensate the impact of most disturbances, time-variant parameters, changes of the set point and changes of the operating points. The high-level supervisory problem is to design a production rate controller for the entire plant, where the high-level controller changes the set points of the low-level control systems. This should guarantee acceptable performance, product quality and safety requirements. The major challenge in this high-level control problem is that some disturbances have a drastic impact on the plantwide dynamic behavior.
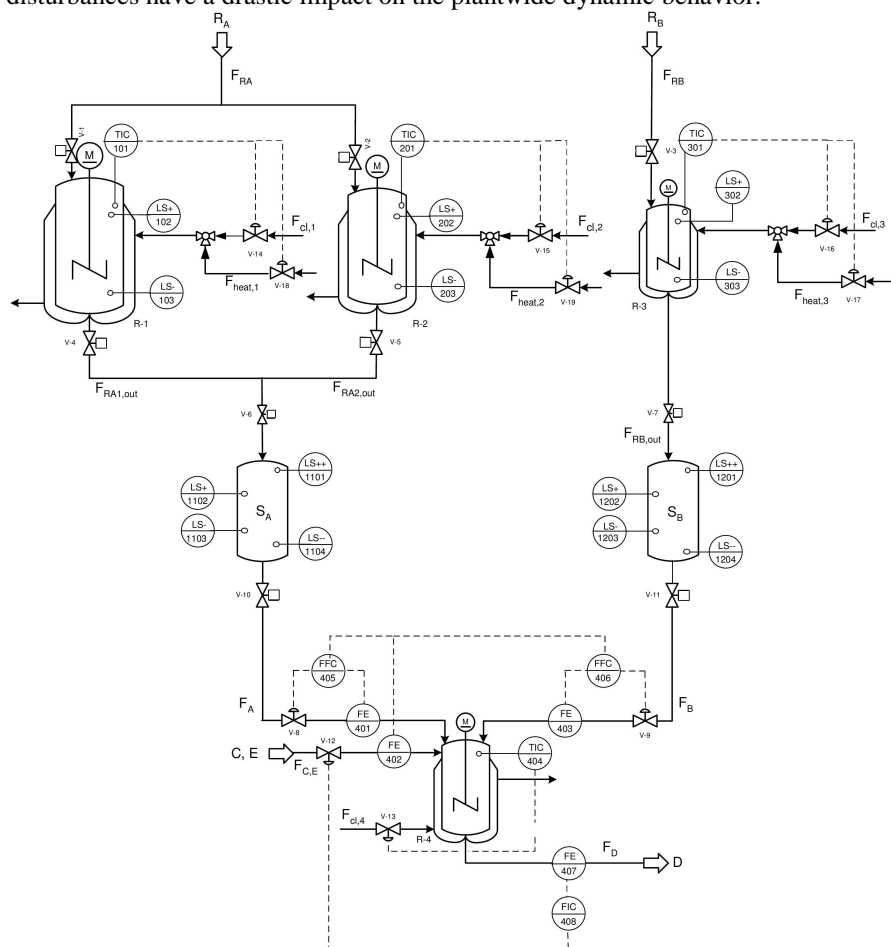
Fig. 1. Flowsheet of the coupled reactors

2. Problem Statement

Discontinuously operated plants are widely used in chemical process industries for the production of fine, or specialty chemicals. The reactors discharge their final product into buffer tanks used to transfer it continuously to the subsequent plant devices. For those units, it is highly desirable to have optimal control and operation schedules (Barton et al. (2000, 2006)). These schedules determine the stand-by times between the batch reactor phases, as well as the flow rates between the different devices in order to maximize the average plant productivity and to have an un-interrupted production without incidents.

The proposed benchmark process consists of a continuous reactor and several batch reactor units which are influenced by continuous and discrete event control inputs. Fig. 1 depicts the flowsheet of the process. Although the model process is less complex than many real world processes it exhibits most of the problems encountered in coupled batch and continuous reactors. The overall process can be divided into two major coupled production steps. In the first step, intermediate products are produced in batch reactors. When the batch has finished and a required purity of the intermediate products is achieved, the reactors are discharged into storage tanks. The intermediate products are then fed into a continuously stirred reactor where they react with another reactant.

The core of the plant is a set of 3 parallel batch reactors (denoted by R-1 to R-3 in Fig. 1). In the first process step, the intermediate product *A* in reactors R-1 and R-2 and the intermediate product B in the reactor R-3 are manufactured. The component A possesses a lower molecular mass than B, and thus, the combined volume of reactors R-1 and R-2 is larger than the volume of reactor R-3. Furthermore, in order to compensate for a larger demand of A during the startup as well as for disturbances, reactors R-1 and R-2 may be used simultaneously. To compensate larger demands of A quickly, reactor R-2 features a smaller volume and consequently a smaller cycle time than reactor R-1.

In the batch reactors R-1 to R-3, two parallel exothermic reactions take place correspondingly. The corresponding reactions are: $R_A \xrightarrow{k_A} A$ and $R_B \xrightarrow{k_B} B$ with the first reaction being faster than the second one. They both are accompanied by the parallel reactions: $R_A \xrightarrow{k_{W_A}} W_A$ and $R_B \xrightarrow{k_{W_B}} W_B$ which produce the waste product W_A and W_B, respectively. If the final concentrations of A and B are lower than $c_{A,min}$ and $c_{B,min}$ at the end of the corresponding batches, the batches are spoilt. When the reactions for the production of A and B are completed under the compliance of the purity requirements, the product of each reactor is discharged and stored in two parallel buffering tanks S_A and S_B. Each tank features a maximal volume, which may not be exceeded. The aim of the buffer tanks is to ensure an un-interrupted supply of the educts to the continuous section of the plant, which, in this case, consists of a continuous reactor R-4. In this reactor, the final product D is produced via the exothermic reaction: $A + B + C \xrightarrow{k_1} D$, also an undesired parallel reaction takes place: $4A + E \xrightarrow{k_2} C$, where the component E represents a contamination of the feed stream of *C* into the continuous reactor. Thus, a contamination with E leads to an increased need of the intermediate product A.

3. Plant under Low-level Control

The major challenge in the presented benchmark process is to coordinate the coupled plant components, whereas control concepts for the single plant components can be easily obtained. In fact, all proposed low-level control concepts are quite simple and widely used in chemical industries. Based on the single control components, we propose a hierarchical control structure which reduces the coordination control problem significantly in terms of complexity.

Batch Reactors

To produce one batch of the intermediate products in one of the batch reactors, a sequence of five production steps has to be performed: filling, heating, cooling while reaction takes place, cooling the reaction products at the end of reaction, and discharging. In the proposed control strategy, every production step can be controlled independently, i.e., at the end of each step an event is generated which enables the next production step. Filling, heating, the cooling process after the reaction has finished and

discharging are controlled by discrete valve positions, while cooling during the reaction can be controlled by continuous coolant inflows $F_{cl,i}$, i = 1; 2; 3 into the reactor shell.

The coolant inflow has been chosen such that the coolant consumption is minimized while temperature constraints and product quality requirements at the end of the reaction are satisfied. The optimization problem can be expressed in the following way:

$$\min_{F_{cl,i}} \int_0^{t_{f,i}} F_{cl,i}(t)dt \tag{1}$$

$$T_{L,i} \le T_{R-i} \le T_{U,i}; c_A(t_{f,i}) \ge c_{A,min}, \text{for } i = 1, 2; c_B(t_{f,3}) \ge c_{B,min}, \text{for } i = 3.$$

This optimization problem has been solved iteratively in two optimization steps. First, the coolant consumption has been minimized for a given reaction duration $t_{f,i}$ such that the reactor temperature T_{R-i}, i = 1,2,3, remains between an upper and lower bound, $T_{L,i}$ and $T_{U,i}$ respectively, and the concentration of the intermediate product at the end of the reaction exceeds a minimal concentration $c_{A,min}$ and $c_{B,min}$ for reactors R-1, R 2 and R-3. Then, the reaction duration has been varied where for smaller $t_{f,i}$ the coolant consumption gets smaller. $t_{f,i}$ has been reduced iteratively until the minimal reaction time still meeting the safety and quality requirements has been reached.

The optimization problem for fixed $t_{f,i}$ has been solved by a NLP approach, where the control profiles has been parameterized by piecewise constant functions. The obtained profile for the coolant inflow $F_{cl,1}(t)$ of reactor R-1 is shown in Fig. 3. The resulting concentration profiles in reactor R-1 are depicted in Fig. 4. In this setting, the only relevant information for the high-level controller is the cycle time $t_{f,i}$ of each reactor.
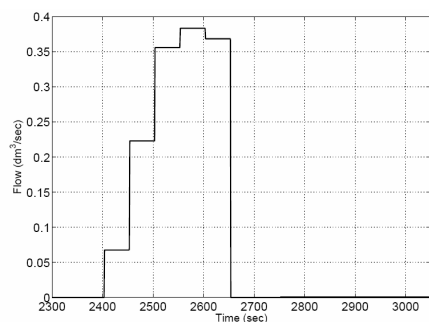
Fig. 3. Coolant flow $F_{cl,1}(t)$ during reaction to produce intermediate product *A* in Reactor R-1.
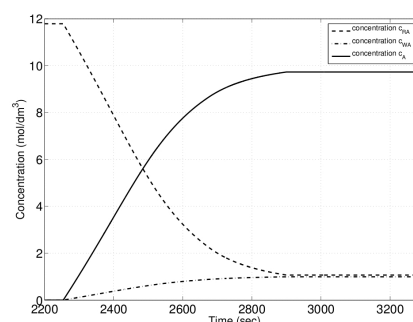
Fig. 4. Concentration profiles of reactant R_A, intermediate product A and waste product W_A in reactor R-1.

3.1. Storage Tanks

The level of both storage tanks, h_{SA} and h_{SB} can be controlled by discrete inlet (V-6,V-7) and outlet valves (V-10, V-11). In the storage tanks, in addition to the upper and lower level thresholds ($h_{uu,j}$, $h_{ll,j}$, j = S_A, S_B) two intermediate measurement thresholds ($h_{u,j}$, $h_{l,j}$, j = S_A, S_B) are introduced (see Fig. 1). Based on this measurement, the task for the low-level controller is to prevent the storage tanks from over- and under filling $h_{ll,j} \le h_{Sj} \le h_{uu,j}$. A simple solution that prevents infinite fast switching is shown in Fig. 5. Closing the outlet valves V-10 and V-11 will cause an interruption of the supply for the subsequent reaction and is, therefore, not allowed. The underfilling of both storage tanks should be prevented by the high-level supervisor by manipulating the production rate of the entire plant. An example profile for the tank level in storage tank S_A is shown in Fig. 6.
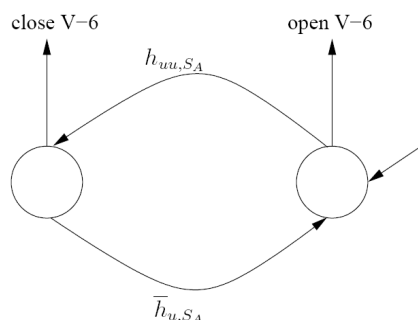
Fig. 5. *Automaton for switching valve V-6 to control filling of storage tank S_A. The event $h_{uu,SA}$ is generated when the tank level in tank S_A has reached the upper threshold and event $h_{u,SA}$ is generated when the intermediate threshold $h_{u,SA}$ has been passed from above.*
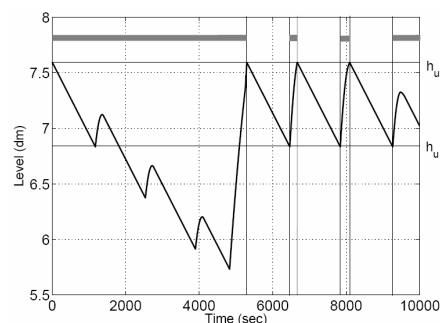
Fig. 6. *Tank level profile in storage tank S_A. The storage tank is completely filled with A in the beginning. First, three batches produced in reactor R-2 are discharged into the storage tank three times in a row. Then, reactor R-1 is discharged. The cycle time of reactor R-1 is ca. three times larger than the cycle time of reactor R-1. Bars mark time intervals when valve V-6 is open.*

3.2. Continuous Reactor

The control problem for the continuous reactor R-4 is to track the reference value for the outflow rate of the final product F_D and the temperature T_{R-4}. The reference values $F_{D,sp}$ and $T_{R-4,sp}$ are adjusted by the high-level controller to be designed. Control inputs are the inflow rate of the cooling liquid $F_{cl,4}$ into the reactor shell and the valve positions controlling the inflow rates of all reactants F_A, F_B and $F_{C,E}$. Measured variables are the outflow F_D and the temperature T_{R-4} in the reactor. The concentration of reactant C in the inflow C,E which is impured with E can be measured with a small sampling rate. The proposed low-level control system for reactor R-4 consists of a temperature control system and an outflow control system and is shown in Fig. 7.
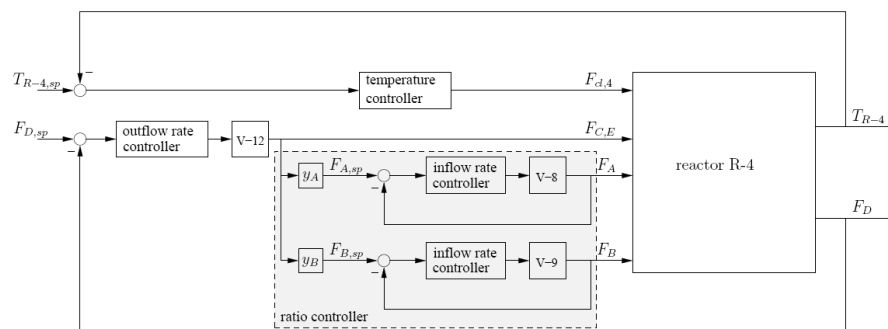
Fig. 7. Low-level control system for reactor R-4.

The control system for the outflow rate F_D consists of a PI controller manipulating the inflow rate $F_{C,E}$ and a ratio control system adjusting the inflow rates F_A and F_B depending on $F_{C,E}$ (see Fig. 8) in order to allow the equimolar reaction $A + B + C \rightarrow D$. The ratio controller is simply formed by two PI controllers for the inflow rates F_A and F_B. The set points for these controllers are calculated by the current inflow rate $F_{C,E}$ multiplied with factors y_A and y_B. The latter depend on the concentration of reactant C in the feed C,E. All PI controllers are enhanced by anti-reset windup structures. Note that

the outflow-rate control system can only work properly if the supply of reactants *A* and B from the previous production steps is guaranteed. The second low-level control goal is to track the reference signal $T_{R-4,sp}$ for the temperature. The temperature controller is synthesized as a PI controller with anti-reset windup.

4. High-Level Problem Statement

On the basis of the low-level control layer, we can now formulate the benchmark plantwide control problem. A high contamination of the feed stream of C,E with E results in a high demand for A. If the production rate F_D were fixed, it could happen that the storage tank S_A runs empty in the case of an unexpected increase of the concentration of E in the feed. To prevent an interruption of the entire process, it is therefore necessary to adjust the reference value $F_{D,sp}$ for the production rate appropriately. Additionally, changes in the production rate also require an adjustment of the temperature reference value $T_{R-4,sp}$ in Reactor R-4 to ensure the specified product quality. This is necessary because the reaction rates k_1 and k_2 of the desired reaction $A + B + C \rightarrow D$ and the undesired reaction $4A + E \rightarrow C$ depend on temperature.

The high-level control inputs are the reference value for the production rate F_D and the temperature T_{R-4} in reactor R-4. The information available for the high-level controller includes the cycle time of each batch reactor and, therefore, the time before the next batch is finished, discrete events issued from the threshold sensors in the storage tanks and the concentration of E in the feed C,E. The high-level control problem is to maximize the average production rate while preventing an interruption of the process, guaranteeing the required product purity and respecting all plant dynamics and constraints. This optimization problem can be formulated in the following way:

$$\max_{T_{R-4},F_{D,sp}} \quad \frac{1}{\tau}\int_0^\tau F_{D,sp}(t)dt \tag{2}$$

$$h_{ll,S_A} \leq h_{S_A} \leq h_{uu,S_A} \,; h_{ll,S_B} \leq h_{S_B} \leq h_{uu,S_B} \,; c_D \geq c_{D,min} \,; T_{R-4} \leq T_{R-4,max} \,,$$

where τ is the overall operating time. The considered high-level problem is, of course, scalable to more complex problems, e.g. if the production costs for the batch reactors are included in the overall cost function.

5. Conclusions

We have presented a hybrid chemical production plant and suggested standard low-level control schemes, which are widely used in industries. The remaining high-level problem has been suggested as a realistic benchmark for hybrid control. A complete and detailed description of all model equations including all process and low-level controller parameters will be provided in a separate document available for download.

References

P. I. Barton, J.R., and S. Galan, (2000). Optimization of hybrid discrete/continuous dynamic systems, Computers & Chemical Engineering, 24(9-10), 2171-2182.

P. I. Barton, C, K. Lee, M. Yunt, (2006). Optimization of hybrid systems. Computers & Chemical Engineering, 30(10-12), 1576-1589.

J. Raisch, T. Moor, (2005). Hierarchical control synthesis and its application to a multiproduct batch plant. In Nonlinear Control and Observer Design, LNCIS, 199-216. Springer-Verlag.

S. Skogestad, (2004). Control structure design for complete chemical plants. Computers and Chemical Engineering, (28).

Resilient control in view of valve stiction: extension of a Kalman-based FTC scheme

Kris Villez[a] Babji Srinivasan[b] Raghunathan Rengaswamy[b] Shankar Narasimhan[c] Venkat Venkatasubramanian[a]

[a] *Laboratory for Intelligent Process Systems, School of Chem. Eng., Purdue University, West Lafayette, IN 47907, USA*
[b] *Dept. of Chem. Eng., Texas Tech University, Lubbock, TX 79409, USA*
[c] *Dept. of Chem. Eng., Indian Institute of Technology, Madras, Chennai 600036, India*

Abstract

In this contribution we propose an active Fault Tolerant Control (FTC) strategy which enables the isolation and identification of valve stiction and valve blocking, in addition to the additive faults like sensor and actuator biases. The developed method is an extension of the original method proposed by Prakash *et al.* (2002). This method is based on the Kalman filter and is developed under the assumption that the monitored system is Linear Time Invariant (LTI). It has been shown to work well for additive faults such as sensor and actuator biases. Within this method the fault isolation and identification task is based on the Generalized Likelihood Ratio (GLR) test by which the most plausible fault type in a library of faults is selected following estimation of fault parameters.

Keywords: Kalman filter, Valve stiction, Fault isolation, Fault diagnosis

1. Introduction

Valve stiction is a problem that has caught the attention of several research groups in the last decade. Valve stiction is considered one of the most common problems in control loops (Shoukat Choudhury et al., 2004). Its presence leads to rather severe non-linear effects which makes its detection, diagnosis and accommodation a challenging problem. In this work, we evaluate an extended Kalman-based method for on-line diagnosis of several faults in control loops with valves. We show promising results for a range of faults and list several opportunities and threats to our approach.

2. Materials and methods

2.1. Simulated system

A buffer tank system model is used for evaluation of our method. The tank level is measured and is affected by an inflow as a disturbance input and gravitational outflow, which is in turn manipulated by a valve. In the original (continuous, non-linear) system, the outflow relates to the tank level as follows:

$$q(t)_{out} = C \cdot v(t) \cdot \sqrt{x(t)}$$

with: $\quad q(t)_{out} \quad$ outgoing flow rate $\quad [m^3/s]$
$\qquad v(t) \qquad$ valve position $\qquad [0-100\%]$
$\qquad x(t) \qquad$ tank level $\qquad [m]$
$\qquad C \qquad$ valve constant $\qquad [0.1414\,m^{2.5}/s]$

(Eqn. 1)

This system is linearized around its equilibrium point corresponding to an ingoing flow rate of 0.1 m^3/s. This corresponds to a volume of 1 m^3 and a valve position of 5%. Furthermore, the model is discretized in time so to obtain the standard discrete state-space model form, including input disturbances and measurement error:

$$\dot{x}(k) = \Phi \cdot x(k) + \Gamma_v \cdot v(k) + \Gamma_w \cdot w(k)$$

$$y(k) = C \cdot x(k) + D \cdot e(k)$$

with: $e(k)$ measurement error   (Eqn. 2)

       $w(k)$ input disturbance

       $\Phi, \Gamma_u, \Gamma_w, C, D$ time-invariant system matrices

This discretized and linearized model is used for all simulations. A PI-controller brings the tank level measurement, y(k), at its set-point by means of manipulation of the valve position, v(k). In nominal operation, the actual valve position is equal to the controller signal, u(k).

2.2. Simulated faults

Four types of faults are simulated. The first type is valve stiction. As soon as this type of fault sets in, the valve is simulated to move only when the difference between the valve position, v(k), and the controller signal, u(k), is larger than a given parameter value. This parameter is called the stiction band. Mathematically, one writes:

$$v(k) = u(k) \quad if \left| (v(k) - u(k)) \right| > b_{stiction}$$

$$v(k) = u(k-1) \quad if \left| (v(k) - u(k)) \right| \leq b_{stiction} \qquad \text{(Eqn. 3)}$$

with: $b_{stiction}$ stiction band

The second type is valve blocking. In this case, the valve does not move at all for any signal sent to the valve as soon as the fault sets in. Now one writes simply:

$$v(k) = v(k-1)$$

$$\text{(Eqn. 4)}$$

The third and fourth types of faults are a bias in the valve position and a bias in the level measurement. Now one writes for the valve bias and sensor bias respectively:

$$v(k) = u(k) + b_u \qquad \text{(Eqn. 5)}$$

$$y_{faulty}(k) = y(k) + b_y \qquad \text{(Eqn. 6)}$$

For the simulation of the system with these faults, one uses equations (1) with equations (4-5) in the case of fault types 1-3 and one replaces *y* with y_{faulty} as measurements in the case of fault type 4. In the case of valve faults, only the desired valve position, *u*, is available for inference (*v* is hidden). For the sensor fault, only y_{faulty} is available (*y* is hidden). Note that fault types 1, 2 and 4 are characterized by their start time and a magnitude parameter (stiction band or bias). Valve blocking (fault type 3) is only characterized by the start time.

2.3. Kalman-filter based Fault Detection and Diagnosis

A Kalman-filter based technique for Fault Tolerant Control exists and has been shown successful for detection, diagnosis and accommodation of process faults (Prakash *et al.*, 2002). Although the framework is general, the method has been tested particularly for additive linear faults such as the valve and sensor bias in our simulations (fault type 3 and 4). Central to the method is the use of a Kalman filter to generate prediction residuals, i.e., the deviations between actual and predicted measurements. In essence, the fault diagnosis part in this method follows from the (deterministic) simulation of

each hypothesized fault after which the fault scenario with the highest likelihood (based on the Kalman-filter) is selected. Several advantages result from the system's linearity and the additive and linear properties of the considered faults. First, the problem of identification of maximum likelihood bias parameters is reduced to a simple linear regression for a given start time of a fault. Also, the likelihood associated with the given fault parameter conditional to the considered time window of observations, follows in one direct step. Therefore, no advanced optimization techniques are necessary and for a given start time of a fault, a unique solution exists.

In the original work, the start time of a fault follows from the fault detection part of the method. The method which is based on a sequential testing is a fast way to obtain a rough estimate of the fault start time. This is not necessary the best to do as the actual fault start time may differ and may affect fault isolation and identification. For this reason, we evaluate a different strategy where every possible (discrete) time within a certain time window before fault confirmation is evaluated as a start time for the fault.

2.4. Extension for valve blocking and valve stiction
The particular problem of detecting and diagnosing valve blocking and valve stiction has not been tackled from the model-based angle described above. Therefore, the 'library' of faults is extended with valve stiction and valve blocking as follows.

Valve stiction and valve blocking are both of a deterministic nature, just like the bias faults. No other parameter than the start time needs to be evaluated for valve blocking. In the case of valve stiction, one needs to estimate the band stiction parameter in addition to the start time. Conditional to a stiction band value and a fault start time, one can evaluate what the true valve position is in a considered time window by applying equations (3) to the series of valve position signals. The expected response of the system is otherwise linear so one can calculate of the likelihood of the observations conditional to the evaluated scenario (stiction time + band parameters) with the Kalman filter.

For band stiction higher than a certain minimal value, the actual simulation will be the same as for a stuck valve. Indeed, if the band stiction is high enough, the valve will not move at all, thus making the two scenarios phenomenologically the same. On the positive side, one can recognize this situation by simply checking whether the valve position changes for the evaluated band stiction value and fault start time during a considered window. The situation that the two faults are not separable is thus detectable.

3. Results
In what follows, simulation results will be shown for faults introduced at sample 76. Figure 1 shows the simulated data for valve stiction. It can be seen that the true valve position fails to follow the demanded valve position. It can also be seen that the valve gets stuck at different positions for periods of time. As a result, control performance degrades as a oscillatory response of the level follows, as is typical for valve stiction problems. Similar oscillations occur in the state estimation errors and the prediction residuals.
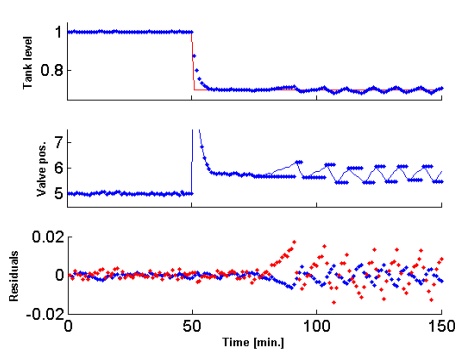
Figure 1: Valve stiction scenario - data. Top: Tank level set-point (─), and measurement y(k) (●). Middle: Valve control signal $u_{signal}(k)$ (─) and position $u_{real}(k)$ (●). Bottom: Kalman estimation and prediction errors ($r_x(k)$ (blue), $r_y(k)$) (red)).
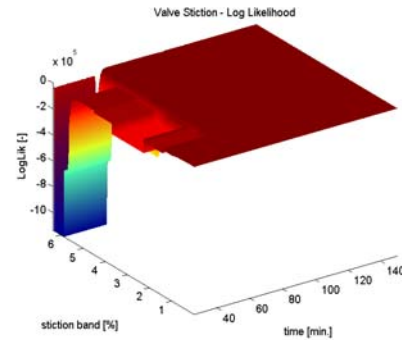
Figure 2: Valve stiction scenario – Log-Likelihood of valve stiction scenario as function of start time and stiction band.

Figure 2 shows the Generalized Likelihood Ratio (GLR) as found for valve stiction, evaluated for a range of band stiction values (resolution 0.01%) and all considered (discrete) fault start times (26 to 150). It can be seen that a large portion of the surface plot is flat, meaning that the likelihood is rather insensitive to the fault parameter values. More importantly, local optima are present, which is typical.

(a)   (b)
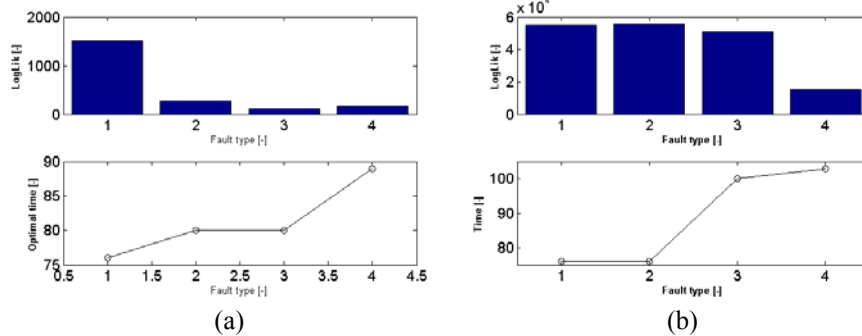
Figure 3: (a) Valve stiction scenario (b) Valve blocking scenario – fault diagnosis. Maximal Generalized Likelihood Ratio values for all considered fault types and corresponding optimal fault start times.

Figure 3a shows the fault diagnostic results. It is seen that the maximal GLR values are large for any fault. In addition, fault type 1 (valve stiction) delivers the highest GLR value found thus leading to a correct identification of valve stiction as the root cause. It is noted that the optimal fault start time for valve blocking (fault type 2) is relatively close to the one for valve stiction (fault type 1). This suggests that the optimal start time for valve blocking may be a good initial guess to start the optimization for valve stiction.

Next, the valve blocking scenario is evaluated. Here, at sample 76, the valve gets stuck and remains at its position for the remainder of the simulation. This leads to an offset in the tank level and increasing discrepancy between desired valve position and increasing state estimation and prediction residual magnitudes (not shown).

Resilient control in view of valve stiction: extension of a Kalman-based FTC scheme

Figure 3b shows the results for the fault diagnosis task in the valve blocking scenario. Also here, the largest GLR is found for the correct fault, namely fault type 2 (valve blocking). In addition, the correct fault start time is found and the optimal start time for valve stiction and blocking are the same. It is noted that the maximal GLR for valve stiction is slightly lower than the one for valve blocking. It is in fact possible to make them equal if one considers scenarios with valve stiction bands so high that the valve doesn't move anymore. Such solutions were automatically discarded. However, valve stiction and blocking remain inseparable without further information. In the discussion section some ideas on how to tackle this issue are provided.

Also valve bias and sensor bias scenarios were investigated. In both cases, state estimation and prediction show performance degradation, the correct fault is found as well as the start time (no results shown). For the sensor bias scenario, the log-likelihood for each fault type is shown as function of time in Figure 4. Plotted values are maximal with respect to other parameters. The most important observation drawn from this graph is that the fault start time is important for correct fault diagnosis. Indeed, at any other time than the correct start time of the fault, fault type 3 (valve bias) would be erroneously preferred over the correct fault type (4, sensor bias). The profiles for fault type 3 and 4 are relatively smooth which may facilitate automated optimization although local optima are present. For presence or evaluation of fault types 1 and 2, the nonlinear estimation problem is more severe.
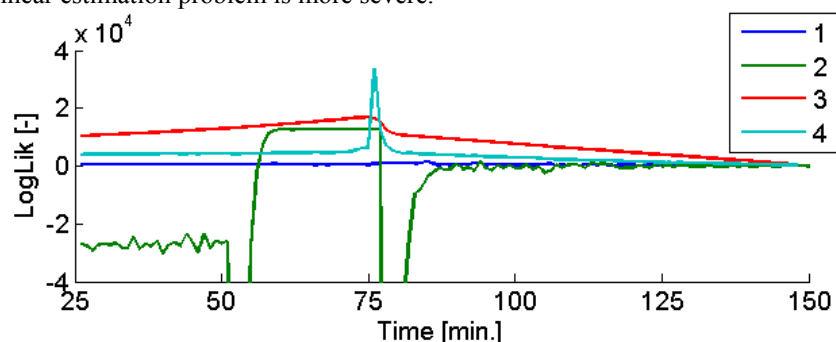
Figure 4: Sensor bias scenario – maximal Log Likelihood as function of time for each fault type (1 to 4).

4. Discussion

In the presented work, a Kalman-based method for fault diagnosis has been evaluated by means of 4 different scenarios with one fault occurrence. Shown results indicate that it is possible to identify the correct faults under certain circumstances. As such, the method is promising for on-line diagnosis of valve-based control loops.

Nevertheless, several remarks are in place with respect to the shown results. First, we have only shown results for fault diagnosis although the original work by the authors provides monitoring as well. In particular, we have assumed that any of the simulated faults is always detected and confirmed at the same, given time instant. This eliminates any effect of fault detection performance on the fault diagnosis performance. However, it may be so that certain fault types are detected faster than others and this may in turn affect the speed of the diagnosis task as well as its accuracy given that the amount of available data may then differ.

For all fault scenarios and all considered faults, the log-likelihood profiles exhibit local optima with respect to time. This is most severe when valve stiction or valve blocking faults are present or evaluated. In our current approach, all possible time instants within a certain window were considered. However, improved optimization strategies for the time parameter may be possible. For one, consider that the optimal start times for valve stiction and blocking were close in both the valve stiction and the valve blocking scenario. Finding the optimal time for either fault may constitute as a fair guess for the other and may therefore reduce computational costs. For valve stiction, the additional band stiction parameter has severe non-linear effects on the likelihood as well. In contrast, for bias faults (valve bias, sensor bias), a unique and global optimum is always found by generalized regression, though conditional to the start time of the fault.

As a last point, consider the problem of separating valve stiction and valve blocking. This is not always possible as a valve stiction scenario with band stiction so high that the valve doesn't move is phenomenologically the same as valve blocking. Additional information is necessary to do so. For this purpose, one may consider to wait for a longer period and collect more data up to a point that valve stiction or valve blocking is ruled out. Such a passive approach to fault diagnosis may be enhanced by modifying the control signal sent to the valve. Srinivasan and Rengaswamy (2008) suggest a two-move strategy by which pulses in the control signal are generated to make the valve move in case of valve stiction. A similar strategy may be taken for diagnostic purposes. Indeed, if valve stiction is present the valve will move if the applied pulse is large enough. If the valve is blocked, it will never move. Such is clearly an active strategy.

5. Conclusion

In this contribution, first results from a study on on-line diagnosis for valve faults are shown and discussed. It is made clear that several faults within a valve-based control loop can be separated under certain conditions. Such faults include valve stiction, valve blocking, valve bias and sensor bias.

Despite the preliminary character of this study, several important remarks were made in view of future research. For example, severe non-linearity of the band stiction estimation problem and of the fault start time was discussed. Further research will therefore be aimed at the search for a better estimation method. In addition, it is worthwhile to further investigate how the separation of valve stiction and valve blocking by means of passive or active collection of informative data can be achieved.

References

Shoukat Choudhury, M.A.A., Thornhill, N.F. and Shah, S.L. (2005). Modelling valve stiction. *Control Eng. Pract.,* **13,** 641–658.

Prakash, J., Patwardhan, S.C. and Narasimhan, S. (2002). A supervisory approach to Fault-Tolerant Control of Linear Multivariable Systems. *Ind. Eng. Chem. Res.,* **41,** 2270–2281.

Srinivasan, R. and Rengaswamy, R. (2008). Approaches for efficient stiction compensation in process control valves. *Comput. Chem. Eng.*, **32,** 218–229.

A Reduced Linear Model Predictive Control Algorithm for Nonlinear Distributed Parameter Systems

Ioannis Bonis and Constantinos Theodoropoulos

School of Chemical Engineering and Analytical Science, University of Manchester, Manchester M60 1QD, UK. E-mail: K.Theodoropoulos@manchester.ac.uk

Abstract

A novel model reduction-based framework for linear Model Predictive Control (MPC) of Distributed Parameter Systems is presented. It exploits the separation of scales exhibited in many systems of engineering interest. It is based on the online, adaptive identification of the dominant modes of the system using the Arnoldi method. The low-dimensional dominant subspace corresponding to those modes is exploited for the linearization of the model. Only low-dimensional Jacobian and sensitivity matrices are involved in this framework. They are projections of the original matrices onto the dominant subspaces, computed efficiently with numerical directional perturbations. The low-order linear model from this procedure is utilized in the context of a MPC scheme. The efficiency of the proposed methodology is illustrated using a temperature tracking control of a tubular reactor which also involves measurement noise and disturbances.

Keywords: model reduction, equation-free, dominant subspace, adaptive linearization, separation of scales.

1. Introduction

Model Predictive Control (MPC) is among the most successful advanced control techniques (Mayne et al, 2000). Many variants of the MPC algorithm exist, all of which implement a control action computed as the solution of a finite horizon open-loop optimal control problem. In the standard MPC formulation, the model of the system is linear, so are the constraints. However, most systems of engineering interest are nonlinear. Nonlinear MPC has significantly higher computational cost, limiting its applications to rather small systems (Henson, 1998). Extracting a linear model for a nonlinear dynamic system is not straightforward. Linearization at the set point may not be satisfactory, as the model may be a poor approximation for the time horizon considered. Hence, several self-tuning control algorithms have been developed to account for the nonlinearity of the system (Zhu, 1991). Even if the model is linear and accurate, computational cost can still be large if the number of states is large. This holds especially for, typically high-dimensional, distributed parameter systems. This work could be considered as part of the equation-free framework. A dynamic simulator of the process is required but is treated in an input/output fashion. A framework for LQR controller design exploiting the dominant subspace of the system by performing a linearization at the stationery point has been presented (Armaou et al, 2005). Depending on the system and the position at the parameter space, a stationery point may not exist or be significantly far from the current state, rendering the linearized model an insufficient approximation of the full one. Here, the nonlinear model is replaced by a linear one in the neighborhood of the current state, so linearization is in the time

domain. It is adaptive, hence it approximates the system better than a single linear model. Furthermore it has an advantage over data-driven model reduction methods, as it eliminates the need to empirically sample the parameter space. The proposed methodology is expected to be particularly useful for the control of multi-scale systems.

2. The proposed Model Predictive Control algorithm

2.1. Successive linearization

A general form of a nonlinear model in state space formulation is

$$\dot{\tilde{x}} = f(\tilde{x}(t), \tilde{u}(t), t), \quad \tilde{x}(t_0) = \tilde{x}_0$$
$$\tilde{y} = g(\tilde{x}(t), \tilde{u}(t), t) \tag{1}$$

Where $\tilde{x}(t), \tilde{u}(t), \tilde{y}(t)$ are the spate, input and output variables correspondingly and *f, g* are vector functions. If we consider a reference point ($\tilde{x}_{ref}, \tilde{u}_{ref}$) then the current state can be considered as perturbation of the reference data:

$$\tilde{x}(t) = \tilde{x}_{ref} + x(t) \quad and \quad \tilde{u}(t) = \tilde{u}_{ref} + u(t) \tag{2}$$

where the increments $x(t)$ and $u(t)$ are small. The nonlinear system can be linearized around the reference point, resulting in a continuous state-space model (Datta, 2003):

$$\left. \begin{array}{l} \dot{x}(t) = \dfrac{\partial f}{\partial \tilde{x}}\bigg|_{(\tilde{x}_{ref}, \tilde{y}_{ref})} \cdot x(t) + \dfrac{\partial f}{\partial \tilde{u}}\bigg|_{(\tilde{x}_{ref}, \tilde{y}_{ref})} \cdot u(t) \\[4mm] y(t) = \dfrac{\partial g}{\partial \tilde{x}}\bigg|_{(\tilde{x}_{ref}, \tilde{y}_{ref})} \cdot x(t) + \dfrac{\partial g}{\partial \tilde{u}}\bigg|_{(\tilde{x}_{ref}, \tilde{y}_{ref})} \cdot u(t) \end{array} \right\} \Rightarrow \begin{array}{l} \dot{x}(t) = Jx(t) + Yu(t) \\[4mm] y(t) = \Psi x(t) + \Theta u(t) \end{array} \tag{3}$$

The discrete equivalent of this system can be calculated for a given discretization time (Franklin et al, 2002). Hence we can retrieve a discrete time, state-space model:

$$\dot{x}(k+1) = Kx(k) + Lu(k)$$
$$y(k+1) = Mx(k) + Nu(k) \tag{4}$$

Usually linearization is around the steady state. However many systems do not exhibit stable steady states for the whole parameter range and/or the current state can be far from the steady state so that the linearization does not hold. In this work we consider successive linearizations around the current state, i.e. the matrices *J* and *Y* (derivatives of *f* and *g* with respect to states and input variables) change in time so they are calculated at the current state, not at the reference one. Since the gradient matrices are typically large, recalculating them at every step is computationally expensive.
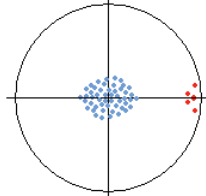
Figure 1. Idealized plot of the eigenspectrum of a dynamic system in the unit circle depicting a clear separation of scales

A Reduced Linear MPC Algorithm for Nonlinear Distributed Parameter Systems

2.2. Model reduction

To reduce the computational cost, we consider linearization on a low dimensional space. Many systems exhibit a separation of scales between the (typically few) dominant modes (corresponding to the eigenvalues of larger magnitude) and the rest of the modes. Fig. 1 shows an idealised picture of a clear separation of scales. Thus, the solution space can be decomposed to two subspaces, the maximal invariant subspace belonging to the *m* eigenvalues with maximum modulus, **P,** and its complement, **Q**. An orthonormal basis for **P**, *Z*, can be calculated efficiently using the Arnoldi method. This is a matrix-free algorithm, which does not require explicit calculation of Jacobians or sensitivities, eliminating the need to provide system equations explicitly and enabling the use of our framework to systems for which a closed form does not exist. The numerical perturbation of f_x to the direction of the vector v is:

$$f_x v \simeq \tfrac{1}{2\varepsilon}\big(f(x+\varepsilon v) - f(x+\varepsilon v)\big), \quad \varepsilon \in \Re_+ \tag{5}$$

Hence, the time integrator can be ran for initial conditions $(x + \varepsilon v)$ and $(x - \varepsilon v)$ for a given reporting time. Since the procedure is successive, a good starting vector for the Arnoldi iteration is the first column of the previous basis. Z is orthonormal, so

$$Z^T Z = I, \qquad ZZ^T = P \tag{6}$$

Where *I* is the identity matrix and *P* the projection matrix onto the subspace **P**. The basis Z can be further exploited for the calculation of a reduced Jacobian for the reporting time involved (Theodoropoulos and Luna-Ortiz, 2006).

$$H = Z^T f_x Z \tag{7}$$

The same methodology is applied for the computations of the other Eq. 3 matrices. It can be shown that the linear system in terms of the reduced state variables, $\xi = Z^T x$ is:

$$\left.\begin{array}{l}\dot{\xi}(t) = Z^T(I - J)Z\xi(t) + Z^T Y u(t)\\ y(t) = \Psi Z\xi(t) + \Theta u(t)\end{array}\right\} \Rightarrow \begin{array}{l}\dot{\xi}(t) = (I - H)\xi(t) + W u(t)\\ y(t) = \Xi\xi(t) + \Theta u(t)\end{array} \tag{8}$$

The linear system of Eq. 8 can be discretized and written in the form of Eq. 4 as:

$$\begin{array}{l}\dot{\xi}(k+1) = A\xi(k) + Bu(k)\\ y(k+1) = C\xi(k) + Du(k)\end{array} \tag{9}$$

2.3. Model Predictive Control

The discrete low-order linear system in state space formulation of Eq. 9 can be used in for MPC over a prediction horizon. For simplicity for the rest of the paper we will write $\xi(t)$ as ξ_t, $u(t)$ as u_t, etc. $\hat{\xi}$ will refer to an estimated state whereas ξ to the real one. For time t = 1, …, N_p, we can estimate the state and output using Eq. 9 as follows:

$$\begin{array}{rclcrcl}
\hat{\xi}_{t+2|t} &=& A\hat{\xi}_{t+1|t} + Bu_{t+1|t} & \qquad & \hat{y}_{t+1|t} &=& C\hat{\xi}_{t+1|t} + Du_{t+1|t} \\[6pt]
\hat{\xi}_{t+3|t} &=& A\hat{\xi}_{t+2|t} + Bu_{t+2|t} & & \hat{y}_{t+2|t} &=& C\hat{\xi}_{t+2|t} + Du_{k+1|t} \\[6pt]
&=& A^2\hat{\xi}_{t+1|t} + ABu_{t+1|t} + Bu_{t+2|t} & & &=& CA\hat{\xi}_{t+2|t} + CBDu_{k+1|t} + Du_{k+2|t} \\[4pt]
\vdots & & & & \vdots & & \\[6pt]
\hat{\xi}_{t+k|t} &=& A^{k-1}\hat{\xi}_{t+1|t} + \displaystyle\sum_{i=1}^{k-1} A^{k-i-1}Bu_{t+k|t} & & \hat{y}_{t+k|t} &=& CA^{k-1}\hat{\xi}_{t+1|t} + C\displaystyle\sum_{i=1}^{k-1} A^{k-i-1}Bu_{t+k|t} + Du_{t+k|t}
\end{array} \tag{10}$$

Eq. 10 can be written in matrix form:

$$Y_t = \Lambda \hat{\xi}_{t+1|t} + \Phi U_t \tag{11}$$

Where:

$$
Y_t = \begin{bmatrix} y_{t+1|t} & \cdots & y_{t+N_p|t} \end{bmatrix}^T \qquad U_t = \begin{bmatrix} u_{t+1|t} & \cdots & u_{t+N_p|t} \end{bmatrix}^T
$$

$$
\Lambda = \begin{bmatrix} C \\ CA \\ \vdots \\ CA^{N-1} \end{bmatrix} \qquad \Phi = \begin{bmatrix} D & & & \\ CB & D & & \\ \vdots & \vdots & & \ddots \\ CA^{N_p-2}B & CA^{N_p-3}B & & D \end{bmatrix} \tag{12}
$$

In MPC, the control move is computed as the solution of a minimization problem. The objective is twofold: to minimize the deviation from the set point and to minimize the control energy applied. This can be formulated in the following objective function:

$$J(\hat{x}_{t+1|t}, U_t) = \tfrac{1}{2} U_t^T H U_t + U_t^T f \tag{13}$$

Where $H = \Phi^T \bar{Q} \Phi + \bar{S}$, $\Gamma = \begin{bmatrix} \Phi^T \bar{Q} \Lambda & -\Phi^T \bar{Q} \end{bmatrix}$ $\tag{14}$

and $f = \Gamma \begin{bmatrix} \hat{\xi}_{t+1|t} \\ R_t \end{bmatrix} - \begin{bmatrix} Su_t \\ 0 \end{bmatrix}$, *with* $R_t = [r_{t+1} \quad \ldots \quad r_{t+N_p}]^T$ the reference output $\tag{15}$

The matrices $\bar{Q}$ and $\bar{S}$ are block diagonal matrices, which contain the weight matrices Q and S. So the optimization subproblem solved in every time interval is:

$$U_t = \arg\min_{U_t} J(\hat{\xi}_{t+1|t}, U_t) \quad s.t. \ L_{eq} U_t = b_{eq}, \ \begin{bmatrix} I \\ -I \end{bmatrix} U_t = \begin{bmatrix} b_u \\ -b_l \end{bmatrix} \tag{16}$$

Where the constraints express physical limitations or design parameters.

Table 1. The proposed Successive Model Predictive Control algorithm

1.	Choose the prediction N_p and control horizons N_c as well as the weight matrices Q and S
2.	Determine the constraints, expressed via the matrices L_{eq}, L_{in}, b_{eq} and b_{in}
3.	For every time interval:
i.	Apply the 1st control decision to the system, as calculated from the previous time step
ii.	Measure the output u_t of the system and estimate the current state $\hat{\xi}_{t+1|t}$
iii.	For the current state linearize the system and retrieve a system of the form of Eq. 9
iv.	Construct the matrices H and Γ (Eq. 14).
v.	Calculate f (Eq. 15).
vi.	Solve the optimization subproblem (Eq. 16) and compute the next N_c control decisions, of which only the first one will be implemented

Table 1 presents the steps involved in our proposed algorithm.

A Reduced Linear MPC Algorithm for Nonlinear Distributed Parameter Systems

3. Case study: temperature control of a tubular reactor with recycle

3.1. Background

To illustrate the features of the proposed scheme, we consider applying it for the control of a tubular reactor with recycle, where an elementary first order irreversible exothermic reaction takes place: A → B (Alonso et al, 2004). The problem is described with two nonlinear partial differential equations (PDEs), which in dimensionless form are:

$$\frac{\partial x_1}{\partial t} = \frac{1}{Pe_1}\frac{\partial^2 x_1}{\partial y^2} - \frac{\partial x_1}{\partial y} - Da(1-x_1)\exp\left(\frac{\gamma x_2}{1+x_2}\right)$$

$$\frac{\partial x_2}{\partial t} = \frac{1}{Pe_2}\frac{\partial^2 x_2}{\partial y^2} - \frac{\partial x_2}{\partial y} - \beta x_2 + CDa(1-x_1)\exp\left(\frac{\gamma x_2}{1+x_2}\right) + \frac{\beta x_{2w}}{Le}$$

$$(17)$$

Here x_1 and x_2 are the dimensionless concentration and temperature correspondingly, x_{2w} is the dimensionless wall temperature and y the dimensionless longitudinal coordinate. All the parameters are described in (Alonso et al, 2004). The boundary conditions are:

$$\frac{\partial x_1}{\partial y} = -Pe_1\left[(1-r)x_1^0 + r\,x_1\big|_{y=1} - r\,x_1\big|_{y=0}\right],\quad at\quad y=0\quad and\quad \frac{\partial x_1}{\partial y}=0\quad at\quad y=1$$

$$\frac{\partial x_2}{\partial y} = -Pe_2\left[(1-r)x_2^0 + r\,x_2\big|_{y=1} - r\,x_2\big|_{y=0}\right]\quad at\quad y=0\quad and\quad \frac{\partial x_2}{\partial y}=0\quad at\quad y=1$$

$$(18)$$

Here r is the recycle ratio. The values of the parameters chosen are $Da = 0.1$, $Pe_1 = Pe_2 = 7.0$, $\gamma = 10.0$, $\beta = 2.0$, $C = 12.0$. The PDEs were discretized using the Finite Element Method, over a mesh of 16 nodes, resulting in a system of 32 ODEs.

3.2. Problem statement

We consider 8 cooling zones on the jacket of the reactor, whose temperature can be controlled independently. The system is stable for $r = 0$ and exhibits sustained oscillations for $r = 0.5$. The objective is to design a controller so that the reactor with $r = 0.5$ behaves the same as $r = 0$. In Fig. 2 the dynamic profiles for the dimensionless concentration and temperature are shown. The output of the model is the dimensionless exit temperature. The size of the basis chosen was 4, therefore from a nonlinear system with 32 variables we extract a liner model of dimension 4 in every time interval. Both control and prediction horizons were set to 7. Identity matrices were used as Q and S. White measurement noise and disturbances of the output were also considered. The values of the cooling zone temperatures can vary from 0 to 1.

3.3. Results and discussion

Numerical experiments showed that the controller designed with the proposed framework successfully stabilized the system and achieved satisfying tracking of the reference trajectory while abiding by the constraints posed. The closed-loop temperature and concentration dynamic profile is illustrated in Fig. 2c. In Fig. 3a, closed loop system profile is presented (solid line) against the reference trajectory (dashed line). The corresponding profile without a controller is shown in dotted lines.

4. Conclusions

We have presented a framework for the design of linear Model Predictive Controllers for nonlinear Distributed Parameter Systems. It is computationally efficient, as it exploits the system's intrinsic separation of scales for model reduction. It relies on an adaptive local linearization of the system in the time domain. Only low-dimensional

Jacobians and sensitivity matrices are employed computed efficiently with numerical directional perturbations and efficient Arnoldi iterations. The resulting linear system is utilized for the calculation of the next control move in a MPC scheme. The advantages of proposed method are simplicity and low computational cost of the controller and at the same time high accuracy of the low order model, as the linearization is adaptive.
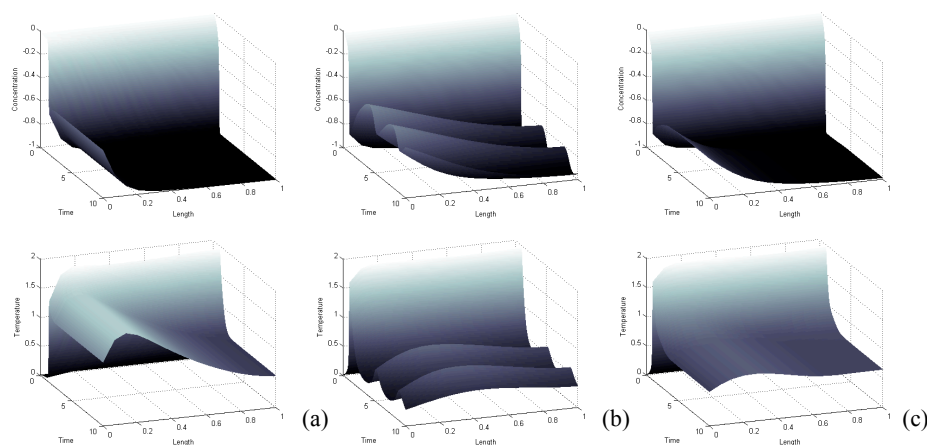
Figure 2. Concentration and temperature dynamic profiles for (a) r = 0 (b) r = 0.5 (c) closed loop.
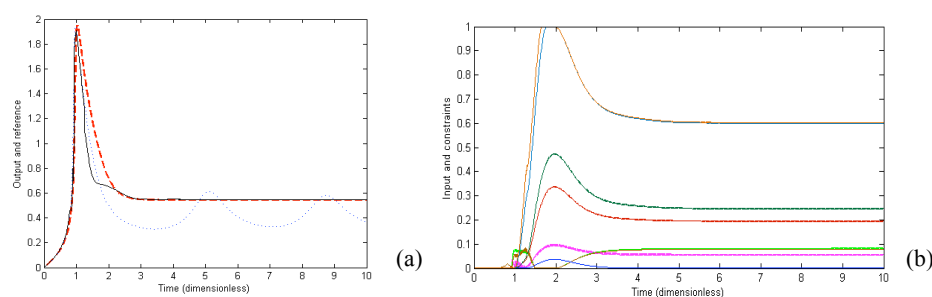
Figure 3. (a) Profiles for the closed loop output (dashed line), the reference (solid line) and the open loop output (dotted line) for r = 0.5 (b) control inputs.

5. Acknowledgements

References

Alonso, A.A., Frouzakis, C.E., Kevrekidis, I.G., 2004, AICHE Journal, 50(7):1438-1452.
Armaou, A., Theodoropoulos, C., Kevrekidis, I.G., 2005, Comput. Chem. Eng., 29(4):731-740.
Datta, B.N., 2003, "Numerical Methods for Linear Control Systems", Elsevier.
Franklin, G.F., Powell, J.D., Workman, M.L., 1990, "Digital Control of Dynamic Systems", 2nd Edition, Addison-Wesley.
Henson, M.A., Computers & Chemical Engineering, 1998. 23(2):187-202.
Mayne, D.Q., Rawlings, J.B., Rao, C.V., P.O.M., 2000, Scokaert, Automatica, 36(6):789-814.
Sunan, H., Kiong, T.K., Heng, LT, 2002, "Applied Predictive Control", Springer.
Theodoropoulos, C. and Luna-Ortiz, E. (2006) in Model reduction and coarse-graining approaches for multiscale phenomena, A. Gorban et al (eds), pp.535-560, Springer.
Zhu, Q.M., Warwick, K., Douce, J.L., 1991, IEEE Proc.-D Control Theory Appl., 138(1):33-40.

Application of non-linear dynamic optimization in advanced process control of product grade transitions of polymerization processes

Laszlo Dobos, Janos Abonyi

Department of Process Engineering, University of Pannonia
H-8200 Veszprém, Egyetem street 10., Hungary
dobosl@fmt.uni-pannon.hu

Abstract

One of the main characteristics of producing synthetic polymers is that the same process is used for the production of different kind of products (various molecular weights, compositions, etc.). Since the producers are forced to satisfy various demands of various costumers, frequent grade transitions are needed. These grade transitions are expected to be short and effective to avoid the production of so-called off-specification products. It became very popular to apply model predictive controllers (MPCs) to reduce the quantity of off-specification products, however most of them use linear models for prediction. Since polymerization reactions are highly non-linear, using linear models may cause significant difference between the response of the model and of the real plant and this can cause problems e.g. in predictive control. The difference appears mainly during grade transitions, hence it is important to tune the appropriate parameters of the regulators to realize the grade transitions as soon as possible. In this article a novel method – in the field of predictive control - is introduced for parameter tuning, although this method is well known in the field of experiment design. The statistical tools like design of experiments (DoE) permit the investigation of the process via simultaneous changing of factors' levels using reduced number of experimental runs. Through a case study the applicability of full factorial design is going to be examined. It will be proven that full factorial design is appropriate for finding the right tuning parameters of MPC controlled polymerization reactor. The aim of the case study is the reduction the time consumption of grade transitions, so applying the tools of design of experiments as a quasi-APC.

1. Introduction

The production of the synthetic polymers represents an important part of chemical industry. In this industrial segment a general practice is that one reactor is used for producing various products (with various molecular weights, compositions, etc.).
During transitions between products, off-specification products are produced. This product is generally worth less than the on-specification material (which fulfill all the commercial and quality requirements), therefore it is crucial to minimize its quantity. The on-specification product can be produced under varying circumstances and at varying operating points, which are more or less sound from an economical point of view, motivating the optimization of the production during production stages.

In these processes a large number of different grades are produced, and the transition times between the productions may be relatively long and costly in comparison with the total amount produced. The demand for reduction of the time and cost of grade transition inspires the researchers to find more innovative solutions (Flores-Tlacuahuac and Biegler, 2008, BenAmor et al., 2004) The optimization of complex operating processes generally begins with a detailed investigation of the process and its control system (Lee et al., 2004). It is important to know how information stored in databases can support the optimization of product transition strategies, and how hidden knowledge an be extracted from stored time-series, which can assure additional possibilities to reduce the amount of off-grade products. The optimization of product grade transition is a typical and highlighted task in process industry (McAuley and MacGregor, 1991).

Unfortunately, it is very difficult to find the right tuning parameters of the controllers in the whole operation range because of the nonlinearity of the polymerization process, and identified models (for MPCs) from input-output data are mostly linear. In the industrial practice commercial Advanced Process Control (APC) systems are installed to handle the problem of tuning parameters of the controllers. In the most of cases the the operation of these systems based on a linear cost function, which usually contains the cost of the production and the price of the raw materials and products. Obviously the goal of the application of APCs is to maximize the quantity of on-specification materials and at the same time minimize the cost of the production. Since these control systems are relatively expensive so limitedly accessible, in some cases the right parameters of the production (e.g. set-points, tuning parameters of controllers, valve positions) is determined experimentally using the intuition of engineers.

One of the common experimentation approaches is One-Variable-At-a-Time (OVAT) methodology, where one of the variables is varied while others are fixed. Such approach depends upon experience, guesswork and intuition for its success. On the contrary, the statistical tools like design of experiments (DoE) permit the investigation of the process via simultaneous changing of factors' levels using reduced number of experimental runs. Such approach plays an important role in designing and conducting experiments as well as analyzing and interpreting the obtained data. These tools present a collection of mathematical and statistical methods that are applicable for modeling and optimization analysis in which a response or several responses of interest are influenced by various designed variables (factors) (Farzaneh and Tootoonchi, 2009).

In this study the applicability of full factorial design is going to be examined through a case study. It will be proven that full factorial design is appropriate for finding the right tuning parameters of MPC controlled polymerization reactor. The aim of the case study is the reduction the time consumption of grade transitions, so applying the tools of design of experiments as a quasi-APC.

The paper is organized as follows: in Section 2 the theoretical background of the applied methodology is going to be presented. In section 3 the applied methodology will be discussed throughout an application example, which is followed by a conclusion.

2. Theoretical background and the applied methodology

The applied methodology can easily be inserted into the scheme of the classical experiment design steps:

I.	As the first step the mean values of the parameters are determined. As the step of experiment design the matrix of factorized variables is constructed which practically contains the values of the factors in the different experiments.

Application of non-linear dynamic optimization in advanced process control of product grade transitions of polymerization processes

II.   In second step the experiments are carried out and the results of the experiments are saved.

III.   In this step, the experiment with the best value of the objective function is chosen and the parameters of this experiment will be substituted into the mean values of initial parameters in the next iteration step/experiment. In this step it is necessary to determine the interval of variation based on the gradient of the fitted linear function estimated by regression based on the results of experiments.

The iteration is continued pre-determined times, because it is important to note that carrying out experiments is a time and cost demanding process.

Experimental-statistical tools permit to reduce the number of experimental runs and to investigate the interaction effects between the designed variables (factors). Therefore, this methodology was employed for optimization of the tuning parameters of linear model predictive controller. These variables are the value of prediction and control horizon and the move suppression coefficient. For statistical calculations the actual variables were coded according to Eq. (1).

$$x_i = \frac{z_i - z_i^0}{h_i} \qquad (i = 1...n) \tag{1}$$

where z denotes the current value of the designed variable, z^0 is the center point (mean) of the designed variable, h is the interval of variation, x is the coded level of the designed variable (dimensionless value) and n is the number of variables. Thus, each variable has two different coded levels (±1). The full factorial design is a set of experimental runs where each level of the designed variable is investigated at both levels ($+1$) and (-1) of all the other factors. It is an orthogonal design, which allows the estimation of a factor effect independently of all other effects.

Based on experimental design matrix the factorial models were developed to ascertain the relationship between responses and factor effects. According to this method, a response y is set as a functional relationship of the designed variables (factors) and for a full factorial design the effects of factors may be estimated by linear regression model:

$$y = b_0 + \sum_{i=1}^{n} b_i x_i \tag{2}$$

where y is the predictor of the response, b_0, b_i are the regression coefficients.

In order to ascertain the regression coefficients of the factorial model, Eq. (2)., the linear regression method was employed. According to this method the least square estimations of the regression coefficients can be written:

$$\bar{b} = (\bar{\bar{x}}^T \bar{\bar{x}})^{-1} \bar{\bar{x}}^T \bar{y} \tag{3}$$

where b is the vector of regression coefficients, x is the matrix of the independent variables levels, y is the vector of the response (experimental values).

To realize the *Step III.* described in 2.1. section the interval of variation has to be determined, which is in this case study represented by Eq. (4):

$$h = \alpha \cdot grad(y) \tag{4}$$

where α constant, *grad(y)* is equal to the parameter values determined in Eq. (3).

3. Application example

In this case study the task is to realize a grade transition between two different kind of grades, called A and B. The main goal is to minimize the amount of the off-grade product, so reduce the grade transition time as much as possible. To qualify the success of this purpose the following linear cost function is applied:

$$E = P_{on-spec} \cdot Q_{on-spec} - P_{off-spec} \cdot Q_{off-spec} \tag{5}$$

The studied polymerization reactor is a SISO (single input-single output) process, a CSTR (Continously Stirred Tank Reactor) where a free radical polymerization reaction of methyl-metacrylate is considered using azobisisobutironitil (AIBN) as initiator, and toulene as solvent. The aim of the process is to produce different kinds of product grades. The number-average molecular weight is used for qualifying the product and process state, where the influenced (control) variable is the flow rate of inlet initiator. When this assumption is considered, and the effect of the temperature is neglected, the multi input-multi output model could be reduced to a SISO process. Because of the isothermal assumption, a four-state model can be obtained. (Maner and Doyle, 1997)

As the next step, the model for MPC is chosen: a linear, step response SRM model was identified, since the designed MPC – a DMC controller – is based on it (Ricker, 1988). As a following step the DMC controller was installed to the reactor to realize the grade transition (Abonyi et al., 2000). Since the set point signal and all the operating parameters - except the tuning parameters of the controller - were pre-determined, the inputs of the controlled system were the value of prediction and control horizon (H_p and H_c), and move supression coefficient (Λ), since these are the main factors of the objective function of the DMC controller:

$$\min_{\Delta u(k+j)} \sum_{j=H_{p1}}^{H_{p2}} (w(k+j) - y(k+j))^2 + \lambda \sum_{j=1}^{H_c} \Delta u^2(k+j-1) \tag{6}$$

The output of the whole system was based on Eq. (5).

To determine the inputs of the system, the steps of classical design of experiments were applied. As Fig.1. shows execution, only 2 iteration steps were enough to reach the maximum value of the cost function (values are normalized).
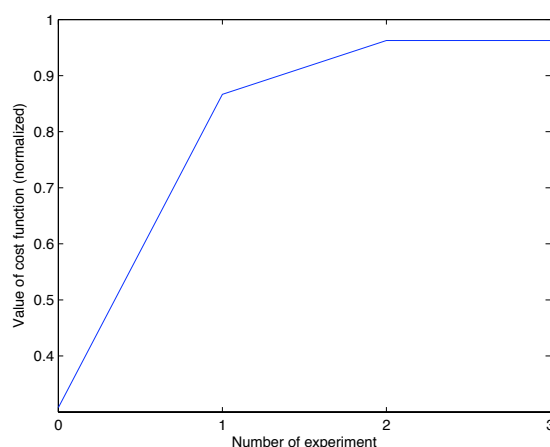
Figure 1. – Change in the value of the cost function during the experiments

Application of non-linear dynamic optimization in advanced process control of product grade transitions of polymerization processes

To express the development in the production of on-specification products the performance of the controller with initial and tuned parameters have been compared, see Fig. 2. (values are normalized).
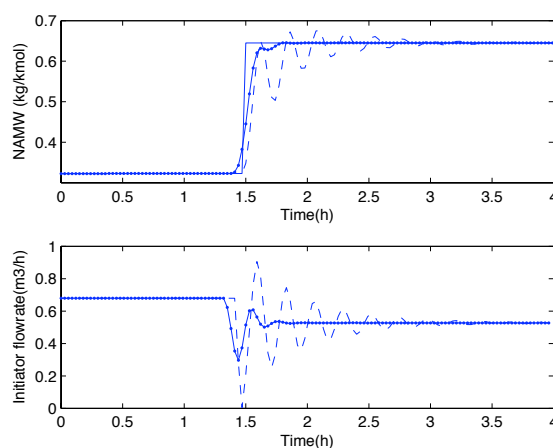
Figure 2. – Comparison of the grade transitions with initial parameters (dashed line) and with the experimentally determined parameters (dotted line)

As Fig. 2. shows, the time demand of the grade transition is significantly shortened. Although a badly tuned MPC is also capable of performing the grade transition, the optimization of tuning parameters is necessary, because in the optimized case the system produces on-spec product in the 91 % of the examined time horizon in contrast to the initial guess where this ratio is only 63 %. To have the opportunity to evaluate the result of the applied methodology, evolutionary strategy (ES) was used to optimize the examined parameters. Evolutionary strategy is a stochastic optimization algorithm that uses the model of natural selection (Madár and Abonyi, 2005). The advantage of ES is that it has proved to be successful in problems that are highly nonlinear and stochastic. Because of the nonlinearity of the polymerization process, this kind of optimization method is worth applying. Deterministic optimization algorithms, like sequential quadratic programming (SQP) have several drawbacks. Since H_p and H_c are discrete variables, it was not possible to use the SQP algorithm in MATLAB Optimization Toolbox successfully, hence ES was applied instead.

In this case study the result of ES and experiment design were almost the same as it is summarized in Table 1.

Table 1. Comparison of the experimentally tuned and the optimized (with ES) values

	H_p	H_c	λ (10^{10})	*Value of obj. fun.*	*On-spec time (%)*	*Function evaluated*
Evolution. strat.	6	3	0.95	0.985	91.79	58
Experiment des.	6	6	1.4	0.963	91.04	16

To sum up the experience of the case study, it can be stated that the method of experiment design is applicable as described in this paper. Although the optimal solution is not reached, its value can be estimated as Table 1 shows. One of the main advantage of this method that only few experiment are needed compared to the optimization method and almost the optimal solution is reached with significantly less effort.

4. Conclusion

In this study an optimization approach incorporating the factorial modeling and analysis approach was applied on optimization of product grade transition of a polymerization process. A full factorial design involving a reduced number of experimental runs to localize the optimal value of tuning parameters of a linear model predictive controller (DMC) which is attached to a benchmark polymerization example. To evaluate the obtained tuning parameters these are also optimized by using evolution strategy. The comparison of results shows that the tools of experiment design are successfully applicable in cases like this, since the solution is pretty close to the optimal solution determined by ES. The main benefits of using the introduced framework are the less number of necessary experiments and there is no need to possess the model of the operating system in contrast to the application of other optimization methods.

5. Acknowledgements

References

Abonyi J., Bódizs Á., L, Nagy L., Szeifert F. 2000, *Hibrid fuzzy convolution model and its' application in predictive control,* Chemical Engineering Research and Design. Vol. 78. 597-604

BenAmor S., Doyle F.J., McFarlane R., *Polymer grade transition control using advanced real-time optimization software*, Journal of Process Control 14 (2004) 349-364

Farzaneh Y., Tootoonchi A.A., 2009, *A novel data reduction method for Takagi-Sugeno fuzzy system design based on statistical design of experiments*, Applied Soft Computing 9 1367-1376

Flores-Tlacuahuac A., Biegler L.T., 2008, *Integrated control and process design during polymer grade transition operations*, Computers and Chemical Engineering, 32 2823-2837

Lee YH, Min KG, Han C, Chang KS, Choi TH, *Process improvement methodology based on multivariate statistical analysis methods*. Control Engineering Practice 12 (2004) 945–961.

Madár J. and Abonyi, J.: *Evolutionary Algorithms*, Chapter 2.10. in Instrument Engineers' Handbook, 4th Edition, Volume 2 - Process Control, editor: B. Liptak, CRC Press, 2005

Maner B.R., Doyle F.J., 1997, *Polymerization reactor control using autoregressive volterra-based MPC*, AIChe Journal 43 1763-1784

McAuley K.B., MacGregor J.F., 1991 *On-line Inference of Polymer Properties in an Industrial Polyethylene Reactor*, AIChe Journal 37, 825-835

Ricker N.L., 1988, *The Use of Biased Least – Squares Estimators for Parameters in Discrete – Time Pulse Response Model*, Ind. Eng. Chem. Res., 27, 343-350

A reactive distillation case study for decentralized control system design using mixed integer optimization

Ganesh Paramasivan[a,*], Achim Kienle[a,b]

[a]*Max Planck Institute for Dynamics of Complex Technical Systems, Sandtorstrasse 1, D-39106, Magdeburg, Germany, *paramasivan@mpi-magdeburg.mpg.de*
[b]*Institut f ür Automatisierungstechnik, Otto-von-Guericke-Universität, Universitätplatz 2, D-39106, Magdeburg, Germany, kienle@mpi-magdeburg.mpg.de*

Abstract

Decentralized control system design comprises the selection of a suitable control structure and controller parameters. In the present paper, a mixed integer optimization is used to determine the optimal control structure and the optimal controller parameters simultaneously. The process dynamics is included explicitly into the constraints using a rigorous nonlinear dynamic process model. Depending on the objective function which is used for the evaluation of competing control systems, we propose two different formulations. These formulations lead to a mixed integer dynamic optimization problems which are solved with mixed integer nonlinear programming methods after discretization of the underlying process dynamics. The proposed methodology is applied to an inferential control of a reactive distillation column as a challenging benchmark problem for chemical process control. Results are compared with previous studies, and conclusions are drawn for future work in this field.

Keywords: decentralized control system, reactive distillation control, mixed integer dynamic optimization, mixed integer optimization for control systems.

1. Introduction

In chemical process control, frequently decentralized linear controllers are used because of their ease of implementation and handling in practice. Usually, the decentralized control system design is done sequentially by first fixing the control structure and then tuning controller parameters. Control structures are often selected using heuristic rules [1] or some simplified interaction measures like Relative Gain Array (RGA) [2]. Controllers for the selected control loops are often designed using some single input single output (SISO) tuning rules in combination with detuning strategies to account for the interaction between the different control loops. This approach is simple and intuitive, but often suboptimal. In particular, hard constraints on the process dynamics cannot be taken into account.

To overcome these limitations, we take an algorithmic approach in the present paper using mixed integer optimization to determine the optimal control structure and controller parameters simultaneously. Depending on how the performance of decentralized control system is measured, this leads to a mixed integer linear programming (MILP) problem or a mixed integer nonlinear programming (MINLP) problem or mixed integer dynamic optimization (MIDO) problem if the process dynamics is explicitly taken into account. In the present paper, we propose two different

formulations based on closed loop response characteristics: (1) minimizing the effort to achieve a specified performance (2) maximizing the overall performance. Both formulations lead to MIDO problems which are solved with MINLP methods after discretization of the underlying process dynamics. The proposed methodology is applied for an inferential control of a reactive distillation column which was proposed by Al-Arfaj and Luyben [3] as a challenging benchmark problem for process control.

2. Mathematical formulation

This section considers the mathematical formulation of the decentralized control system design with an appropriate Proportional Integral and Derivative (PID) controller description,

$$\min_{kp,\tau_I,\tau_D,\delta} J(x_d(t), x_a(t), u(t), y(t), v, \delta, t) \tag{1}$$

s.t., semi-explicit Differential Algebraic Equation (DAE) model:

$$\frac{dx_d(t)}{dt} = f(x_d(t), x_a(t), u(t), v, t)$$
$$h(x_d(t), x_a(t), u(t), v, t) = 0 \tag{2}$$

$$g(x_d(t), x_a(t), u(t), v, t) \leq 0 \tag{3}$$

$$\eta(x_d(t), x_a(t), u(t), y(t), t) = 0 \tag{4}$$

$$e_i(t) = y_{sp,i} - y_i(t), \qquad \forall i = 1,2,.....,N_y \tag{5}$$

$$u_j(t) = \sum_{i=1}^{N_y} kp_{i,j}\left(e_i(t) + \frac{1}{\tau_{I,j}} \int_0^t e_i(t)dt + \tau_{D,j}\frac{de_i(t)}{dt} \right), \quad \forall j = 1,2,......,N_u \tag{6}$$

$$kp_{i,j}^L \delta_{i,j} \leq kp_{i,j} \leq \delta_{i,j} kp_{i,j}^U$$
$$\tau_{I,j}^L \leq \tau_{I,j} \leq \tau_{I,j}^U$$
$$\tau_{D,j}^L \leq \tau_{D,j} \leq \tau_{D,j} \tag{7}$$

$$\sum_{i=1}^{N_y} \delta_{i,j} \leq 1, \qquad \forall j = 1,2,......,N_u \tag{8}$$

$$\sum_{j=1}^{N_u} \delta_{i,j} \leq 1, \qquad \forall i = 1,2,......,N_y \tag{9}$$

$$\delta_{i,j} \in [0,1] \tag{10}$$

Where, $x_d \in \Re^{N_d}$ and $x_d \in \Re^{N_a}$ are the vector of differential state and algebraic variables, $u \in \Re^{Nu}$ is the vector of the manipulated variables, $v \in \Re^{Nv}$ is the vector of disturbances acting on the plant and $y \in \Re^{Ny}$ is the vector of potential measurements (output

A reactive distillation case study for decentralized control system design using mixed integer optimization

variables). The objective function J is an integral over time which is minimized subject to dynamic process model and operating constraints, where the process behaviour is described by the system of differential equations f, and algebraic equations h; g is the vector of inequality constraints that define the feasible operation of the plant; η is the functional relationships between measurements, state variables and input variables. Eq. (6) defines an ideal PID controller written in the continuous time domain form and kp, τ_I and τ_D are controller parameters. Binary variables δ are introduced in the bounds of the controller gain in Eq. (7) in order to restrict the values of the elements of the gain matrix used in the selected pairs and at the same time ensure that the gains of the loops not selected become zero [4]. Equations (8) and (9) are used to enforce the requirements of a decentralized control structure. Because of the presence of continuous and discrete variables, the present optimization problem represents a mixed integer dynamic optimization (MIDO) problem.

2.1. Objective function
This section considers two different formulations to measure the performance of the decentralized control system on the basis of the closed loop response characteristics.

2.2. Formulation I (Minimize the effort to achieve a given performance)
In this formulation, the objective function is a measure of how much effort (manipulated variables movement from the steady state u_{ss}) is required in order to achieve the specified performance on the outputs. The objective function is given as,

$$\min_{kp,\tau_I,\tau_D,\delta} J = \int_0^{t_f} ((u_{ss} - u)^T R(u_{ss} - u))dt \tag{11}$$

The performance specifications are given in terms of overshoot and settling time by inequality constraints as follows,

$$y_i(t) - y_{sp,i} \le y_{ov,i} \quad , \quad \forall t \in [0, t_{set}] \tag{12}$$

$$0.95 y_{sp,i} \le y_i(t) \le 1.05 y_{sp,i} \quad , \quad \forall t > t_{set} \tag{13}$$

In the above constraints, Eq. (12) defines the specified overshoot (y_{ov}), the maximum value by which the output variables are allowed to proceed beyond the set point. Equation (13) defines the desired performance in terms of settling time. This is the time in which the control variables have entered and remained within a specified ε band (error band) around the desired set point. In this study, a value of $\pm 5\%$ is chosen.

2.3. Formulation II (Maximize the overall performance)
In this formulation, the objective function minimizes a weighted sum of the quadratic control error and quadratic control action. The objective function is given as,

$$\min_{kp,\tau_I,\tau_D,\delta} J = \int_0^{t_f} [(y_{sp} - y)^T Q(y_{sp} - y) + (u_{ss} - u)^T R(u_{ss} - u)]dt \tag{14}$$

Q and R are positive definite weighting matrices, which are used here for scaling purposes as discussed in the application section.

2.4. Solution approaches

The inclusion of discrete decisions in the decentralized control system design leads to very challenging mixed integer dynamic optimization (MIDO) problems for complex chemical processes. The case study which is considered for validating the formulations is a reactive distillation column described by nonlinear differential algebraic equations. The simultaneous solution methodology for MIDO problem requires that the state variables are discretized and the MIDO problem is transformed into a finite-dimensional MINLP problem. In the present work, orthogonal collocation on finite elements [5] is used to approximate the differentials at each collocation point. Throughout this work, the resultant MINLP problems are solved using a GAMS/SBB [6].

3. Case study – Ideal reactive distillation

An ideal reactive distillation column with two products and two feeds presented by Al-Arfaj and Luyben [3] is considered as a case study and is shown in Fig 1. The reversible reaction occurring on the reactive trays is given by,

$$A + B \Leftrightarrow C + D \tag{15}$$
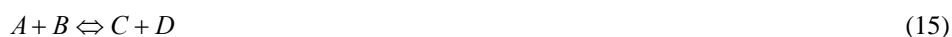
The reactants A and B are intermediate boiling between the products. Therefore, the fresh feed stream F_{0A} containing reactant A is fed at the bottom of the reactive zone, and the fresh feed stream F_{0B} containing reactant B is fed at the top of the reactive zone. The detailed mathematical modeling of the reactive distillation column is found in Al-Arfaj and Luyben [3]. Steady state composition and temperature profiles are given in Fig 2.

3.1. Inferential control of reactive distillation

In this case study, we focus on inferential control, i.e. we use temperature instead of composition measurement for product composition control. The selection of the trays for temperature control loops is the key issue in the inferential control of reactive distillation. Kaymak et al., [7] presented the use of steady state gain and singular value decomposition analysis to choose the trays for temperature control. In the present case study, pressure and level control loops are assumed to be the same as given by Kaymak et al [7]. Further, PI controllers are considered for the temperature control loops in order to compare the results with previous studies, even though the general mathematical formulation was given for PID controllers. Furthermore, the same test scenario is considered, i.e. a ±10% step change in the vapor boil up VB.
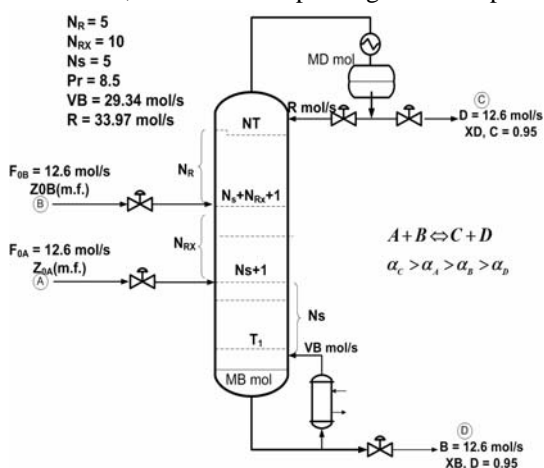
Figure 1 Ideal reactive distillation column   Figure 2 Steady state profiles
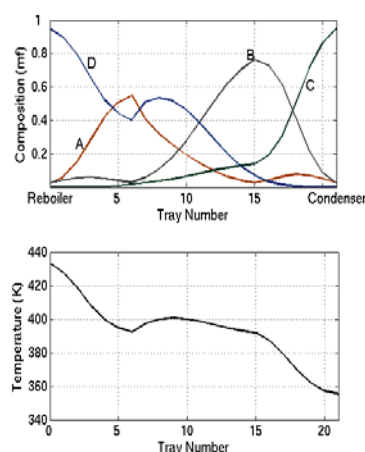
A reactive distillation case study for decentralized control system design using mixed integer optimization
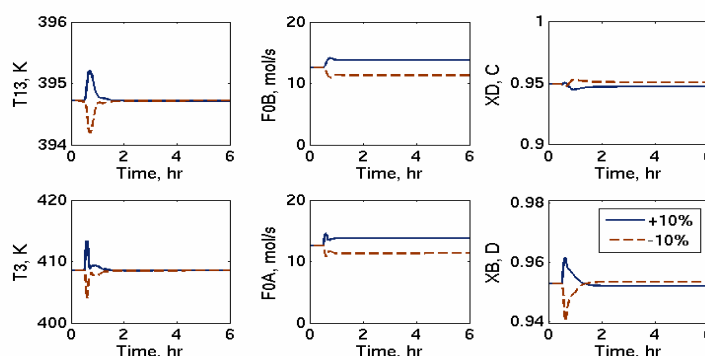
Figure 3 Closed loop response for ±10% change in the production rate VB by formulation I

4. Results

4.1. Decentralized PI controller design using formulation I

In order to apply the formulation I, the performance constraints in terms of the overshoot and settling time are specified for the top and bottom purity. Maximum overshoot of 1% of the set points and a settling time of 1 hr are considered as the performance specifications. These performance specifications provided in the algorithm acts as active constraints and the closed loop response is lying within the constraints. It should be noticed that the problem becomes infeasible if the performance specifications are chosen too tight. The optimal control loops and PI control parameters are shown in Table 1. The closed loop performance of the proposed control structure is shown for a ±10% change in VB in Fig 3 meeting the above requirements.

4.2. Decentralized PI controller design using formulation II

Next, formulation II is applied in order to maximize the overall performance of the control system. The weighting matrices Q and R are diagonal matrices, in which each diagonal entry is the inverse of the steady state values. Alternatively, steady state sensitivities can be used for the weighting matrices Q and R. In the present case, however this leads to very similar results. The optimal temperature control loops and PI control parameters are shown in Table 1. The closed loop performance of the proposed control structure is shown in Fig 4 for ±10% change in VB. The overall response is well behaved and both the temperature loops are fast, less oscillatory and achieved their set points less than an hour compared to the earlier studies by Kaymak et al [7].
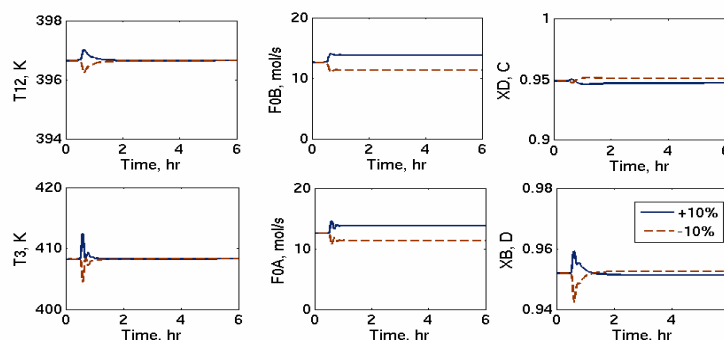
Figure 4 Closed loop response for ±10% change in the production rate VB by formulation II

Table 1. The model statistics and the solutions obtained via the resultant MINLP problems

	Formulation I			Formulation II		
Control structure and PI parameters		kp	τ_I(min)		kp	τ_I(min)
	$F_{0A} - T_3$	0.35	13.5	$F_{0A} - T_3$	0.58	9.5
	$F_{0B} - T_{13}$	2.80	15.0	$F_{0B} - T_{12}$	4.50	14.8
Binary variables	40			40		
Continuous variables	34481			34481		
Equations	34704			34364		
CPU time	85 min			25 min		

5. Discussion and conclusion

A systematic procedure for simultaneous selection of a decentralized control structure and optimal control parameters was developed in view of (1) minimizing the effort to achieve a specified performance or (2) maximizing the overall performance in terms of a quadratic function. Both formulations are constructed as a MIDO problem which is transformed into a MINLP problem using a full discretization approach. For both formulations, the MINLP solution algorithm was capable of successfully solving for the optimal control structure and the optimal controller parameters for a reactive distillation case study. The resulting control structure achieves a better closed loop performance than earlier studies [7] for the same test scenario considered. The same methodology can be applied to multivariable controller design by relaxing the constraints in Eq. (8) and (9). This may give further improvement in the closed loop performance depending on the application. However, it is worth noting, that in the present case not much improvement were found for a multivariable controller compared to the decentralized controllers discussed above.

In the reactive distillation case study, the selection of two-temperature control loops from 20 tray temperature measurements and then combination with 2 manipulated variables has 380 numbers of possible combinations. Table 1 shows the model statistics obtained via both formulations. The resultant MINLP problems required 85 and 25 minutes of CPU solution time for the two formulations respectively, on a Linux workstation with Intel Pentium D CPU 3.0 GHz processor. The MINLP optimization method employed in this paper can only find local optima due to the inherent nonlinearity of the problem. Solution upto global optimality is a challenging task for future research. Further, the control system design was focused on a single particular disturbance scenario. For more complex plants and more complex, i.e. collection of disturbance scenarios, more advanced MIDO solution strategies are required, which will be also the subject of our future research.

References

[1] W.L.Luyben and B.D.Tyreus and M.L.Luyben, Plantwide process control, McGraw-Hill, New York, 1999.
[2] E.H. Bristol, IEEE Trans. Autom. Control., 11 (1966) 133.
[3] M.A. Al-Arfaj and W.L. Luyben, Ind. Eng. Chem. Res., 39 (2000) 3298.
[4] N. Narraway and J.D. Perkins, Comput. Chem. Eng., 18 (1994) S511.
[5] L.T. Biegler, Chem. Eng. Process., 46 (2007) 1043.
[6] A. Brooke and D. Kendrick and A. Meeraus, GAMS: A user's guide, GAMS Software GmbH, Germany, 1998.
[7] D.B. Kaymak and W.L. Luyben, Chem. Eng. Sci., 61 (2006) 4432.

Type-2 fuzzy control of a fed-batch fermentation reactor

Mosè Galluzzo, Bartolomeo Cosenza

Dipartimento di Ingegneria Chimica dei Processi e dei Materiali, Università degli Studi di Palermo, Palermo, Italy, E-mail address: galluzzo@unipa.it

Abstract

The aim of the paper is to present the application of type-2 fuzzy logic controllers (T2FLCs) to the control of a fed-batch fermentation reactor in which the penicillin production is carried out. The performance of the control system using T2FLCs is compared by simulation with that of a control system using type-1 fuzzy logic controllers (T1FLCs). The non linear model used for the simulation study is an unstructured model characterized by the presence of non linearities, parameter uncertainty and measurement noise. Simulation results confirm the robustness of the T2FLC which shows a better performance than its type-1 counterpart particularly when uncertainties are present in the control system.

Keywords: type-2 fuzzy logic controller, uncertainties, non linear system, fed batch fermentation reactor.

1. Introduction

Many industrial chemical processes are characterized by high nonlinear dynamics and by uncertainties and often the control of these processes that makes use of traditional PID controllers does not give satisfactory results. Fuzzy logic controllers provide a relative simple way to control system which have a considerable non-linear behavior, but when the processes to be controlled are characterized by uncertainties, the choice of traditional fuzzy controllers (*type-1* fuzzy controllers), that are usually implemented using type-1 fuzzy sets, may not be the optimal solution. There exist two typologies of fuzzy controllers: the *type-1* and the *type-2* fuzzy controllers. It has been demonstrated that type-1 fuzzy logic controllers cannot handle the uncertainties present in the control system. The type-2 fuzzy logic systems instead, making use of more complex fuzzy sets (type-2 fuzzy sets) with a larger number of parameters, can handle all kind of uncertainties and therefore minimize their negative effects in the control system.

Type-2 fuzzy logic controllers have been applied in anaesthesia control (Castillo et als. 2005), in level control (Wu and Tan, 2006) and more recently in autonomous mobile robot control (Martinez et als., 2009) and in biochemical reactor control (Galluzzo and Cosenza, 2009). The process that is considered in this work is the production of penicillin in a fed-batch reactor. The process model used for the simulations is an unstructured model (Birol et als., 2002) that provides a detailed description of the process, taking in to account many input variables as pH, temperature, aeration rate, agitation power, substrate feed flow rate and the CO2 evolution term. The task of the control system is to maximize the penicillin production, at same time safeguarding the environment for the growth of the biomass by keeping constant the pH and the temperature of the system at the desired values. The performance of type-1 FLCs was first compared with that of traditional PI controllers, developed by (Birol et als., 2002),

and then with that of type-2 fuzzy controllers. The process model was made more realistic by the introduction of measurement noise and dead time.

2. Type-2 Fuzzy Logic

2.1. Type-2 fuzzy sets

The concept of higher order fuzzy sets, as a natural development of type-1 fuzzy sets, was firstly introduced by Zadeh (1975). The effective development of type-2 fuzzy logic takes place only twenty years later (Karnik and Mendel, 1998). An extended presentation and discussion of type-2 fuzzy sets can be found in Mendel (2001).

While membership functions of type-1 fuzzy sets are crisp functions, membership functions of type-2 fuzzy sets are fuzzy numbers in the range 0-1; each fuzzy number represents the uncertainty that is associated with each value of the set.

A type-2 fuzzy set $\widetilde{A}$ is mathematically defined (Karnik and Mendel, 2002) as:

$$\widetilde{A} = \int_{x \in X} \int_{u \in J_x} \mu_{\widetilde{A}}(x,u)/(x,u) \qquad \text{(Eq. 1)}$$

where $x \in X$, $u \in J_x \subseteq [0,1]$ and $\mu_{\widetilde{A}}(x,u) \in [0,1]$ is a type-2 membership function (Fig.1a).

The domain of this secondary membership function (third dimension in Fig. 1a) is the *primary membership function* (FOU = Footprint of Uncertainty) of x (blurred triangular areas of Fig. 1a). At a specific value of *x*, say *x'*, there is not a single value for the membership function as for type-1 fuzzy sets; instead, the membership function takes on values (in the third dimension) wherever the vertical line intersects the blur.

For computational reasons only interval type-2 fuzzy sets, a particular case of type-2 fuzzy sets, are used. An interval type-2 fuzzy set $\widetilde{A}_I$ (Fig1b) is defined as follows:

$$\widetilde{A}_I = \int_{x \in X} \int_{u \in J_x \subseteq [0,1]} 1/(x,u) \qquad \text{(Eq. 2)}$$

In this case the membership grades of all elements in the FOU are all equal to 1 (Fig. 1b).
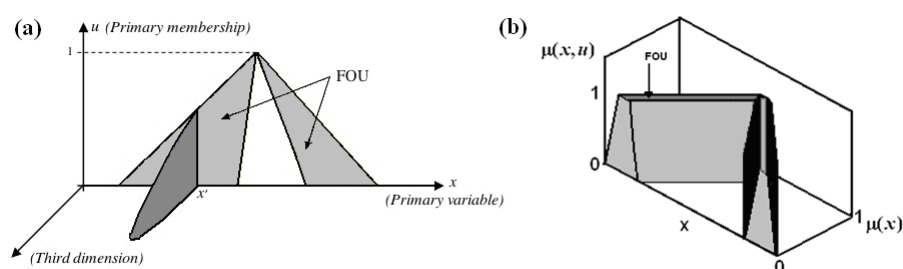
Fig.1. (a) General type-2 fuzzy set and (b) Interval type-2 fuzzy set. FOU (shaded region).

2.2. Uncertainty

Fuzzy logic based on type-2 fuzzy sets shows all its potential if applied to systems characterized by uncertainties. It is well known that uncertainty is an inherent part of real control systems. The measurement noise, the coarse estimation of process parameters, the system parameters changes, even the ambiguity in the meaning of words used in the rules are all source of uncertainties (Klir and Wierman, 1998). The FOU, the main characteristic of type-2 fuzzy sets that corresponds to the shaded region in Fig.1, is used to capture the uncertainties present in the system, therefore minimizing their negative effects.

3. Process control

3.1. Control strategy

The model of the process for the penicillin production is that described by (Birol et als., 2002). The main aim of the fed-batch fermentation control is to maximize the penicillin production, minimizing the presence of not desired compounds. This condition must be compatible with the safeguard of the biomass or the process will not give any good result. The cellular consumption of glucose produces acid substances which decrease the pH in the system, consequently the controller that manipulates the basic stream has a fundamental role in keeping a suitable environment for the penicillin production. The controller that manipulates the acid stream operates only for excessive actions of the first controller.

The set point of the pH control loop is set at a value of 5.0. It is also necessary to control the temperature of the system in order to have a good running of the process, since with the growth of the biomass the reaction heat, produced by the cellular metabolism, increases the temperature of the system, risking to wipe out all cellular population.

The set point of the temperature control of the culture medium is set at 298 K. Two fuzzy controllers were designed to control the temperature of the culture by manipulating the heating/cooling water flow rates.

In conclusion the performances of three control systems (PID, type-1 FLC and type-2 FLC) were analyzed and compared by simulation. Each control system is composed by 4 controllers of the same kind: traditional PID, traditional type-1 FLCs and the new type-2 FLCs respectively.

3.2. Fuzzy controllers

Each fuzzy controller was designed using two input variables: the error (e) and the integral of the error (inte) and one output (control variable). For each variable of type-1 FLCs, Gaussian membership functions were chosen and Sugeno inference method (with constant output) was used. Also Type-2 FLCs use the Sugeno inference method (with constant output): the type-2 Gaussian internal and external membership functions,. the gain, the integral actions, the operative range, the shape of membership functions and the values of constant outputs of type-1 FLCs were chosen minimizing the ISE (Integral-Square-Error) index. The structure of type-2 FLCs is the same of type-1 FLC (same operative range, rules, membership functions layout, Sugeno outputs, gain and integral actions); the only difference regards the Gaussian membership functions amplitude. Each type-1 fuzzy Gaussian membership functions is in fact characterized by an amplitude value that is the average of type-2 fuzzy internal and external Gaussian membership functions amplitude values.

The rules of a type-2 FLS just represent a type-2 relation between the input space and the output space. Their structure is the same of type-1 rules, the only difference consists

in the membership functions nature. The lth-rule of a type-2 FLS with p inputs and l outputs is given by:

$$R^l: \text{IF } x_1 \text{ is } \tilde{F}_1^l \text{ and } \ldots \text{and } x_p \text{ is } \tilde{F}_p^l, \text{ THEN } y \text{ is } \tilde{G}^l \quad l = 1, \ldots M$$

4. **Results and discussions**

4.1. *Type-1 FLC vs PID*

Birol et al. (2002) developed a software, available at the web site (http://www.chee.iit.edu/_/control/software.html), for the simulation of the temperature and pH control of the penicillin process, that makes use of four traditional PID controllers.

In Fig. 2 the performances of traditional controllers of the simulator software were compared with that of type-1 fuzzy logic controllers for the control of pH and temperature respectively. The type-1 FLC leads the system from a disturbed point to the desired value of pH faster than PID controller, attenuating the acidity change which occurs at 45 hours after the start. Moreover with type-1 FLC the system keeps stable at the set-point value without oscillations, that are present with the PID control.
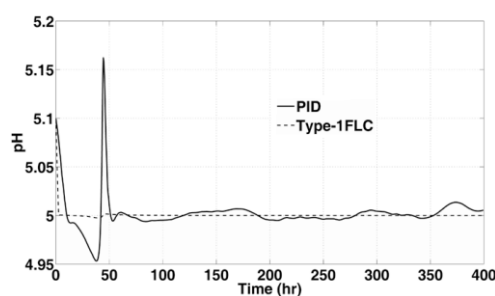
Fig. 2. Control of pH by PID control and type-1 starting from a disturbed point with pH set point equal to5 - System without uncertainties.

Also for the temperature control the performance of PID controller is inferior to that of type-1 FLC.
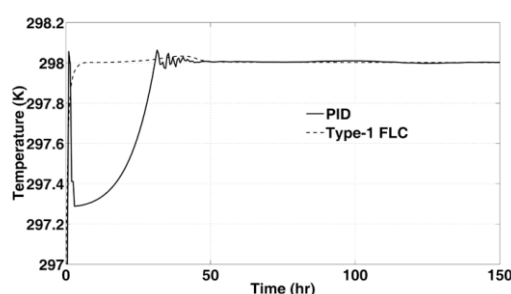
Fig. 3. Temperature control by a traditional PID controller and a type-1 FLC, starting from a disturbed with temperature set-point equal to298 K.

The PID controller leads the system to set-point slowly and with an oscillatory behavior at the begin of the simulation. The fuzzy controller dynamics is much faster than and also in this case the set-point value is reached without abrupt changes.

Type-2 fuzzy control of a fedbatch ractor

4.2 Type-1 FLC vs Type-2 FLC

In a real context control must take into account the inaccuracies and the dynamics of measurement instruments. Dead times and uncertainties may destabilize a system making it difficult to be controlled. The measurement of pH is more difficult and more subjected to the risks of uncertainties and in these simulations only in the feedback control ring of the pH were introduced the uncertainties. represented by measurement noise.The dynamics of the pH measurement instrument was considered by adding a dead time (18 seconds) and a noise. Type-1 and type-2 FLCs performances were compared in this situation. In the last figures (Fig.4-6) only the performances of type-1 and type-2 FLCs for the pH control are shown. For the temperature control where uncertainties are not introduced into the control system, the performances of type-1 and type-2 FLCs are very similar.

From Fig. 4 it is easy to note that without the presence of the noise, the behaviour of pH in the systems controlled by type-1 and type-2 FLCs is very similar. Each fuzzy controller leads the system to the set-point value highly reducing the oscillation amplitude that occurs around 40 hr after the start.
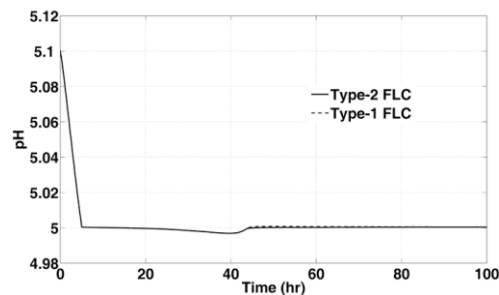
Fig. 4. pH control by type-1 FLC and type-2 FLC, with set-point = 5, starting from an initial offset. - System without uncertainties.
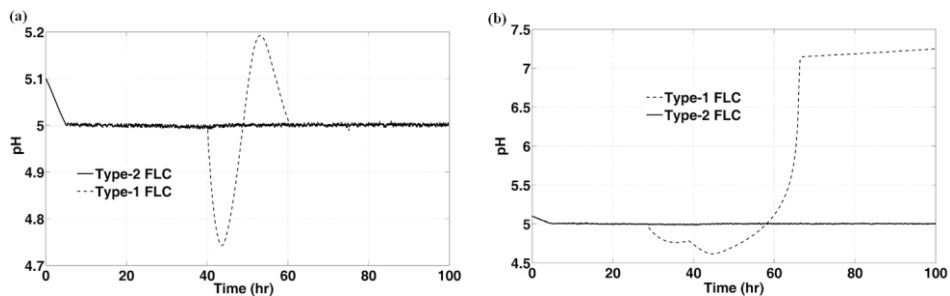
Fig. 5. pH control by type-1 FLC and type-2 FLC, with set-point = 5, starting from an initial offset. – (a) Noise in the control variable measurement with low amplitude value. (b) Noise with high amplitude in the control variable measurement.

Fig. 5a shows the performance of type-1 and type-2 FLCs with the introduction of noise in the measurement of the controlled variable. The system controlled by type-1 FLCs shows a large oscillation amplitude around 40 hours, at this time in fact there is a passage to the fed-batch operation. Large variations of pH from the desired value (= 5) erase the favorable environment for penicillin production. The maximum and minimum peaks of pH value reached by the system controlled by type-1 FLCs are around 5.19

and 4.75 respectively. After 60 hours the behavior of the system becomes regular and similar to that controlled by type-2 FLC. If the amplitude noise is further increased (Fig. 5b), the type-1 FLC cannot control the system and the value of pH increases enormously at t = 60 h reaching the value of 7.25 at t=100 h. The increase of the noise amplitude does not affect instead the performance of the Type-2 FLC that is able to assure the desired constant value of pH, and to reduce the oscillations.

5. Conclusions

In this paper a feedback control configuration with traditional PID controllers, type-1 FLCs and type-2FLCs were considered for the pH and temperature control of a fed-batch reactor for the production of penicillin and their performances were compared.

The results of simulations show that PID controllers cannot compare with type-1 FLCs. The non-linearities present in the system can be better handled using fuzzy control.

By use of type-2 fuzzy sets is possible to increase the performance of fuzzy controllers, in fact the simulation results for the pH control show that type-1 FLCs performance is outclassed by its type-2 counterpart when the system is characterized by uncertainties. Only in this environment characterized by uncertainty the type-2 FLC can show all its potential. Increasing the uncertainty degree of the control system, the difference between the performances of type-1 and type-2 FLC becomes even more evident. In an environment characterized by noise, as in the real world it occurs, the type-2 fuzzy logic control shows all its robustness and effectiveness.

References

G. Birol, C. Undey, A. Cinar, 2002, A Modular Simulation Package for Fed-batch Fermentation: Penicillin Production, Computers and Chemical Engineering, 26, 11, 1553-1565.

O. Castillo, G. Huesca, F. Valdez, 2005, Evolutionary Computing for Optimizing type-2 Fuzzy Systems in Intelligent Control of Non-Linear Dynamic Plants. Proc. North American Fuzzy Info. Processing Society (NAFIPS), Ann Arbor, MI, 247-251.

M. Galluzzo and B. Cosenza, 2009, Control of the biodegradation of mixed wastes in a continuous bioreactor by a type-2 fuzzy logic controller, Computers & Chemical Engineering, 33 (9), 1475-1483.

http://www.chee.iit.edu/_/control/software.html

N.N. Karnik and J. M. Mendel, 1998, Introduction to type-2 fuzzy logic systems, in Proc. 1998 IEEE FUZZ Conf., pp 915-920, Anchorage, AK (USA).

N.N. Karnik and J.M. Mendel, 2002, Operations on Type-2 Fuzzy Sets, Fuzzy Sets and Systems, 122 327-348.

G.J. Klir and M.J. Wierman, 1998, Uncertainty-based information, Physica-Verlag, Heidelberg, Germany.

R. Martinez, O. Castillo, L.T. Aguilar, 2009, Optimization of interval type-2 fuzzy logic controllers for a perturbed autonomous wheeled mobile robot using genetic algorithms, Journal of Information Sciences, 179, 13, 2158-2174.

J.M. Mendel, 2001, Uncertain Rule-Based Fuzzy Logic Systems. Prentice Hall PTR, Upper Saddle River, NJ, USA.

D. Wu and W.W. Tan, 2006, Genetic learning and performance evaluation of interval type-2 fuzzy logic controllers, Engineering Applications of Artificial Intelligence 19 829-841.

L. A. Zadeh, 1975, The concept of a linguistic variable and its applications to approximate reasoning-1. Information Sciences, 8, 199–249.

Dynamic Process Optimisation in Free-Radical Multicomponent Polymerisation: butyl methacrylate and butyl acrylate case study

Bruno Amaro,[a] Charles D. Immanuel,[a] Efstratios N. Pistikopoulos,[a] Andreas Daiß,[b] Klaus Hungenberg,[b] Pedro A. Saraiva[c]

[a]*Centre for Process Systems Engineering, Department of Chemical Engineering and Chemical Technology, Imperial College London, Exhibition Road, London SW7 2AZ, United Kingdom, bruno.amaro05@imperial.ac.uk*
[b]*BASF, Polymer Technology - Process Development and Modelling, BASF Aktiengesellschaft, GKE/M - B1, D-67056 Ludwigshafen, Germany, andreas.daiss@basf.com*
[c]*Departamento de Engenharia Química, Faculdade de Ciências e Tecnologia da Universidade de Coimbra, Rua Sílvio Lima, Pólo II, 3030-790 Coimbra, Portugal, pas@eq.uc.pt*

Abstract

In the present work, an effective mathematical model for the calculation of entire molecular weight distribution (MWD) in free-radical copolymerisation is presented. The distribution is obtained by discretising the infinite chain length domain into a finite number of intervals. The model makes use of a pseudo-homopolymerisation approximation, thus allowing for a significant model simplification. An optimisation formulation is developed and tested for the butyl methacrylate-butyl acrylate (BMA-BA) copolymerisation system with the purpose of identifying optimal reactor conditions using the above model, revealing the ability to attain target MWD and copolymer composition, by manipulating process variables such as monomer and initiator feed rate profiles.

Keywords: MWD, modelling, optimisation, BMA-BA, copolymerisation

1. Introduction

Polymerisation reactions involving two, or more, types of monomer units open a wide range of possibilities, regarding product end-use properties. This provides an extremely powerful synthetic route to a diverse variety of materials since copolymers exhibit properties combining those from the parent homopolymers.

Apart from some particular cases of initial monomer compositions (azeotropic mixtures), the copolymer composition of a typical free-radical copolymerisation often shows fluctuations along the polymerisation process. This phenomenon results both from the different reactivities of the monomers and the growing polymer radicals. Together with copolymer composition the molecular weight distribution (MWD) can also show drifts that can become critical when reactions are carried to high conversion (Sharma and Soane, 1988). There is, therefore, a need for modelling such kind of polymerisation reactions providing a means for building a description of how certain polymer molecular properties (such as the MWD) can control the end-use characteristics of the desired products.

In the present work, a general mathematical model is presented for the calculation of MWD in multicomponent polymerisation systems in a simple, yet effective, fashion. The model accounts for diffusion limitations and makes use of a pseudo-homopolymerisation approximation in order to lower its mathematical complexity. Supported by the developed model, an optimisation approach is formulated comprising an objective function that involves both reaction time and total initiator added to the process.

2. Model development

For a linear addition polymerisation, the equations must consider initiation or catalytic site activation, chain propagation, chain termination and chain transfer reactions. The kinetic scheme proposed is a typical one, including chain transfer to monomer, solvent, modifier and polymer molecules, as well as termination both by disproportionation and combination.

In order to obtain complete representations of molecular weight distributions, a methodology based in finite weight fractions is applied (Crowley and Choi, 1997). It consists of dividing the entire polymer population into discrete intervals and calculating the weight fraction of polymer in each of the discrete intervals. For every interval the polymer weight fraction is calculated with Eq. (1) and Eq. (2).

$$
\begin{aligned}
\frac{df(m,n)}{dt} = \frac{1}{\mu_1} & \left\{ VP\left(\frac{k_p M}{\alpha} - \frac{k_{tc} P}{(1-\alpha)}\right) \Big\{ [m(1-\alpha)+\alpha]\alpha^{m-1} \right. \\
& - [(n+1)(1-\alpha)+\alpha]\alpha^n \Big\} \\
& + \frac{VP^2 k_{tc}}{2\alpha} \left\{ \alpha^{m-1}\left[\frac{\alpha^2(m^2-3m+2)-2m\alpha(m-2)+m(m-1)}{(1-\alpha)}\right] \right. \\
& \left. \left. -\alpha^n\left[\frac{\alpha^2(n^2-n)-2\alpha(n-1)+n(n+1)}{(1-\alpha)}\right]\right\}\right\} - \frac{f(m,n)}{\mu_1}\frac{d\mu_1}{dt}
\end{aligned}
\tag{1}
$$

$$
\begin{aligned}
\frac{1}{V}\frac{d\mu_1}{dt} = & -k_{fp}\frac{\mu_1\lambda_0}{V} + k_{fp}\frac{\lambda_0\mu_0}{V(1-\alpha)} \\
& + \left[\frac{\alpha(2-\alpha)}{(1-\alpha)}\right]\left(k_{fm}M + k_{fs}S + k_{fx}X\right)P \\
& + \left[\frac{\alpha(2-\alpha)}{(1-\alpha)}\right]k_{td}P^2 + \frac{1}{(1-\alpha)}k_{tc}P^2
\end{aligned}
\tag{2}
$$

MWD calculations performed in this work were supported by a kinetic model, whose mathematical complexity was reduced by using a pseudo-homopolymerisation approximation. This enabled us to utilise models already developed for free-radical homopolymerisation, just by modifying the rate constants to account for the particular kinetics of copolymerisation reactions. Storti et al. (1989) proposed this method for reducing a set of population balances, in emulsion polymerisation, to a smaller one, where each population is characterised only by the overall number of radicals of any type contained inside each particle. After lumping the two radical types of a copolymerisation into a single one the model then assigns the polymer chains to one type or the other based on a pseudo-homopolymer probability (Eq. (3) and Eq. (4)). This

Dynamic Process Optimisation in Free-Radical Multicomponent Polymerisation: butyl methacrylate and butyl acrylate case study

methodology is similar to the pseudo-kinetic rate constant method, first used by Hamielec and MacGregor (1983) for the molecular weight calculation in linear chains and copolymer chains with long branches.

$$p_1 = \frac{(k_{p21} + k_{fm21})M_1}{(k_{p21} + k_{fm21})M_1 + (k_{p12} + k_{fm12})M_2} \tag{3}$$

$$p_2 = 1 - p_1 \tag{4}$$

Regarding diffusion limitations, Buback and Kowollik (1999); Buback and Feldermann (2002) demonstrated that plotting the experimental $log(k_{t,copol})$ against monomer mole fraction shows an almost linear correlation for many (meth)acrylate copolymerisation systems, described by Eq. (5), used in the derived model.

$$log(k_{t,copol}) = f_{m1}log(k_{t11}) + f_{m2}log(k_{t22}) \tag{5}$$

For the calculation of polymer macromolecules composition along reaction time we make use of Eq. (6), proposed by Mayo and Lewis (1944). This equation was later confirmed by Hocking and Klimchuk (1996), after deriving a refinement for the established Alfrey-Goldfinger terpolymer equation (Alfrey and Goldfinger, 1944).

$$F_{p1} = \frac{r_1 f_{m1}^2 + f_{m1} f_{m2}}{r_1 f_{m1}^2 + 2 f_{m1} f_{m2} + r_2 f_{m2}^2} \tag{6}$$

3. Dynamic process optimisation

An optimisation approach is formulated for multicomponent polymerisation systems, as depicted in Fig. 1, and implemented in the gPROMS platform. A target molecular weight distribution is given to the gPROMS optimiser, together with a target polymer composition that is required to be kept close to constant throughout the polymerisation reaction.
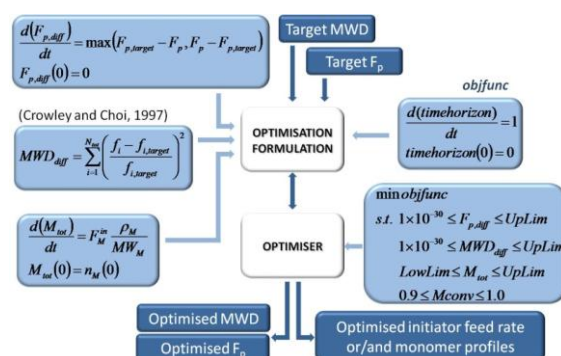
Figure 1 Representation of optimisation procedure for multicomponent polymerisation.

For that effect, an objective function is embedded in the model and set to minimise total reaction time. The total monomer conversion, is constrained in order to get close to completion and not to fall below 90%, while desired MWD (at final reaction time) is

achieved by using Eq. (7) that accounts for the differences between target weight fraction and the one being calculated by the optimiser for each of the chosen chain length intervals. In an industrial process, the total final amount of polymer is usually pre-determined and consequently there is a limit for the total monomer to be added in a semi-batch fashion. For these cases, when monomer feed rate is chosen as a control variable, a constraint is set in order to guarantee that the total quantity of monomer added to the process (M_{tot}) lies within a narrow interval.

$$MWD_{diff} = \sum_{i=1}^{N_{int}} \left(\frac{f_i - f_{i,target}}{f_{i,target}} \right)^2 \tag{7}$$

Adding the most reactive monomer along the reaction time is an effective strategy to minimise polymer composition fluctuations that are often observed in copolymerisation processes.

The mathematical model presented before was tested for the butyl methacrylate-butyl acrylate (BMA/BA) solution polymerisation, in a system also formed by initiator(*tert*-butyl peroxyacetate) and solvent (xylene). For the simulations leading to the results described next the initiator was fed to the process along the reaction time, at the constant flow rate of 5.737×10^{-7} *L s^{-1}* and BMA was partly added in a semi-batch fashion, following the initial profile from Fig. 2 a).
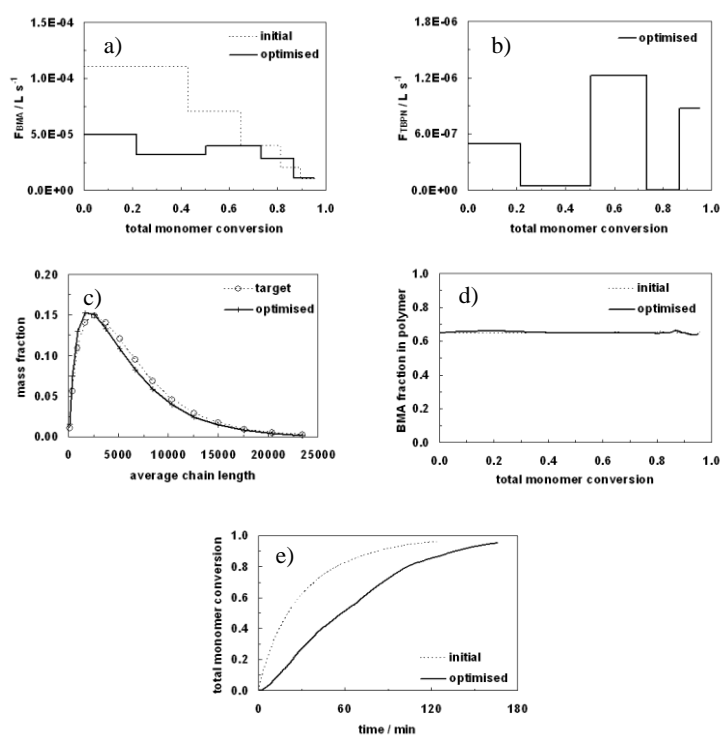
Figure 2 a) Initial and optimised monomer feed rate profiles; b) Optimised initiator feed rate profile; c) Target and optimised MWD; d) Initial and optimised polymer composition; e) Initial and optimised total monomer conversion

Dynamic Process Optimisation in Free-Radical Multicomponent Polymerisation: butyl methacrylate and butyl acrylate case study

In view of such conditions, the optimiser was expected to find the ideal combination of profiles (both initiator and monomer) leading to same predefined MWD and F_p. Although only one result is presented here, different combinations of boundary conditions for the constraints were tested, all leading to different final outcomes. Therefore, there is scope for the user to define how rigid the constraints should be and, to assign different levels of importance to both MWD and F_p, according to product specifications or other process requirements.

The monomer feed rate profile selected by the optimiser is shown in Fig. 2 a), where it can be compared to the initial one. Fig. 2 b) holds only the optimised initiator feed rate profile since in the initial situation all of it was added to the reactor right from the start. These profiles generate a molecular weight distribution quite similar to the target one (Fig. 2 c)), whereas polymer composition shows more mild fluctuations along reaction time than in the initial simulation but still at a reasonable level (Fig. 2 d)). An identical total amount of initiator is employed in the optimised solution, although it takes more time to reach similar monomer conversions (Fig. 2 e)). Reaction time could be reduced by relaxing the constraints but MWD would start to significantly deviate from the target and polymer composition would exhibit more pronounced oscillations.

4. Concluding remarks

A comprehensive mathematical model was developed for bulk and solution copolymerisation systems, using a pseudo-homopolymerisation approximation that allowed for a significant model simplification. Calculation of MWD was achieved by using finite weight fractions, as originally developed for homopolymerisation systems, an advantageous consequence of using the above mentioned simplification.

An optimisation formulation was presented comprising an objective function, to be minimised, accounting for both reaction time and total initiator added to the process. The results obtained for BMA-BA showed that it was possible to obtain a MWD identical to the one predefined, at constant polymer composition, with a slightly different monomer feed rate profile, identified by the optimiser. Although there were no improvements in process time and the same total amount of initiator was added to the process in this particular configuration, the potential of such optimisation formulation was here clearly demonstrated.

5. Nomenclature

BA	Butyl acrylate
BMA	Butyl methacrylate
F	feed volume flow rate, $L\ s^{-1}$
F_p	number-average copolymer composition
f(m,n)	polymer weight fraction in chain length interval *[m,n]*
f_m	monomer fraction
I	initiator concentration, $mol\ L^{-1}$
k_d	initiator decomposition rate constant, s^{-1}
k_{fm}	chain transfer to monomer rate constant, $L\ mol^{-1}\ s^{-1}$
k_{fp}	chain transfer to polymer rate constant, $L\ mol^{-1}\ s^{-1}$
k_{fs}	chain transfer to solvent rate constant $L\ mol^{-1}\ s^{-1}$
k_{fx}	chain transfer to modi_er rate constant, $L\ mol^{-1}\ s^{-1}$
k_I	initiation rate constant, $L\ mol^{-1}\ s^{-1}$
k_p	propagation rate constant, $L\ mol^{-1}\ s^{-1}$
k_{tc}	combination termination rate constant, $L\ mol^{-1}\ s^{-1}$
k_{td}	disproportionation termination rate constant, $L\ mol^{-1}\ s^{-1}$
M	monomer concentration, $mol\ L^{-1}$

MWD	Polymer molecular weight distribution
P	total live polymer chains concentration, *mol L^{-1}*
p$_i$	pseudo-homopolymer probability
r	reactivity ratio
t	time, *s*
V	system volume, *L*

Greek letters

α	probability of propagation
λ$_0$	zeroth moment of the "live" polymer distribution
μ	moments of the "dead" polymer distribution
ρ	density, *kg dm^{-3}*

Subscripts

0	initial conditions
1	monomer unit of type 1 or polymer chain ending in type 1 monomer
2	monomer unit of type 2 or polymer chain ending in type 2 monomer
I	initiator

6. Acknowledgements

References

Alfrey, T., Goldfinger, G., 1944. Copolymerization of systems of three and more components. Journal of Chemical Physics 12 (7), 322–322.

Buback, M., Feldermann, A., 2002. Termination kinetics of free-radical methyl methacrylate-dodecyl acrylate and dodecyl methacrylate-methyl acrylate copolymerizations. Australian Journal of Chemistry 55 (6-7), 475–481.

Buback, M., Kowollik, C., 1999. Termination kinetics in free-radical bulk copolymerization: The systems dodecyl acrylate dodecyl methacrylate and dodecyl acrylate methyl acrylate. Macromolecules 32 (5), 1445–1452.

Crowley, T. J., Choi, K. Y., 1997. Calculation of molecular weight distribution from molecular weight moments in free radical polymerization. Industrial & Engineering Chemistry Research 36 (5), 1419–1423.

Free-Radical Polymerization: Kinetics and Mechanism II Ciocco. Wiley-V C H Verlag Gmbh, Lucca, Italy, pp. 149–168.

Hamielec, A. E., MacGregor, J., 1983. Hanser Publishers, New York, p. 21.

Hocking, M. B., Klimchuk, K. A., 1996. Refinement of the terpolymer equation and its simple extension to two- and four-component systems. Journal of Polymer Science Part a-Polymer Chemistry 34 (12), 2481–2497.

Mayo, F. R., Lewis, F. M., 1944. Copolymerization i a basis for comparing the behavior of monomers in copolymerization, the copolymerization of styrene and methyl methacrylate. Journal of the American Chemical Society 66, 1594–1601.

Sharma, D. K., Soane, D. S., 1988. High-conversion diffusion-controlled copolymerization kinetics. Macromolecules 21 (3), 700–710.

Storti, G., Carra, S., Morbidelli, M., Vita, G., 1989. Kinetics of multimonomer emulsion polymerization - the pseudo-homopolymerization approach. Journal of Applied Polymer Science 37 (9), 2443–2467.

Dynamics of Knowledge Flow in Research on Distillation[*]

Sitarz R.[a,b,*], Kraslawski A.[b], Jezowski J.[a]

[a]*Department of Chemical and Process Engineering,Rzeszów University of Technology
Al. Powstańców Warszawy 6, 35-959 Rzeszów, Poland,*
[b]*Department of Chemical Technology, Lappeenranta University of Technology, POBox
20, 53851 Lappeenranta, Finland*

Abstract

One of the crucial factors influencing the decision on funding R&D are the development trends in the given field of research. The dynamics of the research as well as patenting activity in a given period of time are one of the basic factors in building the portfolio of the research projects when profit and risk have to be balanced. The existing bibliographic approaches for the structuring of the given field of research exclusively on the basis of keywords, are far from being acceptable. The paper presents an original method for the identification of the thematic clusters and dynamics of their change for a given research field as well as the application of financial analysis techniques for prediction of the development trends. . The proposed approach is applied to research in old and well-developed field of chemical engineering – distillation research conducted between 1990-2008. The identified 43 thematic clusters of research have shown the diversified patterns of development like stagnation, revival, slow development or intensive growth.

Keywords: distillation**,** technology forecasting**,** knowledge management

1. Introduction

The university and industrial R&D activities are characterized by growing specialization in research, exponential growth of the number of publications as well as increasing costs of research and shortening life time of the products and processes. Such a situation requires a very careful and broad analysis of the development of the given research field and operational practices before any decision on the allocation of the R&D funds would be made [1, 2].

One of the crucial factors influencing the decision on funding R&D are the development trends in the given field of research. It is extremely important for the decision makers to analyze the dynamics of the development of the research in a well-defined area. The number of the scientific publications and patents is an indirect hint illustrating the importance and the potential of the specific area of expertise [3]. Another important factor is a change in time with the respect of the research intensity¸ measured by the number of publications per year. The dynamics of the research as well as patenting activity in a given period of time are one of the basic factors in building the portfolio of the research projects when profit and risk have to be balanced.

The existing bibliographic approaches for the structuring of the given field of research exclusively on the basis of keywords, are far from being acceptable due to the lack of the required granularity of information [4].The needs of the R&D decision makers and the lack of the effective methods and tools for the determination of the existing and

---

* E-mail The Corresponding Author: robs@prz.edu.pl

emerging trends in a given research field is a major motivation for the analysis of the dynamics of knowledge flow presented in this paper.

The paper presents an original method for the identification of the thematic clusters and dynamics of their change for a given research field as well as the application of financial analysis techniques for prediction of the development trends. The presented material is limited exclusively to the research papers and covers the publications from ISI Web of Science for the years 1990-2008. The proposed approach is applied to old and well-developed field of chemical engineering – distillation research

2. Method description

The method for the identification of the development trends in the given research field is composed of three main blocks: Articles Identification, Words Processing and Trend Identification. The detailed structure of the method is given in Fig 1.

I Articles Identification

- Retrieving articles from ISI Web of Science Database

- Creation of Article_Data database

- Selection of the articles

II Words Processing

- Creation of fingerprints table

- Creation of words co-occurrence table

- Generation of four and three word's sets

- Identification of thematic clusters

- Determination of changes in the number of articles in every cluster

III Trend Identification

- Base line trend analysis

- Oscillators method

- K%D stochastic oscillator

Figure 1. The structure of the method for identification of the development trends in a given field of research

In the first block, Articles Identification, the papers are downloaded from ISI Web of Science database using, as a topic of query, the specific name of the research field. The information about the papers contain: authors, title, source name, publication year, volume, page key- words, and abstract if available. Next, the downloaded data are introduced into "Article_Data", Microsoft Access database, in order to facilitate their use in the consequent steps of the analysis. In the last step of the first block, the selection of the articles is continued and only the papers with abstracts published after 1990 are contained in the database.

The first step of the next block, Words Processing, is aimed at the identification of the fingerprint of each article contained in the database. The fingerprint is a set of characteristic words which constitutes the meaning of a given article. It is created by the statistical analysis of the article text and removal of the non-specific words like "the", "be", "can" etc. The creation of the fingerprints table is realized by a dedicated tool Rapidminer [5] . The input data to Rapidminer are taken from "Article_Data" database. This data consisted of title, keywords, and abstract. The columns of the generated table are composed of the fingerprint's words and the row are the numbers corresponding to the identification numbers of the particular articles. The values in each cells show how many times a given word was used in a given article. Next, the binary table is received

in such a way that in every cell where the number of the words is greater than 0, a number 1 is introduced. The creation of the binary table is motivated by the fact that the subject of the analysis is not the frequency of the use of the words but only a fact that a given word exist in an article.

In the second step, "WordSet " program is used to build a words co-occurrence table. The columns and rows of the words co-occurrence table are composed of the fingerprint's words. The values in each cells show how many times a given words pair co-occur in considered articles.

In the following step, "WordSet " program is used to generate four and three word's sets, which determine a thematic field of research. Every possible combination of four or in the special cases three words' sets are checked, and only these one were used further, which satisfy two conditions. The first condition is that the number of the articles, where all of word in the four word's set occur, is greater than The A threshold value. The second one is that ratio, number of articles, where all of word in the four word's set occur, to the arithmetical average number of articles where each of this words occur, is greater then the B threshold value. The meaning of the A threshold is, that if the considered thematic field will be taken into account than there should be a minimal number of the articles belonging to this cluster-thematic field. In order to identify the essential research fields this threshold value should be kept at the level ensuring the recognition of the existing fields as well as the emerging ones. The B threshold takes into account the fact, that only the significant words should be considered. It means that the high co-occurrence words constituting the same thematic field will be grouped together and simultaneously they should have very low co-occurrence with the others one.

In the fourth step of the second block, the identified sets were grouped into thematic groups by analyzing them case by case.

In the last step, the Makro programme was used, to determine the number of the papers belonging to a given thematic field cluster as well as the yearly distribution of those papers in the period 1990-2008.

The last block, Trend Identification, is composed of three methods for technical analysis of financial markets [6]. It is used for the determination of the prospects of the given field of research basing on the analysis of the dynamics of publishing. In the first step a trend line approach is applied. The trend line is a straight line which pass through the specific points described in more details below. It shows trend's rate and direction of change. If the trend is growing (falling), then the trend line pass through points, which are minimums (maximums) of the relative number of publications. When this number of publications in the next year, crosses the trend's line, it could be signal, that trend is changing the rate, or direction.

In the next step, the oscillators method is used. The momentum is defined as a difference of the relative number of publications in given year and x years back. The x parameter is a measure of the oscillator changeability. The value of momentum is calculated for every year. If the momentum value is positive, than the trend is increasing and if negative then the trend is decreasing. If the value of momentum is very high, or very low, relatively to the other points, then it could be a signal that trend probably will change its direction.

The last method is stochastic oscillator (K%D). The oscillator shows the relation of the actual relative number of publications to the difference between the max and min values of the relative number of publication in the period of y years. The oscillator consists of two lines(%K and %D).

The K line is calculated as :

$$K = 100[(C-L)/(H-L)] \qquad (1)$$

where:

C   actual relative number of publications

L   the lowest C value from y previous years,

H   the highest C value from y previous years.

The second line (%D) is calculated as the average of 3 predestining points situated in K line. If the lines intersect, than it could be a signal, that trend is increasing (if the lines go up), or decreasing (if the lines go down).

The presented three methods are used simultaneously to achieve more credible results.

3. Results

The presented method was used for the determination of the dynamics of the development of major fields of research related to distillation.

In the first step there were identified 13450 articles, in ISI Web of Science, where the word "distillation" was mentioned in the full text. These data were saved to .txt files (500 articles in each file). Next the files were loaded to "Article_Data" database, and using limitation, relating to abstract accessibility, 11402 articles data was used for the further analyze, which were saved as files to further analyzing by Rapidminer program. In this program, the words which were used by less then five articles were deleted from the fingerprints table. Finally, there were saved 7389 words for 11402 articles. In "WordSet" program, after fingerprints table from CSV file was loaded, the following non-informative words, or their fragments, were removed: distil, elsevi, scienc, alpha, beta, right, reserve, us, ga, mass, channel, quantum, entangle, wiley, copyright. After this operation, the articles for which the fingerprints consisted of less then four words were removed from the fingerprints table. The number of the articles and the words which were 11263 and 7374 respectively. In this research, "WordSet" program was used as follows. In the first step 2156 four word's sets were generated for the A threshold equal 15, and the B threshold equal 0,1 (10%). In the second step, the additional 993 three word's sets were generated which were not contained in previous four word's sets. The A and B thresholds in this step were accordingly 15 and 0,15. The proposed number of the words was determined as a more detailed analysis has shown that four and three word's sets are the best descriptors of the thematic field. In the course of the analysis of the word's sets, some of them were removed as their relations to the research on distillation seemed to be rather distant (e.g. chromatography, spectroscopy or sedimentation).

The obtained sets were grouped into 43 thematic clusters. The examples of the clusters are: Reactive distillation, Membrane distillation, Equilibrium aspect in distillation, Modeling the system of distillation column, Optimization in distillation, etc. Next, the number of the articles existing in every cluster in years 1990-2008 was identified. The number of the papers in a given cluster was divided by a total number of publications in a given year. The relative number of the papers belonging to a the given cluster was presented graphically. In Fig 2 a graph is given representing a change of the relative number of papers, belonging to the cluster "reactive distillation", in 1990-2008, and the first two analyzing methods described in the previous section (the base line and oscillator methods). The solid straight lines (from I to IV) are the application of trend line method, which show trend directions in a given periods of time. The dashed lines show oscillator method for 5 (oscillator 5) and 3 (oscillator 3) years parameters. Two kind of oscillators were used in order to get more credible results. The trustworthy results are obtained when both oscillators show the same kind of behavior.

Dynamics of Knowledge Flow in Research on Distillation
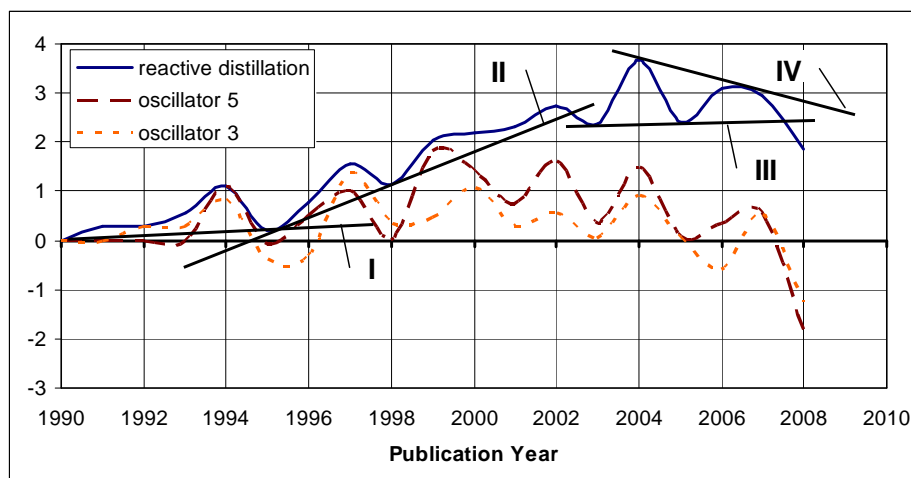
Figure 2. The cluster "reactive distillation" – relative number of publication, oscillator method, and base line methods

It means both are positive or negative, otherwise the obtained information could be misleading. The use of K%D oscillator is presented in Fig. 3. The solid curve show %K line calculated for y parameter equal 5, the dashed one %D line.
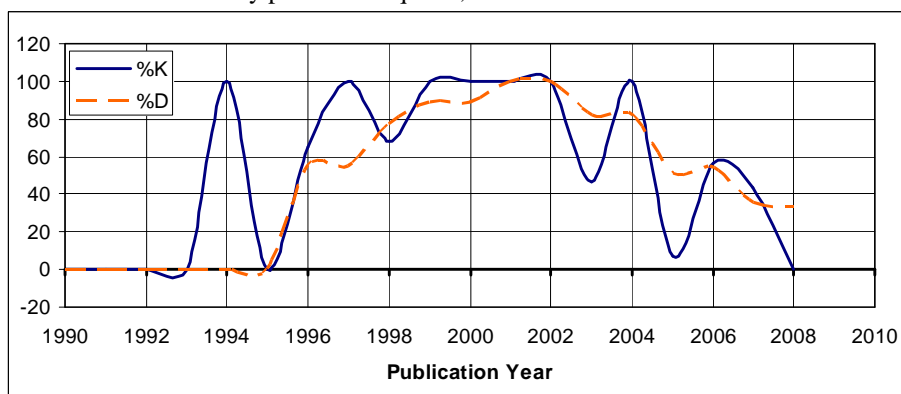
Figure 3. The cluster "reactive distillation" - K%D oscillator method

The following trends can be identified when analyzing the results given in Fig 2 and 3,. The analyzed period of time (1990-2008), was divided into three periods of characteristic trend direction for the research on reactive distillation. In the period of time from 1990 to 1995 the publishing activities were at the top in 1994, and increasing trend was observed, when applying the trend line I. The both oscillator methods (oscillators and K%D) could not be applied for the given period of time as 1994 is the first year when the oscillator calculations are feasible. For the period 1995-2002 the trend line II shows an increasing trend (stronger then in the previous period). The values of oscillators 3 and 5 for this period are positive and increasing. It means that there is an increasing trend. However, both oscillators have the diminishing values at the end of the analyzed period what could be a visible sign of the future trend change. K%D method for this period shows that trend is increasing, however in the 2002 there is the signal of trend's change manifested by the decreased values of both lines and their intersection.

For the period 2002-2008, the trend's lines III and IV built a descending triangle pattern. At the end of this period, there is breaking down of the horizontal trend line (III)**.** It means that the decreasing trend in publishing activities is expected to continue after 2007. The oscillators 3 and 5 for this period show decreasing trend starting from 2007 when both of them have the negative values.

It can be observed that there are a few intersections of %K and %D lines, caused by the strong oscillations of the publishing activities. In consequence the K%D method shows in this period a strongly unstable trend with a tendency to decrease.

Based on the presented results it could be predicted that in the next 2-3 years, there will be a decreasing trend in the number of publications dealing with "reactive distillation".

4. Summary

The major achievements of the presented research are:
- Identification of the thematic clusters of the research in the field of distillation
- Determination of the dynamics of research in the field of distillation,
- Prediction of the short-term trends in the development of the specific thematic fields

The identified 43 clusters have shown the diversified patterns of development. Some clusters, e.g. (Modeling of system of distillation columns, or Optimization in distillation) show clear signs of stagnation and "dying out" . The others (like Membrane distillation, or Extractive distillation) present the visible signs of the revival after a several years of "silence" or slow development. There are also the examples of the emerging, new fields of research, e.g. Artificial neural networks application from year 2000, or Direct contact membrane distillation also starting around 2000.

The presented results have been obtained using the own tools for the words co-occurrence and the clusters identification as well as the methods of trends analysis applied in the financial applications. The trend analysis method is based on the studies of the past publishing activities. The used data are limited to nineteen data points (1990-2008). The limited number of the data and the restricted character of the database ISI Web of Science clearly shows a need for the use of other powerful databases as well as the extension of the analyzed time interval.

References

[1] Juan Camilo Zapata, Vishal A. Varma, G.V. Reklaitis
Impact of tactical and operational policies in the selection of a new product portfolio
Computers & Chemical Engineering, Volume 32, Issues 1-2, January 2008, Pages 307-319

[2] Anuradha Rajapakse, Nigel John Titchener-Hooker, Suzanne S. Farid
Modelling of the biopharmaceutical drug development pathway and portfolio management
Computers & Chemical Engineering, Volume 29, Issue 6, 15 May 2005, Pages 1357-1368

[3] Bernd Fabry, Holger Ernst, Jens Langholz, Martin Köster
Patent portfolio analysis as a useful tool for identifying R&D and business opportunities—an empirical application in the nutrition and health industry
World Patent Information, Volume 28, Issue 3, September 2006, Pages 215-225

[4] David Hull, Salah Aït-Mokhtar, Mathieu Chuat, Andreas Eisele, Éric Gaussier, Gregory Grefenstette, Pierre Isabelle, Christer Samuelsson, Frédérique Segond
Language technologies and patent search and classification
World Patent Information, Volume 23, Issue 3, September 2001, Pages 265-268

[5] Rapidminer Internet link : http://rapid-i.com/

[6] John J. Murphy : Technical Analysis of the Financial Markets, 1999, New York Institute of Finance

Towards energy-based dynamic optimization of monoclonal antibody producing GS-NS0 Cultures

Alexandros Kiparissides, Efstratios N. Pistikopoulos and Athansios Mantalaris

Biological Systems Engineering Laboratory, Centre for Process Systems Engineering, Imperial College London

Abstract

Mammalian cell culture systems produce clinically important high-value biologics, such as monoclonal antibodies (mAb). Cell lines transfected with the *Glutamine Synthetase* (GS) gene are amongst the most industrially significant mAb production systems due to the high yields they achieve. Metabolic models of GS culture systems presented thus far take into account only glucose as a growth limiting nutrient, neglecting the fact that in the absence of glutamine in the media, glutamate becomes a necessary dietary component in GS systems. Previously, we have presented the development of a systematic framework for modelling of mammalian cell bioprocesses. Herein, we present, for the first time, the development of a dynamic model describing growth and monoclonal antibody formation in GS-NS0 cell cultures that interlinks cellular growth rate with the availability of both glucose and glutamate. This is the first step, of many, towards the derivation of a dynamic model that interlinks the availability of ATP, through the dietary intake of the cell, to its growth and productivity characteristics. Such a model would facilitate the derivation of an optimal feeding profile, constraining the amount of provided energy through the feed to the required minimal, hence avoiding the excessive feeding of glucose which in turn shifts metabolism towards energy inefficient pathways.

Keywords: Mammalian Cell Modeling, ATP/ADP balance, energy based optimization, monoclonal antibodies, GS-NS0 cell line

1. Introduction

The advancements in molecular biology and analytical techniques over the last 20 years have significantly elevated the biological industry in the economical scale. mAbs alone have a projected market of \$49bn by 2013 (Monoclonal antibodies Report, 2007). mAbs are primarily produced in batch or fed-batch processes, however the control of such processes in the biotechnological industry still remains fundamentally manual. In previous work (Kiparissides *et al.* 2009) we have shown the advantages of using a systematic model development framework from conception to validation and how such a framework paves the way towards model based optimization and control. Moreover we have showcased the benefits of using a hybrid approach to modeling, by coupling structured models describing the assembly and secretion of mAbs to unstructured growth/metabolic models thus reducing computational and experimental costs.

The common denominator of all metabolic models of GS systems thus far in the literature is that they disregard a number of vital metabolites for the growth of cells in culture. More specifically, to the extent of our knowledge, none of the presented models monitors the concentrations of essential amino-acids in the extracellular environment and how their depletion affects growth and mAb productivity. Furthermore studies for the derivation of optimal feeding profiles for fed-batch cultures presented thus far in the

literature merely take into account the cells' needs on glucose. Even though the results are indeed an improvement to heuristic or empirical feeding strategies they lead to an excessive amount of glucose being fed to the culture. According to the work of Xie and Wang (1994) however, the presence of glucose in abundance in the culture media shifts cell metabolism towards more energy inefficient pathways.

There is an imminent need to update existing metabolic models of GS culture systems so that they account for the effects of essential amino acid concentrations and available ATP levels on growth and productivity. A model able to identify the cells' minimal requirements of ATP for proliferation and mAb production could lead to the derivation of a feeding profile that would maximise final antibody titre, whilst supplying the culture with merely the required amount of nutrients. Such an approach would be beneficial in multiple ways. First and foremost by eradicating the excessive presence of glucose fed to the culture, the amount of accumulating lactate will be significantly reduced, allowing for prolonged culture viability. Moreover, when glucose is not fed in excessive concentrations, cellular metabolism is limited to energy efficient pathways (Xie and Wang, 1994).

Herein we present the first step towards an energy orientated model that would allow the derivation of such an optimal feeding profile. We present for the first time, the coupling of cellular growthrate of GS systems with glutamate concentration. This is of great importance, since for GS systems which grow in glutamine free media, glutamate becomes a necessary dietary component and its depletion would inhibit growth.

2. Mathematical model and experimental setup

2.1 Experimental Setup

GS-NS0 cells (kindly provided from Lonza biologics) where cultured in triplicate 1L Erlenmeyer flasks (Corning) with 200mL working volume. The media contained Advanced-DMEM X1 (Invitrogen Ltd.), MEM-Vitamins (Gibco) X2, GS-Supplement (SAFC) X2, Penicillin/Streptomycin (Gibco) X1, 4.5 g/L MSX (Sigma-Aldrich) and 10% Fetal bovine serum (Gibco). Samples were taken on 24h intervals and stored in -20^0C prior to analysis.

Metabolite Measurments

1.5mL samples were taken from each flask on 24h intervals and centirfuged at 800 RPM for 5minutes. The supernatant was stored in -20^0C prior to analysis. Extracellular Glucose, Glutamate and Lactate concentration were measured using a Nova BioProfile 400 Analyser.

Extracellular Antibody Quantification

The extracellular antibody concentration was determined using a sandwich-based Enzyme- Linked Immunosorbent Assay (ELISA). A 96-well plate was first coated with an anti-human gamma Fc antibody (Jackson immunoresearch, US) in a coating buffer (15 mM Na2CO3, 35 mM NaHCO3, pH 9.6) at a concentration of 2 µg/mL overnight in a 4^0C refrigerator. The coating solution was then removed and the wells blocked with a solution consisting of the coating buffer with 0.5% (w/v) casein hammerstein (VWR) for 1 hour at room temperature. Subsequently, the wells were rinsed 6 times with 300 µL of washing solution (PBS with 0.05% Tween). Known standard concentrations of the cB72.3 IgG antibody (kindly provided by Lonza Biologics, UK) and cell free supernatant samples diluted in sample-conjugate buffer (12.1 g/L Tris, 5.84 g/L NaCl, 2.0 g/L Casein Hammerstein (VWR) and 0.2 mL Tween) were added next to the wells

Towards energy-based dynamic optimisation of monoclonal antibody producing GS-NS0 cultures

(100 µL/well) and incubated for 1.5 hours at room temperature on an orbital shaker. Sample-conjugate buffer was also added to at least two wells to serve as negative (background) controls. The standards, samples and sample-conjugate buffer were discarded and the wells washed with the washing solution described above. An antihuman kappa chain Fab antibody fragment conjugated to horseradish peroxidase (Sigma) was then added at a dilution of 1:8000 (in sample-conjugate buffer) to each well (100 µL/well) and incubated for a further 1 hour with shaking at room temperature. After the incubation period, wells were washed with the washing solution before substrate solution (100 µL/well) was added to the wells. The substrate solution consisted of a TMB tablet (1mg/tablet of 3,3′,5,5′-Tetramethylbenzidine, Sigma), which was dissolved in 10 mL of 50 mM phosphate-citrate buffer (pH 5.0). Immediately prior to use, 2 µL of 30% (w/v) hydrogen peroxide solution was added to the mixture. The reaction was allowed to proceed in the dark at room temperature for 15 to 30 minutes before being stopped by the addition of 50 µL of 2.5 M H2SO4 solution to each well. The OD450 of each well was measured using an ELISA microplate reader (BioTek, US). OD450 values for standards and samples were normalised by subtracting the average OD450 reading of the negative control wells. Each sample was assayed at least in quadruplicates.

2.2 Mathematical model
The proposed model is the first to couple the cellular growth rate of GS systems to glutamate concentration. The model works under the standard operating assumption of perfect mixing within the bioreactor and furthermore assumes the presence of a homogeneous culture of "average" cells. The model will be presented in its batch operation mode in accordance to the experiments described above.

2.2.1 Unstructured metabolic model
The total balances on viable (X_V, cells mL^{-1}) and dead (X_D, cells mL^{-1}) cells is given by:

$$\frac{d(V * X_V)}{dt} = (\mu - k_d) * X_V * V \qquad (2.2.1.1)$$

and

$$\frac{d(V * X_D)}{dt} = k_d * X_V * V - k_{lys} * X_D * V \qquad (2.2.1.2)$$

respectively, where μ denotes the specific growth rate (h^{-1}) and k_d the specific death rate (h^{-1}). k_{lys} (h^{-1}) is the specific cell lysis rate.
The specific growth rate has been modelled using standard Monod kinetic expressions for the growth affecting nutrients, namely glucose and glutamate and is given by:

$$\mu = \mu_{MAX} \frac{[GLC]}{K_{GLC} + [GLC]} \frac{[GLU]^2}{K_{GLU}^2 + [GLC]^2} \qquad (2.2.1.3)$$

Where, [GLC] and [GLU] are the extracellular concentrations of glucose and glutamate respectively, measured in mM and μ_{MAX} denotes the maximum specific growthrate (h^{-1}). The specific death rate has been adapted from the work of deTrembley *et al.* (1992), as presented by Ho (2006) for the same cell line, and is given by:

$$k_d = \frac{K_{d,1}}{\mu_{MAX} - K_{d,T}[LAC]} \qquad (2.2.1.4)$$

Where, $K_{d,1}/\mu_{MAX}$ is the minimal specific death rate in the absence of lactate in the media and $K_{d,T}$ ($h^{-1}mM^{-1}$) is the death rate associated with the toxicity of lactate ([LAC], mM). The specific lysis rate has been adapted from the work of Ho (2006) and is modelled as a linear function of the specific death rate.

$$k_{lys} = k_{l,1} * k_d - k_{l,2} \qquad (2.2.1.5)$$

Where, $k_{l,1}$ (dimensionless) and $k_{l,2}$ (h^{-1}) are the associated constants.
The nutrient uptake rates are given by:

$$\frac{dS_i}{dt} = -\left(\frac{\mu}{Y_{X,i}} + m_i\right) * X_V * V \qquad (2.2.1.6)$$

Where, S_i denotes nutrient (*i*) and is measured in mM. $Y_{X,i}$ denotes the yield on biomass when nutrient (*i*) is consumed by the cells, and m_i (Mm h^{-1}) is the non-growth associated consumption rate of nutrient (*i*) for housekeeping purposes. Similarly the accumulation of the metabolism's by-products is given by:

$$\frac{dP_j}{dt} = \left(\frac{\mu}{Y_{j,i}} + m_j\right) * X_V * V \qquad (2.2.1.7)$$

Where, P_j denotes metabolite (*j*) and is measured in mM. $Y_{j,I}$ denotes the yield on metabolite (*j*) from the consumption of nutrient (*i*), and m_j (mM h^{-1}) is the non-growth associated metabolite accumulation term.

2.2.1 Structured model of covalent mAb assembly
The unstructured metabolic model was coupled to a structured model describing the covalent assembly of the IgG4 antibody that is produced by the GS-NS0 cells studied. According to the work of Percy (1970), IgG4 antibodies are assembled in the endoplasmic reticulum of the cells from heavy and light polypeptide chains following mechanism of equation (2.2.2.1):

$$[H] + [L] \longrightarrow [HL]$$
$$[HL] + [HL] \longrightarrow [H_2L_2] \qquad (2.2.2.1)$$

The model has been successfully adapted and applied to describe the accumulation mAb in GS-NS0 cultures in previous studies from our group (Ho *et al.*, 2006). For a detailed derivation of the structured model and its complete set of equations the reader should refer to the original work of Ho *et al.* (2006).

3. Results and Discussion
Parameter estimation experiments and model simulations where carried out on an Intel® Core™2 Duo (E4600 – 2.4, 2.39) personal computer with 3.24 GB of RAM memory and all model simulations and parameter estimation experiments were implemented in the advanced process modelling environment gPROMS® (Process Systems Enterprise, 2009). gPROMS is an equation-oriented modelling system used for

Towards energy-based dynamic optimisation of monoclonal antibody producing GS-NS0 cultures

building, validating and executing first-principles models within a flow sheeting framework. The experimental results and model simulations can be seen in figures (1) and (2).

(a)   **(b)**
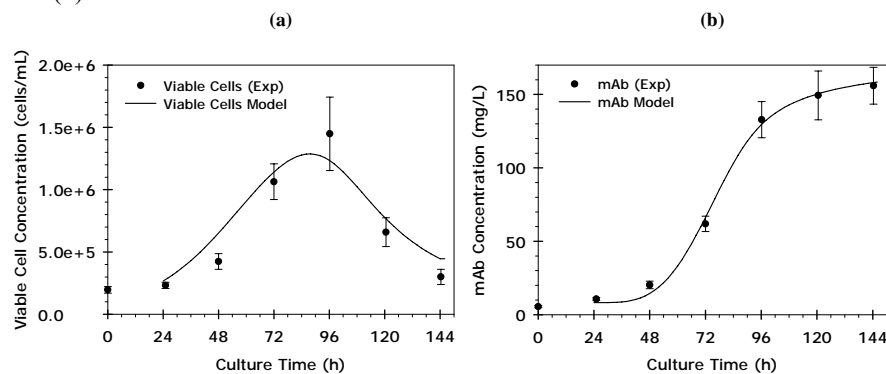
Figure 1: a) Viable Cell concentration and b) mAb concentration over time for batch cultivation of GS-NS0 cells.

During the first 24 h of culture the cells remained completely idle; therefore no effort to model the cellular behavior during that period was made. From that point onwards the cells start growing and a peak in viable cell concentration can be observed (Figure 1) after roughly 96 h of culture time. The model successfully predicts the time point of the peak although it slightly underestimates the actual magnitude of the peak. What is interesting is that after this peak is reached, nutrient uptake is completely halted (Figure 2). Both glucose and glutamate consumption stop abruptly even though both nutrients are still in abundance. This is a significant observation that leads us to the conclusion that neither glucose nor glutamate is the growth limiting substrate. Therefore, in order to derive a truly optimal feeding profile, identification of the growth limiting nutrient is required.

(a)   **(b)**
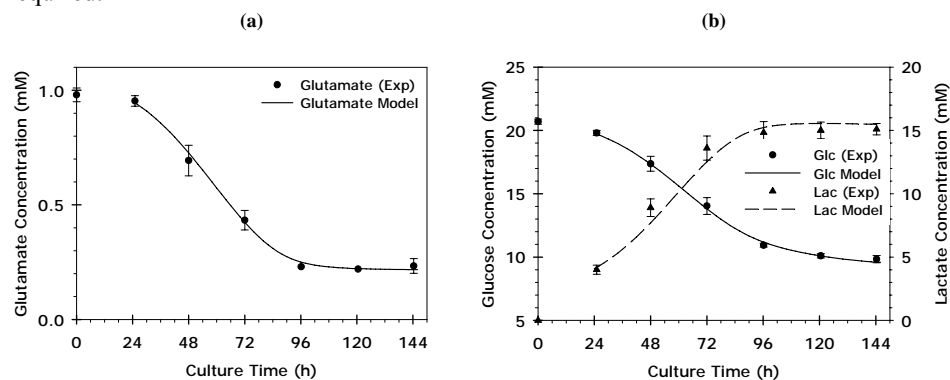
Figure 2: a) Glutamate concentration and b) Glucose & Lactate concentrations over time for batch cultivation of GS-NS0 cells.
`

In the absence of glutamine from the culture medium, it is synthesized through a GS catalyzed reaction involving glutamate, ATP and a NH_3 group usually provided by asparagine. We therefore presume that growth might be halted at 96 h of culture time due to the depletion of a viable NH_3 source for the production of glutamine. The next

step in our work towards the energy based derivation of an optimal feeding profile is the identification of the growth limiting nutrient and its inclusion in the current model. None the less the model is able to successfully capture the trends observed experimentally and is a solid first step towards the derivation of a dynamic model that successfully captures the most significant elements of GS-NS0 metabolism, facilitating the derivation of a truly optimal feeding profile.

4. Conclusions & Future Work

We have successfully presented, for the first time, a dynamic model that couples both glucose and glutamate concentration to the cellular growth rate of GS-NS0 cultures. Moreover we have identified that neither glucose nor glutamate are the nutrients that limit growth in the particular system, leading to the assumption that one or more of the essential amino acids are depleted after 96h of culture time. Therefore the next step is to include the uptake rates of the most significant essential amino-acids in our model prior to coupling ATP availability to growth and productivity characteristics for the first time. This will enable us to derive optimal feeding profiles for GS-NS0 cultures that maximize mAb titer whilst avoiding the excessive feed of glucose, thus maintaining cellular metabolism in energy efficient pathways and avoiding the overproduction of lactate.

5. Acknowledgements

References

De Tremblay M., Perrier M., Chavarie C., Archambault J., 1992, *Optimization of fed-batch culture of hybridoma cells using dynamic programming: single and multi feed cases*, Bioprocess Eng., 7, 229-234

Ho Y., Varley J., Mantalaris A., 2006, *Development and analysis of a mathematical model for antibody-producing GS-NS0 cells under normal and hyperosmotic culture conditions*, Biotechnol Prog, 22 (6), 1560-9

Kiparissides A., Kucherenko S., Mantalaris A., Pistikopoulos E.N., 2009, *Global Sensitivity Analysis Challenges in Biological Systems Modeling*, Ind. Eng. Chem. Res., 48 (15), pp 7168–7180

Process Systems Enterprise, 1997-2009, *gPROMS*, www.psenterprise.com/gproms

Xie L., Wang, D.I.C., 1994, *Applications of improved stoichiometric model in medium design and fed-batch cultivation of animal cells in bioreactor*, Growth Factors, *1*, 17-29

A nonlinear approach to the design of gain-scheduled controllers

Altimari Pietro,[a] Erasmo Mancusi,[b] Lucia Russo,[c] and Silvestro Crescitelli[d]

[a]*Dipartimento di Ingegneria Chimica Alimentare Università di Salerno, Via Ponte Don Melillo, 84084, Fisciano (SA), Italy*
[b]*Facoltà di Ingegneria, Università del Sannio, Piazza Roma , 82100, Benevento, Italy*
[c]*CNR, Istituto sulle ricerche sulla combustione, P.le Tecchio 80, Napoli, 80125, Italy*
[d]*Dipartimento d'Ingegneria Chimica Università "Federico II" Piazzale Tecchio 80, 80125, Napoli, Italy*

Abstract

Gain scheduling is a method widely applied in industrial practice to control processes where large changes of the operating conditions can occur. In its standard implementation, this technique requires to compute a family of steady states covering the operating region of interest and then to design a family of linear feedback controllers ensuring stability and desired output behavior about the selected steady states. In this contribution, a novel approach to design gain-scheduled controllers of nonlinear processes is presented. Parametric continuation and optimization techniques are implemented to compute a parameterized family of steady states covering the output range of interest and, at the same time, fulfilling a prescribed set of control requirements. Then, bifurcation analysis is performed to design a family of linear feedback controllers guaranteeing desired output behavior around the selected steady states and preventing the occurrence of state multiplicity. The method is validated on the problem of controlling a continuous exothermic reactor exhibiting state and input multiplicity.

Keywords: Gain-scheduling, bifurcation analysis.

1. Introduction

Gain-scheduling (GS) is a method widely applied in industrial practice to control nonlinear processes where large changes of the operating conditions can occur (Rugh and Shamma, 2000). In its standard application, this technique relies on the design of a family of linear feedback controllers, each of them guaranteeing desired output behavior around a different steady state. Plant steady states are, in this framework, parameterized by a set of reference variables, typically function of output variables. Hence, GS is performed by implementing the feedback controller identified by the reference signal issued to the closed-loop system.

Since this approach makes use of established linear control tools, it guarantees significant flexibility and simplicity during the design of the closed-loop system. However, GS-controllers are typically synthesized based on the interpolation of feedback controllers designed at isolated steady states. As a result, poor controller performance or even instability can occur the due to hidden coupling terms (Stilwell and Rugh, 2000). Moreover, the global performance evaluation of GS-controllers is generally performed by simulation studies. This approach likely fails to detect the presence of multiple stable solution regimes or identify robustness margins of the

desired steady states from instability boundaries. Under these conditions, sudden variations of the reference signal can cause undesired transition regimes as, for example, plant shut-down or thermal runaways (Zhang et al., 1996).

It must be also observed that no systematic methodologies have been developed to compute parameterized families of steady states covering the operating region of interest. In GS practice, steady states are typically parameterized by the output variables to be controlled and then local feedback controllers are designed around steady states corresponding to isolated values of such variables. In accordance with this design philosophy, infinite admissible parameterizations of steady states can be found as the number of manipulated inputs is greater than the number of controlled outputs. Finding the optimal parameterization is, however, an unsolved problem.

In this contribution, a novel approach to design GS-controllers tracking scalar outputs of nonlinear processes is presented. Our analysis focuses on the case where multiple inputs are simultaneously varied. Hence, parametric continuation and optimization techniques are implemented to compute a parameterized family of steady states fulfilling covering the output range of interest and fulfilling a prescribed set of control requirements. Then, bifurcation analysis is performed to design a family of linear feedback controllers guaranteeing desired transient behavior about the computed steady states and, at the same time, preventing the occurrence of state multiplicity. The method is validated on the problem of controlling a continuous exothermic reactor exhibiting state and input multiplicity.

2. Control structure design

Let the plant to be controlled be described by the following nonlinear system:

$$
\begin{cases}
\dot{x}(t) = f(x(t), u(t)), \\
y(t) = g(x(t), u(t)),
\end{cases}
\tag{1}
$$

where $x(t) \in \mathbb{R}^n$ is the state vector, $y(t) \in \mathbb{R}$ is a scalar output, $u(t) \in \mathbb{R}^m$ is the vector of manipulated inputs. The control objective is to track the output y over a prescribed range $[y_a, y_b]$. The first step to design a GS-controller performing this task is to compute a parameterized family of steady states $(x(r), u(r))$, r being the desired (or reference) output value, such that $g((x(r), u(r)) - r = 0$ as r varies in $[y_a, y_b]$. Once this family is obtained, system (1) can be linearized about $(x(r), u(r))$ leading to the following linear parameter varying model:

$$
\dot{\delta x} = \left[\frac{\partial f(x(r), u(r))}{\partial x} \right] \delta x + \left[\frac{\partial f(x(r), u(r))}{\partial u} \right] \delta u,
\tag{2}
$$

where δx, δu denote deviations from the steady state. Then, linear feedback control can be implemented upon system (2) to enforce stability and desired output behavior as r varies. In this paper, we consider the following class of state feedback GS-control laws:

$$
u(x, r) = u(r) - G(r) \cdot (x - x(r))
\tag{3}
$$

where $G(r)$ is a parameterized family of feedback gain matrices.

A nonlinear approach to the design of gain-scheduled controllers

3. Constructing optimal families of steady states

Application of the described controller design procedure requires to compute a parameterized family of steady states $(x(r), u(r))$ covering the output range of interest. If a unique manipulated input is employed, that is $m = 1$, such a problem results in computing at each $r \in [y_a, y_b]$ the steady state (x, u) fulfilling the following set of nonlinear constraints:

$$f(x,u) = 0, \ g(x,u) - r = 0, \ r \in [y_a, y_b] \tag{4}$$

However, no feasible steady state fulfilling (4) might be found as r varies due to the nonlinearity of f and g. In this case, further manipulated inputs can be varied, if possible, to compute feasible steady states spanning the output range of interest. It is not uncommon, in industrial practice, the case where several inputs are varied to cover the desired output range. It should be also noted that even if varying a unique input is sufficient to get a parameterization $(x(r), u(r))$ fulfilling (4), system linearization about such steady states might exhibit unsatisfactory controllability characteristics. Steady states producing large right-half plane eigenvalues of the open-loop system Jacobian might be, for example, detected. Under these conditions, large control efforts are needed to achieve stability. On the other hand, infinite families $(x(r), u(r))$ fulfilling (4) can be, in general, found when $m>1$. In such a case, we recast the problem of computing $(x(r), u(r))$ as follows:

$$\Psi(x(r),u(r)) = \min_{x(r),u(r)} \int_{y_a}^{y_b} L(x(r),x(r))dr \tag{5}$$

$$f(x(r),u(r)) = 0, \ g(x(r),u(r)) - r = 0$$

where L is a scalar penalty function and should be chosen to accommodate process objectives and prevent the selection of steady states around which system (1) exhibits unsatisfactory controllability characteristics (Altimari et al., 2009a).

The approach we take to solve problem (5) is geometric in nature. Let, indeed, assume that f is continuously differentiable with full-rank Jacobian, then the set of steady states of system (1) is an m-dimensional manifold (or hypersurface) and a locally invertible map $\varphi(s)=(x(s), u(s))$ exists such that $f(\varphi(s))=0 \ \forall s \in D \subseteq \mathbb{R}^m$ (Isidori, 1995). From this perspective, problem (5) can be also thought as the problem of computing a curve $(x(\alpha), u(\alpha))$ with $\alpha \in [0,1]$ constrained on the steady state manifold of system (1), minimizing the functional Ψ and such that the output function $g(x(\alpha), u(\alpha))$ is invertible. The latter condition indeed ensures that the computed curve can be parameterized in terms of r by constructing $(x(r), u(r)) = (x(\alpha(r)), u(\alpha(r)))$ with $\alpha(r) = g^{-1}(\alpha(r))$. On the other hand, a curve of steady states can be identified, based on the map φ, by a curve in $D \subseteq \mathbb{R}^m$ (see Figure-1). This allows to reduce (5) to the problem of computing a parameterization $\kappa(\alpha):[0,1] \longrightarrow D$ such that:

$$\Psi(\alpha) = \min_{\gamma(\alpha)} \int_0^1 L(\alpha) \left(\frac{\partial g}{\partial x} \frac{dx}{ds} + \frac{\partial g}{\partial u} \frac{du}{ds} \right) \frac{d\kappa}{d\alpha} d\alpha \tag{6}$$

where $L(\alpha) = L(\varphi(\kappa(\alpha)))$. It should be stressed that with such approach there is no need to embed into the optimization problem steady state equality constraints since the use of the map φ guarantees that $f(\varphi(\kappa(\alpha)))=0 \ \forall \alpha \in [0,1]$.
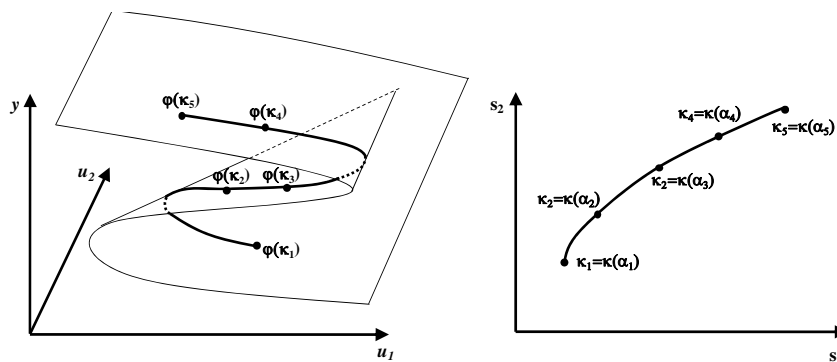
Figure-1. Parameterization of a curve constrained on the steady state manifold.

Problem (6) can be solved by replacing $\kappa(\alpha)$ by the sequence $K_N = \{\kappa(\alpha_1),.., \kappa(\alpha_N)\} = \{\kappa_1,.., \kappa_N\}$ where $\alpha 1 < .. < \alpha_N$ is a grid for the interval $[0,1]$ (see Figure-1). This gives:

$$\Psi_N(K_N) = \min_{\kappa_N} \sum_{i=1}^{N-1} L(x(\kappa_i),u(\kappa_i))(g(\kappa_{i+1}) - g(\kappa_i)) \tag{7}$$

where $g(k_i) = g(\varphi(k_i))$. It is important to note that the number of unknowns of problem (7) only depends on N and on the number m of inputs while it is independent of the dimension n of system (1). This element along with the guaranteed fulfillment of the steady state equality constraints makes the proposed approach effective to solve (5). It must be, however, stressed that these advantages are achieved at the expense of computing the map φ. An effective numerical implementation of such a map can be achieved by means of parametric continuation methods (Seydel, 1991). Further details on how to apply such methods to compute φ can be found in (Altimari et al., 2009b).

4. Application to a continuous jacketed tank reactor

As representative example, we report in this section the results of application of the proposed method to the problem of controlling a jacketed continuous tank reactor where the series of two irreversible exothermic reactions $A{\rightarrow}B{\rightarrow}C$ takes place. The model employed for such system and its parameter values can be found in (Altimari et al., 2009a) and will not be here reported. Such model consists of four differential equations describing the evolution of the concentrations x_1 and x_2 of the species A and B, of the reactor temperature x_3 and of the coolant temperature x_4. The control objective is to track the reactor temperature x_3 over the range [305, 350] by employing the feed flow rate u_1 and the coolant flow rate u_2 as manipulated inputs.

A parameterized family of steady states $(x(r), u(r))$ covering the output range of interest is first computed by applying the optimization scheme described in section 3. To this aim, the penalty function L is structured as weighted sum of the right-half plane eigenvalues of the open-loop system Jacobian to prevent the selection of steady states requiring large control efforts to be stabilized and of the yield of the intermediate product B to accommodate process performance (see (Altimari et al., 2009) for a

detailed discussion). Then, the considered output range is partitioned into three sub-domains and problem (7) is sequentially solved on each sub-domain with N = 10. The obtained results are reported in Figure-2. Here, we show the evolutions of $x_2(r)$, $u_1(r)$ and $u_2(r)$.
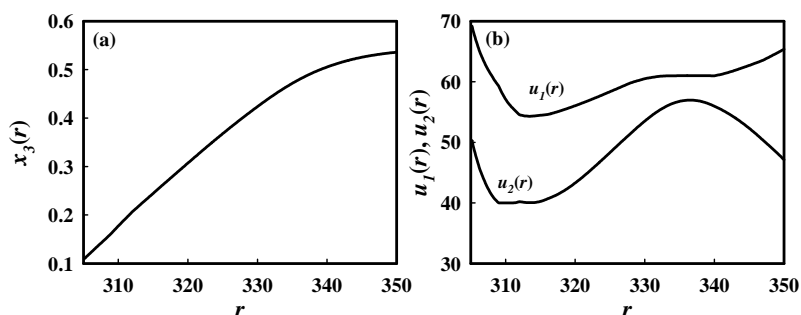
Figure-2. Optimal parameterization; (a) concentration of B; (b) feed and coolant flow rates.

It should be noted that the computed $u(r)$ represents the feedforward control law to be implemented in order to track the computed family of steady states. Besides such control law, feedback is needed to guarantee stability as r varies. To this aim, the only coolant flow rate is employed to perform feedback control. Hence, a family of feedback gain matrices $G(r)$ is computed by applying the linear quadratic regulation (LQR) algorithm (Kailath, 1996). This gives the following GS-control law:

$$u_2(r,x) = u_2(r) - 0.5B^T(r)P(r,\lambda)(x - x(r)) \tag{8}$$

where $B(r)=[D_u f(x(r),u(r))]$, $P(r,\lambda)$ is the solution of the Riccati equation $A^T P + PA - PBB^T P - R(\lambda) = 0$, $R(\lambda)$ being the diagonal matrix defined by $v = \lambda(10,10,1,1)$.
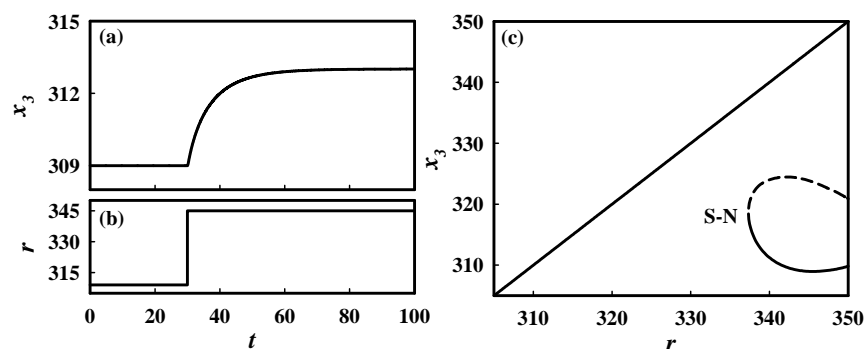
Figure-3. (a) Reactor temperature response to a step change (b) of the reference signal; (c) Steady states of the GS-closed-loop as varies.

Results of the implementation of GS-control law (8) are reported in Figure-3a. Here, we show the reactor temperature response to a step-change of the reference r from 305 to 345 K. It is apparent that the closed-loop system cannot reach the issued reference value. The reasons of such a behaviour are illustrated in Figure-3b. Here, we report the solution diagram of the closed-loop system as r varies. The desired branch $(x(r), u(r))$ is entirely stable. However, further steady states are found due to the saddle-node bifurcation S-N.

In order to examine the effects of modifying controller parameters on such coexistence, the evolution of the bifurcation S-N is analyzed as λ varies. Results of such analysis are reported in Figure-4a. The bifurcation S-N vanishes when λ>4.3. Therefore, selecting λ values in such a range enables to remove state multiplicity. However, we note that large λ values can cause undesired transient behaviour. To this regard, we report in Figure-4b reactor temperature and feedback controller responses observed, when *r* is varied according to Figure-3b, for increasing λ values. The required control efforts become larger as λ grows. On the other hand, large λ values are only needed when *r*>333 since no undesired solutions are detected, independently of λ, at lower *r* values. According to this idea, different λ values can be selected as *r* varies to optimize the controller performance and enforce prescribed margins from the saddle-node bifurcation points (Monnigmann and Marquardt, 2003).
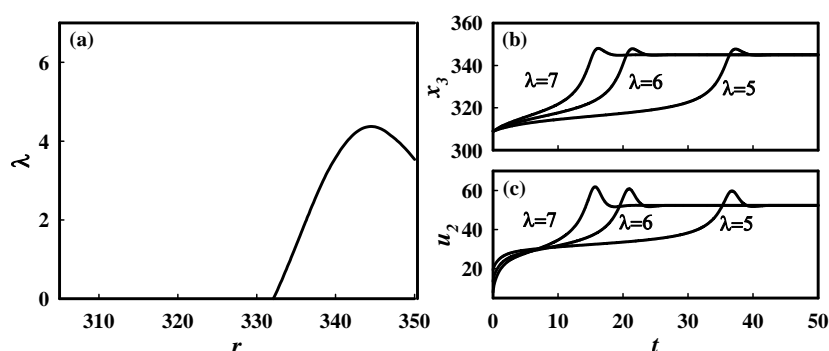
Figure-4.(a) Saddle-node bifurcations of the closed-loop system; (b) Reactor temperature and feedback controller responses (c) to a step-change in *r* from 305 to 345 K.

5. Conclusions

A method to design GS-controllers of nonlinear processes was presented. Parametric continuation and optimization techniques were implemented to compute a family of steady states covering the output range of interest and fulfilling a prescribed set of control requirements. The proposed approach was discussed thorough the application to the problem of controlling a continuous exothermic reactor. In this framework, the use of bifurcation analysis to select feedback gains enforcing desired closed-loop behavior and, at the same time, preventing the occurrence of state multiplicity was thoroughly discussed.

References

P. Altimari, L. Russo, E. Mancusi, M. di Bernardo, S. Crescitelli, (2009a), *Industrial Engineering and Chemistry Research*, 48, 9128.
P. Altimari, E. Mancusi, L. Russo, M. di Bernardo, S. Crescitelli, (2009b), *International Journal of Bifurcation and Chaos*, in press.
T. Kailath, (1980), Linear Systems, Prentice Hall, Englewood Cliffs, NJ.
M. Monnigmann, W. Marquardt, (2003), *AIChE Journal*, 49, 3110.
W. J. Rugh, J. S. Smamma, (2000), *Automatica*, 36, 1401.
R. Seydel, (1994), Practical Bifurcation and Stability Analysis, Springer-Verlag, New York.
D. J. Stilwell, W. J. Rugh, (2000), *Automatica*, 36, 665.
S. X. Zhang, N. K. Read, W. H. Ray, (1996) *AIChE Journal*, 42, 2911.

ARX-Model based Model Predictive Control with Offset-Free Tracking

Jakob Kjøbsted Huusom,[a] Niels Kjølstad Poulsen,[b] Sten Bay Jørgensen,[a] and John Bagterp Jørgensen[b]

[a]*Department of Chemical and Biochemical Engineering, { jkh,sbj}@kt.dtu.dk*
[b]*Department of Informatics and Mathematica modelling, {nkp,jbj}@imm.dtu.dk*
Technical University of Denmark, DK-2800 Lyngby, Denmark

Abstract

ARX models, is a suitable model class for linear control implementations. The parameter estimation problem is convex and easily handed for both SISO and MIMO system in contrast to ARMAX or State Space model. Model predictive control implementations insuring offset-free tracking are discussed and related. Special attention is given to an adaptive disturbance estimation method with time-varying forgetting which is shown to be less sensitive to the nature of the disturbance.

Keywords: Model Predictive Control, Offset-Free Tracking, Adaptive Estimation,

1. Introduction

Model Predictive Control (MPC) is a state of the art control technology which utilizes a model of the system in order to predict the process output over some future horizon. It solves an open loop quadratic optimization problem with the manipulated variable as decision variable. The first of the controls is implemented and after retrieving the next process output, the problem is solved again for the next control to achieve feedback. Inequality constraints can be formulated for both manipulated variables and the process outputs.

Early achievements and industrial implementations in MPC include IDCOM by Richalet et al. (1978) and Dynamic Matrix Control by Cutler and Ramaker (1980). These early algorithms were based on step or impulse response models. More general linear input-output models structure were used by Clarke et al. (1987) in Generalized Predictive Control, but an interest in MPC implementations based on state space models were created by the seminal paper by Muske and Rawling (1993). The state space approach provides a unified framework for discussion of the various predictive control algorithms and is well suited for stability analysis (Mayne et al. 2000). Other types of linear model representations, which may be convenient for system identification, can be converted to state space form for the MPC implementation. This paper will focus on the following linear, discrete time, single input/single output ARX model representation

$$A(q^{-1})y(t) = B(q^{-1})u(t) + d + \varepsilon(t), \qquad \varepsilon \in N(0, \sigma^2) \qquad (1)$$

Where A and B are polynomials of order n in the backwards shift operator q^{-1}.

$$A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} + \cdots + a_n q^{-n}$$

$$B(q^{-1}) = b_1 q^{-1} + b_2 q^{-2} + \cdots + b_n q^{-n}$$

This paper advocates the advantages in MPC based on ARX models and discus closed loop performance of the controller in case of unmeasured step disturbances. In order to

reject such types of disturbances, the basic MPC formulation needs to be expanded or designed including integrators to achieve offset-free closed loop performance. The paper is organized with an introduction to ARX MPC as a control paradigm in section 2. Implementations insuring offset-free tracking are presented in Section 3. A basic simulation case in Section 4 demonstrates the performance of the controllers and conclusions are drawn in Section 5.

2. ARX MPC as a Control Paradigm

Most industrial MPC implementations are currently based on linear model representations of the underlying system dynamics. Linear MPC is becoming a mature control technology with well established conditions for stability and robustness (Rawlings and Mayne 2009). Since MPC calculate the control based on an optimization over a prediction horizon, closed loop performance will depend strongly on the predictive capabilities of the system model. Hence typically model parameters are regressed based on prediction error methods of either the one-step-ahead prediction error or as a multistep method (Rossiter and Kouvaritakis 2001, Haber et al 2003). Estimation of parameters in ARMAX or State Space models by prediction error methods is a nonlinear and non-convex optimization problem. Parameter estimation in the ARX model structure is a convex problem. Furthermore the ARX model structure provides a much simpler estimation problem of multivariable system than the ARMAX model. Zhu (1998) and Hjalmarsson (2003) identify high order ARX models that are reduced before used in control design. Qin and Badgwell (2003) survey vendor MPC implementations and report that Honeywell's Robust Model Predictive Control Technology (RMPCT or profit-controller) as well as Invensys' model predictive control technology (Connoisseur) are based on ARX models.

Implementation of an MPC with input constraints based on an ARX model of the system is fairly simple as outlined in Huusom et al. (2009a). The ARX model is written as a State Space system on innovation form and optimal predictions of future output is given by the stationary Kalman filter where the data update is based on the innovation

$$\varepsilon_k = y_k - \hat{y}_{k|k-1} \tag{2}$$

The performance is given as the following quadratic cost

$$\phi = \frac{1}{2} \sum_{j=0}^{N-1} \left\| \hat{y}_{k+1+j|k} - r_{k+1+j|k} \right\|_2^2 + \rho \left\| \Delta u_{k+j|k} \right\|_2^2 \tag{3}$$

with penalty on the tracking error and the control move. Based on the observability matrix and the impulse response matrices a vector/matrix description the process output for the entire prediction horizon can be written. This can be used to formulate the minimization of the performance cost over the horizon, *N,* as a quadratic program which can be handled by a standard solver.

3. Offset-Free Tracking

A requirement which has to be provided by any industrial control implementation is offset-free tracking. Offset from the set point may occur for an MPC implementation if an unmeasured sustained disturbance is entering the system. It may also be the result of a mismatch between the true system and the model used for predictions by the MPC. A classical approach to avoid offset is by introducing an integrator in the control loop, as in case of PI control. In integrator can be included in the ARX-model based MPC, if the

noise term in Eq. (1) is modeled as integrated white noise. This approach was used in GPC (Bitmead et al. 1990). If we model the random noise and a constant disturbance as integrated white noise we get

$$\frac{1}{(1-q^{-1})}e_k = d + \varepsilon_k \Rightarrow e_k = (1-q^{-1})(d+\varepsilon_k) = \varepsilon_k - \varepsilon_{k-1}$$

where the constant disturbance term disappears. It is seen that the price is that the variance of *e(t)* is twice that of the actual noise. We use the following model, labeled the ΔARX model, in the MPC design, thereby introducing integration and eliminating offset. The term $(1-q^{-1})$ is also known as a Δ operator.

$$A(q^{-1})y(t) = B(q^{-1})u(t) + \frac{1}{1-q^{-1}}e(t) \Rightarrow$$

$$(1-q^{-1})A(q^{-1})y(t) = (1-q^{-1})B(q^{-1})u(t) + e(t) \tag{4}$$

The system order in this model is one order higher than the original system. An alternative approach which also extents the model order is by augmenting the system model with a disturbance state model as

$$d_k = d_{k-1} + \xi, \qquad \xi \in N(0, Q_\xi) \tag{5}$$

In this way the unknown disturbance can be estimated together with the system states by the state observer. This method was introduced by Davison and Smith (1971) and analyzed for use in linear MPC design by Pannocchica and Rawlings (2003) with conditions for detectability of the augmented system. This approach to offset-free MPC offers the disturbance state variance as a tuning parameter. If the variance approaches zero no ability to detect the disturbance is given in the state estimator design. Choosing the variance very large gives a high sensitivity to the prediction error in the disturbance state update. In the limit this approach is equivalent to having the integration. Choosing an appropriate value for the disturbance variance is not a trivial task, also in view of this tuning parameter being unbounded. An alternative approach advocated in Huusom et al (2009a) models the noise as an integrated moving average process with one lag

$$\frac{1-\alpha q^{-1}}{(1-q^{-1})}e_k = d + \varepsilon_k \Rightarrow (1-\alpha q^{-1})e_k = (1-q^{-1})(d+\varepsilon_k) = \varepsilon_k - \varepsilon_{k-1}$$

This lead to the EΔARX model used for the MPC design

$$(1-q^{-1})A(q^{-1})y(t) = (1-q^{-1})B(q^{-1})u(t) + (1-\alpha q^{-1})e(t) \tag{6}$$

It is clear that when the tuning parameter α is changes between 0 and 1 this approach also have the nominal ARX model and the ΔARX model as the extremes. Huusom et al. (2009a) show that this approach is equivalent to augmenting the system with a disturbance state since a state transformation will bring one formulation into the other. The advantage of tuning α rather than the variance of the disturbance state, is that this parameter is bounded between 0 and 1. Furthermore the variance of the disturbance state dependents on, where the disturbance is modelled to enter the system, i.e. as input, output or state disturbance. All the approaches discussed so fare suffers from a tradeoff between fast disturbance estimation versus noise sensitivity which is affected by the tuning. An attempt to get the best of both worlds was presented in Huusom et al. (2009b). Here the disturbance is estimated using adaptive techniques discounting old measurements. The forgetting is time-varying according to the prediction error and its variance as proposed in Fortescue et al. (1981). The idea is that when the level of the disturbance is known, the estimation uses a small gain from the prediction error in the

disturbance estimation, making it insensitive to noise. When large prediction errors are observed, the method increases the gain and adapt faster to the new level. In the MPC formulation the disturbance level is used in the predictions and when optimizing the control signal. The recursive algorithm is based in the following set of equations

$$e_k = y - C\hat{x}_{k|k-1}$$

$$\hat{d}_k = \hat{d}_{k-1} + \kappa_k e_k$$

$$\kappa_k = P_{k-1}(\lambda_k + P_{k-1})^{-1} \tag{7}$$

$$\lambda_k = \max\left\{1 - \frac{e_k^2}{N_\infty \sigma^2(1+P_{k-1})}, \lambda_{\min}\right\}$$

$$P_k = (1-\kappa_k)P_{k-1}\lambda_k^{-1}$$

Where κ_k is the gain from the prediction error e_k in the disturbance update, λ_k is the forgetting factor which is bounded from below, σ^2 is the process noise variance of the model in Eq. (1) and P_k is an approximation of the variance of the prediction error which is distributed as

$$e_k \in N(0, \sigma^2(1+P_k))$$

N_∞ is the equivalent horizon, which is the tuning parameter of the method. It is seen from Eq. (7) that when the system know the disturbance level, the forgetting factor is approximately $1 - 1/N_\infty$ and P_k approximates the noise variance. I.e. the gain gets very small and reduces the effect of noise in the prediction error on the update of the disturbance. If the disturbance changes abruptly to a different level, the forgetting factor will decrease, making both P_k and κ_k larger, and render the method able to follow the change. The main result in the analysis in Huusom et al. (2009b) is that closed loop performance is less sensitive with respect to the nature of the disturbance by this method than the classical approach, augmenting the system with an extra state. Hence tuning is less dependent on knowing the true size and frequency of a series of step disturbances.

4. An Example

A simulation study is performed to show the characteristics of the methods for offset-free MPC based on ARX models. The model in Eq (1) is simulated with the following set of parameters which gives a pole in 0.9 and a pair of complex poles in 0.75±0.37i.

$$a_1 = -2.4, \quad a_2 = 2.05, \quad a_3 = -0.63, \quad b_1 = 0.5, \quad \sigma^2 = 0.1$$

For a simulation horizon of 10.000 samples and an unconstraint implementation of the input in the MPC, a series of simulations are performed for a range of the tuning parameters, ρ in Eq. (3) and α and Q_ξ for the EΔARX model and the disturbance model respectively. The following implementations are tested with different models: The true ARX model, the ΔARX model, the EΔARX and finally the augmented system with a disturbance model. The results are shown in a Pareto plot for the input and output variance on Fig. 1where $\alpha \in \{0, 0.1, .., 1\}$ for the EΔARX model and $Q_\xi \in [10^{-7}; 10^2]$ in the disturbance modeling approach. It is clearly seen that the ARX and ΔARX MPC implementations is achieved in the limit for the two other methods.

ARX-Model based Model Predictive Control with Offset-Free Tracking
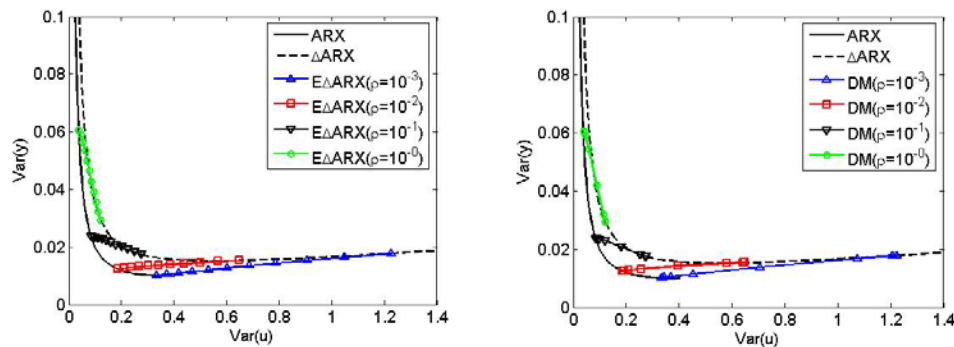
Figure 1. Pareto plot for four MPC implementations of ARX model based MPC.

In order to see the advantage of using the ARX MPC with a disturbance estimation algorithm, a set of simulation was performed over a horizon of 1000 samples. In the Base case no disturbance enters the system. For the case Small step and Large step a sustained disturbance enters at time 50 with a magnitude of 0.25 and 1 respectively. Finally a disturbance which drifts as integrated white noise with the same variance as the process noise is used. The results are shown in Fig. 2 as the closed loop performance versus the sensitivity to the prediction error in the disturbance update. By converting the actual tuning parameters like this, the plots are more easily compared. L_d is the observer gain to disturbance state which is a function of the variance for the disturbance model.
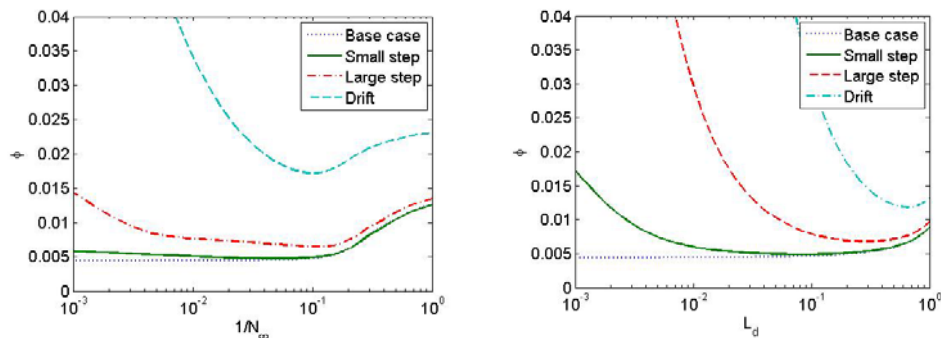
Figure 2. MPC performance versus sensitivity to the prediction error in the disturbance update.

While the minimum value for the performance cost for the four cases are very close for both methods, disturbance estimation with time-varying forgetting clearly provides better performance for different types of disturbances over a broad range of the tuning parameter. This is not the case for the disturbance modelling, which is clearly superior when the disturbance drifts, but this method requires detailed knowledge of the disturbance to provide good performance.

5. Conclusion

ARX models are well suited for control design since this linear model class is associated with a convex parameter estimation problem for both SISO and MIMO systems. Several methods for ensuring offset-free tracking by manipulation of the system model is presented and related in a simulation study. The disturbance model approach or the EΔARX model gives an extra degree of freedom for tuning, compared

to including an integrator as in the ΔARX model. This tuning needs to balance fast convergence of the disturbance state against high noise sensitivity in the estimate. These methods are identical in performance, but the tuning parameter, α, in the EΔARX model is bounded. Adaptive disturbance estimation with time-varying forgetting adjusts the speed of adaptation for the disturbance estimate, according to the prediction errors. Hence it provides fast estimation when needed, while low noise sensitivity when the disturbance is known. It was shown in a simulation study that this formulation is less sensitive to the nature of the disturbance and must be preferred if the disturbance consist of infrequent steps of changing size.

6. Acknowledgements

References

Bitmeat, R. R., Gervers, M. and Wertz, V. (1990) Adaptive Optimal Control – The Thinking Man's GPC. Prentice Hall.

Clarke, D., Mohtadi, C. and Tuffs, P.S. (1987). Generalized predictive control – part i. the basic algorithm. *Automatica* 23(2), 137-148.

Cutler, C. and Ramaker, B. (1980). Dynamic matrix control – a computer control algorithm. In *Proceedings of the Joint Automatic Control Conference*.

Davison, E. J. and Smith, H. W. (1971) Pole assigment in linear time-invariant mutivarible systems with constant disturbances. *Automatica* 7(4) 489-498

Fortescue, T. R., Kershenbaum, L. S. and Ydstie, B. E. (1981) Implementation of self-tuning regulators with variable forgetting factors. *Automatica* 17(6) 831-835.

Haber, R., Schmitz, U. and Bars, R. (2003) Long-range optimal model and multi-step-ahead pediction identification for predictive control. *In 13th IFAC Symposium on System Identification*. 501-506.

Hjalmarsson, H. (2003) From experiments to closed loop control. *In 13th IFAC Symposium on System Identification*. 1-14.

Huusom, J. K., Poulsen, N. K., Jørgensen, S. B. and Jørgensen, J. B. (2009a) Offset-Free Model Predictive Control based on ARX models. *Submitted for Journal of Process Control*.

Huusom, J. K., Poulsen, N. K., Jørgensen, S. B. and Jørgensen, J. B. (2009b) Adaptive Disturbance Estimation for Offset-Free Model Predictive Control. *Submitted for the 9th International Symposium on Dynamics and Control of Process Systems – DYCOPS9*.

Mayne, D.Q., Rawlings, J.B., Rao, C.V. and Scokaert P.O.M. (2000). Constrained model predictive control: Stability and optimality. *Automatica* 36(6), 789-814.

Muske, K.R. and Rawlings J.B. (1993). Model predictive control with liner models. *AIChE Journal*, 39(2), 262-287.

Pannocchia, G. and Rawlings, J. B. (2003) Disturbance models for offset-free model-predictive control. *AIChE Journal*, 49(2) 426-437

Qin, S. J. and Badgwell, T. A. (2003) A survey of indutrial model predictive control technology. *Control engineering Practice* 11(7) 733-764.

Rawlings, J. B. and Mayne, D. Q. (2009) Model Predictive Control: Theory and Design. Nob Hill Pulications.

Richarlet, J., Rault, A., Testud, J.L. and Papon, J. (1978) Model predictive heuristic control: Application to industrial processes. *Automatica* 14(5), 413-428.

Rossiter, J. A. and Kouvaritakis, B. (2001) Modelling and implicit modeling for predictive control. *International Journal of Control* 74(11) 1085-1095

Zhu, Y. (1998) Multivariable process identification for MPC: The asymptotic method and its applications. *Journal of Process Control* 8(2), 101-115.

Simultaneous State Estimation and Model Predictive Control by Multi-Parametric Programming

Anna Voelker[a], Konstantinos Kouramas[a], Efstratios N. Pistikopoulos[a]

[a]Centre for Process Systems Engineering, Chemical Engineering, Imperial College, London, SW7 2AZ, UK
E-mail:avoelker@imperial.ac.uk, e.pistikopoulos@imperial.ac.uk

Abstract

In this work, we demonstrate how Moving Horizon Estimation (MHE) and Model Predictive Control (MPC) can be simultaneously addressed via multi-parametric programming. In particular, we present a method for obtaining the error dynamics, the error bounds, and the dynamics of the estimated states of the MHE which are essential for robust multi-parametric explicit MPC. The explicit incorporation of the estimation error bounds guarantees robustly the satisfaction of constraints on the states and the control inputs.

Keywords: MHE, explicit/multi-parametric MPC, multi-parametric programming, robust control

1. Introduction and problem statement

We consider the explicit/multi-parametric Model Predictive Control (MPC) problem of a linear discrete-time system with state and input constraints

$$\min_{u} \left\| x_{N_{MPC}} \right\|_{P_{MPC}}^2 + \sum_{i=1}^{N_{MPC}} \left\| x_i \right\|_{Q_{MPC}}^2 + \sum_{i=0}^{N_{MPC}-1} \left\| u_i \right\|_{R_{MPC}}^2$$

$$\text{s.t.} \quad x_{i+1} = Ax_i + Bu_i, \quad y_i = Cx_i \tag{1}$$

$$x \in \mathbb{X} = \left\{ x \in R^n \mid D_x x \le d_x \right\}, u \in \mathbb{U} = \left\{ u \in R^m \mid D_u u \le d_u \right\}$$

where x, y, u are the states, outputs and inputs of the system, $\mathbb{X}$, $\mathbb{U}$ are the sets of the state and input constraints that contain the origin in their interior, Q_{MPC}, P_{MPC} are symmetric semi-positive definite matrices and R_{MPC} is a symmetric positive definite matrix, N_{MPC} is the horizon of the MPC. The problem of explicit/multi-parametric MPC (1) has received significant attention in the open literature (see [3-6] and references within). The key idea of explicit/multi-parametric MPC is to solve the optimization problem (1) by multi-parametric programming and derive the control inputs $u_0 = f(x_0)$ as a set of explicit functions of the current system states x_0 [3-6]. The main advantage of explicit/multi-parametric MPC is that it obtains the control variables u_0 with simple function evaluations by implementing the control law $u_0 = f(x_0)$, instead of solving online the optimization problem (1).

The main issue in the implementation of the explicit/multi-parametric MPC, as with any other state-space model based control method, is that it relies on the availability of the system states to derive the control variable. However, for many real systems the state has to be estimated from the measurements using estimation techniques, such as Moving Horizon Estimation (MHE) [1]. The estimator is used to obtain an estimate $\hat{x}$

of the real state x. This estimate $\hat{x}$ then replaces the unknown real value of x in (1) to obtain the control variable u. However, due to the estimation error $e_x = x - \hat{x}$, the state constraints for the real system are described by $D_x(\hat{x} + e_x) \leq d_x$. It is obvious then that variations of the estimation error may result in constraint violations and that the effect of the estimation error e_x on the state constraints has to be explicitly accounted for. This is achieved by considering e_x as an unknown disturbance in the MPC problem and by employing robust MPC methods for the control of the system [9]. In order to achieve this, the dynamics and bounds of the estimation error have to be obtained [8]. In this work we present for the first time a method for obtaining the estimation error dynamics and the error bounds for a general formulation of an unconstrained Moving Horizon Estimator (MHE). The design of robust explicit MPC that accounts for the effect of the estimation error is then demonstrated by presenting an example from the relevant literature. The tube-based MPC method of [9] is used together with the error dynamics and the bounds obtained by the proposed method of this work. The robust MPC is then solved using multi-parametric programming methods.

2. Moving horizon estimation

The objective of this work is to develop a method for obtaining the error dynamics and error bounds for the unconstrained MHE. The MHE obtains the estimates of the system states by solving the following optimization problem

$$\min_{\hat{x}_{T-N|T}, \hat{W}_{T-N|T}} \left\| \hat{x}_{T-N|T} - \underline{x}_{T-N|T} \right\|_{P^{-1}}^2 + \sum_{k=T-N}^{T-1} \left\| \hat{w}_k \right\|_{Q^{-1}}^2 + \sum_{k=T-N}^{T} \left\| \hat{v}_k \right\|_{R^{-1}}^2 \tag{2}$$

$$\text{s.t.} \quad \hat{x}_{k+1} = A\hat{x}_k + Bu_k + G\hat{w}_k, \quad \hat{y}_k = C\hat{x}_k + \hat{v}_k \tag{3}$$

where T is the current time, Q, R, P are the covariances of w, v, x assumed to be symmetric and positive definite. Note, that the disturbance and noise variables w, v are either artificial variables usually employed in MHE for estimation purposes or can be used to describe the real system disturbances and noise. Moreover, $(\cdot)$ denotes the variables of the real system $x_{k+1} = Ax_k + Bu_k + Gw_k$, $y_k = Cx_k + v_k$, $(\hat{\cdot})$ the estimated variables of system (3), $(\hat{\cdot})^*$ the optimizer of problem (2), N is the horizon of the MHE, i.e. the amount of past data taken into account and $\underline{x}_{T-N|T}$ is the estimated mean value of $\hat{x}_{T-N|T}$, also referred to as cost to arrive or arrival cost. The estimated mean $\underline{x}_{T-N|T}$ is updated by the so called 'smoothed' update [1] at each step by $\underline{x}_{T-N|T} = A\hat{x}^*_{T-N-1|T-1} + Bu_{T-N-1|T-1} + G\hat{w}^*_{T-N-1|T-1}$. Note that the arrival cost is updated only after $T > N$ while for time $T \leq N$, the full information estimator is solved and the arrival cost is not updated [1]. The MHE is stable if Q, R, P are positive definite, $N \geq n$, and (A, C) is observable [1]. Note: the MHE (2), (3) is unconstrained if (3) is substituted into (2). *The result of the minimization at time T is the system state* $\hat{x}^*_{T-N|T}$ *at time T-N, and the estimated noise sequence* $\hat{W}^*_{T-N|T} = \{\hat{w}^*\}_{T-N}^{T-1}$ *from time T-N up to time T-1.*

The MHE has received significant attention in the literature of state estimation (see [1] and reference within). However, the simultaneous use of MHE and MPC still remains an unresolved problem [8]. In [8] the error dynamics and bounds were obtained for the

unconstrained MHE, however without considering any estimates of the disturbances. For reasons of optimality, the additional estimates of the disturbance W can only improve the estimation result. Hence, we consider here the more general case of (2). Our objective is to determine the dynamics of the estimation error

$$e_T = \begin{bmatrix} \hat{x}_{T|T} \\ \hat{W}_{T-N|T} \end{bmatrix} = \begin{bmatrix} x_T - \hat{x}_{T|T} \\ W_{T-N} - \hat{W}_{T-N|T} \end{bmatrix} \tag{4}$$

as well as the bounds of this error that can then be used with robust explicit/multi-parametric MPC methods to ensure that the state constraints of the system are not violated due to the presence of the error.

Dynamics of the estimation error

We now present a method for deriving the error dynamics for the unconstrained MHE (2), (3). In order to obtain the error dynamics, the following three steps are applied [7,8]: **1.** obtain algebraic solution of the MHE by partial differentiation of the objective function (2) with respect to $\hat{x}$ and $\hat{W}$, **2.** recall that the MHE estimates $\hat{x}^*_{T-N|T}$ and formulate $\hat{x}_{T|T} = A^N \hat{x}_{T-N|T} + b_T U + g_T \hat{W}_{T-N|T}$ where b_T, g_T are the last rows of b, g (see below) and U is a column vector with the N most recent inputs, **3.** from the definition of the estimation error (4) and after some relatively straight forward algebraic manipulations such as substituting the arrival cost and the state-space description of the system (3), the error dynamics are given as:

$$e_{T+1} = MH^{-1}F_e M^{-1} e_T + \begin{bmatrix} MH^{-1}F_W & MH^{-1}F_V \end{bmatrix} \begin{bmatrix} W \\ V \end{bmatrix} \tag{5}$$

where

$$M = \begin{bmatrix} A^N & g_T \\ 0 & I \end{bmatrix}, \quad H = \begin{bmatrix} 2 \cdot P^{-1} + 2 \cdot ca^T \cdot \mathrm{diag}(R^{-1}) \cdot ca & 2 \cdot ca^T \cdot \mathrm{diag}(R^{-1}) \cdot cg \\ 2 \cdot cg^T \cdot \mathrm{diag}(R^{-1}) \cdot ca & 2 \cdot \mathrm{diag}(Q^{-1}) + 2 \cdot cg^T \cdot \mathrm{diag}(R^{-1}) \cdot cg \end{bmatrix},$$

$$F_e = \begin{bmatrix} 2 \cdot P^{-1} \cdot A & \begin{bmatrix} 2 \cdot P^{-1} \cdot G & 0 & \dots & 0 \end{bmatrix} \\ 0 & 0 \end{bmatrix}, F_W = \begin{bmatrix} 0 \\ 2 \cdot \mathrm{diag}(Q^{-1}) \end{bmatrix}, F_V = \begin{bmatrix} -2 \cdot ca^T \cdot \mathrm{diag}(R^{-1}) \\ -2 \cdot cg^T \cdot \mathrm{diag}(R^{-1}) \end{bmatrix},$$

$$a = \begin{bmatrix} I \\ A \\ \vdots \\ A^N \end{bmatrix}, \quad g = \begin{bmatrix} 0 & 0 & \dots & 0 & 0 \\ G & 0 & \dots & 0 & 0 \\ A \cdot G & G & \dots & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ A^{N-1} \cdot G & A^{N-2} \cdot G & \dots & G & I \end{bmatrix}, \quad b = \begin{bmatrix} 0 & 0 & \dots & 0 \\ B & 0 & \dots & 0 \\ A \cdot B & B & \dots & 0 \\ \vdots & & \ddots & \vdots \\ A^{N-1} \cdot B & A^{N-2} \cdot B & \dots & B \end{bmatrix}, \qquad H$$

$$ca = \mathrm{diag}(C) \cdot a, \quad cg = \mathrm{diag}(C) \cdot g, \quad W \in \mathbb{W}_N = \underbrace{\mathbb{W} \times \mathbb{W} \times \dots \times \mathbb{W}}_{N-1\,\text{terms}}, \quad V \in \mathbb{V}_N = \underbrace{\mathbb{V} \times \mathbb{V} \times \dots \times \mathbb{V}}_{N\,\text{terms}}$$

and M are positive definite matrices and hence the inverses exists, $\mathrm{diag}(Q)$ is a matrix of appropriate size with Q on its main diagonal and zero everywhere else. Note on step **1.** of the method: The MHE (2) is posed as a dynamic optimization problem. Taking the partial derivatives becomes significantly easier if (2), (3) is first reformulated as a static quadratic programming problem [2].

Eq. (5) describes the error dynamics with a linear autonomous dynamic system. The error dynamics (5) can then be used to derive the set $\mathbb{E}$ in which the estimation error is bounded by using set theoretic methods and specifically the minimal robust positively invariant (mRPI) set theory [9,10].

In addition, the dynamics of the estimated state $\hat{x}$ can now be obtained from (5) as follows [8]:

$$\hat{x}_{k+1} = A\hat{x}_k + Bu_k + \underbrace{\left(\begin{bmatrix} A & 0 \end{bmatrix} - \begin{bmatrix} I & 0 \end{bmatrix} MH^{-1}F_e M^{-1}\right)}_{A_{\text{err\_x}}} e_k +$$

$$\underbrace{\left(\begin{bmatrix} 0 & \ldots & 0 & G \end{bmatrix} - \begin{bmatrix} I & 0 \end{bmatrix}\right) MH^{-1}F_W W}_{g1} \quad \underbrace{-\begin{bmatrix} I & 0 \end{bmatrix} MH^{-1}F_V}_{g2} V = A\hat{x}_k + Bu_k + t \qquad (6)$$

Eq. (6) describes the effect of the disturbances, (i.e. the noise w, v and the estimation error) on the state estimate given the past estimate. The disturbances are expressed by the new variable t. Since the disturbances w, v are not known, the bounds on the estimation error can be derived only from the knowledge of the bounds on w and v, by obtaining the error set $\mathbb{E}$. The bounds of t can be obtained by using set theoretic methods [9,10] as $t \in \mathbb{T} = A_{\text{err\_x}} \cdot \mathbb{E} \oplus g1 \cdot \mathbb{W}_N \oplus g2 \cdot \mathbb{V}_N$ where $\oplus$ denotes the Minkowski sum.

In conclusion, we have presented above a method for describing the error dynamics as the linear autonomous dynamic system (5). It is now possible to employ a variety of robust control design methods that can utilize the description (5) of the error dynamics and (6) of the estimated state as it is demonstrated in the following section.

3. Illustrative example

In order to demonstrate how the error dynamics and bounds can be used with robust explicit MPC methods, we use the tube-based MPC method of [9] where we incorporate the state dynamics (6) and error bounds expressed by the set $\mathbb{E}$ as follows:

$$\min_{\bar{x}_0, \bar{u}_0} \left\| \bar{x}_{N_{MPC}} \right\|_{P_{MPC}}^2 + \sum_{i=0}^{N_{MPC}} \left\| \bar{x}_i \right\|_{Q_{MPC}}^2 + \sum_{i=0}^{N_{MPC}-1} \left\| \bar{u}_i \right\|_{R_{MPC}}^2$$

s.t. $x_{i+1} = Ax_i + Bu_i + Gw_i$ (actual system), $\bar{x}_{i+1} = A\bar{x}_i + B\bar{u}_i$ (nominal system),

$$\hat{x}_{i+1} = A\hat{x}_i + Bu_i + t \text{ (estimated system Eq. (6))}, u_T = \bar{u}_T^* + K\left(\hat{x}_T^* - \bar{x}_T^*\right), \bar{u} \in \mathbb{U} \sim K\bar{\mathbb{S}}, \qquad (7)$$

$$\bar{x}_i \in \bar{\mathbb{X}} = \mathbb{X} \sim \mathbb{S}, i = 1 \ldots N_{MPC} - 1, \mathbb{S} = \mathbb{E}_X \oplus \bar{\mathbb{S}}, \hat{x}_0 \in \bar{x}_0 \oplus \bar{\mathbb{S}}, \bar{x}_f \in \bar{\mathbb{X}}_f$$

Where $\sim$ denotes the Pontryagin set difference [9, references within], K is a stabilizing state-feedback controller, $\mathbb{E}_X = \begin{bmatrix} I & 0 \end{bmatrix} \mathbb{E}$ is the error in the estimation of the system states only (recall that the MHE also estimates W) and $\bar{\mathbb{S}}$ is the mRPI set to the system $\hat{x}_{i+1} - \bar{x}_{i+1} = (A + BK)(\hat{x}_i - \bar{x}_i) + t$. The terminal set $\bar{\mathbb{X}}_f$ is chosen to be the maximal invariant set of the nominal system [9]. The objective of the controller is to: (a) satisfy the constraints of the actual system in the presence of the noise and the estimation error and (b) ensure that the state at time N_{MPC} lies within the terminal set $\bar{\mathbb{X}}_f$. Note that (7)

explicitly accounts for the estimation dynamics (6) as well as the error bounds $\mathbb{E}$. The state constraints are tightened to account for the effect of the estimation error, by using $\mathbb{X} \sim \mathbb{S}$, where $\mathbb{S}$ includes the error bound set $\mathbb{E}$. It is obvious that without the results from section 2 the MPC formulation (7) could not be used. The MPC (7) is a quadratic optimization problem with linear constraints that can be solved as a multi-parametric quadratic programming problem [3-6] to obtain $\overline{x}_0, \overline{u}_0$ as an explicit solution of $\hat{x}_T^*$: $\left[\overline{x}_0, \overline{u}_0\right] = K_i \ \hat{x}_T^* + c_i$ if $\hat{x}_T^* \in CR^i$ where CR^i is the critical regions where the above expression of $\left[\overline{x}, \overline{u}\right]$ as function of $\hat{x}_T^*$ is valid. For illustration purposes, we apply the robust explicit MPC (7) for the following system [8,9]:

$$x_{k+1} = \begin{bmatrix} 1 & 1 \\ 0 & 1 \end{bmatrix} x_k + \begin{bmatrix} 1 \\ 1 \end{bmatrix} u_k + \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} w_k, \quad y_k = \begin{bmatrix} 1 & 1 \end{bmatrix} x_k + v_k,$$

$$x_k \in \mathbb{X} = \left\{ x \in \mathbb{R}^2 \mid x_1 \in [-50,3], x_2 \in [-50,3] \right\}, \quad u_k \in \mathbb{U} = \left\{ u \in \mathbb{R} \mid |u| \leq 3 \right\}$$

$$w_k \in \mathbb{W} = \left\{ w \in \mathbb{R}^2 \mid |w|_\infty \leq 0.1 \right\}, \quad v_k \in \mathbb{V} = \left\{ v \in \mathbb{R} \mid |v|_\infty \leq 0.05 \right\}$$

The set-up of the MHE is: $N=4$, $Q=0.1 \cdot I$, $R=0.05 \cdot I$, P is the solution to the algebraic Riccati equation of the actual system. The configuration of the MPC is: $N_{MPC}=13$ $Q_{MPC}=I$, $R_{MPC}=0.01$, $K=[-1\ -1]$, P_{MPC} is the solution of the algebraic Riccati equation for the nominal system [9]. Fig. 1 depicts the 1444 critical regions of the explicit/multi-parametric solution of the MPC, Fig. 2 shows the trajectories of the actual system for the initial state $x_0=[-3,-8]$ and Fig. 3 shows the corresponding control input. As it is obvious, the MPC manages to drive the system trajectories into the terminal set while satisfying all constraints.
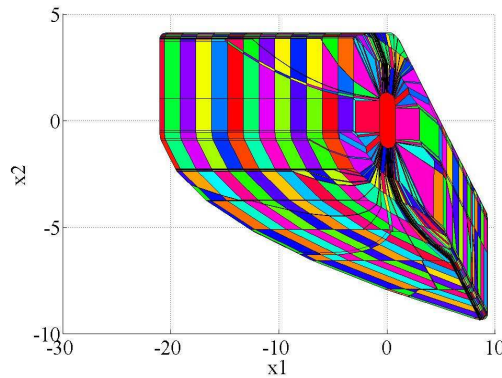
Fig. 1: Multi-parametric solution of the MPC.
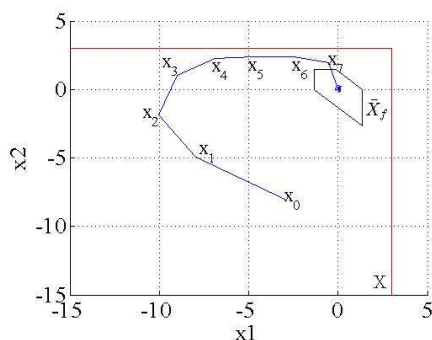
Fig. 2: Trajectories of the actual system.
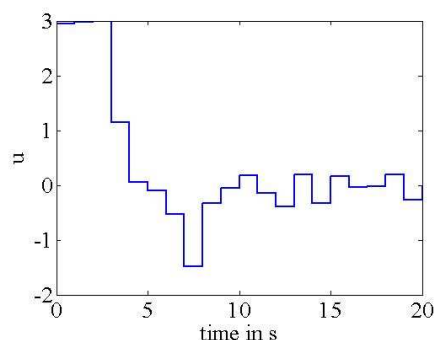
Fig. 3: Control input.

4. Conclusions and future work

The objective of this work is to present a method for obtaining the error dynamics and bounds for the unconstrained MHE, which are of fundamental importance for the integration of MHE and explicit/multi-parametric MPC. Based on previous work on tube-based MPC, an explicit/multi-parametric MPC controller was shown that integrates in its design the information of the error dynamics and bounds. Future work on the simultaneous MHE and explicit/multi-parametric MPC will focus on further extending the results of this work for the more general case of constrained MHE.

5. Acknowledgements

References

[1] C.V. Rao, *Moving Horizon Strategies for the Constrained Monitoring and Control of Nonlinear Discrete Time Systems*, Ph.D. thesis, University of Wisconsin-Madison, 2000

[2] M.L. Darby, M. Nikolaou, *A parametric programming approach to moving-horizon state estimation*, Automatica, 2007, 43, 885-891

[3] E.N. Pistikopoulos, M.C. Georgiadis, V. Dua, *Multi-Parametric Programming*, WILEY-VCH, 2007

[4] E.N. Pistikopoulos, M.C. Georgiadis, V. Dua, *Multi-Parametric Model-Based Control*, WILEY-VCH, 2007

[5] A. Bemporad, M. Morari, V. Dua, E.N. Pistikopoulos, *The explicit linear quadratic regulator for constrained systems*, Automatica, 2002, 38, 3-20

[6] E.N. Pistikopoulos, *Perspectives in multiparametric programming and explicit model predictive control*, AIChE Journal, 2009, 55, 1918-1925

[7] A. Alessandri, M. Baglietto, G. Battistelli, *Receding-horizon estimation for discrete-time linear systems*, IEEE Transactions on Automatic Control, 2003, 48, 473-478

[8] D. Sui, L. Feng, M. Hovd, *Robust output feedback model predictive control for linear systems via moving horizon estimation*, ACC, 2008, 453-45

[9] D. Mayne, D. Rakovic, R. Findeisen, F. Allgöwer, *Robust output feedback model predictive control of constrained linear systems*, Automatica, 2006, 42, 1217-1222

[10] S. Rakovic, E. Kerrigan, K. Kouramas, D. Mayne, *Invariant approximations of the minimal robust positively Invariant set*, IEEE Transactions on Automatic Control, 2005, 50, 406-410

Control of Process Operations and Monitoring of Product Qualities through Generic Model-based in Batch Cooling Crystallization

Noor Asma Fazli Abdul Samad, Ravendra Singh, Gürkan Sin, Krist V. Gernaey, Rafiqul Gani

Department of Chemical and Biochemical Engineering, Søltofts Plads, Building 229, Technical University of Denmark, DK-2800 Lyngby, Denmark, rag@kt.dtu.dk

Abstract

A generic model-based framework has been developed for crystallization processes, with applications aiming at the control of process operations and the monitoring of product quality. This generic model-based framework allows the systematic development of a wide range of crystallization models for different operational scenarios. This enables the design and control engineers to analyze various crystallization operations and conditions, thus facilitating the development of process control and monitoring systems (PAT systems) for crystallization processes. The generic framework has been implemented in the ICAS-PAT software which allows the user to design and validate PAT systems through a systematic computer-aided framework. The application of the framework is highlighted for batch cooling crystallization of paracetamol where the framework was applied for design of a process monitoring and control system to obtain a desired crystal size distribution (CSD).

Keywords: crystal size distribution (CSD), PAT, process monitoring and control, crystallization, paracetamol

1. Introduction

Crystallization is an important operation when manufacturing fine chemicals or pharmaceuticals. It is a widely used technique in solid-liquid separation processes to obtain solid products of high purity at relatively low costs. Requirements for crystal products are usually high purity, a specific crystal size distribution and a desired crystal shape [1]. Consequently many efforts have been made to model the crystallization process to support the development of appropriate process operations and control scenarios to meet specific end product demands. So far, the published crystallization process models have been problem specific, meaning that the models were developed with a certain crystal product in mind. Hence it is not surprising to notice that research on crystallization modeling emphasizes different issues such as crystal size distribution (CSD) or crystallization kinetics, depending on the aim of the specific modeling study. Furthermore, specific models employ numerous underlying assumptions, for example, on agglomeration and crystal breakage factors. As a consequence, there are many specific models available in the literature with different degrees of complexity, which makes their selection and use for a specific problem difficult if not confusing. There is therefore a need for the development of a generic crystallization model to assist the systematic and efficient development of appropriate models for specific crystallization processes.

Once an appropriate crystallization model has been developed, it can be used as a tool for process design, and for design of control and monitoring systems to ensure the desired end product quality. Such a process control and monitoring system is required for a pharmaceutical production process that is operated according to the Process Analytical Technology (PAT) guidance [2]. The objective of this work is thus to develop a generic model-based framework that allows the study of different crystallization operational scenarios, and which also supports the design, comparison and validation of process control and product monitoring systems. Since this framework will need to rely on process models, the framework will be extended with a tool to systematically develop crystallization models.

The use of the generic crystallization model, the PAT design framework and the associated tools is highlighted through the ICAS-PAT software. The application of the model-based framework is highlighted using a paracetamol batch cooling crystallization process as a case study, where the objective is to obtain a desired CSD.

2. Generic Model-based Framework

An overview of the extended framework for design of process control and product monitoring systems is shown in Fig. 1 where the generic model options have been added to the original [2]. The starting point for the design methodology is the problem definition in terms of process specifications and product quality specifications that is usually provided by the manufacturer or PAT system designer. A model library and a knowledge base have been developed and act as the supporting tools for the design of the process control and product monitoring system.
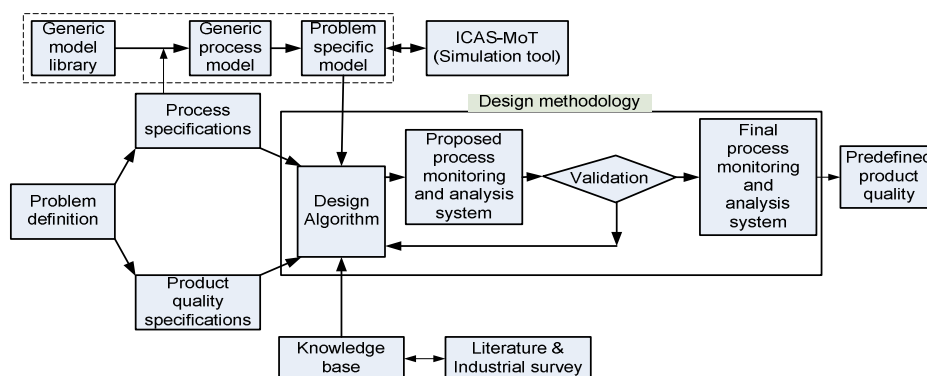
Figure 1. Extended schematic representation of the PAT design framework [2]

A systematic modeling framework (see Fig. 2) has been developed and implemented in the ICAS-PAT model library to create the various crystallization process/operation models from a generic batch cooling crystallization model. This modeling framework starts with the selection of the chemical system that needs to be investigated and the associated known information about its production scenarios. Then, the necessary balance equations and constitutive equations are extracted from the generic model library. The balance equations consist of population balance, overall mass balance and energy balance equations for the defined crystallization volume plus energy balance equations for the cooling jacket. The constitutive equations library contains a set of models of nucleation, crystal growth rate, supersaturation, saturation concentration,

metastable concentration and physical properties corresponding to different types of crystallization processes. Subsequently a problem specific model is created which is verified through model analysis and solution. Finally the problem specific model is transferred to the ICAS-PAT model library through ICAS-MOT. In this way, based on the process and product quality specifications supplied by the user (Fig. 1), the generic model is adapted to reflect a specific case study and it allows the user to consider the necessary operational scenarios enabling thereby analysis of crystallization operations and conditions.
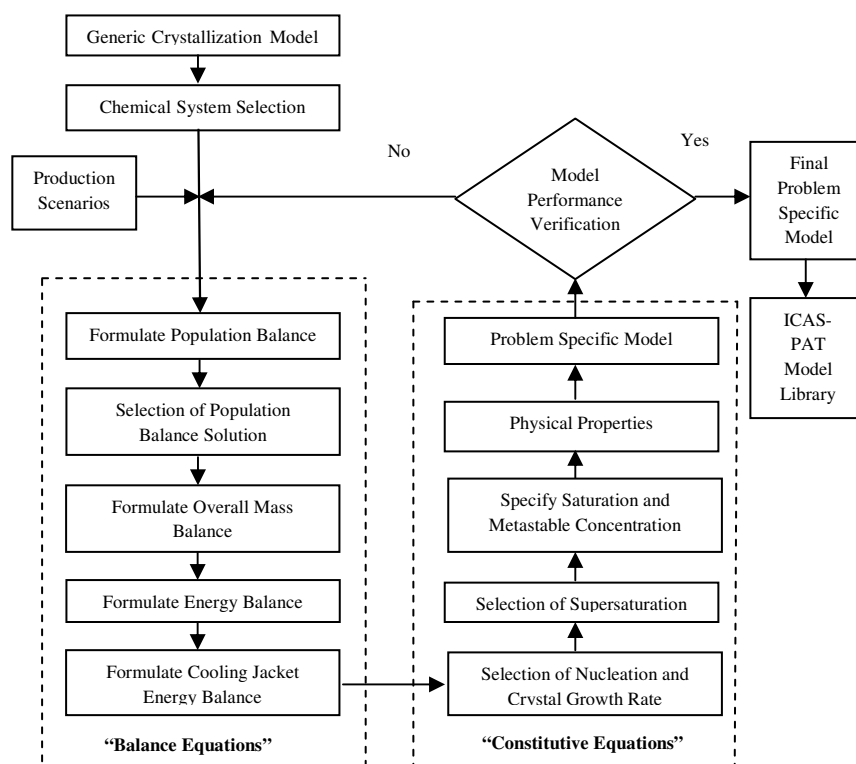
Figure 2. Generic crystallization modeling framework

The developed design algorithm in the PAT design framework (see Fig. 1) relates the product and process specifications to the available supporting tools and subsequently generates a design proposal for the process monitoring and analysis system. If the obtained PAT system satisfies the requirements then it is selected as the final design of the process control and product monitoring system, which can then be subsequently implemented and used in practice to obtain the predefined product quality consistently.

3. Case study: Paracetamol crystallization process

The paracetamol crystallization process is adopted from the literature [1, 3]. The objective here is to design a PAT system for this process using the extended ICAS-PAT software. An overview of the features available in ICAS-PAT is shown in Fig. 3 [4].
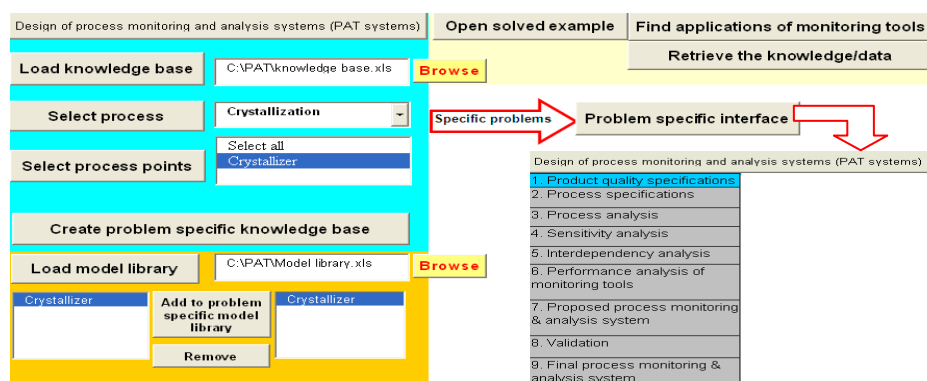
Figure 3. ICAS-PAT software overview [4]

3.1. Problem specific knowledge base: A knowledge base containing the information/data required for design of a PAT system for a crystallization process is created from a generic knowledge base. (see Fig. 3, top, left).

3.2. Problem specific process model: First a generic crystallization process model is selected from the model library (see Fig. 3, bottom), and then problem specific process models (paracetamol crystallization) are created from the selected generic crystallization process model (methodology illustrated in Fig. 2).

3.3. Design of a PAT system: The problem specific user interface (see Fig. 3, right) is used to design a PAT system for the paracetamol crystallization process. The design procedure consists of 9 hierarchical steps [2, 4] (see Fig. 3, bottom, right)

Step 1. Product property specifications: The desired product is paracetamol with the following predefined qualities: paracetamol concentration: 0.012 g/g; mean crystal size: 100 µm; total crystal mass: 10 g.

Step 2. Process specifications: The basic raw materials required include: Water as a solvent and paracetamol as a solute assuming that the pure paracetamol has been isolated with water during the organic synthesis step. The process equipment used is a jacketed batch crystallizer.

Step 3. Process analysis: The process analysis provides a list of process points and corresponding process variables. A batch crystallizer is the only process point considered in this case and involves the following variables: solute concentration, mean crystal size, temperature, supersaturation, crystal growth rate and nucleation rate.

Step 4. Sensitivity analysis: A sensitivity analysis based on open loop simulations is performed next to identify the variables that need to be monitored and controlled in order to assure the predefined end product quality. The process variable, solute concentration, is here considered as an example for the sensitivity analysis. As shown in fig. 4, the solute concentration profile was found to violate the operational limits, indicating thereby that this variable needed to be monitored and controlled. Repeating this procedure for all variables yielded a list of critical process variables: solute concentration and temperature.

Step 5. Interdependency analysis: Interdependency analysis is performed for each critical process variable identified in Step 4 to select a suitable actuator.

Control of process operations and monitoring of product qualities through generic model-based in batch cooling crystallization
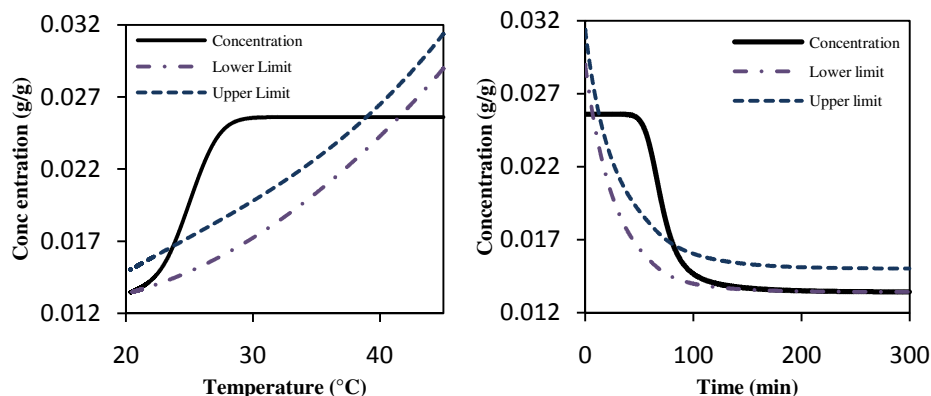
Figure 4. Sensitivity analysis for the solute concentration

As shown in Fig. 5, a critical process variable (solute concentration) and the corresponding actuator candidates (coolant flow rate and inlet coolant temperature) were selected for analysis. The analysis indicated that inlet coolant temperature is more sensitive. Therefore, it was selected as an actuator to control the solute concentration in the batch crystallization (see Fig. 4). Repeating the procedure for all critical control variables yielded the same corresponding actuators (inlet coolant temperature).

Step 6. Performance analysis of monitoring tools: The available monitoring techniques and tools for each identified critical process variable are retrieved from the knowledge base. The performance of these monitoring tools (obtained from the knowledge base) is then compared (based on the selected specifications). ATR-FTIR was selected to monitor the concentration and the temperature is monitored by a thermocouple.

Step 7. Proposed PAT system: A feasible alternative of a PAT system is proposed based on the outcomes of the Steps 3-6. The critical process variables are solute concentration and temperature while the corresponding actuator candidate is inlet coolant temperature. The monitoring tools are ATR-FTIR for monitoring the concentration and thermocouple for temperature monitoring.

Step 8. Validation: A closed-loop simulation is performed to validate the proposed PAT system. This step involves controller configuration, control-monitor verification, sensitivity verification and product properties verification. The solute concentration will be controlled within the concentration set-point trajectory. The temperature will serve as an input to concentration set-point since the set-point trajectory is temperature dependant.
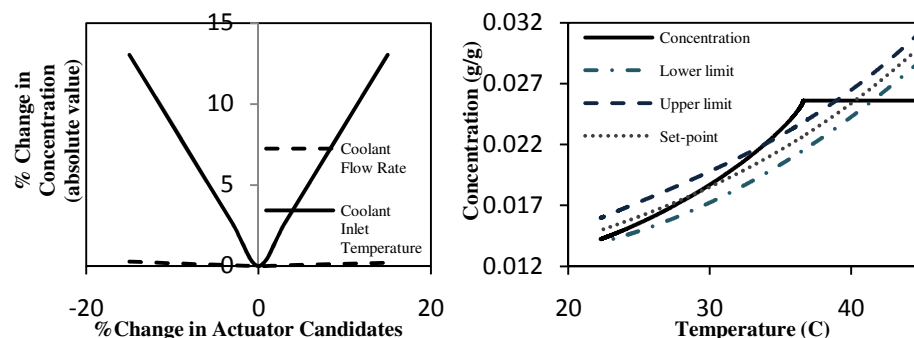
Figure 5. Interdependency analysis and control-monitor verification (closed loop simulation)

Fig. 5 shows the closed loop response (PI controller with 2 inputs) of the solute concentration in the batch crystallizer. It can be concluded that although the solute concentration was beyond the upper limit initially, it stayed within the operational limits by the end of the operation. The concentration profile also satisfied the predefined product qualities specified in Step 1. Fig. 6 shows that the mean crystal size of 96 µm which is closed to design target (100 µm) and the total crystal mass of approximately 9.5 g was also achieved.

Step 9. Final PAT system: A feasible alternative of the PAT system as shown in Fig. 6 was obtained. A cascade control system was used to control the solute concentration. The concentration is monitored by ATR-FTIR and the temperature is monitored by a thermocouple. The inlet water temperature is manipulated by blending hot and cold water.
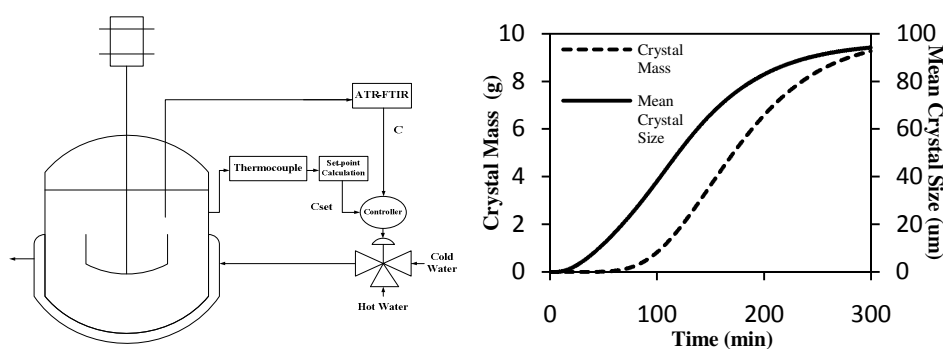
Figure 6. Paracetamol crystallization process flowsheet with designed PAT system and product property verification

4. Conclusions

A generic model for crystallization process/operation has been developed and implemented in the ICAS-PAT software. This generic crystallization process model provided the means to generate the necessary crystallization process operational models for different production scenarios and thereby increased the model (re)usability. The application of the model-based framework in ICAS-PAT is highlighted through the paracetamol case study. The designed process monitoring and control system ensured that the critical process variables are measured and maintained within the operational limits and therefore predefined end product quality can be achieved precisely and consistently.

5. Acknowledgements

References

[1] M. Fujiwara, Z.K. Nagy, J.W. Chew, R.D. Braatz, Journal of Process Control, 15(2005) 493
[2] R. Singh, K. V. Gernaey, R. Gani, Computers & Chemical Engineering, 33(1) (2009) 22
[3] Z.K. Nagy, J.W. Chew, M. Fujiwara, R.D. Braatz, Journal of Process Control, 18(2008) 399
[4] R. Singh, K. V. Gernaey, R. Gani, Computers & Chemical Engineering, (2009),
    doi:10.1016/j.compchemeng.2009.06.021

A simple approach for on-line PI controller tuning using closed-loop setpoint responses

Mohammad Shamsuzzoha[a],*Sigurd Skogestad[a],Ivar J. Halvorsen[b]

[a]*Norwegian University of Science and Technology, Trondheim, Norway,*
*shamsuzz@chemeng.ntnu.no, *skoge@ntnu.no*
[b]*SINTEF ICT, Applied Cybernetics, N-7465 Trondheim, Norway,*

Abstract: A simple and new procedure has been developed for the PI controller tuning of an unidentified process using closed-loop responses. The method requires only one step test in the closed-loop system to obtain the proportional gain and integral time. The step test is a setpoint change performed with a proportional only controller while disabling any integral and derivative action. From the setpoint response one observes the overshoot and the corresponding time to reach the peak. In addition one observes the proportional gain (k_{c0}) and the steady-state offset. Based on a range of first-order with delay test processes, a simple analytical correlation has been developed for the controller gain (k_c/k_{c0}) as a function of the overshoot. The integral time setting is mainly a function of the time to reach the peak. The settings were derived to match the SIMC tuning rule (with $\tau_c=\theta$) which gives good robustness with a gain margin of about 3 and sensitivity peak (M_s-value) of about 1.6. The proposed tuning method, originally derived for first-order with delay processes, has been tested on a broad range of other stable and integrating processes. The results using the closed-loop data are comparable with the SIMC tuning rule using the open-loop model.

Keywords: PI controller, step test, closed-loop, SIMC, Shams's setpoint method

1. Introduction

The proportional integral (PI) controller is widely used in the process industries due to its simplicity, robustness and wide ranges of applicability in regulatory layer. Several papers have reported that a large number of PI controllers are poorly tuned and one reason is that quite tedious plant tests are needed for getting process parameters to finally obtain the appropriated controller setting. The classical method of Ziegler-Nichols [1] has the great advantages of requiring very little information about the process and testing under closed-loop conditions. However, it is well known that the Ziegler-Nichols [1] settings are aggressive for lag dominant (integrating) process and slow for delay dominant process. The other and more significant disadvantage of the Z-N method is that the system is brought at the limit to instability and that a number of trials may be needed to obtain the ultimate gain. An alternative is to induce sustained oscillation by using an on-off controller, i.e. relay tuning (Åström and Hägglund, [5]), but this is a bit difficult to use in practice because one needs to switch to an on/off-controller.

The original IMC-PID tuning method of Rivera et al. [2] and other related direct synthesis [3] methods provide very good performance for setpoint changes but give poor responses for input (load) disturbances in lag time dominant processes. To improve the input disturbance rejection, Skogestad [4] proposed the SIMC tuning rules where the integral time is reduced for lag-dominant (integrating) processes. The SIMC rule has

one tuning parameter, the closed-loop time constant τ_c, and for "fast and robust" control is recommended to choose $\tau_c = \theta$, where θ is the effective time delay. The SIMC tuning rule requires that one first obtains a first-order plus delay model of the process, which involves approximations. Often, an open-loop experiment is used for getting the model parameters which may be time consuming and may upset the process and even lead to process runaway.

Therefore, there is need of an alternative closed-loop approach for plant testing and controller tuning which reduces the number of trails, avoids the instability concern during tuning experiment and works for a wide range of processes. The proposed new method satisfies these concerns:

1. The proposed method requires only a single experimental closed-loop test instead of a trial-and-error procedure under closed-loop condition.

2. The process is not forced to the stability limit, unlike Ziegler-Nichols [1] cycling method.

3. The method is applicable for both integrating and delay dominating process and gives satisfactory disturbance rejection performance.

4. The method is simpler in use than existing approaches and allows the process to be under closed-loop control.

2. SIMC tuning rules

A first-order process with time delay is a common representation of dynamics for process control and is given as:

$$g_{\mathrm{p}} = \frac{ke^{-\theta s}}{\tau s + 1} \tag{1}$$

where k is the process gain, τ the dominant (lag) time constant and θ is the effective time delay and PI controller is given as:

$$u(t) = k_c e(t) + \frac{k_c}{\tau_I} \int_0^t e(t)\,dt \tag{2}$$

k_c is the proportional gain, τ_I is the integral time and the ratio $k_c/\tau_I = K_I$ is known as the integral gain. The SIMC tuning rule (Skogestad, [4]) for the process (1) gives

$$k_{\mathrm{c}} = \frac{\tau}{k(\tau_c + \theta)} \tag{3}$$

$$\tau_I = \min\{\tau,\, 4(\tau_c + \theta)\} \tag{4}$$

This study is based on the "fast and robust" setting $\tau_c = \theta$, which gives a good robustness with a gain margin of about 3 and sensitivity peak (M_s-value) of about 1.6. On dimensionless form, the SIMC tuning rules become

$$k'_{\mathrm{c}} = k_c k = 0.5\tau/\theta \tag{5}$$

$$\tau_I/\theta = \min\left(\frac{\tau}{\theta},\, 8\right) \tag{6}$$

Note that we have scaled time with respect to the delay θ which is approximately the same as the closed-loop time constant (with $\tau_c = \theta$). It is also of interest to consider the integral gain (K_I) on dimensionless form,

$$K_I' = \frac{k_c k}{\tau_I/\theta} = \min\left(0.5,\, \frac{1}{16},\, \frac{\tau}{\theta}\right) \tag{7}$$

The dimensionless gain k'_c and K'_I are plotted as a function of τ/θ in Figure 1.

3. Closed-loop experiment

As mentioned, the objective is to use closed-loop data as basis for the controller tuning. For practical purpose, the simplest closed-loop experiment is a setpoint step response.

A simple approach for on-line controller tuning for closed-loop response
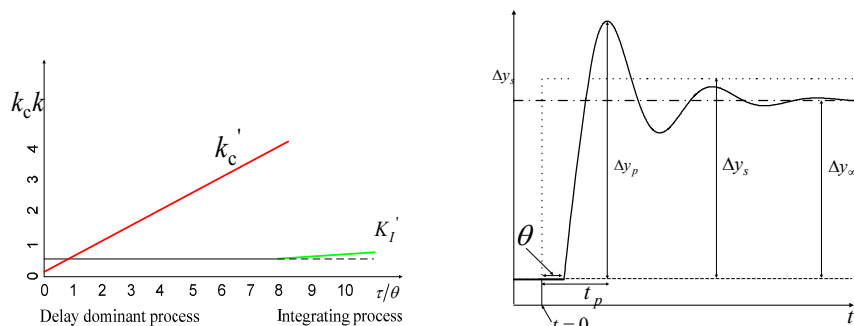
Figure 1. k_c and τ_I for SIMC tuning rule.

Figure 2. Setpoint response with P-control

1. Switch the controller to P-only mode (for example, increases the integral time τ_I to its maximum value or set K_I close to zero). In an industrial system, with bumpless transfer, the switch should not upset the process.

2. Make a setpoint change with an overshoot between 0.10 and 0.60 (about 0.30 is a good value) Most likely, unless the original controller was quite tightly tuned, one will need to adjust (increase) the controller gain to get a sufficiently large overshoot. From the closed-loop setpoint response, see Figure 2, record the following values

Δy_s: Setpoint change; Δy_p: Peak output change; t_p: Time from setpoint change to reach peak output; Δy_∞: Steady-state output change after setpoint step test; k_{c0}: Controller gain used in experiment.

From this data compute the following parameters

$$\text{Overshoot}=\frac{\Delta y_p - \Delta y_\infty}{\Delta y_\infty}, \quad b=\frac{\Delta y_\infty}{\Delta y_s}, \quad kk_{c0} = \frac{b}{1-b} \tag{8}$$

Note that a P-controller is used and (1-b) is the resulting relative steady-state offset. The expression for the overall loop gain (kk_{c0}) is derived from the expression for the closed-loop transfer function, $b = kk_{c0}/(1+kk_{c0})$.

From Figure 1 we note that the integral term (K'_I) is most important for delay dominant processes ($\tau/\theta<1$), but for other processes the proportional term k'_c is most significant. For close-to integrating process ($\tau/\theta>8$), the SIMC rule is to increase the integral term to avoid poor performance (slow settling) to disturbance at the plant input (''load disturbance'').

4. Correlation between setpoint response and SIMC-settings

One could use the closed-loop setpoint response data to first determine the open-loop model parameters (k, τ, θ) and then use the SIMC-rules (or others) to derive PI-settings. A more direct approach is to directly compute from the data the PI-settings as proposed in this study. The goal is then to derive a correlation, preferably as simple as possible, between the setpoint response data (Figure 2) and the SIMC PI-settings in Eq. (3) and (4). For this purpose, we considered 15 first-order with delay models parameterized to cover a range of processes; from time delay dominant to lag-dominant (integrating).

τ/θ=0.1,0.2,0.4,0.8,1.0,1.5,2.0,2.5,3.0,7.5,10.0,20.0,50.0,100.0

For each of the 15 process we obtained the value of k_c and τ_I using the SIMC-setting in Eq. (3) and (4) for $\tau_c=\theta$. Furthermore, for each of the 15 processes we generated 6 step

setpoint responses (Figure 2) using P-controllers that give different fractional overshoots. (Overshoot= 0.10, 0.20, 0.30, 0.40, 0.50 and 0.60)

In total we then have 90 setpoint responses. Note that small overshoots, less than 0.10, were not used. One reason is that it is difficult in practice to obtain from experimental data accurate values of the overshoot and peak time if the overshoot is too small.
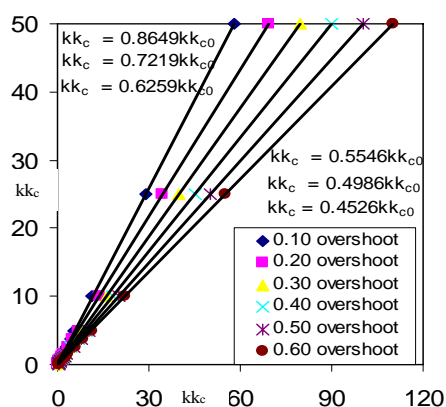
Figure 3. kk_c vs. kk_{c0} for different overshoot   Figure 4. A vs. overshoot
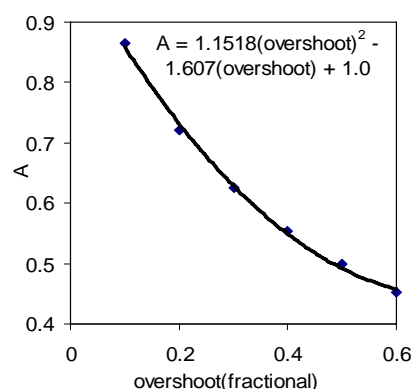
We first seek a relationship for the controller gain k_c. Interestingly, for a fixed value of the overshoot, the ratio k_c/k_{c0} is approximately constant,

$$k_c/k_{c0} = A \tag{9}$$

Note that A is almost independent of the value of τ/θ. This is illustrated in Figure 3 where we plot kk_c (SIMC) as a function of kk_{c0} for the 90 setpoint responses. A is the slope of the line for each overshoot, and is plotted in Figure 4 as a function of the overshoot. The following equation (solid line in Figure 4) fits the data very well,

$$A=[1.152(\text{overshoot})^2 -1.607(\text{overshoot})+1.0] \tag{10}$$

where the correlation is based on data with fractional overshoot between 0.1 and 0.6. Note that a good fit of k_c is not so important for delay-dominant processes ($\tau/\theta<1$), in the lower left corner in Figure 3, where the integral contribution is the most important. Next, we want to find a correlation for the integral time. Since the SIMC tuning formula in Eq. (4) uses the minimum of two values, it seems reasonable to look for a similar relationship, that is, to find one that matches processes with a relatively large delay ($\tau_I = \tau$) and one that works well for integrating process ($\tau_I =8\theta$), and then take the minimum.

First, consider processes with relatively large delay ($\tau/\theta<8$ or $\theta>0.125\tau$), where the SIMC-rule is to use $\tau_I =\tau$. From Figure 1, it is clear that for a delay-dominant process ($\theta>\tau$) the integral term (K_I) is most important. This means that it is particularly important to obtain a good value of $K_I=k_c/\tau$ in this region. In other words, it is not so important that k_c and τ_I are correct individually, but rather that their ratio K_I is close to the SIMC-value. Inserting $\tau=\tau_I$ in the SIMC rule for k_c in Eq. (5) and solving for τ_I gives

$$\tau_I=2kk_c\theta \tag{11}$$

To get K_I correct, we here must use the actual value for the controller gain k_c. From (9) we have obtained the correlation $k_c/k_{c0}=A$, where A is given as a function of the overshoot in Eq. (10). However, we also need the value of the process gain k, and to this effect, write

$$kk_c=kk_{c0}\cdot k_c/k_{c0} \tag{12}$$

Here from Eq. (8), $kk_{c0}=b/(1-b)$ where b is obtained from the steady-state value of the setpoint response. In summary, we have following equation for τ_I for a delay dominant process

$$\tau_I = 2A\frac{b}{(1-b)}\theta \tag{13a}$$

where θ is the effective time delay. Similarly, for a lag-dominant (integrating) process ($\tau > 8\theta$) the SIMC rule gives

$$\tau_I = 8\theta \tag{13b}$$

Equations (13a) and (13b) for the integral time have all known parameters except the effective time delay θ. One could obtain the effective time delay directly from the closed-loop setpoint response, but this may be difficult. Fortunately, as shown in Table 1, there is a good correlation between θ and the peak time t_p which is easier to observe.

Case-a: For processes with a relatively large time delay ($\theta > \tau/8$), the ratio θ/t_p varies between 0.27 and 0.5 (depending on the overshoot and value of τ/θ). We select to use the value $\theta = 0.43t_p$, (note that a large value is more conservative as it increases the integral time). This gives

Process with relatively large time delay: $\tau_I=0.86\,A\dfrac{b}{(1-b)}t_p$ $\hspace{2cm}$ (14a)

Case-b: For a lag-dominant process ($\tau > 8\theta$) we find that θ/t_p varies between 0.25 and 0.36 (depending on the overshoot and value of τ/θ). We select to use the average value $\theta = 0.305t_p$ and get

Integrating process: $\tau_I=2.44t_p$ $\hspace{4cm}$ (14b)

In conclusion, the integral time τ_I is obtained from the minimum of the above two values and becomes

$$\tau_I = \min\left(0.86A\frac{b}{(1-b)}t_p, 2.44t_p\right) \tag{15}$$

5. Analysis and simulation

This section presents only three typical cases to show the effectiveness of the proposed tuning rule. The results of the step test for the controller tuning and corresponding PI setting with M_s value are listed in Table 2. The peak of maximum sensitivity is a measure of robustness and is defined as $M_s=\max \left| 1/[1+g_p g_c(i\omega)] \right|$; a small M_s value indicates that the stability margin of the control system is large.

The resulting closed-loop PI-response for case E2 is shown in Figure 5. A unit step setpoint change is made at t=0 and a unit step change for a load disturbance at the process input is made at t=100. The recommended PI settings vary somewhat with the overshoot as seen in Table 2. From Figure 5, it is clear that the responses are close to those with the SIMC settings. The resulting PI tunings depend on the overshoot used in the experiment and based on the fitting and the results in example processes (Table 2) we recommend that an overshoot around 0.3 is used in practise.

6. Conclusion

The proposed tuning method is based on a single closed-loop setpoint step test using a P-controller. The PI-controller settings are then obtained directly from three characteristic numbers from the setpoint step test: The overshoot, the time to the first peak t_p and the relative steady state change b. The tuning formulas for the proposed **"Shams's setpoint method"** method are:

$$k_c = k_{c0}A/F; \quad \tau_I = \min\left(0.86A\frac{b}{(1-b)}t_p, 2.44t_p F\right)$$

where A=[1.152(overshoot)^2 -1.607(overshoot)+1.0]

The above settings give a robust and reasonable fast response for F=1 (corresponding to a closed-loop time constant $\tau_c=\theta$ in the original SIMC method). If one wants to detune the controller to get a smoother response with more robustness and less input usage then one may introduce the detuning parameter F>1. One may in some cases choose F<1 to speed up the response but the system will then be less robust.

The new method works for a wide variety of the processes, except unstable and highly oscillating system. The novelty of the proposed method is that only one experiment in closed-loop is sufficient for getting significant information for controller tuning. We believe that it could be the simplest and easiest approach for PI controller tuning to use in process industries.

Table 2: PI controller setting for proposed and SIMC method

Case	Process model	k_{c0}	Over-shoot	t_p	b	k_c	τ_I Eq.(14)	$K_I = k_c/\tau_I$	M_s
E1	$\dfrac{0.2e^{-7.4s}}{s}$	0.40	0.110	29.33	1.0	0.338	71.58(b)	-	1.68
		0.55	0.315	24.09	1.0	0.338	58.77(b)	-	1.70
		0.75	0.610	21.47	1.0	0.342	52.38(b)	-	1.74
		SIMC	-	-	-	0.338	59.2	-	1.70
E2	$\dfrac{(-s+1)e^{-s}}{(6s+1)(2s+1)(2s+1)}$	0.70	0.119	16.63	0.412	0.583	8.33(a)	-	1.42
		1.30	0.311	13.93	0.565	0.805	9.65(a)	-	1.57
		2.2	0.610	12.27	0.687	1.00	10.60(a)	-	1.74
		SIMC	-	-	-	0.70	7.0	-	1.63
E3	$\dfrac{e^{-s}}{(0.05s+1)^2}$	0.10	0.10	1.99	0.091	0.086	0.147(a)	0.59	1.70
		0.30	0.30	1.99	0.231	0.189	0.321(a)	0.59	1.62
		0.60	0.60	2.0	0.375	0.275	0.473(a)	0.58	1.65
		SIMC	-	-	-	0.037	0.075(a)	0.49	1.60
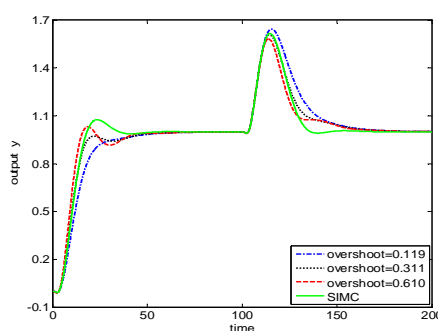
Table: 1 variation of θ/t_p with τ/θ and overshoot.

τ/θ	θ/t_p Overshoot =0.1	θ/t_p Overshoot =0.3	θ/t_p Overshoot =0.6
0.1	0.50	0.50	0.50
1.0	0.36	0.41	0.44
8.0	0.27	0.32	0.36
100	0.25	0.30	0.34

Figure. 5 Response for case E2 process

7. References

[1]. Ziegler, J. G.; Nichols, N. B. Optimum Settings for Automatic Controllers. *Trans. ASME* (1942), *65*, 433-444.

[2]. Rivera, D. E., M. Morari and S. Skogestad; Internal Model Control. 4. PID Controller Design, *Ind. Eng. Chem. Process Des. Dev.*, (1986), 25, 252–265.

[3]. Seborg, D. E., T. F. Edgar and D. A. Mellichamp; Process Dynamics and Control, 2nd ed., John Wiley & Sons, New York, U.S.A.,2004.

[4]. Skogestad, S.; Simple Analytic Rules for Model Reduction and PID Controller Tuning, *J. Process Control,* (2003), 13, 291–309.

[5].Åström K.J. and Hägglund, T., Automatic tuning of simple regulators with specifications on phase and amplitude margins, *Automatica,* (1984), 20, pp. 645–651.

Multi-Scale Modelling And Coarse-Grained Analysis Of Triglycerides Dynamics

Brasiello A.[a], Russo L.[b], Siettos C.[c], Milano G.[d], Crescitelli S.[e]

[a] *Dipartimento d'Ingegneria Chimica ed Alimentare, Università di Salerno, via Ponte don Melillo, 84084, Fisciano (SA), Italy, abrasiello@unisa.it*
[b] *Istituto Ricerche sulla Combustione, CNR, Piazzale V. Tecchio 80, 80125, Napoli, Italy, lucrusso@unina.it*
[c] *School of Applied Mathematics and Physical Sciences, National Technical University of Athens, GR 157 80, Athens, Greece, ksiets@mail.ntua.gr*
[d] *Dipartimento di Chimica, Università di Salerno,via Ponte don Melillo, 84084, Fisciano (SA), Italy, gmilano@unisa.it*
[e] *Dipartimento d'Ingegneria Chimica, Università di Napoli „Federico II", Piazzale V. Tecchio 80, 80125, Napoli, Italy, crescit@unina.it*

Abstract

The paper is focused on Molecular Dynamics (MD) simulations of liquid triglycerides. Triglycerides are the major constituents of edible lipids and therefore represent a key topic in the field of medical science and food engineering. The main drawback of using MD for any practical application is the computational cost needed for long run simulations. Here we present a first effort to overcome this drawback by considering the synergy of two approaches: (a) a coarse-grained modelling of triglycerides molecules which allows the reduction of the equivalent simulation time for more than one order of magnitude with respect to the atomistic simulations; (b) the so-called coarse time-stepper approach (Kevrekidis et al., 2003), which can be used to accelerate the atomistic time evolution computations directly.

We show how, using a coarse-grained model, one may reach simulation times of one order of magnitude lower with respect to the atomistic simulations. Moreover, applying the time stepper approach we are able to reduce the simulation time one more order of magnitude, giving an overall gain of two orders of magnitude.

Keywords: Molecular dynamics, multi-scale modelling, time- stepper, triglycerides

1. Introduction

Triglycerides are important constituents of oils and fats. Triglycerides are tri-esters of glycerol with fatty acid and are widely used in foods, cosmetics, and medicine as nutrients or matrix materials. In medicine, for example, triglycerides can be profitably used as carrier for controlled release drugs (Pardeike et al., 2009). In this field it is important to understand structural properties of the solid material and transport properties of drugs in liquid (Yuan et al., 2007). Both the issues depends on molecules conformations. Triglycerides can in fact crystallize according to several solid shapes called polymorphs, identified and named α, β, β' phases (Wille and Lutton, 1966). They differ from each other in molecules conformations and crystal packing. The α polymorph, the least stable, is characterized by a hexagonal cell packing, the *β'* polymorph by an orthorhombic cell packing, while the *β* polymorph, the most stable, by a triclinic cell packing. Melting points grow from the least to the most stable. Each

crystal packing is characterized by precise molecular conformations, which could strongly affect the quality of products (Himawan et al., 2006). The relationship between crystal packing and molecules conformations is highlighted in several experimental works (e.g. Yano and Sato, 1999). According to these works, the *α* polymorph is characterized by tuning-fork conformation while *β* and *β'* by chair conformations. Bunjes et al. (1996) demonstrate that liquid triglycerides also hold some characteristics of the conformational arrangement found in crystals. From this, it is clear that also concerning with liquid triglycerides, the conformational arrangement cannot be left out of consideration. The understanding of the effects of macroscopic process conditions on conformational properties of triglycerides at the molecular level is then the key topic to improve industrial managements and can be pursued using microscopic simulations. In the last years, efforts in this direction have been made in the field of molecular dynamics simulations of triglycerides. For example, Chandrasekhar and Van Gusteren (2002) analyze GROMOS96 united-atoms force field developed for aliphatic alkanes with which simulate, for few nanoseconds, a box of trioctanoin molecules in α phase. It is an approximated atomistic model in which a methyl or a methylene group is represented by a single interaction site. The main drawback of such molecular dynamics of triglycerides is the huge computational costs in term of calculation time. To overcame this drawback a coarse-grained model was exploited using the concept of coarse time-stepping (Kevrekidis et al., 2003).

The construction of the new coarse-grained force field starts from the definition of a more coarse representation of triglycerides molecules. We study tripalmitin molecules as they are quite long chains molecules commonly present in natural fats and used in medical applications. Model parameters are obtained through a suitable optimization procedure based on the comparison of suitable statistical functions related to the atomistic model and the coarse-grained-one.

The coarse-grained model is able to perform a less detailed simulation but for a longer time. The time stepper technique relies on the developed coarse-grained model. It is a projective multi-step iterative numerical integration algorithm which links detailed MD or atomistic simulations with macroscopic dynamics. The coarse time-stepper approach treats the detailed atomistic simulator as a black box input-output map of the coarse-grained observables, sidestepping the need of obtaining explicit coarse-grained models; the approach accelerates the simulations through appropriately initialized, short "bursts" of atomistic MD dynamics simulations (Kevrekidis et al., 2003). In our case we choose density as the macroscopic parameter since it is strictly related to the whole force field and thus represents a key parameter for model validation. We show that the developed model is able to reproduce quite accurately physical properties experimentally derived and to provide an insight into molecular conformational trends.

2. Model development

2.1. Coarse-grained model development

For the purpose of coarse-grained model force field development, GROMOS96 united-atoms model will be considered as an optimal approximation of the real system, since it provide good results in triglycerides properties calculations (Chandrasekhar and Van Gusteren, 2002). In order to built a more coarse particles system representing quite faithfully the united-atoms model, we define two coarse bead type. The first type named C type particle represents three united-atoms interaction sites of methylene/methyl type. The second type named N type represent a $CH_n - COO-$ group of the united-atoms

model relative to the glycerol backbone and the carboxylic group. The equation of motion solved during a coarse-grained molecular dynamics reads:

$$m_i \frac{\partial^2 \mathbf{r}_i}{\partial t^2} = -\frac{\partial U}{\partial \mathbf{r}_i} \qquad (1)$$

The right-hand side of (1) represents the coarse-grained force field and it is given by:

$$U = \underbrace{\sum_{\text{all bonds}} \frac{k_{ij}^b}{2}\left(r_{ij} - r_{ij}^{eq}\right)^2 + \sum_{\text{all angles}} \frac{k_{ijk}^\theta}{2}\left(\cos\left(\theta_{ijk}\right) - \cos\left(\theta_{ijk}^{eq}\right)\right)^2}_{U^{\text{inter}}} + \underbrace{\sum_{i>j} 4\varepsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^6\right]}_{U^{\text{intra}}} \qquad (2)$$

The coarse-grained force field parameters to be determined are $k_{ij}^b, r_{ij}^{eq}, k_{ijk}^\theta, \theta_{ijk}^{eq}$ for the intermolecular harmonic type potentials and $\varepsilon_{ij}, \sigma_{ij}$ for the intramolecular Lennard-Jones type potentials. Intermolecular potentials take into account bond and angle interactions. No torsional interaction are considered in order to reduce further the computational cost. Intramolecular potentials take into account Lennard-Jones interactions. Force field parameters are obtained through an optimization procedure by matching suitable statistical distributions and radial distribution functions of the coarse beads with those related to the corresponding centers of mass in the atomistic model. Complete discussions on model hypotheses, computational details and whole force field parameters values are reported elsewhere (Brasiello, 2009).

The model validation is performed on the basis of two macroscopic properties at several temperatures: density, as it is a static property deriving from all force field interactions and self-diffusion, determined through the calculation of mean squared displacement as it is strictly related to motion properties of the whole system (Allen and Tildesley, 1987). Some properties calculations at several temperature are reported and compared with experimental data in table 1. As can be observed, numerical data are in good agreement with the experimental ones in the range of temperatures of interest, the maximum deviation lying within a range of about 6% respect to experimental data.

Table 1. Physical properties calculation: comparison between experimental and numerical data

T (K)	Density (SI)		Self-Diffusion (cm^2/s)	
	Numerical	Experimental	Numerical	Experimental*
373	868.92	854	$1.62 \cdot 10^{-5}$	$4.0 \cdot 10^{-5}$ (available at
353	895.16	868	$1.33 \cdot 10^{-5}$	343K)
339	912.73	877	$1.19 \cdot 10^{-5}$	*Caboi et al. (2005)

2.2. The coarse time-stepper approach

The main assumption behind the methodology is that a coarse-grained model for the fine-scale dynamics in principle exists and closes in terms of a few coarse-grained variables (observables), but due to the overwhelming complexity is not analytical available in a closed form. Typically these few coarse-grained variables are low-order moments of microscopically evolving distributions, and the –in principle- existence of a coarse-grained model implies that the higher order moments of the distributions become, relatively quickly over the time scales of interest, "slaved" to the lower, few, "master" ones. What the Equation-Free approach does, in fact, is providing a closure on demand ("just in time"); relatively short bursts of the fine scale simulator naturally establish this slaving relation (Kevrekidis et al., 2003; Siettos et al., 2003).

The computation methodology consists of the following steps:

(a) Choose the statistics of interest for describing the long-term behavior of the system and an appropriate representation for them. For example, this could be the density and velocity fields. We call this continuum description **u**. These choices determine a restriction operator M from the individual-based description **U** to the continuum description **u**= M **U**.

(b) Choose an appropriate lifting operator μ from the continuum description **u** to the individual based description **U**. For example, μ could make random position and velocity assignments consistent with the continuum statistics. Note that μ M =I, that is, lifting from the continuum to the individual-based description, then restricting down again has no effect, apart from round off effects.

(c) Prescribe a continuum initial condition **u**(t_o).

(d) Transform this initial condition through lifting to one (or more) consistent individual-based realizations **U**(t_o)=μ**u**(t_o).

(e) Evolve this realization using the individual-based model for a desired time T, generating the value **U**(T).

(f) Obtain the restrictions **u**(T)= M **U**(T).

This constitutes the *coarse time-stepper*, or *coarse time-T map*.

If the coarse time-stepper is accurate enough, we can use it to perform coarse projective integration. The basic idea is that coarse time-stepper can be used to approximate the time derivatives of the corresponding continuum formulation, even if the continuum equations are not known in closed form. Specifically, we execute the following steps:

(g) Repeat step **(f)** over several time steps, giving several **U**(t_i), as well as their restrictions **u**(t_i)= M **U**(t_i), i=1,2,…,k+1.

(h) Use the chord connecting these successive time-stepper output points to estimate the derivative of the continuum variables. Note that this does *not* require that we know the explicit continuum equations.

(i) Use this derivative in an outer integrator (such as forward Euler) to estimate the continuum state **u**(t_{k+1+m}) much later in time.

(j) Go back to step **(d)**.
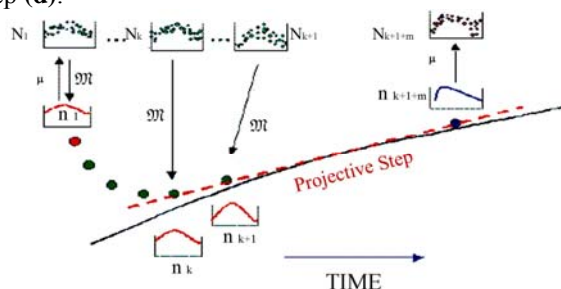
Figure 1 Schematic representation of projective integration (Kevrekidis et al., 2003).

It is important to note that one must integrate the individual-based equations for some time before estimating the time derivative of the continuum variables. This allows higher moments of the continuum description to become slaved to the statistics of interest. See Figure 1 for a schematic representation of this computational procedure. The relation is used in the propagation step to project density in time. According to the new density value a reconstruction of the molecular box is carried out and a new coarse-grained simulation is performed and so on. The length of density propagation is strictly related to the length of the coarse-grained simulation. The coarse-grained simulation on the other side should last for a time during which density fluctuations around the mean

value became irrelevant. In our case, a coarse-grained simulation lasting *3 ns* should support a density propagation of *7 ns*. Box reconstruction constitutes a serious issue concerning the coarse timestepper procedure as wrong reconstructions could cause the coarse-grained simulations to fail because of high internal forces generated by non-equilibrated system. For this reason, starting from the last generated coarse-grained box, the reconstruction consist in a series of three cycles of box dimension modification, minimization and NVE coarse-grained simulation, until internal forces reaches order of magnitude of the previous coarse-grained simulation.

3. Temperature annealing

As a key study, we consider a temperature annealing simulation. The objective is to analyze changes in conformational trends consequent to a linear temperature reduction.
In Figure 2 the temporal evolution of density is shown. Temperature is decreased linearly from *335 K* to *200 K* in a time interval of *400 ns* using the time stepper algorithm. In this case, molecular dynamics simulations lasting *3 ns* are performed and a propagation step of *7 ns* is adopted.
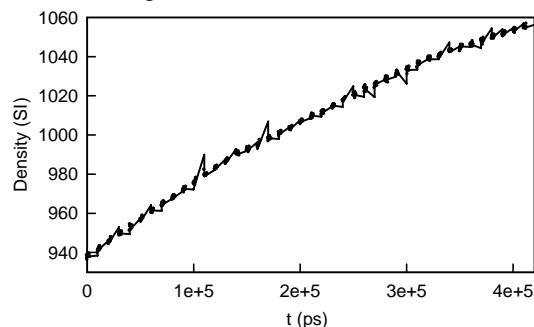
Figure 2 Temporal evolution of density due to linear reduction of temperature. Molecular dynamics are indicated by dots while time stepper operations are indicated by solid lines.

Time stepper procedure allows a significant computational time reduction. In the case of Figure 2, the coarse-grained simulation with time stepper lasting *400 ns* is performed in one day real time. This is a good result if compared with *120ns/day* of a simple coarse-grained simulation or with *1.18 ns/day* of a united-atoms simulation of the same system. Better results can be also obtained adopting further optimization procedures to fix the molecular simulation time and propagation one.
Despite the question concerning simulation time, time stepper algorithm allows the investigation of structural properties of the system occurring on time scale of hundreds or thousands of nanosecond, as for example, the changes in conformational trends of molecules due to temperature annealing. In Figure 3, for example, the effect of temperature annealing of the performed simulation on the molecular distribution is highlighted through the radial distribution functions of triglyceride molecules at *335 K* and *200 K*. In particular, the radial distribution function at *200 K* shows a higher peak at about *0.5 nm* and a secondary peak at about *1nm*. This is a clear indication of the trend of a system moving towards more ordered conformational arrangements.

4. Conclusions

In this paper, the synergy of two techniques was adopted to overcome computational burden: a coarse-grained model and the coarse timestepping scheme. The development of the coarse-grained model was carried out through the comparison between suitable
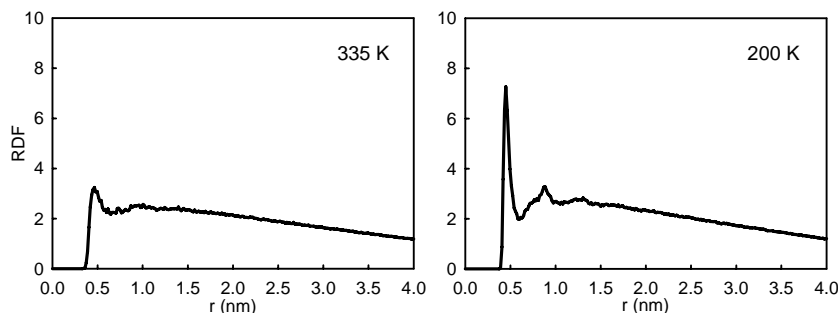
Figure 3 Comparison between radial distribution functions of the tripalmitin system at *335 K* and *200 K*.

statistical functions derived from the coarse-grained model and from a united-atoms model adopted in literature. The developed coarse-grained model is able to predict with good accuracy physical properties of liquid triglycerides. The concept of the coarse timestepper was exploited to accelerate simulation in molecular conformational trends during temperature annealing. As an example, the case of tripalmitin was studied showing system tendency to go towards more ordered conditions. We are currently investigating the possibility of analyzing, with the proposed procedure, a crystal formation from pure liquid.

References

Allen, Tildesley,1987, Computer Simulation of Liquids, Clarendon Press – Oxford.

Brasiello A., 2009, Molecular dynamics of triglycerides: atomistic and coarse-grained approaches, Ph.D. Thesis, Fedriciana Editrice Universitaria Dottorati.

Bunjes H., K. Westesen, M. H. J. Koch, 1996, Crystallization tendency and polymorphic transitions in triglyceride nanoparticles, *International Journal of Pharmaceutics*, 129, 159-173.

Caboi F., P. Lazzari, L. Pani, M. Monduzzi, 2005, Effect of 1-butanol on the microstructure of lecithin/water/tripalmitin system, *Chemistry and Physics of Lipids*, 135, 147–156.

Chandrasekhar I. and W.F. van Gusteren, 2002, A comparison of the potential energy parameters of aliphatic alkanes: Molecular dynamics simulations of triacylglycerols in the alpha phase, *Eur. Biophys. J.*, 31, 89–101.

Himawan C., V.M. Starov, A.G.F. Stapley, 2006, Thermodynamic and kinetic aspects of fat crystallization, *Advances in Colloid and Interface Science,* 122, 3–33.

Kevrekidis I. G., Gear C. W., Hyman J. M., Kevrekidis P. G., Runborg O., Theodoropoulos C., 2003, Equation-free coarse-grained multiscale computation: enabling microscopic simulators to perform system-level tasks, *Comm. Math. Sciences*, 1(4), 715-762.

Pardeike J., A. Hommoss, R. H. Müller, 2009, Lipid nanoparticles (SLN, NLC) in cosmetic and pharmaceutical dermal products, *International Journal of Pharmaceutics*, 366, 170–184.

Siettos C. I., Graham M., Kevrekidis I. G., 2003, Coarse Brownian Dynamics for Nematic Liquid Crystals: Bifurcation Diagrams via Stochastic Simulation, *J. Chem. Phys.*, 118, 10149-10156.

Wille R.L., E.S. Lutton, 1966, Polymorphism of Cocoa Butter, J. Am. Oil Chemists'Society, 43, 491–496.

Yano J. and K. Sato, 1999, FT–IR Studies on Polymorphism of Fats: Molecular Structures and Interactions, *Food Res. Int.*, 32, 249–259.

Yuan H., LL. Wang, YZ. Du, J. You, FQ. Hu, S. Zeng, 2007, Preparation and characteristics of nanostructured lipid carriers for control-releasing progesterone by melt-emulsification *Colloids and Surfaces B: Biointerfaces,* 60, 174–179.

POD reduced order dynamical model of non-isothermal Circulating Fluidized Bed Combustor

Katarzyna Bizon[a], Gaetano Continillo[b]

[a]*Istituto di Ricerche sulla Combustione – C.N.R., Via Diocleziano 328, 80124 Napoli, Italy, katarzyna.bizon@irc.cnr.it*
[b]*Dipartimento di Ingegneria, Università del Sannio, Piazza Roma 21, Benevento, Italy, continillo@unisannio.it*

Abstract

This paper presents the spectral reduction of the 1-D distributed dynamic model of a non-isothermal Circulating Fluidized Bed Combustor (CFBC). The continuum model is first approximated by a finite-difference method and integrated in time by means of Adams-Moulton method to provide a "reference" solution. Then, Proper Orthogonal Decomposition (POD) is introduced, coupled with a spectral penalty method, to derive a reduced order model (ROM). The penalty method deals with the variable boundary conditions which characterize the original model. The numerical performance of the proposed approach is evaluated.

Keywords: Fluidized Bed Combustors Modeling, Model Reduction, Proper Orthogonal Decomposition, Spectral Penalty Method

1. Introduction

Empirical methods such as Proper Orthogonal Decomposition (POD) have been successfully implemented for model reduction of dynamical systems described by partial differential equations. POD delivers an optimal set of orthogonal basis functions from the observation of the system, and coupled with spectral method allows to reduce considerably the order of the PDE's system. This property is highly desirable in view of the dynamical analysis of chemically reactive systems, which usually demand a large amount of numerical computations.

In this work, an empirical orthogonal set of basis functions obtained by means of POD is used to build a reduced order model of dynamic non-isothermal Circulating Fluidized Bed Combustor (CFBC) for solid fuel combustion. In order to emphasize the dynamical aspects of the model, resulting from the high complexity of CFBC system, the model details are kept to a minimum. The model integrates simplified hydrodynamics treated as steady-state problem, fuel combustion and attrition, and heat transfer – both internal (between fuel/inert particles and gas phase) and external (i.e. heat transfer to the wall). In order to account for time-variable boundary conditions, resulting from the solid recycle, a specific treatment based on the modification of the original Galerkin spectral method is used, namely spectral penalty method.

2. Circulating Fluidized Bed Combustor model

The Circulating Fluidized Bed Combustor system is idealized as a 1-D distributed, dynamical plug flow tubular reactor [1-3], operated in non-isothermal conditions, followed by an external gas-solid separator (cyclone) and continuously stirred tank reactor (CSTR) to loop-seal simulation (Figure 1).

Under the assumption that fluid dynamic changes occur much faster than thermal or concentration changes, fluid dynamics is treated as a steady-state problem. The fluid dynamic model reported in [3] is used to calculate voidage profile and gas and solid velocities. Both volatile matter and ash content is neglected in the present formulation. Population balance equation on char is simplified by lumping fixed carbon into two classes: coarse and fine char. The coarse phase is depleted both by combustion and attrition, whereas the fine phase is enriched by attrition and depleted by combustion. Combustion is modeled as a single one-step heterogeneous reaction ($C+O_2\rightarrow CO_2$).

Mass balance equations, in dimensionless form, for the coarse phase, fed at the bottom into the riser, and the fine phase, are given respectively by:

$$\frac{\partial \alpha_c}{\partial \tau} + \frac{\partial (v_c \alpha_c)}{\partial \zeta} = -\sigma_a \alpha_c - \sigma_c \alpha_c$$

$$\frac{\partial \alpha_f}{\partial \tau} + \frac{\partial (v_f \alpha_f)}{\partial \zeta} = \sigma_a \alpha_f - \sigma_f \alpha_f \qquad (1)$$

The mass balances for gas phase, for O2 and CO_2 are:

$$\frac{\partial (\varepsilon \alpha_{O_2})}{\partial \tau} + \frac{\partial (v_g \varepsilon \alpha_{O_2})}{\partial \zeta} = -k_1 (\sigma_c \alpha_c + \sigma_f \alpha_f)$$

$$\frac{\partial (\varepsilon \alpha_{CO_2})}{\partial \tau} + \frac{\partial (v_g \varepsilon \alpha_{CO_2})}{\partial \zeta} = k_1 (\sigma_c \alpha_c + \sigma_f \alpha_f) \qquad (2)$$

The energy balances for solid phases and gas phase are:

$$\frac{\partial (\alpha_c \theta_c)}{\partial \tau} + \frac{\partial (v_c \alpha_c \theta_c)}{\partial \zeta} = \beta \sigma_c \alpha_c - \phi_{c-g} h_c \alpha_c (\theta_c - \theta_g) - \sigma_a \alpha_c \theta_c$$

$$\frac{\partial (\alpha_f \theta_f)}{\partial \tau} + \frac{\partial (v_f \alpha_f \theta_f)}{\partial \zeta} = \beta \sigma_f \alpha_f - \phi_{f-g} h_f \alpha_f (\theta_f - \theta_g) + \sigma_a \alpha_c \theta_c$$

$$\frac{\partial ((1-\varepsilon) \theta_b)}{\partial \tau} + \frac{\partial (v_b (1-\varepsilon) \theta_f)}{\partial \zeta} = -\phi_{b-g} h_b (1-\varepsilon)(\theta_b - \theta_g) - \phi_{b-w} h_w (\theta_b - \theta_w) \qquad (3)$$

$$\frac{\partial (\varepsilon \alpha_g c_{pg} \theta_g)}{\partial \tau} + \frac{\partial (v_g \varepsilon \alpha_g c_{pg} \theta_g)}{\partial \zeta} = k_3 \phi_{c-g} h_c \alpha_c (\theta_f - \theta_g) + k_3 \phi_{f-g} h_f \alpha_f (\theta_f - \theta_g)$$

$$+k_3 \phi_{b-g} h_b (1-\varepsilon)(\theta_b - \theta_g)$$

Mass balances for the CSTR in the recycle loop, for coarse and fine, are given respectively by:

$$\frac{\partial \alpha_{c,rec}}{\partial \tau} = k_2 \left[\eta_c v_c (1,\tau) \alpha_c (1,\tau) - F_{rec} \alpha_{c,rec} \right]$$

$$\frac{\partial \alpha_{f,rec}}{\partial \tau} = k_2 \left[\eta_f v_f (1,\tau) \alpha_f (1,\tau) - F_{rec} \alpha_{f,rec} \right] \qquad (4)$$

POD reduced order dynamical model of non-isothermal CFBC
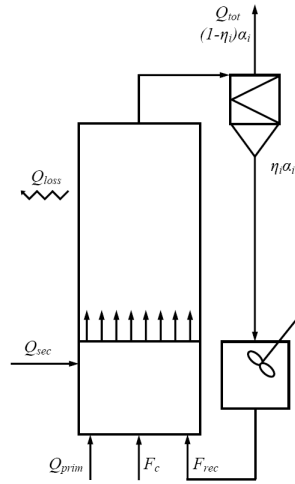
Figure 1. Scheme of the idealized CFBC

Table 1. Main parameter values of CFBC model

Riser height	H [m]	10
Riser diameter	D [m]	1
Pressure	P [kPa]	101.325
Gas velocity @293K	u_0 [m/s]	1.5
Lambda	λ [-]	1.2
Inert particle diameter	d_b [μm]	300
Inert material density	ρ_b [kg/m^3]	2600
Coarse diameter	d_c [mm]	3
Fine diameter	d_f [μm]	100
Char density	ρ_c [-]	1500
Cyclone efficiency for coarse	η_c [-]	1
Cyclone efficiency for fine	η_f [-]	0.9
Slip multiplier for coarse	γ_c [-]	0.5
Slip multiplier for fine	γ_f [-]	1

3. Mathematical theory

3.1. Proper Orthogonal Decomposition

The objective of the POD approach is to determine an empirical set of optimal orthogonal functions based on the spatiotemporal simulation data $u_t(x)$ [4]. Sampled data from the simulation can be represented in the matrix form as:

$$U = \begin{bmatrix} u_1(x_1) & u_2(x_1) & \cdots & u_M(x_1) \\ u_1(x_2) & u_2(x_2) & \cdots & u_M(x_2) \\ \vdots & \vdots & \ddots & \vdots \\ u_1(x_N) & u_2(x_N) & \cdots & u_M(x_N) \end{bmatrix} \tag{5}$$

where N is the number of positions in the spatial domain, and M is the number of samples taken in time. The POD basis $\{\varphi_i\}_{i=1}^{N}$ can then be obtained by solving the eigenvalue problem:

$$C\varphi = \lambda\varphi \qquad \text{where} \qquad C = \langle U, U^T \rangle. \tag{6}$$

When $N>M$ it is more convenient to use the so-called method of snapshots proposed by Sirovich [5], which transforms the original eigenvalue problem of the dimension of N into a computationally less expensive M-dimensional eigenvalue problem.

3.2. Spectral penalty method

In a classical spectral approach it is required that the projection of the residual onto basis be orthogonal [6], i.e.

$$\int_{\Omega} R_N(x,t)\,\varphi_m(x) = 0 \tag{7}$$

Hence, for the general problem in the form:

$$\frac{\partial u}{\partial t} = Du(x,t) \tag{8}$$

with D being a non-linear operator that involves spatial derivatives of the dependent variable, after the substitution of u by the spectral approximation of order K, i.e.:

$$u_K(x,t) = \sum_{n=0}^{K} a_n(t)\varphi_n(x) \tag{9}$$

and making use of the orthogonality of the basis functions, one gets:

$$\frac{\partial a_m}{\partial t} = D(a_n, \varphi_n, \varphi_m) \quad n,m = 1,\ldots,K \tag{10}$$

where index n corresponds to the expansion coefficients and functions, while index m correspond to the POD function on which the equation is projected.

In order to account for a variable or non-homogenous boundary condition $g(t)$, the problem defined by Eq. (8) is approximated by employing the so-called penalty method proposed in Reference [7]:

$$\frac{\partial u}{\partial t} = Du(x,t) - \tau\left[u_K(x,t) - g(t)\right] \tag{11}$$

where τ is the penalty parameter. Then, similarly as in the classic Galerkin approach, performing the orthogonal projection of the approximated system leads to the ODE's in the form (10).

4. Results

A finite difference method with staggered grid, employing 500 spatial nodes was used for the approximation of the PDE's system. Then, the Adams-Moulton implicit method was used to solved resulting ODE's, together with 2 ODE's for the CSTR in the solid recycle loop, in order to build the reference solution and collect data for the determination of the POD basis for model reduction.
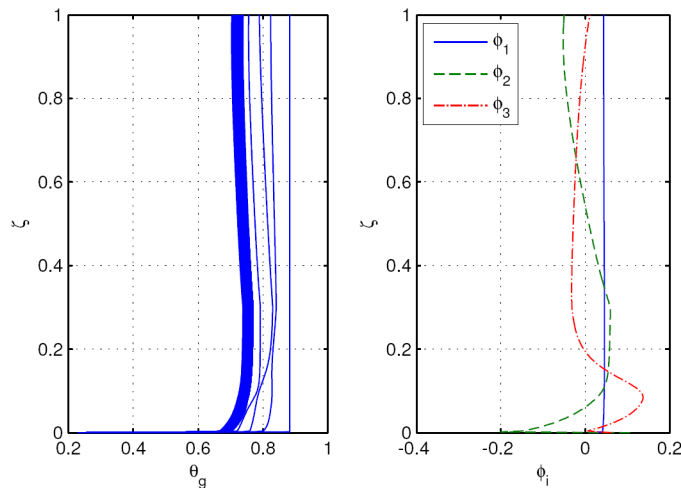
Figure 2. Set of snapshots collected for gas temperature (a) and its three leading POD modes (b).

POD reduced order dynamical model of non-isothermal CFBC

A set of equally spaced snapshots of gas temperature from transient and steady state, together with the corresponding POD modes (3 leading modes) are reported in Figure 2. It has to be underlined that separate sets of POD basis functions were determined for each state variable, using appropriate snapshots.

Then, the reduced order model was determined by using the spectral penalty method, and the resulting ODE's for modal coefficient were integrated in time, again using the Adams-Moulton method as done with the full model. Results from the reduced order models ROM1, ROM2 and ROM3, obtained by projection of the PDE's onto 1, 2 and 3 POD modes respectively (for each state variable), are presented in Figure 3 and Figure 4. Figure 3 reports the comparison of the concentration of the coarse char particles along the riser both in the early transient and at steady-state obtained with the full order model (FOM) and the reduced order models. Similarly, in Figure 4 FOM and ROM solutions of the corresponding temperature of particles are compared.
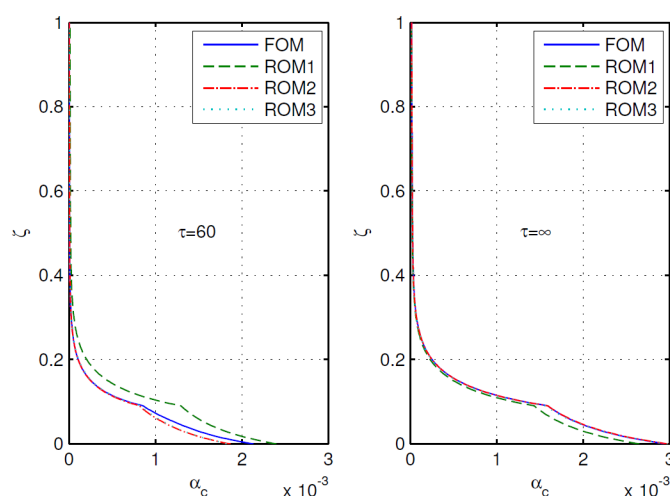
Figure 3. Comparison of vertical concentration profile of coarse char in early transient (a) and in steady-state (b) (FOM – full order model, ROM – reduced order model).

It can be seen that reduced order models are not able to predict correctly the transient behavior of the system. ROM1, i.e. the model obtained by projection onto 1 POD mode, underestimates significantly the temperature of the coarse (Figure 4), which results in higher, on average, values of the concentration of coarse along the riser (Figure 3). On the other hand, ROM2 and ROM3 overestimate the temperature along the riser: this however does not influence so much the concentration profiles which, in this case, follow quite closely the full order solution, even in the early transient. In fact, despite the large error in the approximation of the temperatures in the upper zone of the riser, temperature approximation in the dense bed, where mostly the combustion process proceeds, is improved significantly by introducing higher order modes. The accuracy of the steady-state solution obtained by employing all determined ROMs is acceptable, obviously higher for higher order ROMs i.e. ROM2 and ROM3.

Despite its complexity (many non-linear terms lead to complication of the spectral scheme), the constructed ROM's still give significant computational savings. Namely, while the simulation of the full order model takes around 100,000 s, this time drops to 3,000 s with the reduced order model. These results are related to the model projected onto only one POD mode (per state variable), composed of 7 ordinary differential

equations (mass balances for coarse, fine, and CO_2, and energy balances for coarse, fine, bed inert, and gas). Computation savings can be large even with more modes, when the model only consists of mass balance equations with at most bilinear terms (i.e. isothermal CFBC model). In this case, for an accurate ROM with 24 ODE's the computation saving is still of more than three orders of magnitude [2].
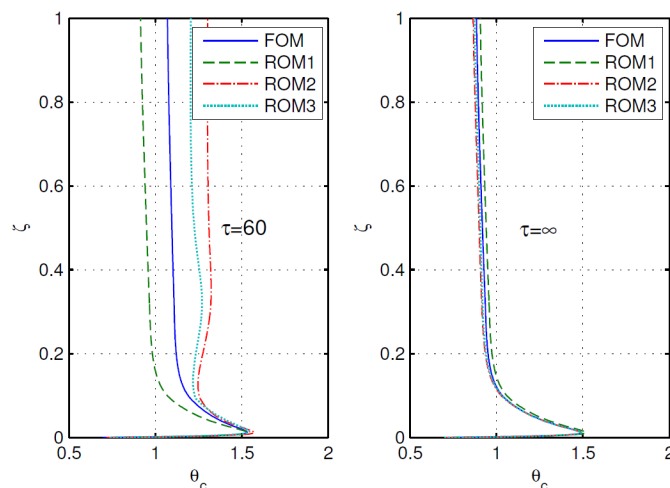
Figure 4. Comparison of vertical temperature profile of coarse char in early transient (a) and in steady-state (b).

5. Conclusions

Application of POD method coupled with spectral penalty method for reduction of CFBC model has been reported. We have shown that the non-isothermal complex non-linear system can be successfully projected onto only one POD mode for each state variable, while retaining qualitative and quantitative features of the spatial distribution. This substantial reduction of the order of the original model results also in significant computational savings.

References

[1] D. Barletta, A. Marzocchella, P. Salatino, S.G. Kang, P.T. Stromberg, Modeling fuel and sorbent attrition during circulating fluidized bed combustion of coal, 17th Int. FBC Conf., Jacksonville, USA, 2003.

[2] K. Bizon, G. Continillo, Spectral reduction on empirically derived orthogonal basis of the dynamical model of a Circulating Fluidized Bed Combustor, Comp. Aided Chem. Eng. 32 (2008) 1305.

[3] K. Bizon. G. Continillo, Formulation and spectral reduction of the dynamical model of a circulating fluidized bed combustor, Chem. Proc. and Product Model. 2 (2009).

[4] P. Holmes, J.L. Lumley, Turbulence, coherent structures, dynamical systems and symmetry, Cambridge University Press, 1996.

[5] L. Sirovich, Turbulence and the dynamics of coherent structures ,Quart. of App. Math. 45 (1987) 561.

[6] J.S. Hesthaven, S. Gottlieb, D. Gottlieb, Spectral methods for time-dependent problems, Cambridne University Press, 2007.

[7] S. Sirisup, G.E. Karniadakis, Stability and accuracy of periodic flow solutions obtained by a POD-penalty method, Physica D 202 (2005) 218.

Hybrid Simulation-Optimization Algorithms for Distillation Design

José A. Caballero [a], Ignacio E. Grossmann [b]

[a]*Department of Chemical Engineering, University of Alicate. Apartado de correos 99, 03080 Alicante Spain., E-mail Caballer@ua.es*
[b]*Department of Chemical Engineering, Carnegie Mellon University, 5000 Forbes Av. 15213. Pittsburgh, PA USA. E-mail grossmann@cmu.edu*

Abstract

This work addresses the rigorous design of distillation columns using a mixed approach that combines mathematical programming with explicit equations, and the rigorous models that are available in commercial Chemical Process Simulators. A superstructure that has embedded all the potential configurations is proposed. Based on this superstructure representation the problem is formulated as an optimization problem using generalized disjunctive programming (GDP) to minimize the total cost of the process, subject to design specifications. The method determines the optimal number of equilibrium stages and operation conditions to obtain the specified product separation.
The model is solved at different levels. Mass balances, purity specifications and other constrains are included as explicit equations in the model. Properties, like specific enthalpies, and flash equilibrium are calculated by implicit models at the level of process simulator (Simulis Thermodynamics ^{TM}). In this way we increase the robustness of pure equation based models because we take advantage of tailored numerical methods maintaining the flexibility of mathematical programming based approaches for distillation design.
Two simple examples (easy to reproduce) are used to show the performance of the model.

Keywords: Distillation, optimization, implicit models, process simulators, MINLP-

1. Introduction

The optimal synthesis of distillation columns has remained a major challenge since the pioneering work by Sargent and Gaminibandara in 1976. This interest is plenty justified due to the high investment and operating cost involved in this systems. The recent trends in this area have been to address models of increasing complexity through the use of mathematical programming. However, the high degree of non-linearity and non-convexity of these models have prevented these methods to become in tools that can be readily used by the industry.

The first approaches tried to select the optimal feed tray location for a fixed number of trays. The feed is split in as many streams as candidate trays (that can be all trays except condenser and reboiler or a subset of selected trays). This is the essence of the pioneering work by Sargent and Gaminibandara (1976). Latter, Viswanathan and Grossmann (1990) showed that this model can be transformed in a MINLP model by considering all the mass, energy and phase equilibrium equations (MESH, but most of the times the optimal solution was obtained in the initial relaxed NLP.

Viwanathan and Grossmann (1993) proposed the first extension to optimize not only the feed tray position, but also the number of trays. The basic idea was to consider a fixed

feed tray with an upper bound of trays above and below the feed. The reflux is then returned to all trays above the feed, and the reboil to all trays below the feed. The actual number of trays is determined by assigning 0-1 variables to the existence of each of the reflux and reboil return options, resulting in a MINLP model. (Figure 1a shows a scheme of the superstucture). The major difficulty with this model is that trays not selected above the feed only handle vapor flow since the liquid flow is zero. A similar situation arises with trays not selected below the feed. However, vapor liquid equilibrium equations, in non-selected trays, although redundant must be satisfied, introducing important numerical problems. In spite of the difficulties this model has been successfully used by different researchers (Bauer and Stichlmair 1998, Dunnenier and Pantelides, 1999; Jackson and Grossmann, 2001;).

To avoid the numerical problems in MINLP models Yeomans and Grossmann (2000) proposed a Generalized Disjunctive Programming model by allowing the bypass of those trays that are not selected. Figure 1b shows the column representation for this approach. For each existing tray the mass transfer task is accounted for and modeled with the MESH equations. For a non-existing or inactive tray the task considered is simply an input-output operation with no mass transfer. Because the MESH equations include the solution for trivial mass and energy balances, the only difference between existing and non-existing trays is the application of the equilibrium equations. As for the permanent trays, all the equations for an existing tray apply. The advantage of the disjunctive modeling approach is that the MESH equations of the non-existing trays do not have to be converged, and no flows in the column are required to take values of zero, making the convergence of the optimization procedure more reliable. Also, by using Generalized Disjunctive Programming (GDP) as the modeling tool, the computational expense of solving the problem can be reduced. Barttfeld et al., (2003) considered different representations for the GDP model, with fixed and variable. However, the computational results showed that the most effective structure was the one originally presented by Yeomans and Grossmann (2000). Numerical results studies for nonsharp separation of ternary mixtures in a single column suggests that the GDP formulation requires less solution time but is more sensitive toward local optima than MINLP formulations.
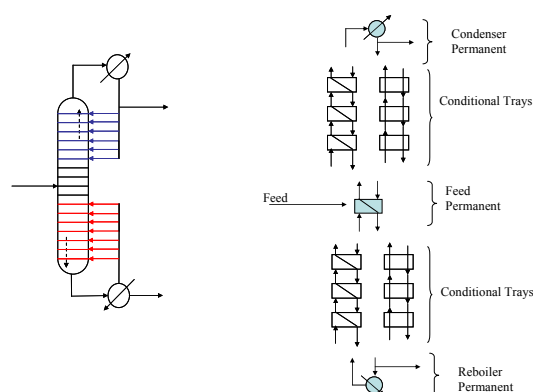
Figure 1. a Superestructure by Viswanathan and Grossmann, b superstructure by Yeomans and Grossmann

Due to the complexity, nonlinearities and nonconvexities involved in both, the MINLP and GDP models, good initial values and bounds are essential in order to achieve

convergence. Barttfeld et al., (2003) used the theory of reversible separation to generate initial values for the rigorous optimization. Their method is mainly limited by the drawbacks of this so-called "preferred separation", because, for azeotropic mixtures, usually nonsharp splits are generated. The extension to the most common sharp split is not trivial. Kossack et al (2006) proposed to use the Rectification Body Method (RBM) that can be used in all the cases. However, the initialization procedure is rather complex with a three step algorithm previous the rigorous optimization.

A non-trivial and remarkable aspect of the implementation presented by Kossac el al (2006) is that activity coefficient models are calculated as implicit blocks of equations in that way the robustness of the model increases and the flexibility to chose different thermodynamic models increases.

On other side, process simulators are commonly used tools in both academy, even by undergraduate students, and industry to accurately reproduce and forecast the behavior of complex distillation systems. Process simulators include complex thermodynamic and transport models that allow accurately calculate properties of pure components, mixtures and state of the art algorithms to calculate from equilibrium (i.e. flash equilibrium using different independent variables: temperature and pressure, pressure and enthalpy, etc) to a complete distillation column. Using optimization algorithms with these types of models is a challenging problem because some variables cannot be accessed or modified directly by the user; which sometimes introduces non-differentiabilities. Furthermore, integer variables also cannot be modified during the optimization process (i.e. number of distillation trays in a distillation column). Even though these difficulties, Caballero et al. (2005) have proposed an algorithm that allows the use of process simulators for designing distillation columns. The method has proved to be very robust and reliable. Its major drawback is that relies on a heuristic to generate a Master problem

2. Disjunctive hybrid simulation-optimization model

In this work, we propose an intermediate solution that tries to take advantage of both approaches: The flexibility of equation based methods and the robustness (and partially of the specially tailored algorithms) of modular approaches. The idea consists of developing a superstructure that uses as basic unit the distillation tray. The distillation tray is considered as a completely implicit block including all thermodynamic calculations and mass and energy balances in terms of its inputs (liquid coming from upper trays and vapour from lower trays). The type of 'flash' performed on each tray (and the algorithm used to solve it) varies depending on the model needs (i.e. flash at constant pressure and enthalpy, or at constant Pressure and Temperature, etc).

A distillation column can be represented as a condenser followed by a set of LV flash units that finish in a reboiler (that can be considered as well as a LV flash unit). If we are interested in the economic optimization of a distillation column, we need the total number of distillation trays and feed tray location. However, the numerical details of how a given equilibrium tray is calculated is not relevant and we can take advantage of this fact letting a process simulator to perform the calculations, so we can compare different thermodynamics and we do not have to worry about estimations of enthalpies, vapour pressures or any other physical properties.

The disjunctive model can be conceptually represented as follows
Variables

V_i, y_i Vapor flow and molar fractions of vapor stream entering in tray i
D, x_D Flow rate and molar fractions in distillate

$\min : Total\ cost$

$$\begin{bmatrix} W_i \\ (V_i^{out}, y_i^{out}, T_i, x_i, L_i, h_{Li}) = \\ f(V_i, y_i, RR, D, x_D) \end{bmatrix} \underline{\vee} \begin{bmatrix} \neg W_i \\ flows,\ compositions\ and\ other \\ specific\ properties\ in\ output = \\ flows,\ compositions\ and\ other \\ specific\ properties\ in\ input \end{bmatrix}$$

$$\begin{array}{l} y_{i,j}^{out} = y_{i+1,j}\ \ j \in Components \\ \quad {}_i^{out} = V_{i+1} \\ \quad \sum_j x_{D_j} = 1 \\ \quad \sum_j y_{i,j} = 1 \forall i \end{array} \left.\begin{array}{l} \\ Column\ especifications \\ \\ Logic\ relationships\ to\ assure\ that\ in\ a\ given \\ section\ all\ trays\ are\ consective\ (no\ redundant\ solutions) \end{array}\right.$$

Implicitly is assumed in previous GDP formulation a superstructure like that presented by Yeomans and Grossmann (2000).

Note that for a fixed configuration of trays (existing and non existing trays are known) it is possible perform the tray calculations sequentially, i.e. starting at condenser and with the values of reflux ratio, distillate flow rate and concentration calculate the heat in the condenser and the flows and compositions of vapor exiting from first tray and liquid entering in the first tray. Then with values of Vi and yi,j calculate the flows and compositions of vapor exiting from next tray and liquid entering into the next tray and so forth. Finally the reboiler can be also calculated. Remark that only flows and compositions in vapor streams are used as variables, considerably reducing the number of variables seen by the external solver. (Liquid flows and compositions could also be used)

The main difficulty with the calculation procedure commented above, is that we need a fixed configuration of Boolean variables to solve the NLP problem. In other words, we cannot solve a relaxed problem in which binary variables could take fractional values. Another way of seen this difficulty is that if we want to create a Master problem we would need initialization values for all the existing and non existing terms in the disjunction.

One way of overcoming that difficulty consists of reformulating the problem using a big M formulation that transforms the problem into a MINLP problem, and then instead of solving the initial relaxed problem, the outer approximation algorithm start for a given fixed configuration of trays (all trays exist). In this way, the right part of the disjunction is not active, but it is present in the NLP optimization and those equations can be used to generate a valid Master problem.

There are different ways of calculating a tray depending on which are the independent variables selected. A procedure that has proved to be very robust consists of using a flash at fixed pressure and enthalpy. This is a two step procedure: first the enthalpy of the vapour stream entering in a given tray, assumed at its dew point, is calculated. Then the enthalpy of the mixture entering to a tray can be calculated by the mixture of liquid coming from previous tray and the enthalpy of the vapour stream. With this information a flash at fixed pressure and enthalpy can be performed. With that flash we also obtain the flows, compositions and enthalpies of the liquid and vapour streams exiting from that tray.

Hybrid simulation-optimzation algorithms for distillation design

To model the implicit block "distillation tray" the Simulis Thermodynamics [TM] software was used, due to its easy access to different flash calculations. However, any process simulator (Hysys, Aspen Plus, etc) could also be used. MATLAB[TM]-TOMLAB controls all the optimization process using state of the art solvers (SNOPT for the NLP optimization, CPLEX in LP optimization and proprietary implementations of decomposition algorithms for MINLP and disjunctive models.

The model has proved to be very robust, and in general no complex initialization procedures are needed. The major drawback is that most of the CPU time is spent in interchanging information between the implicit model and the optimization solver, and in calculating accurate derivatives. In the actual implementation derivatives are calculated numerically by perturbing independent variables. However, it is possible that at the same time the module performs the flash calculations also estimate the derivative information, and therefore it is expected that the CPU time can be reduced even in orders of magnitude, but this approach is still under development.

3. Examples

The proposed method is illustrated with two applications to single distillation columns:

1. Mixture Benzene-Toluene-p-Xylene-Ethylbenzene. (BTXE)
2. Mixture of Methanol, Acetone and Water. (MeOH-Ac-W)

In the first case the objective is separate Benzene to obtain a distillate with at least 0.95 mol fraction in benzene and at least a benzene recovery of 95%. In the second case, there is a minimum boiling azeotrope between Acetone and Methanol. In this example the objective is to obtain a mixture of acetone and methanol with a combined mol fraction greater than 0.99 and a recovery greater than 99%.

To facilitate the reproducibility of the example (without loosing generality) instead of a rigorous calculation of the total annual cost we assume that the cost is given by a simple expression in terms of the heat loads and number of column trays.

$$\min : \; Qreb \, (kW) + 0.2 \, Cond \, (kW) + 80(N^\circ \, Trays)$$

In both cases, the total number of trays was fixed at 20. All the relevant data of these examples are showed in Table 1. and Figures 2 and 3 show the results.

The solution is equivalent to that obtained using a commercial process simulator, when the optimal parameters are introduced as specifications.

Table 1. Data for the examples

Mixture	BTXE	Mixture	BTXE
Feed	100 kmol/h	Feed	100 kmol/h
Feed mol fraction	0.3-0.4-0.2-0.1	Feed mol fraction	0.4-0.3-0.3
Feed thermal state	Saturated liquid	Feed thermal state	Saturated liquid
Thermodynamics	SRK	Thermodynamics	NRTL(default parametes)
Specifications		Specifications	
Benzene mol frac.	>0.95	MeOH+Ac mol frac.	>0.99
Benzene recovery	> 95%	MeOH+Ac recovery	>99%
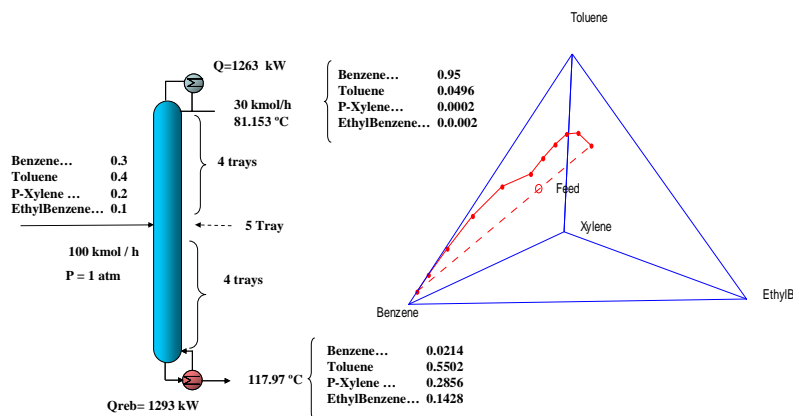
Figure 1. Results of example 1
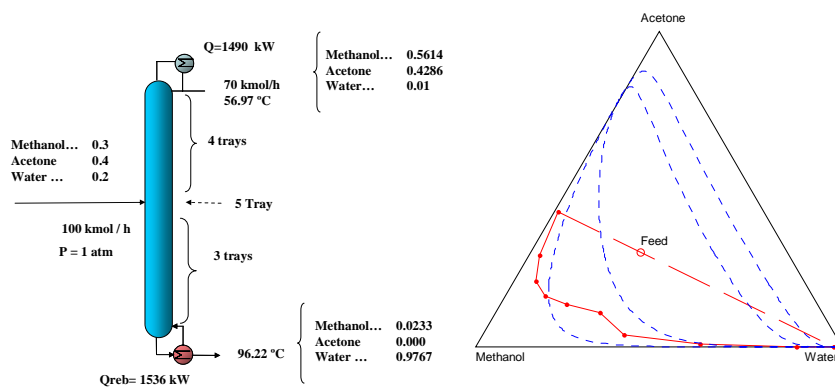
Figure 2. Results of example 2

4. Acknowledgements

References

Barttfeld, M., P. A. Aguirre and I. E. Grossmann, 2003. Computers & Chemical Engineering 27 (3): 363.

Bauer, M. H. and J. Stichlmair, 1998 Computers & Chemical Engineering, 22 (9): 1271.

Caballero,J.A. and Grossmann, I.E. 2005. Ind. Eng. Chem. Res. 44 (17) 6760

Dunnebier, G. and C. C. Pantelides, 1999. Ind. Eng. Chem. Res, 38 (1) 162.

Jackson, J. R. and I. E. Grossmann, 2001.Computers & Chemical Engineering 25 (11-12): 1661.

Kossack, S.; Kraemer, K.; Marquardt, W. 2006. Ind. Eng. Chem. Res. 2006, 45, 8492

Sargent, R. W. H. and K. Gaminibandara 1976. Optimization in Action. L. W. C. Dixon, Academic Press, London: 267

Viswanathan, J and I.E. Grossmann, 1990. Computer & Chemical Engineering, 14(7) 769

Viswanathan, J and I.E. Grossmann, 1993. Computer & Chemical Engineering, 17(9) 949

Yeomans, H. and I. E. Grossmann, 2000. Ind. Eng. Chem. Res39 (11): 4326.

Numerical Determination of Distillation Boundaries for Multicomponent Homogeneous and Heterogeneous Azeotropic Systems

Juan A. Reyes-Labarta, Jose A. Caballero, Antonio Marcilla

Department of Chemical Engineering University of Alicante. Ap. Correos 99, Alicante E-03080, Spain, E-mail: ja.reyes@ua.es

Abstract

As it is very well known, especially when dealing with the planning and operation of distillation processes and equipment, it is essential to characterise the intrinsic limitations due to the VL equilibrium of the system under consideration. This paper addresses a new approach to directly and numerically calculate the distillation boundaries present in ternary and quaternary azeotropic systems, simply involving the concept that such boundaries are distillation trajectories passing through specific singular points. To generate the tested distillation trajectories, that after the optimization process will be the searched boundary, we use cubic splines and polynomial equations (although any other adequate equation could be used). The commercial software package Simulis Thermodynamics (ProSim) has been used to calculate the VLE using the available thermodynamic models such as NRTL and UNIFAQ.

Keywords: Distillation boundary, azeotropic distillation, liquid-vapour equilibrium, VLE.

1. Introduction

Many strategies have been searched in the past to define the distillation boundaries present in a multicomponent system with azeotropic compositions. In this sense, basic theory and different necessary but not sufficient properties have been published [1-4]. In addition, several topological and approximated methods have been developed to predict the general location of these boundaries [5-8], having in mind their importance when dealing with azeotropic distillation processes.

As very well known [9-11], distillate and residue curve maps present a set of singular or critical points related to the distillation trajectories that can be classified (by checking the boiling temperatures or the signs of the derivatives of the function (yi-xi) in the singular points), as follows:

- stable node: critical point with all paths approaching
- unstable node: critical point with all paths departing
- saddle: singular point with finitely many paths both approaching and departing.

In the present work, we suggest a new algorithm to directly calculate the distillation boundaries present in ternary and quaternary azeotropic systems, simply applying the concept that define such boundaries as distillation trajectories passing through specific singular points, previously classified.

2. Mathematical treatment

As commented before, the proposed algorithm is based on the concept that distillation boundaries are distillation trajectories that pass through specific singular points such as points that represent azeotropic mixtures (binary or ternary) or pure components. Additionally, the concept the tie lines are chords of the distillation curves is also used (Fig. 1).
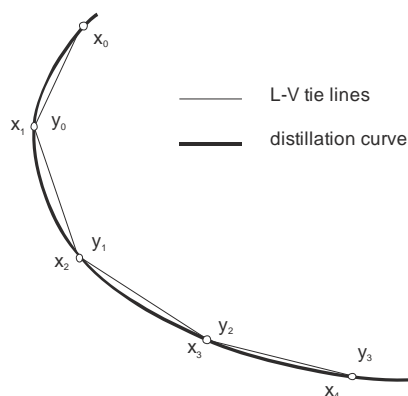
Figure 1. Relationship between distillation curves and L-V tie lines.

To generate the probe distillation trajectories, that after the optimization process will be the searched boundary, we use cubic splines or n^{th}-order polynomial equations (although any other adequate equation could be used).

Fig. 2 shows the general scheme of the proposed calculation algorithm. This proposed procedure consists in the following steps:

1. Classify the different singular points of the composition diagram, i.e. the azeotropic compositions and pure components present in the system, as stable, unstable or saddle points.

2. Define the origin and the end of the boundary *t*rajectory searched, its *n*umber of *i*ntermediate *p*oints to be calculated (*nipt*) and the *n*umber of *i*ntermediate *p*oints to be used for the *c*ubic *s*plines (*nipcs*) or the number of parameters (*n*) in the case of using a nth-order polynomial function or any other algebraic function.

3. Define the composition of the component that will be the independent variable (e.g. x_2)

4. Select the independent variable values for each intermediate point homogeneously distributed, i.e.: $x_{2,k}$ (*k= 1,2,...,nipt*) and $x_{2,k'}$ (*k'= 1,2,...,nipcs*).

5. Guess initial values for $x_{1,k'}$ (*k'= 1,2,...,nipcs*), to be used in the cubic spline calculations or the initial values for the parameters A_j (*j= 1,2,...,n*) of the algebraic function used.

Numerical determination of distillation boundaries for multicomponents homogeneous and heterogeneous azeotropic systems

```
┌─────────────────────────────┐
│       Singular points       │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ Trajectory extremes, nipt    │
│      and nipcs (or n)        │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ Independent variable (e.g. x₂)│
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ Compositions: x₂,ₖ (k=1,2,...,nipt) │
└─────────────────────────────┘
              │
              ▼
┌─────────────────────────────┐
│ Initial values for the dependent │
│ variable: x₁,ₖ' (or Aⱼ)      │
└─────────────────────────────┘
```

$$x_{2,k} \ (k=1,2,...,nipt)$$

$$x_{1,k'} \ (\text{or } A_j)$$

Compositions: $x_{1,k}^{cal.} \ (k=1,2,...,nipt)$

Compositions: $y_{i,k}^{eq.} \ (k=1,2,...,nipt)$

Compositions: $y_{1,k}^{cal.} \ (k=1,2,...,nipt)$

¿ $y_{1,k}^{cal.} = y_{1,k}^{eq.}$?   yes → End

no

New values for the dependent variable $x_{1,k'} \ (k'=1,2,...,nipcs)$
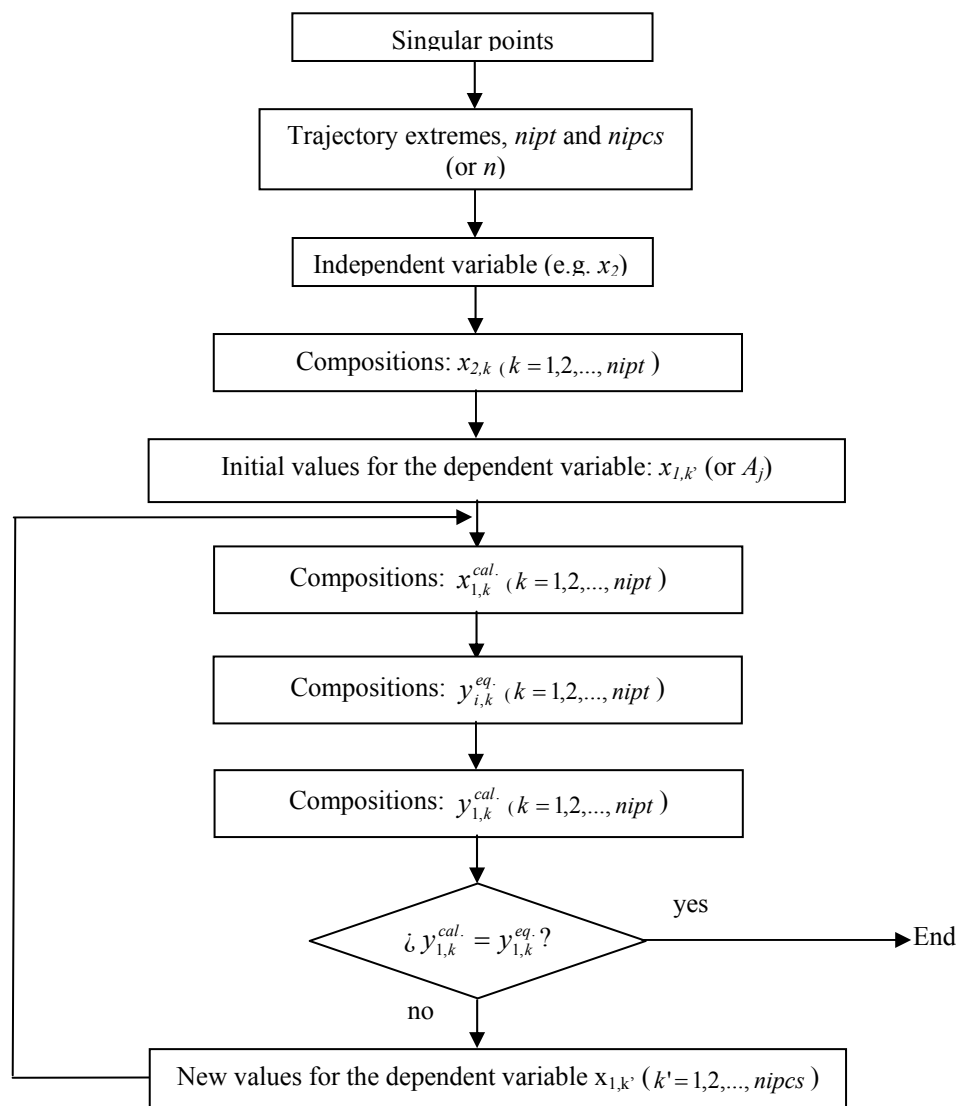
Figure 2. General scheme of the proposed method to calculate distillation boundaries using cubic splines (or algebraic equations) to generate distillation trajectories.

6. Calculate $x_{1,k}^{cal.}$, using the cubic spline (*cs*), the nth-order polynomial or any other adequate algebraic function (*af*) and $x_{2,k}$, i.e.: $x_{1,k}^{cal.} = f^{cs}(x_{2,k}, x_{1,k'}, x_{2,k'})$ with k= 1,2,...,nipt and k'= 1,2,...,nipcs or $x_{1,k}^{cal.} = f^{af}(x_{2,k}, A_j)$ with j= 1,2,...,n.

7. Calculate, the LV equilibrium of each liquid point $x_{i,k}$, using an adequate thermodynamic model: $y_{i,k}^{eq.}$

8. Calculate $y_{1,k}^{cal.}$, using the cubic spline (*cs*), the nth-order polynomial or any other adequate algebraic function (*af*) and $y_{2,k}^{eq.}$, i.e.: $y_{1,k}^{cal.} = f^{cs}(y_{2,k}^{eq.}, x_{1,k'}, x_{2,k'})$ with $k = 1,2,...,nipt$ and $k' = 1,2,...,nipcs$ or $y_{1,k}^{cal.} = f^{af}(y_{2,k}^{eq.}, A_j)$ with $j = 1,2,...,n$.

9. Using an optimization solver over the cubic spline or the algebraic function, calculate the trajectory ($x_{1,k'}$ with $k' = 1,2,...,nipcs$, passing through the selected singular points of origin and ending) that satisfies the following objective function:

$$\sum_{k=1}^{nipt}\left(y_{1,k}^{eq.} - y_{1,k}^{cal.}\right)^2 = 0.$$

In the present work, the commercial software package Simulis Thermodynamics (ProSim) has been used to calculate the VLE using the available thermodynamic models such as NRTL and UNIFAQ, although any other empirical equation could be used [12].

3. Results and discussion

To validate the proposed algorithm, the following systems have been studied using the suggested procedure:

System A: Chloroform + Methanol + Acetone, 760 mmHg with 3 binary azeotropes and 1 homogeneous ternary azeotrope.

System B: Benzene + Isopropanol + Water, 760 mmHg with 3 binary azeotropes and 1 heterogeneous ternary azeotrope.

System C: Acetone + Methanol + Propanol + Water, 760 mmHg with 2 binary azeotropes.
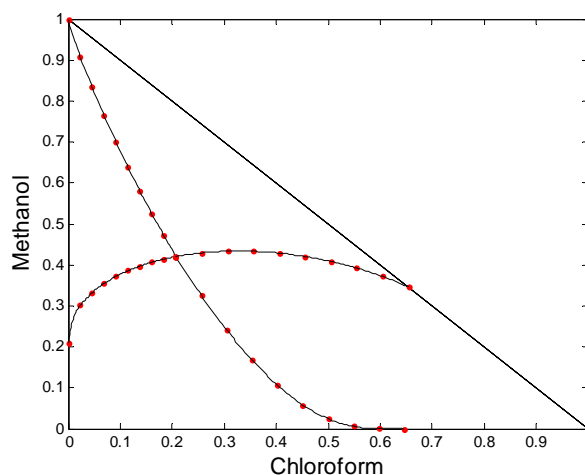
Figure 3. Results for the studied system A: Chloroform + Methanol + Acetone at 760 mmHg. a) distillation boundaries (□ cubic spline nodes, ◊ distillation boundary).

Figure 3 shows the results for the calculated distillation boundaries of the homogeneous system A, following the procedure previously described. As we can observe, this system

Numerical determination of distillation boundaries for multicomponents homogeneous and heterogeneous azeotropic systems

shows 3 different distillations boundaries, starting all of them at the ternary azeotrope and ending at each of the binary azeotropes present in the system. Figure 4 and 5 also show the results of the distillation boundaries and surfaces for the heterogeneous ternary and homogeneous quaternary studied systems, respectively.
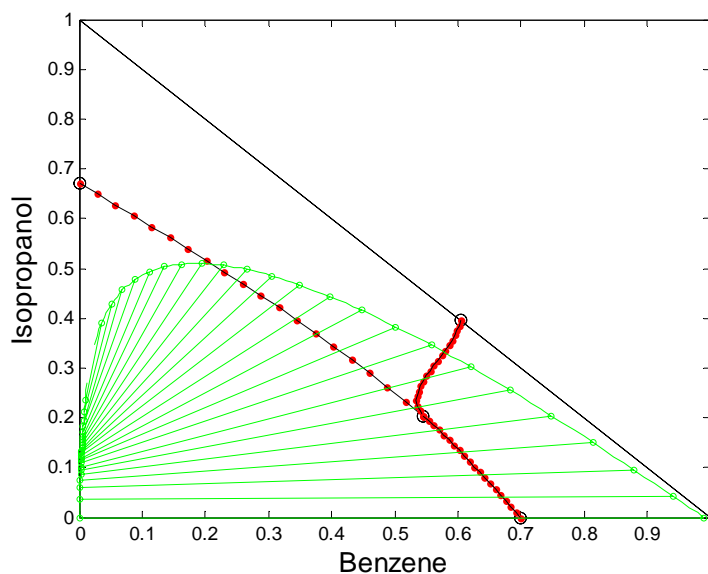
Figure 4. Results for the studied heterogeneous ternary system: Benzene + Isopropanol + Water at 760 mmHg, including the liquid-liquid equilibrium at 298 K (☐ cubic spline nodes, ✎ distillation boundary, ○ azeotropic mixtures, ✎ LL tie-lines).
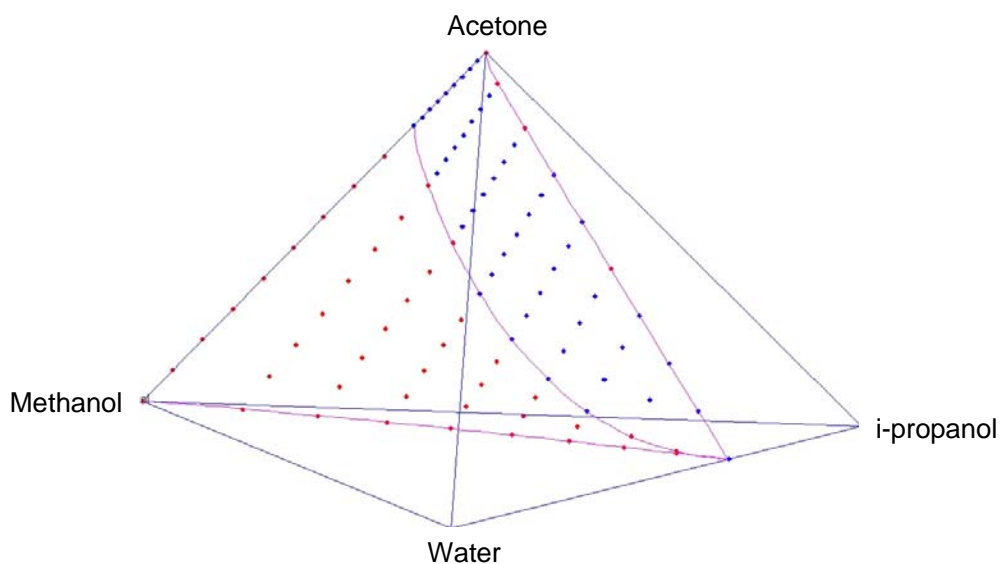
Figure 5. Results for the studied homogeneous quaternary system: Acetone + Methanol +Propanol + Water at 760 mmHg showing the corresponding distillation boundary surfaces.

4. Conclusions

The algorithm proposed to calculate distillation boundaries, using thermodynamic models in the VLE calculations, produces very goods results in the homogeneous and heterogeneous azeotropic ternary and quaternary studied systems, having in mind that if we increased the number of intermediate points for the cubic spline (nipcs) or in same cases the type of algebraic function used, the results would be improved.

5. Acknowledgements

References

[1]   A. Gómez, F. Ruiz, A. Marcilla, J. Reyes, S. Menargues. Ingeniería Química, 377 (2001) 219.

[2]   A. Gómez, F. Ruiz, A. Marcilla, J. Reyes, S. Menargues. Ingeniería Química. 379 (2001) 253.

[3]   D.B. Van Dongen, M.F. Doherty. Chem. Eng. Sci., 39 (1984) 883.

[4]   A. Lucia, R. Taylor. AIChE J. 52 (2006) 582.

[5]   E.R. Foucher, M.F. Doherty, M.F. Malone. Ind. Eng. Chem. Res., 29 (1991), 760.

[6]   E.J. Peterson, L.R. Partin. Ind. Eng. Chem. Res. 36 (1997) 1799.

[7]   R.E. Rooks, V. Julka, M.F. Doherty, M.F. Malone. AIChE J., 44 (1988), 1382.

[8]   T. Pöpken, J. Gmehling. Ind. Eng. Chem. Res. 43 (2004) 777.

[9]   V.N. Kiva, E.K. Hilmen, S. Skogestad. Chem. Eng. Sci, 58 (2003), 1903.

[10] S. Widagdo, W.D. Seider. AIChE J., 42 (1996) 96.

[11] M.F. Doherty, J.D. Perkins. Chem. Eng. Sci. 34 (1979) 1401.

[12] A. Marcilla, J.A. Reyes-Labarta, R. Velasco, M.D. Serrano, M.M. Olaya. EQUIFASE'09, VIII Iberoamerican Conference on Phase Equilibria and Fluid Properties for Process Design (2009), 84.

Modelling of Direct Contact Membrane Distillation for Desalination

Edward Close,[a] Eva Sørensen,[b]

[a]*Department of Chemical Engineering, University College London (UCL), Torrington Place, London WC1E 7JE, U.K, e.close@ucl.ac.uk*
[b]*Department of Chemical Engineering, University College London (UCL), Torrington Place, London WC1E 7JE, U.K, e.sorensen@ucl.ac.uk*

Abstract

Direct Contact Membrane Distillation is a new process with exciting opportunities for use in desalination. With 60% of the world's population expected to be experiencing severe water shortages by 2025, and unhealthy water estimated to be responsible for 90% of all disease in developing countries, the production of clean, affordable water is becoming increasingly important in order to sustain life on Earth. In this work, a mathematical model of membrane distillation was developed, taking account of all aspects of the process, more so than anything previously published. Its use can enable membrane distillation to fulfill its potential, and provide clean water worldwide.

Keywords: Membrane distillation, desalination, modelling, direct contact, water

1. Introduction

With dramatic increases in the human population, changes in life-style, increased economic activities, and pollution hindering the use of natural fresh water resources, it is predicted that 60% of us will be suffering from serious water shortages by 2025 (1). Furthermore, the common use of unhealthy water in developing countries is estimated to be causing 80-90% of all diseases and 30% of all deaths (1). Desalination of brackish or salt water is therefore becoming increasingly important, not just to support industry, but to sustain life.

Desalination is the removal of excess salt and minerals from water and is typically used to provide clean water from seawater or brackish water. The main desalination processes are currently multi-stage flash desalination and reverse osmosis. For both, the processing system is simple and their modularity allows for a large degree of flexibility. As the existing technologies are so well established, membrane distillation has so far received limited acceptance in the desalination industry. However, there are exciting opportunities for its use on a smaller scale in developing countries or in disaster scenarios with limited infrastructure, as low grade energy sources such as waste heat or solar radiation can be applied resulting in low operational costs and high mobility. As the operating costs are already low, membrane distillation may become a serious alternative to existing technologies in the longer term.

Membrane distillation is a technology whereby solutions of dissolved substances (such as salt water) are concentrated by a micro-porous membrane. This membrane does not allow liquid water to pass through it, but is permeable for water vapour. In the case of brackish or salt water desalination, fresh water passes through the membrane leaving the salt on the seawater side. This fresh water can then be collected as it exits the process. The distillation is operated at atmospheric pressure and a maximum temperature of around 80°C, resulting in low operating costs.
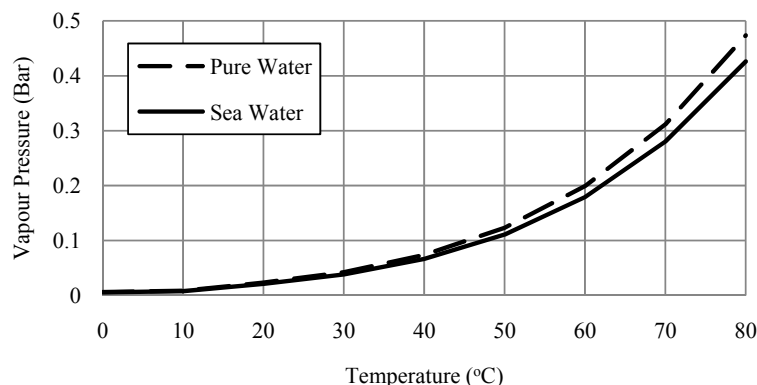
Figure 1. Vapour pressure of pure water and seawater as a function of temperature

2. Mechanism of Separation

The movement of water across the membrane is the result of an imbalance in the vapour pressure between the different sides of the membrane due to a corresponding temperature difference, *i.e.* water evaporates at an entrance to a membrane pore, enters into a region of high water vapour concentration in the air within the membrane, and diffuses down a concentration gradient. The water vapour moves through the membrane until it reaches the region of low water vapour concentration at the exit of the pore. Here it condenses, and can be collected as pure water. This driving force is generated by heating the source water, and thereby increasing its vapour pressure above that of the desalted water collected as distillate on the other side, which is cooled.

In order to understand the basis of the process it is helpful to see how the vapour pressure of pure and sea water change with temperature. As shown in Figure 1, seawater entering a membrane distillation unit at 60°C has a corresponding vapour pressure of approximately 0.18 Bar whilst pure water exiting the unit at 20°C has a corresponding vapour pressure of approximately 0.02 Bar. Thus a pressure difference of 0.16 Bar would exist over the membrane, providing the driving force for the separation.

3. Modelling of Membrane Distillation

The objective of this work was to explore the potential of membrane distillation in desalination, particularly in terms of its use of low grade energy, by developing a detailed mathematical model of the process which allowed the investigation of membrane distillation in much greater detail than had previously been considered by other authors. In order to facilitate the development of this model, it was necessary to consider the three sections of the process individually before bringing them together (see Figure 2.). The process was divided into i) the vapour-liquid equilibrium section, ii) the mass transfer section, and iii) the heat transfer section. (See appendix).

i) In membrane distillation the membrane is permeable to water vapour, but impermeable to water. Simply put, the membrane allows the passage of water vapour, but not water. This results in a vapour-liquid interface at the entrances and exits of the membrane pores, *i.e.* these are the regions where liquid and vapour are in contact with each other.

Modeling of Direct Contact Membrane Distillation for Desalination
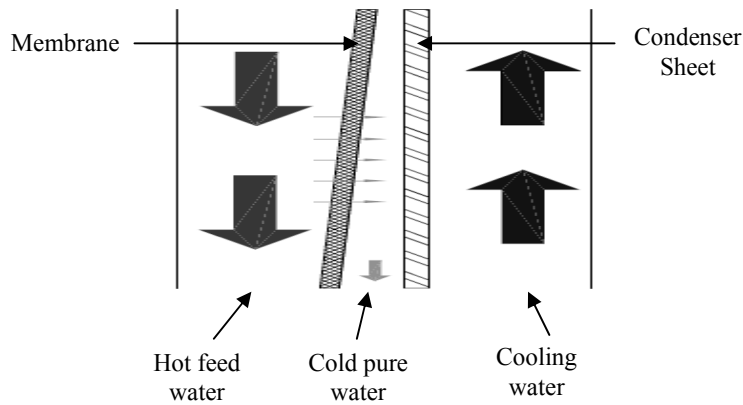
Figure 2. A typical Direct Contact Membrane Distillation

The evaporation of liquid to vapour, and the condensation from vapour to liquid, occur at a rate dependent on the temperature, pressure, and respective concentrations of water and salt in the region. The mathematical equations representing this part of the process were developed based on Antoine equation and activity coefficients.

ii) Mass transfer in membrane distillation consists of a series of consecutive steps: 1) Water from the bulk flow of the saltwater must move through a very thin boundary layer on the feed side of the membrane, before 2) evaporating at the surface of the membrane, 3) water then moves through the membrane pores, and 4) condenses at the membrane surface on the permeate side of the membrane. It must then 5) move across another boundary layer before it then joins the permeate bulk flow to be collected as a product. Mathematical equations describing this part of the process were developed based on boundary value theory and the dusty gas model.

iii) Heat transfer in membrane distillation was broken down according to heat transfer between different sections of the membrane distillation unit. This part of the process was modeled by developing mathematical equations describing the amount of heat transferred between each section and then combining them. The equations were based on heat transfer coefficients and temperature boundary conditions.

3. Simulation

The equations representing each section of the model were modelled within gPROMS, an equation based modelling and optimization software package developed by Process Systems Enterprise.

In order to confirm that the model simulates the operation of Direct Contact Membrane Distillation accurately, the model was verified against experiments presented in the literature and excellent agreement was found as shown in Figure 3.

The model was then used in an extensive sensitivity study to predict the effect of changing individual parameters, such as the temperature of the seawater, on the amount of pure water produced. The results of these simulations also showed excellent agreement with the trends reported in the literature.

E. Close and E. Sørensen
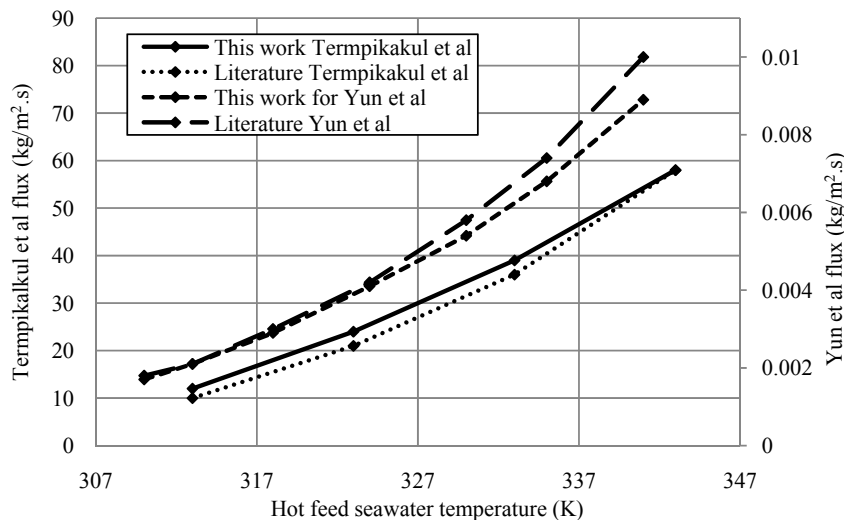
Figure 3. Comparison of the fluxes found by the model developed in this work, and that found in the literature (2) (3)

In addition, design parameters that have previously seen little or no work, such as membrane thickness and porosity, were investigated. As Figure 4 shows, the model predicts an optimal membrane thickness which maximizes the flux (for the set of operating/design conditions specified). This may be the result of competition between the distance that the water has to diffuse across the membrane and the magnitude of the driving force, *i.e.* the smaller the membrane thickness gets, the less distance the water has to diffuse. However, as the membrane thickness decreases, so does the driving force (due to a smaller temperature difference across the membrane).

Where there were conflicting results in the literature regarding the effect of different variables on the flux, the model was able to provide an explanation. For example, Banat (4), found that increasing the cooling water flow rate had no effect, whilst Ohta et al. (5), found that the flux increased with flow rate. As figure 5 shows, the
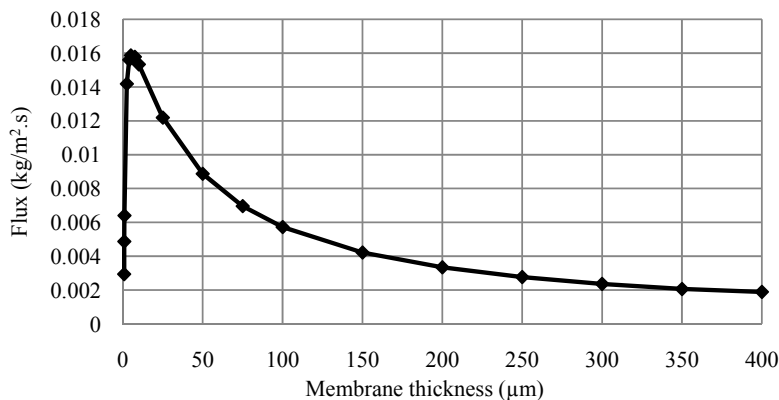
Figure 4. Flux as a function of membrane thickness

Modeling of Direct Contact Membrane Distillation for Desalination
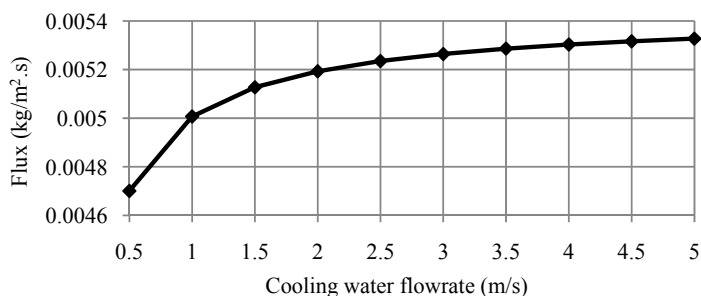
Figure 5. Flux as a function of cooling water flowrate

model developed in this work finds that the flux increases with increasing cooling water flow rate. However, the rate at which the flux increases decreases as the cooling water flow rate increases. Thus, the reason that Banat found that the cooling water flow rate had no effect on the flux, is likely to have been the result of the nature of the experimental set-up and operating conditions, *i.e.* the range of cooling water flow rates investigated did not extend low enough for any difference in flux to be observed.

4. Conclusion

The model of Direct Contact Membrane Distillation developed in this work takes all aspects of the process into account, going further than any model previously published by any other authors. Excellent agreement was found between the results given by the model and actual experiments presented in the literature. Where there were conflicting results in the literature regarding the effect of the variables on the flux, the model was able to provide an explanation. Using the model, researchers can now determine the optimal operation and design of these units, something not previously achievable, and the model will provide engineers with a proven framework in order to design membrane distillation plants. In a wider context, the uptake of membrane distillation in the form of low cost processing units utilizing low grade energy sources such as waste heat or solar energy, should ensure that we can meet the future needs for clean, safe, low-cost water, worldwide, for those that need it the most.

References

(1) Ettouney, H M and El Dessouky, H T. *Fundamentals of Salt Water Desalination*, Elsevier, 2002.
(2) Termpiyakul, R., Jiraratananon, R., & Srisurichan, S. (2005). Heat and mass transfer characteristics of a direct contact membrane distillation process for desalination. *Desalination, Vol 177* , 133-141.
(3) Yun, Y. Ma, R. Zhang, W. Fane, A. G. & Li, J. (2006). Direct contact membrane distillation mechanism for high concentration NaCl solutions. *Desalination, Vol 188* , 251-262.
(4) Banat. (1994). Membrane distillation for desalination and removal of volatile organic compounds from water, PhD Thesis, McGill University, Canada.
(5) Ohta, K., Kikuchi, K., Hayano, I., Okabe, T., Goto, T., Kimura, S., et al. (1990). Experiments on sea water desalination by membrane distillation. *Desalination*, Vol 78, 177-185.
(6) Habib, K. (1989). Predictability of a Membrane Transport Model; A General Approach. *Desalination,* Vol *75* , 41-53.

Notation

A, B, C	Antoine coefficients
C_{pw}	Heat capacity of water (J/(kg·K))
D_{ij}	Diffusion coefficient (m²/s)
D_{ije}^0	Membrane structural parameter
D_{ie}^k	Knudsen diffusivity coefficient (m².s)
d	Effective diameter (m)
h	Heat transfer coefficient (W/m²·K)
H	Mass Transfer Coefficient (m/s)
K_1	Membrane structural parameter
K_0	Membrane structural parameter
k	Thermal conductivity (W/m·K)
L	Module length (m)
l	Distance between parallel plates (m)
M	Molecular weight (kg/mol)
N	Flux (mol/m²·s)
N^V	Viscous molar flux (mol/(m².s))
N^D	Diffusive molar flux (mol/(m².s))
p^{sat}	Components saturation pressure (Pa)
p	Partial pressure (Pa)
P	Total pressure (Pa)
Q	Heat Flux (W/m²)
R	Universal gas constant (J/(K·mol))
r	Pore radius (μm)
T	Temperature (K)
u	Flow velocity (m/s)
x	Mole fraction
Z	Thickness (m)

Greek Letters

β_0	Membrane structural parameter
γ	Activity coefficient
∇	Grad
ΔH_{vap}	Heat of vaporisation of water (J/kg)
ϵ	Membrane porosity
μ	Liquid viscosity (Pa·s)
ρ	Liquid density (kg/m³)
τ	Membrane Tortousity

Subscripts

i	Component i
j	Component j
f	Feed
m	Membrane
1	Membrane surface feed side
2	Membrane surface permeate side
p	Permeate
3	Cooling plate permeate side
cp	Cooling plate

Appendix

A1. The vapour-liquid equilibrium section

$$p_i^{sat} = exp\left[A - \frac{B}{C+T}\right]$$
$$p_i = \gamma_i \cdot x_i \cdot p_i^{sat}$$
$$\gamma_i = 1 - 0.5x_j - 10x_j^2$$

A2. The mass transfer section (6)

$$N_i = N_i^D + N_i^V$$
$$\frac{N_i^D}{D_{ie}^k} + \sum_{j=1 \neq i}^n \frac{p_j \cdot N_i^D - p_i \cdot N_j^D}{D_{ije}^0} = -\frac{1}{R \cdot T} \cdot \nabla p_i$$
$$N_i^V = -\frac{p_i \cdot \beta_0}{R \cdot T \cdot \mu} \cdot \nabla P$$
$$D_{ije}^0 = K_1 \cdot P \cdot D_{ij}$$
$$D_{ie}^k = K_0 \cdot \left(\frac{8 \cdot R \cdot T}{\pi \cdot M_i}\right)^{\frac{1}{2}}$$
$$\beta_0 = \frac{\epsilon \cdot r^2}{8 \cdot \tau}$$
$$K_0 = \frac{2 \cdot \epsilon \cdot r}{3 \cdot \tau}$$
$$K_1 = \frac{\epsilon}{\tau}$$
$$N_i = H_f \cdot \rho \cdot ln\left(\frac{x_f}{x_1}\right)$$

A3. The heat transfer section

$$Q_f = h_f \cdot (T_f - T_1)$$
$$Q_m = \left[N_i \cdot C_{pw} + \frac{k_{c,m}}{Z_m}\right] \cdot (T_1 - T_2) + N_i \cdot \Delta H_{vap}$$
$$Q_p = h_p \cdot (T_2 - T_3)$$
$$Q_{cp} = \frac{k_{c,cp}}{Z_{cp}} \cdot (T_3 - T_4)$$
$$Q_c = h_c \cdot (T_4 - T_c)$$
$$Q_f = Q_m = Q_p = Q_{cp} = Q_c$$

A4. Heat and mass transfer coefficient correlations (2)

$$\frac{H_f \cdot d}{k} = 0.023 \left(\frac{d \cdot \rho \cdot u}{\mu}\right)^{0.8} \left(\frac{\mu}{\rho \cdot k}\right)^{\frac{1}{3}} \left(1 + \frac{6d}{L}\right)$$
$$d = 2l$$

$$\frac{h \cdot d}{D_{ij}} = 0.023 \left(\frac{d \cdot \rho \cdot u}{\mu}\right)^{0.8} \left(\frac{\mu}{\rho \cdot D_{ij}}\right)^{\frac{1}{3}} \left(1 + \frac{6d}{L}\right)$$

Dividing wall column application for platformate splitter – A case study

Igor Dejanović[a], Ljubica Matijašević[a], Žarko Olujić[b]

[a]*Faculty of Chemical Engineering and Technology, University of Zagreb, Savska cesta 16, HR-10000 Zagreb, Croatia, ideja@fkit.hr*
[b]*Process and Energy Laboratory, Delft University of Technology, Leegwaterstraat 44, 2628 CA Delft, the Netherlands, Z.Olujic@tudelft.nl*

Abstract

This paper presents results of a simulation study indicating that a DWC would be a beneficial alternative to conventional two columns in series configuration for a platformate splitter application in a petroleum refinery. The DWC was designed using commercial software, complemented by a fundamentally sound shortcut method proposed by Halvorsen et al. [1-3] which allows generation of reliable initial values for rigorous simulations. Compared to actual, two-columns-in-series configuration, a DWC requires approximately 33.6 % less energy to deliver three fractions at required product specifications.

Keywords: distillation, thermally coupled columns, dividing wall columns

1. Introduction

Energy used for continuous distillation processes comprises approximately 40% of total energy use in chemical process industry [4]. This leads us to conclude following: distillation is still the most common method for separation of liquid mixtures and it is very energy intensive. Together with a fact that many successful separation systems were invented by experience alone [5], this makes distillation, which is considered to be the most mature among separations technologies, still an amenable subject for process intensification.

Most impressive development in this respect is implementation of so called dividing wall column (DWC), which not only leads to energy saving but also to capital saving. Upon establishing itself as packed column in various chemicals separations performed mainly under vacuum, DWC is now making an inroad into refinery world dominated by large tray columns [6], where in many situations multicomponent feeds are separated into three products. In case of sharp product specifications, this often means use of two columns connected in series. Figure 1 shows schematically a typical (indirect sequence) two-columns-in-series arrangement for obtaining three pure products from a three component feed. Where appropriate, this sequence can be realized as a so called Petlyuk column, which, as shown in Fig.2, consists of a prefractionator column fully thermally coupled to the main column. This configuration implies that only one condenser and one reboiler are sufficient, which means a reduction in the number of main equipment and accompanying investment costs. A DWC, shown schematically in Fig.3, is a single shell thermodynamic equivalent of a fully thermally coupled column, which allows separation of three or more components into high purity products within one shell. This is achieved by using a vertical partition wall that divides mainly central part of the column into prefractionator and main column sections. In each section, two components with greatest difference in volatility are separated, while others are allowed to distribute.
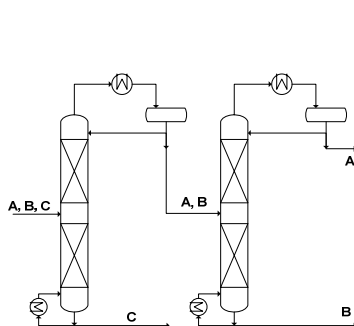
Figure 1. Conventional three component separation arrangement
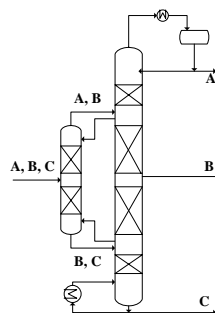
Figure 2. Petlyuk arrangement
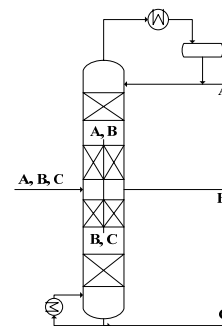
Figure 3. DWC arrangement

The overall thermodynamic efficiency of Petlyuk column and its thermodynamic equivalent DWC is significantly larger than that of conventional configuration. This is mainly because of its configuration, which allows avoidance of the entropy of mixing formation caused by composition differences between feed stream and feed stage, as well as minimization of re-mixing effect of middle boiling components in separation of mixtures with more than 2 components [7], which is effectively a direct loss of separation work. Although the idea of thermal (heat) coupling is older, F. Petlyuk was instrumental in its implementation and therefore, this and more complex fully thermally coupled columns are known as Petlyuk columns. Most of the work, Petlyuk and co-workers have devoted over the years to distillation theory and its application to arrive at optimal design of columns and column configurations or sequences, which was not readily available because it was described in publications written mainly in Russian, can be found completely and consistently presented in English in a recently published book by Petlyuk [8].

Recent advances in design of DWCs including non-welded partition walls that can be placed in off-centre positions to allow optimal design for unsymmetrical (with respect to composition), and partly vaporized feeds, are described elsewhere [9].

The objective of this paper is to demonstrate the usability of commercial simulation software for performance evaluation and conceptual design of a DWC column for a refinery application, concerning separation of a multicomponent aromatics rich mixture into three fractions, according to given specifications.

2. Problem statement

Platforming process product (platformate) comprises approximately 40 components, which however have been reduced for the purposes of this simulation study into a 15 component feed mixture with composition shown in Table 1. This table contains also the compositions of three product streams, i.e. fractions:

    1. C5-C6 gasoline containing no more than 1.5 mass % of benzene;
    2. Benzene rich cut (BRC) containing 68 mass % benzene;
    3. Heavy reformate containing no more than 0.5 mass % of benzene.

These represent specifications as encountered in actual refinery case employing a conventional configuration shown in Fig.1.

3. Base case

To have a reference point for comparison of non-conventional distillation configurations, base case model of the process was developed in Chemcad. Compositions of feed stream and resulting product streams are shown in Table 1. Results are in good agreement with real plant data. Total required reboiler duty is 5548 kW, and condenser duty -5028 kW.

Table 1. Base case feed and product compositions

Stream Name		Platformate	C5-C6	BRC	Heavy
Temperature	[°C]	37.0	40.0	115.3	45.0
Pressure	[bar]	3.01	5.70	3.17	5.00
Total flow	[kg h^{-1}]	31730	6938	3697	21095
Component mass frac.					
N-Butane	[-]	0.019	0.089	0.000	0.000
Isopentane	[-]	0.064	0.293	0.000	0.000
N-Pentane	[-]	0.045	0.207	0.000	0.000
2-Methylpentane	[-]	0.080	0.349	0.035	0.000
N-Hexane	[-]	0.043	0.049	0.273	0.000
Benzene	**[-]**	**0.086**	**0.013**	**0.680**	**0.005**
3-Methylhexane	[-]	0.020	0.000	0.013	0.028
Toluene	[-]	0.247	0.000	0.000	0.372
Ethylbenzene	[-]	0.035	0.000	0.000	0.053
P-Xylene	[-]	0.042	0.000	0.000	0.064
M-Xylene	[-]	0.122	0.000	0.000	0.183
O-Xylene	[-]	0.055	0.000	0.000	0.083
M-Ethyltoluene	[-]	0.047	0.000	0.000	0.071
1-3-5-Trimethylbenzene	[-]	0.077	0.000	0.000	0.116
1-4-Diethylbenzene	[-]	0.017	0.000	0.000	0.025

4. DWC design

Regarding the simulation approach, as indicated in Fig.4, a DWC has significantly more degrees of freedom than a conventional distillation column. That is why a good initialization method (shortcut model) is essential to give rough design estimates, which can be used to set up rigorous simulation. According to own evaluations, this appeared to be a rather simple but in this respect very effective method introduced recently by Halvorsen and Skogestad [1-3].

4.1. Halvorsen's V_{min} diagram method

Halvorsen and Skogestad [2] proposed a design method based on Underwood's equations to determine minimum vapour and liquid flows needed for binary separation occurring in particular column sections, with infinite number of stages. Their basic postulate is that minimum vapour flow required for separation of n components feed into n pure products, corresponds with that required for the most difficult binary split. Halvorsen's method is available as a commercial software package, which can be used to gain insight into requirements to separate n-component feed mixture into n-products. However, V_{min} diagram can also be developed using rigorous simulation [1].
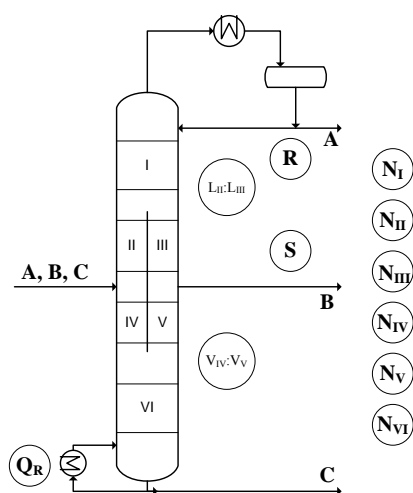
Figure 4. DWC design parameters
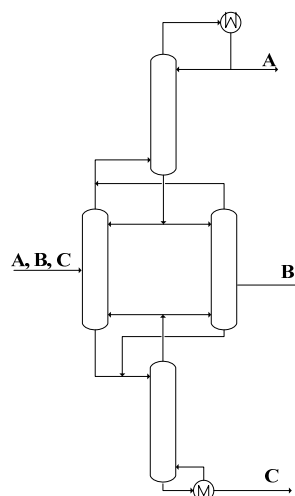
Figure 5. 4 Colum DWC model

For the purpose of 3-product DWC design, 15 components from the feed were grouped in three representative key components: n-pentane, benzene and toluene. Then, V_{min} diagram was constructed by rigorous simulation of possible binary splits, with stage number being at least $4N_{min}$, and key component's mass recoveries set to 0.999. From that diagram, initial design parameters for rigorous simulation with infinite number of stages can easily be obtained. According to Halvorsen's method, energy saving estimate for this three component - three pure products case compared to conventional arrangement were around 14%, and by rigorous simulation around 18%. Present actual case with fractions as products, specified to maximize recovery of main components, allows more in this respect.

5. Rigorous performance simulation

Commercial process simulators such as Chemcad, Aspen or HYSYS allow a DWC to be modelled as a sequence of simple columns. Several approaches can be used, ranging from one to four column models, each exhibiting some good and bad sides. As a general rule, one-column models with pump-arounds have better convergence properties than models with more columns. Four-column model is more difficult to initialize and exhibits a slower convergence, but it offers most flexibility. In this study, four-column model, i.e. the configuration shown in Fig 5 was used in conjunction with Chemcad.

In simulating DWC according to initial design parameters from V_{min} diagram, first step was to restore feed composition to full 15 components. Benzene mass fractions in distillate and side draw liquid flow rate were set to be the same as in base-case simulation, while initial value for reboiler duty was set to give required minimal vapour flow.

Liquid and vapour split ratios were adjusted, until minimal possible mass fraction of benzene in bottoms was achieved, which was approximately the same as in base case.

Next step was to determine actual number of stages in each section. This was done by reducing number of stages in sections, using composition profiles as guidelines, keeping mass fractions of benzene in distillate and bottoms, as well as liquid side draw flow rate constant.

Dividing wall column application for platformate splitter – A case study

For every converged case, an optimization was performed using optimization tool built in Chemcad, objective function being Min Qr, and independent variables liquid and vapour split ratios.

To determine the optimal ratio between energy cost and number of stages, empirical objective function, Min $N(R+1)$ was used, which approximates minimum of total annualized cost of a conventional distillation column.

6. Results

Dependence of proposed objective function used to approximate minimum of total annualized cost of DWC on total number of stages is shown in Fig. 6. It can be seen that minimal value is obtained with total of 64 equilibrium stages. This is well bellow 78, i.e. the total number of equilibrium stages contained in two actual sieve tray columns connected in series. Comparison of simulation results of base-case and DWC design is shown in Table 2, indicating that a DWC would require approximately 33.6 % less energy, based on reboiler duties.
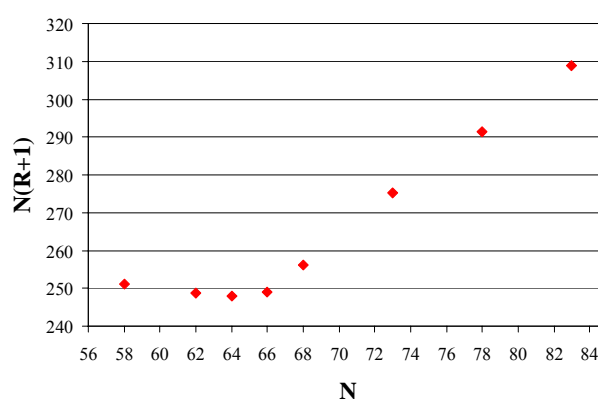
Figure 6. Dependence of objective function on total number of stages

Table 2. Base case and DWC simulation results

		Base case		DWC
		Column 1	Column 2	
Product name				
C5-C6	mas %	1.31		1.30
BRC	mas %	68.00		68.12
Heavy reformate	mas %	0.51		0.50
Reflux ratio	-	1.57	1.78	2.87
Reboiler duty	kW	3381	2900	3685
Total	kW	5548		3685
Condenser duty	kW	-2986	-2042	-2839
Total	kW	-5028		-2839
Number of stages	-	40	38	64

This clearly indicates that an increase in thermodynamic efficiency of distillation leads not only to a significant saving in energy consumption but also in significant reduction

of overall column dimensions. The financial aspects of designing a DWC for present case will be elaborated in detail in forthcoming full scale paper.

One interesting option in this particular industrial case is to consider retrofit of first, larger platformate column, with a tangent to tangent height of 40 m into a DWC, by employing high performance structured packings. This could even enable a considerable capacity increase with minimum investment. Namely, originally both columns have been designed to operate at two operating pressures (approximately 1 bar difference) to be able to accommodate different product purity requirements. This means that the diameter of first column is much larger than required for present separation task, and thus may be larger than needed to accommodate internal vapour flows as associated with DWC designed for this purpose. If this proves to be feasible, the second column would be available for reassignment. This would lead to a maximized gain compared to a new design situation. For that purpose, however a detailed sizing of the proposed configuration needs to be performed, which is not in the scope of present work.

7. Conclusions

In this paper we have demonstrated that a commercial simulator can be used in conjunction with initial guesses for governing variables obtained from a simple but theoretically founded short-cut method, to generate without computation difficulties an optimized internal configuration of a DWC. Compared to conventional two-column-in series configuration for obtaining benzene and toluene rich fractions from a 15-component feed, a DWC requires 33.6 % less energy to get the same product specifications. This as well as the fact that less capital and space will be needed to install new DWC makes this option highly interesting for implementation.

References

[1] I.J. Halvorsen and S. Skogestad, Ind. Eng. Chem. Res. 42 (2003) 596.
[2] I.J. Halvorsen and S. Skogestad, Ind. Eng. Chem. Res. 42 (2003) 605.
[3] I.J. Halvorsen and S. Skogestad, Ind. Eng. Chem. Res. 42 (2003) 616.
[4] J.L. Humphrey and G.E.Keller II, Separation Process Technology, McGraw-Hill, New York, 1997.
[5] M.F. Doherty and M.F. Malone, Conceptual Design of Distillation Systems 1st ed., McGraw-Hill, New York, 2001, pp. 568.
[6] Ž. Olujić, M. Jödecke, A. Shilkin, G. Schuch, B. Kaibel, Chem. Eng. Proces. 48 (2009) 1089.
[7] C. Triantafyllou and R. Smith, Trans IChemE -Chem. Eng. Res. Des. 70 (1992) 118.
[8] F.B. Petlyuk, Distillation Theory and Its Application to Optimal Design of Separation Units, Cambridge Press syndicate of the University of Cambridge, 2004.
[9] B. Kaibel, Dividing Wall Columns. Encyclopedia of Separation Science, I.D.W. Colin and F. Poole (eds.), Elsevier Science Ltd., Oxford, 2007.

Non-linear model approximation using new multivariate Hammerstein block structure model for large processes

Omar Naeem and Adrie E.M. Huesman

"Delft Centre for Systems & Control, Mekelweg 02, Delft 2628CD, the Netherlands" ,
E-mail: O.Naeem@tudelft.nl

Abstract

In this paper, a new approach is discussed to achieve a full order approximation model which leads to reduced order models for the large scale industrial process models. The model class used for this purpose is "block structure" models. The final aim is to achieve the reduced order model which is derived from the full order block structure model. Previously, Hammerstein structures have been used to approximate a mathematical non-linear model of a process by black box identification mainly. Input-Output Hammerstein structure, defined as classical Hammerstein model has been extended to new Hammerstein structure making use of states and inputs, hence called as *input-state (IS) Hammerstein* structure. During the research, it is proven that expansion of Taylor series leads to IS Hammerstein structure. The accuracy of the approximation is improved by including higher order approximation. IS full order Hammerstein model has been implemented on test case (high purity distillation column). It gives good approximation of the original non-linear system in full order. Over an operating domain of a process, the Input-State Hammerstein structure provides opportunities to achieve the reduced order model which is final goal of the research.
Keywords: NL model approximation, Block structure models, Hammerstein structure, high purity distillation column

1. Introduction

Rigorous models of the processes are typically stiff and large which normally lead to computationally inefficient models. It has been noticed that the nonlinear (NL) rigorous models are not always exact match of real processes. There is mismatch at some point between the real process and first order principle models. This gives opportunity for the approximated or identified models which can be very useful if the approximated models match the first principle NL model over a certain operational window. Moreover, full order approximated models may lead to reduced order models can be used for the real-time applications (e.g.; control and optimization purposes). If the full order approximation model not only provides the better approximation of the nonlinear process but also provides *handles* to get the reduced order model, will be highly beneficial and advantageous for the process modeling. Nonlinear rigorous models available for industrial processes can be characterized as a set of differential and algebraic equations (DAE) which is capable to express the majority of process nonlinearities. This means that the methodology to achieve a reduced model should be capable of handling DAE models. Typically the transformation from DAE to ordinary differential equation (ODE) format is regarded as a major model reduction step because the algebraic equation calculation disappears in that case. A model approximation methodology, which can approximate process represented in DAE as ODE, should be a beneficial because it helps in the next step of model reduction. Hence the development of full order approximation model for NL processes in ODE format from DAE

mathematical model will be considered as reduction step itself; of course it will be highly valuable if such full order approximation model provide opportunities to get reduced order model for NL processes.

Though model approximation of first order principle models has been addressed in the past but the approximation of physical model, keeping in view the model reduction has not been addressed previously in literature, specifically dealing with the computational load. Hahn [1] elaborated on model approximation and later on the reduction by balancing empirical Grammians and showed model order reduction but it was limited in computational effort and time. Later Aling [2] used POD to get identified model for rapid thermal processing system. Reduction of computational load by a factor ten was reported, which was one of very rare report about computational load. Ref [3] developed a reduced order model of batch distillation column using travelling waves. The closed loop simulation of this reduced model was six times faster than original model in closed loop. Jogchem [4] reported that if computational load has to be reduced, not only model order reduction is to be targeted but the complexity (and stiffness) of reduced model has to be lower. Scherpen [5] found an approximate model by nonlinear balancing approach which resulted in model order reduction; however results for reduction of computational load were not presented.

The literature review shows that there is not much research material available on full model approximation subject, which later on leads to model reduction. There are not many model approximation methodologies which have addressed the problem of simplification of approximation model which leads to simple reduced order model. The field is open for research for this specific area. Block structure models have an advantage over other model approximation methodologies; the structure of approximation model gives insight to the complexity of the process and breaks down the complexity of the NL process. Use of block structure gives an advantage of approximating the nonlinear processes with higher accuracy and enhances the chances to get a reduced model, which is uncomplicated and is computationally efficient.

2. Block structure models and Taylor series

Block structures are preferred for the approximation, because these block structures are simple representation of the NL processes. The block structure models are simple and straight forward representation, which gives chance for the model reduction. There are different block structures which are known for model reduction (and empirical modeling); Wiener and Hammerstein etc. (References [11, 12])

Wiener and Hammerstein block structure models are most widely used structures in literature for the representation of nonlinear physical processes [12,13] and will be shortly discussed here. Wiener models have limitations (specifically for chemical processes) which mean that identification by Hammerstein structure has advantage over Wiener structure. Wiener models not only limit the nonlinearity measure to be approximated, but they also increase the complexity involved in identifying or approximating the process. Harnishmacher [7, 8] reported that Wiener models restrict the dynamic NL behavior that can be identified and approximated by it, while in comparison, Hammerstein structure has advantages (such as ease of identification, input directional dynamics representation) in model approximation/identification. Hammerstein structure is used for the approximation of NL processes in this study. The methodology is extended further to IS Hammerstein structure to improve the approximation.

Classical Hammerstein model can be seen as nonlinear static gain, followed by linear dynamics. For a continuous system, classical Hammerstein block's mathematical form is given as:

$$\dot{x} = Ax + g(\text{u})$$
$$y = Cx \tag{1}$$

2.1. Derivation of full order IS Hammerstein model:

The input-output (I/O) Hammerstein model can be modified to Input-state (I/S) Hammerstein model under few assumptions I/S Hammerstein model can be derived from expansion of Taylor series [8].

Since typically, the mathematical process models of large processes are in DAE form, the transformation to ODE form will be considered as reduction step. Given an ODE modeled in any suitable environment, $\dot{x} = f(x,u)$, the Taylor series expansion in the neighborhood of point (x^*, u^*) is given as follows:

$$\dot{x} = f(x,\text{u}) = f(x*,u*) + J_x\big|_{x*,u*}(x - x*) + J_u\big|_{x*,u*}(u - u*) + \ldots \tag{2}$$

The eq.(2) can be expressed in block format shown in fig.(1).

This Taylor series can be extended to IS Hammerstein structure under following assumptions:

a. It is assumed, within the operating domain that every 'u*' leads the system finally to steady-state 'x_{ss}' which means a stable process is considered. Moreover, it is assumed that steady-state is calculated by 'u'. Mathematically, $x_{ss} = g(\text{u})$. Setting $x^* = x_{ss}$ results in output of the (constant) block (f) zero (the system is being evaluated at steady-state x*).

b. Input 'u' is chosen freely, but is chosen such that it is equal to the input at steady-state, mathematically; u = u*; this implies that gradient input to block (J_u) becomes zero; (since u-u* = 0).

Under above assumptions, adding '*g(u)*', removing blocks 'f' and 'J_u' and rearranging the block structure in fig.(1) , we get the block structure shown in fig.(2). Observing this structure it can be considered *IS Hammerstein structure*; it has two blocks, a NL steady-state mapping block, followed by linear dynamic block.

The IS Hammerstein block structure shown in fig.(2) used for the approximation of NL processes. The accuracy of approximation of I/S Hammerstein structure is improved by estimating Jacobian online.

Jacobian is estimated (and updated) based on information of Jacobian basis J_b, reduced state z and input u. Jacobian basis J_b are calculated by SVD analysis of Jacobian data. Jacobian data is collected by taking snapshots of Jacobians over the operating domain ('*input design*') by exciting the NL system with inputs to acquire most information in operating envelope [9].
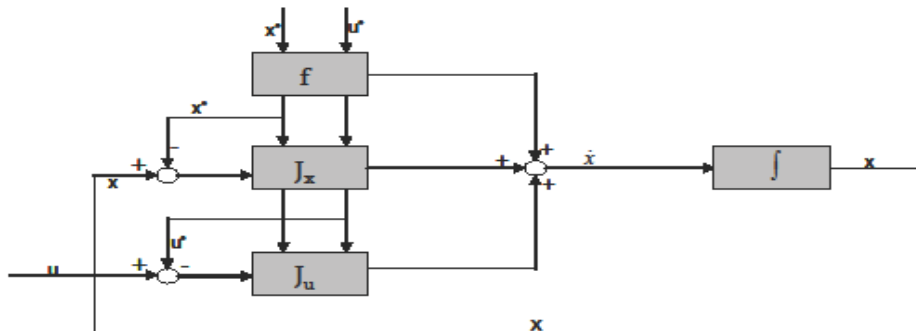
Fig. (1): Taylor series block structure
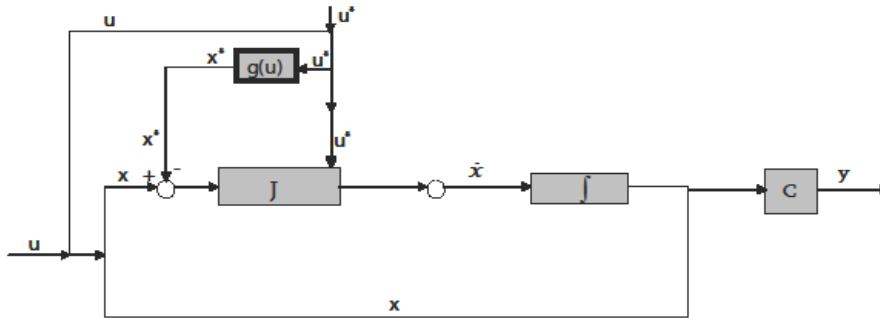
Fig. (2): IS Hammerstein block derived from Taylor series block

Similarly the reduced order states *z* is calculated by transformation matrix (U_1), obtained by SVD analysis of steady-state and dynamic state (snapshot) data, taken over the operating domain. IS Hammerstein structure with update Jacobian (first order approximation) is shown in fig.(3).

Mathematically, for NL case, I/S Hammerstein structure can be given as:

$$\dot{x} = J(x - x_{ss}) + g(u)$$
$$y = Cx$$

(3)

where, J = Jacobian; C = output state matrix ; y = output; xss= g (u) is steady-state, scheduling (implemented by lookup table)

First order approximation of NL system by expansion of Taylor series to I/S Hammerstein structure is proved above.

The approximation accuracy of this block structure can be improved by including the higher order terms in Taylor series. Mathematical form of higher order approximation is given as Eq. (4) while the detailed derivation can be seen in [9]:

$$f(x,u) = f(x*,u*) + \frac{1}{2}\left[\left.\frac{\partial f}{\partial x}\right|_{x*,u*} + \left.\frac{\partial f}{\partial x}\right|_{x,u} \right](x - x*)$$

(4)

Eq. (4) is the approximation of NL system/process *f(x,u)* making use of higher order terms of Taylor series. There are two Jacobian evaluations involved in this approach, an approximation using knowledge at steady-state (*x**) and approximation using current state (*x*) knowledge. The block structure representation of this approximation is shown in fig. (4).

3. Application to High Purity Distillation column

A high purity distillation column is used as benchmark, on which the approximated and reduced model is applied. The benchmark distillation column has following properties; the high purity distillation column has *72* trays, a total condenser and partial reboiler. The case study is a nonlinear system. The thermodynamics of the column is governed by relative volatility constant. The relative volatility for this specific system is *1.33*. The distillation benchmark model has been explained in detail by [10].

There are two types of input signals tested for the validation of approximated full order model.

   i. 'Separation index' (SI) is phenomenon in distillation, when both the input variables (reflux rate (L) and vapour boilup (V)) are given step at the same time.
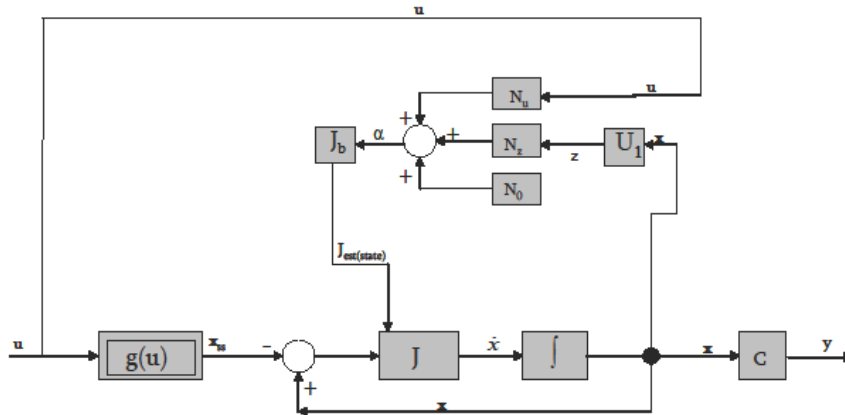
Fig. (3): IS Hammerstein block with Jacobian update (first order)

Where, J =Jacobian; U_1 =transformation matrix to transform full state x to redu ced state z; J_b = Jacobian basis obtained by SVD analysis of s napshots of Jaco bians (within operating domain) ; N_0, N_1, N_2 = the parameters relating Jacobian with reduced-state z, input u and constant.

   ii. 'Effective Cut Point' (ECP) is phenomenon in distillation, when one input variable (reflux rate (L) or vapour boilup (V)) is kept constant and step is introduced to the other input variable. The phenomenon is known to be highly non-linear for high purity distillation column (in process industry).

Both these test signals are used for the validation of full order approximated model. The results of approximation for high NL process (ECP) are shown in Fig (5). The results show that full order approximation model approximates the original NL process model very well. The accuracy of this specific approximation has been calculated, using root mean square (RMS) given by following formula:

$$\varepsilon_{RMS} = \frac{1}{\sqrt{n}} \left\| y - \tilde{y} \right\|_2 \tag{5}$$

where, n = le ngth of the test sequence; y = measured output; $\tilde{y}$ = predicted output by approximated model;

The approximation/prediction error for the IS Hammerstein model structure for ECP change is 0.00021 for the purity of light component at the top (output 1) and is 0.000009 for the light component at the bottom (output 2).
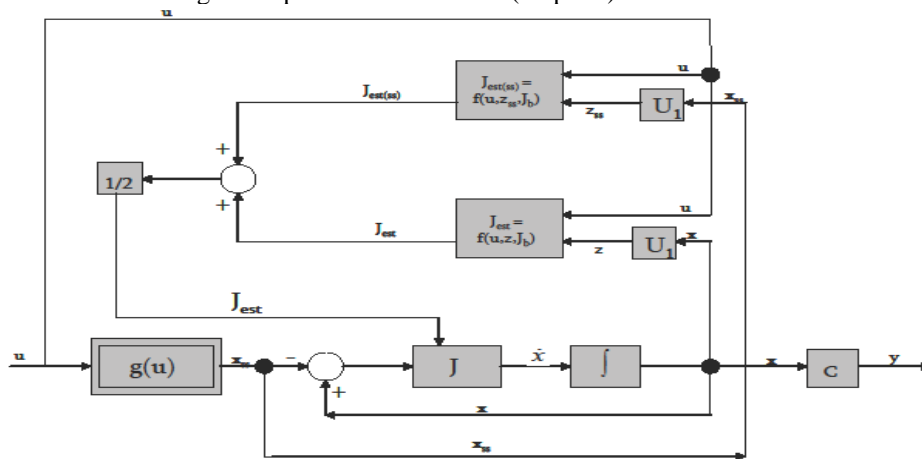
Fig. (4): IS Hammerstein block with Jacobian update at steady & current state (higher order)

4. Conclusions & Future work

In this paper, it is shown that IS Hammerstein structure can be derived from the Taylor series under few assumptions. The approximation model's accuracy is improved by including higher order terms. The approximation results were shown for a high purity distillation benchmark, which are acceptable for the kind of application. Moreover, the full order approximation model provides the handles and opportunity to achieve a reduced order model by reduction in Jacobian and state order. The next step in this research approach is to achieve a reduced order model for the NL processes.
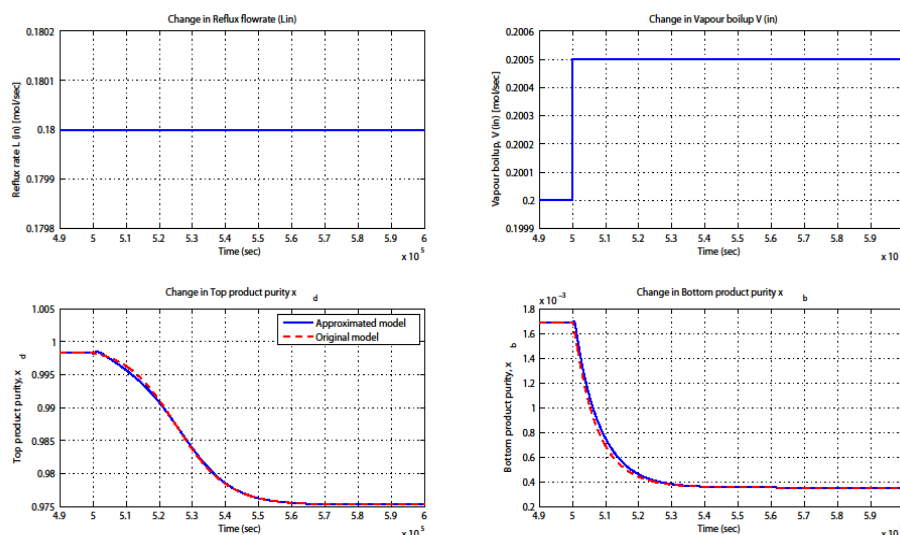
Fig. (5): Results for approximation by full order IS Hammerstein structure (ECP change)

5. Acknowledgements

References

[1]   J. Hahn and T.E. Edgar, In Proceedings of the ACC, (2000), 2864
[2]   H. Aling and S. Banerjee, In Proceedings of the ACC, (1997), 2233
[3]   L.S. Balasubramhanya and F.J. Doyle, Journal of Process Control, No. (10), (2000), 209
[4]   Jogchem Van den Berg, Model reduction for dynamic real-time optimization of chemical processes, Optima, Rotterdam, The Netherlands, (2005).
[5]   J. Scherpen, Systems and Control Letters, No.(21), (1993) 143.
[6]   G. Harnischmacher and W. Marquardt, Journal of Process Control, No. (17), (2007) 539
[7]   G. Harnischmacher and W. Marquardt, Block structured modeling for control, VDI, Dusseldorf, Germany, (2007).
[8]   O. Naeem, A.E. Huesman and O.H. Bosgra, In the proceedings of REDUCIT Symposium, (2008), Frankfurt, Germany.
[9]   O. Naeem, A.E. Huesman and O.H. Bosgra, In the proceedings of ADCHEM, (2009), Istanbul, Turkey.
[10] J. Levine and P. Rouchon, Chemical Engineering Science, No. 27, (1991), 463
[11] E. Eskinat and S.H. Johnson and W. L. Luyben, Use of Hammerstein models in identification of nonlinear systems, AlChE Journal, No. (37), (1991) 255.
[12] S. Norquay and A. Palazoglu and J. Romangnoli, Application of Wiener model predictive control (WMPC) to an industrial C2-splitter, J. Proc. Control, No. (9), (1999), 461.
[13] S.A. Billings and S. Fakhouri, Identification of Nonlinear Systems- A survey, IEEE Proc. Pt.D, (1977), 127.

Process simulation as a domain-specific OPC Unified Architecture information model

Paolo Greppi

Università di Genova Dipartimento DICAT, Via Opera Pia 15 16145 Genova, ITALY,
paolo.greppi@unige.it

Abstract

Currently one of the key challenges for the process industries is to react quickly to market changes; also the process automation and information technology infrastructure should follow seamlessly these changes without requiring frequent reconfiguration or manual adaptation of custom interfaces.

This driver for flexibility also applies to the model-based applications in use in the chemical, petrochemical and polymer industries: soft-sensors, inferentials, reconciliation, model predictive control and operator training systems. It is widely recognized that a Service-Oriented Architecture can match these requirements; in particular the recently released OPC Unified Architecture (OPC UA) specification provides a service-oriented architecture for interoperability in industrial automation.

In this work the prototype implementation of a simple process model as an OPC UA Server is presented; the model is a first-principle soft sensor that calculates a few useful physical properties from the composition of a process stream measured for example by a process gas-chromatograph.

The prototype demonstrates that an OPC UA soft sensor can be self contained, performant, transparent and easy to configure. Furthermore OPC UA is a very general technology that can be used to interface process models with other software components and implement distributed simulation systems; in this sense it challenges the existing interoperability standard CAPE-OPEN.

Keywords: Object-oriented, service oriented architecture, soft sensors, on-line

1. Introduction

According to a recent review [1] the current challenge in the process industry business environment is to react quicker and quicker to the market dynamics and to smoothly adapt the business and technological processes to these changes; but the current information technology infrastructure and in particular industrial process automation is not exactly flexible.

Service-Oriented Architecture (SOA) can bridge the gap, allowing the industry to use existing IT investments and achieve the flexibility required. With regards to the process industry, the recently released OPC Unified Architecture (OPC UA) specification [2] builds on the successes of the established Classic OPC standard [3] for interoperability between enterprise information systems and industrial automation. It provides platform-independence, the capability to expose the semantics of the data and a service-oriented architecture (SOA).

If a standard domain-specific information model based on OPC UA exists for an application, data and functionalities can be made available as services in a vendor-independent fashion; work is under way [4] to standardize information models for

physical device information, analyser devices, plant operation and maintenance, batch control, PLC programming.

Of course also process simulation capabilities can be encapsulated as OPC UA services; this can be done for simple models (such as a material stream or a unit operation) as well as for more comprehensive models (for example model predictive control or plant-wide mass balance reconciliation). What is more, not only the mere calculation of the model can take place through the OPC UA interface, but also discovery of services and configuration of the models.

In this work an existing soft sensor interfacing with the process via Classic OPC was upgraded to OPC UA; to our knowledge this is the first working prototype of an OPC UA server for process simulation.

In the following the modelling capabilities of the soft sensor and the approach used for the implementation are described, and the significance of this technology is commented.

2. The soft sensor

The simple process model which has been wrapped as an OPC UA Server in our prototype is a first-principle soft sensor that calculates the following useful properties for a material stream composed of short-chain hydrocarbons:

- Lower / higher heating value and Wobbe-Index;
- Explodibility-related quantities such as LEL / UEL (lower and upper explosive limits) and LOC (limiting oxygen content);
- Density, compression factor and temperature / pressure dew-point and bubble-point with an equation of state.

The typical application of this soft sensor is to provide inferential measurements when a process gas-chromatograph or other process analytic device is available and measures the composition of a process stream at regular intervals, such as in natural gas processing, transport and distribution, polymers and petrochemicals production.

This kind of application in the natural gas sector is well-known and regulated by international standards [5] [6], with an ongoing evolution towards the application of more sophisticated equations of state [7]. In the polymers and petrochemical sectors the interest is more in traditional cubic or statistical equations of state [8].

3. OPC UA server implementation

A key element of OPC UA is the capability to expose the semantics of the data: a temperature is different from a pressure; a variable reserved for the storage of a temperature measured in the field is conceptually different from a variable used to store a calculated temperature; the temperature, pressure and composition for a given process fluid belong together; the capability to compute a model is common to all different models.

To express these relationships, Variable and Objects types can be defined extending the built-in OPC UA types. The transposition of the above examples in our implementation are: we included an EngineeringUnit property to the base variable type to distinguish different physical quantities; we use the AccessLevel field to make a calculated variable read-only; temperature, pressure and composition for a process fluid are grouped in a "stream" abstract object type; all concrete stream models are subtyped from an abstract modelBase object which has pure virtual methods `Calculate`, `Reset` and `ResetErrors`.

Process simulation as a domain-specific OPC UA information model

For the implementation of the OPC UA server the first and currently only commercially available high-level C++ software development kit (SDK) [9] has been used. The thermodynamic calculations and the empirical correlations are implemented using LIBPF (LIBrary for Process Flowsheeting), a process flowsheeting and modelling tool arranged as a C++ library [10].

The object-oriented design and portability of the LIBPF library fit with the philosophy of OPC UA. This reduced the development time and resulted in an extremely lean implementation: coding the LIBPF / OPC UA bridge required a mere 5000 lines of C++ code. Key used LIBPF functionalities were:

- units of measurements: all quantities in LIBPF carry a unit field which was used to populate the EngineeringUnit property of the corresponding OPC UA variables
- reflection: the ability to programmatically inspect and use types at run-time)
- dynamic type manipulation via the object factory without reference to Windows-specific object broker mechanisms such as COM or .NET
- object persistency.

Both the OPC UA SDK as well as LIBPF library support Windows and Linux, so the prototype server was successfully tested in both environments. If desired the soft sensor could even be integrated directly into the automation device, whatever real-time operating system it is running, thanks to the portability of both libraries.
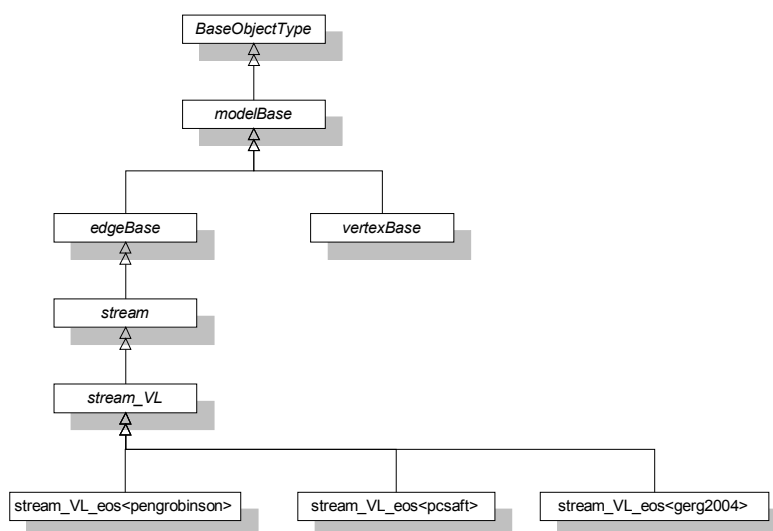
Figure 1 - OPC UA types from the implemented process simulation information model

An exemplificative subset of the implemented type hierarchy, expressed in the OPC UA (UML-derived) graphical notation is shown in figure 1.

Shaded boxes indicate object types; object type names in italic indicate that the respective objects are abstract i.e. they express an incomplete set of characteristics, and can not therefore be instantiated but only contribute to the definition of a more complex type. The double-arrows connection indicate the types are in a HasSubtype relation.

The instantiated objects exist in some form in the memory of the OPC UA server, as "live", instantiated objects of certain types. OPC UA clients can browse and access the objects, and execute methods on them. Either the server has a fixed set of live object instances, determined at compile-time, or it can expose a node creation service (note: this functionality is not yet implemented in our prototype). In this latter case clients can browse the hierarchy of available types known to the server, and request that one of them should be instantiated at run-time; this scenario makes sense for a simulation information model, where the types match virtual, non-physical entities, or during sensor/controller/actuator configuration for an hardware-related information model.

If the server has to serve node creation service requests, and the clients require that server-side objects "survive" a restart of the server process, the server needs a facility to persist the live instances. For this purpose LIBPF offers a built-in persistency mechanism to an external relational database.

The persistency mechanism is also useful if the server exposes a large number of live instances, which are infrequently accessed: with persistency unused objects can be cached to disk reducing the consumption of resources.

3.1. Comparison with Classic OPC

A similar soft sensor had been previously implemented with a Classic OPC DA interface. The main difference with the new OPC UA server prototype is that the old-style soft sensor was an OPC client: the OPC tags required to store the results of the calculations had to be manually created on the Classic OPC server of the DCS; next the tag names had to be manually configured with the soft sensor using an XML file. Any change to the OPC server configuration would require a manual update of the client configuration; this is where the limited flexibility of a non-service oriented architecture becomes evident.

Another difference is that the Classic OPC client contained a bare-bone state machine to schedule the repeated execution of the calculations, but this state machine was opaque and again only configurable with a custom interface (in that case, values stored in the operating system registry). The orthodox way to achieve this in OPC UA (although not yet implemented in our prototype) would be to use the State Machine information model and the Program abstraction provided by the specification.

In the resulting OPC UA soft sensor there would be then no need for configuration files, registry settings or a maintenance user interface: the configuration could be performed using any OPC UA client, requiring no client-side configuration thanks to the discover capabilities.

The preferred use case would be to have a higher level OPC UA server, i.e. the one provided by the DCS vendor, controlling the soft sensor according to the "chained server" pattern. The elected OPC UA client to perform the set-up in this case would be the generic control system maintenance user interface itself.

3.2. Off-line application

Once the process simulation domain is exposed as an OPC UA information model, and objects such as material streams, unit operations and flow-sheets can be instantiated, the models could be invoked also for off-line use, without any interaction with plant data.

OPC UA would then be used as a communication protocol between a calculation kernel and any client, be it a standalone user interface or a web server or a light client on a mobile device. Although the OPC UA specification has been created for distributing applications, and designed for efficient communication when the client and the server

sit on different machines, the architecture can be simplified with the two coexisting on the same workstation.

3.3. Comparison with CAPE-OPEN

It is interesting to compare the OPC UA off-line application scenario above with CAPE-OPEN [11], the de-facto standard for interoperability of CAPE (Computer-Aided Process Engineering) applications or components based on COM middle-ware. The supported operating systems, architecture and network layouts, configuration requirements and performance issues are reviewed for both technologies.

The equivalent to the CAPE-OPEN simulation executive would be a specially designed OPC UA client. This client would act as host for complete flow-sheets, unit operation or thermo models that would be discovered, browsed and controlled on either local or remote servers using OPC UA service oriented architecture. Any complex model could in turn be a host for a thermo model pulled from another server, using the chained OPC UA server architecture.

CAPE-OPEN is based on COM middle-ware and therefore bound to the Windows platform, whereas OPC UA is not tied to a specific operating systems, which is very desirable both for low-end (real-time, embedded systems) as well as for high-end (servers) environments.

CAPE-OPEN can connect to remote objects using the distributed COM (DCOM) technology and its proprietary protocol, but not across a WLAN since a transparent http-based SOA protocol is not supported. OPC UA on the other hand can operate as a web service using SOAP/HTTP transport protocol, making it possible to access services across a firewall.

CAPE-OPEN requires that all object types (unit operation, reaction or thermo) are assigned a Class ID and registered with the operating system as COM objects - the instantiation of the objects is performed by calling the operating system CreateObject function. OPC UA on the other hand is zero-configuration at the client side, as all server capabilities can be discovered and browsed directly from the server.

CPU performance issues are very important in particular for the thermo interface, which lies at the heart of any process model. Real-world, quantified profiling data are not available at the time of writing for our prototype, apart from the qualitative observation that the overhead caused by OPC UA was not significant in this specific case.

On this topic there are certain general considerations that apply. The recent benchmark data in [12] indicate that the worst-case communication latency (round trip) for an OPC UA client-server configuration on a typical laptop is about 0.5 ms for 10 floating-point values (comparable to a thermo remote procedure call) and about 50 ms for 10000 values (comparable to a unit operation remote procedure call). Based on these figures a 10% performance degradation bound for off-line CAPE applications based on OPC UA can be achieved provided the design of the interface makes sure that whenever 10 variables are transferred, the associated calculation overhead is in the range of 5 ms, while whenever 10000 variables are transferred the calculation overhead is in the range of 500 ms. The technique of grouping remote procedure calls is well-known in the automation world and is integral already to the legacy, Classic OPC DA (data access) standard. From a process modeling point of view, it is perfectly feasible for unit operations which are simultaneously solved to call the thermo engine once to perform a number of flashes and physical property computations in a single remote procedure call.

4. Conclusions

We have presented a first-of-a-kind soft sensor prototype implemented as an OPC UA server. To fully exploit the OPC UA promise for portability, flexibility and performance the modelling calculation kernel used to implement the process simulation information model should be modern, designed around an object-oriented paradigm and portable; the C++ LIBrary for Process Flowsheeting satisfies these requirements and allowed fast prototyping of the application.

The soft sensor calculates certain properties such as heating value, explodibility, density, compression factor and dew-/bubble-points for hydrocarbon mixtures. The latter either with industry-standard cubic equations of state, or with more sophisticated and accurate equations of state for specific mixtures (GERG-2004) or finally with the PC-SAFT statistical equation of state. To make the prototype complete would require implementing the node creation interface, the State Machine information model and the Program OPC UA interface; in this way it would be possible to completely steer the soft sensor using OPC UA for set-up and configuration as well as for communication at run-time.

OPC UA, if accepted as an industry-wide standard, is a potentially revolutionary technology that could reshape the automation industry and its business model.
The challenge for the CAPE community now is to leverage CAPE-OPEN and define a basic, standard, expandable, designed-for-performance OPC UA information model for process simulation.

References

[1] Credle R., Akibola V., Karna V., Pannerselvam D., Pillai R. and Prasad S. "Discovering the Business Value Patterns of Chemical and Petroleum Integrated Information Framework" IBM redbooks 2009 ISBN 0738433136
[2] OPC Unified Automation Specification 1.01 Parts 1-9, OPC Foundation 2009
[3] OPC Data Access Specification, OPC Foundation 1996
[4] Wolfgang Mahnke, Stefan-Helmut Leitner and Matthias Damm "OPC Unified Architecture" Springer-Verlag 2009
[5] INTERNATIONAL STANDARD ISO 6976:1995 "Natural gas — Calculation of calorific values, density, relative density and Wobbe index from composition"
[6] INTERNATIONAL STANDARD ISO 12213:2006 "Natural gas -- Calculation of compression factor -- Parts 1, 2 and 3"
[7] O. Kunz, R. Klimeck, W. Wagner and M. Jaeschke, "The GERG-2004 Wide-Range Equation of State for Natural Gases and Other Mixtures", GERG TM15 2007
[8] J. Gross and G. Sadowski, "Perturbed-chain SAFT: An equation of state based on a perturbation theory for chain molecules", Industrial & Engineering Chemistry Research, vol. 40, pp. 1244-1260, Feb 21 2001
[9] C++ based OPC UA Server/Client SDK V1.0.0, Unified Automation 2009
[10] P. Greppi, "LIBPF: A LIBRARY FOR PROCESS FLOWSHEETING IN C++" Proceeding of the International Mediterranean Modelling Multiconference 2006, pp. 435-440
[11] CAPE-OPEN Laboratories Network, CAPE-OPEN Open Interface Specifications - Thermodynamic and Physical Properties Version 1.0, 2002
[12] Intel, AscoLab, UnifiedAutomation, Reducing Product Development Effort for OPC Unified Architecture (white paper) (2009)

A New Modeling Environment Based on Internet-Standards XML and MathML

Stefan Kuntsche[a], Harvey Arellano-Garcia[a], Günter Wozny[a]

[a]*Chair of Process Dynamics and Operation, Berlin Institute of Technology
Sekr. KWT-9, Str. des 17. Juni 135, D-10623 Berlin, Germany*

Abstract

A new modeling environment called MOSAIC is presented that tries to further narrow the gap between model presentation in literature and model implementation in a computer-readable form. The new modeling approach focuses on the notation element and allows simpler reuse of equations and equation systems

Keywords: Object-oriented, re-use, flow-sheeting, notation

1. Introduction

There are many mathematical models for the description of phenomena, units and processes of process engineering. Such models differ according to the application in the degree of detail. Their is a large quantity of models, however, that is based on the conservation laws of mass, energy, and impulse, the thermodynamic equilibrium correlations, the equations of state and the approaches for transport phenomena. Models that use such equations often result in large differential algebraic equation systems that have to be solved by the use of computer software. To cope with the complexity, and to shift the focus away from numerical programming to the model equations, a great number of computer languages and mathematical environments have been developed.

Object-oriented languages such as Modelica [1] and gProms® [2] aim to avoid reimplementation of model parts by means of inheritance and modularization. Powerful mathematical environments, such as MapleSim® allow working with two dimensional symbolic equations to minimize the visual gap between the model presentation in literature and the model specification for the evaluation by a computer.

The exchange of knowledge in publications is based on model equations given in symbolic two dimensional mathematic expressions as shown in figure 2-top. Usually the publication provides a notation that contains a description for the symbols used in the model equations. In this work a new modeling environment called MOSAIC is proposed that tries to further narrow the gap between literature and calculable model by considering the notation as separate and mandatory model element belonging to equations and equation systems.

2. Motivation for a new approach

2.1. Notations and Variable Names

Notations are the fundamental modeling element in MOSAIC. They are used to give the symbols appearing in the equation systems a meaning that will be visible to the user throughout the entire modeling process. Notations in MOSAIC are a direct projection of notations found in literature: The notation belongs to a mathematical model. All symbols are listed systematically. Each symbol is followed by a short description providing its meaning in the context of the mathematical model that both the symbol

and the notation belong to. As it is good practice in literature, the notation in MOSAIC divides up the symbols into groups according to their position and their functionality. These groups are base names, superscripts, subscripts, and indices.
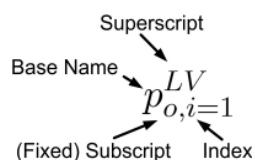
Figure 1. Elements of a variable name in MOSAIC

2.2. Proximity to literature

As mentioned above, in literature, the model equations are presented in two-dimensional mathematic expressions. Thus, a variable name can consist of several symbols including subscripted and superscripted elements. To reflect this in MOSAIC, the variables are distinguished by their name which consists of symbols on the base line and which may have several superscripts, one fixed subscript, and several indices taking values from 1 to a maximum value, see figure 1. It is assumed here that this way of naming variables is more readable than a straight sequence of characters as it is used in conventional programming languages like Fortran, Modelica or Matlab®.

2.3. Modeling on a web application

The proposed modeling environment is implemented as a Java Applet. Thus, it is possible to use it from a conventional web browser. The calculation of the models can be expensive. To avoid bottlenecks in calculation and while providing a full calculation capability at the same time, the calculation can either be done on the modeling server or at the local computer.

2.4. Motivating Example

In the following it is shown how a model given in [3] is transferred from a paper into MOSAIC, and solved using the BzzMath numerical library [4].
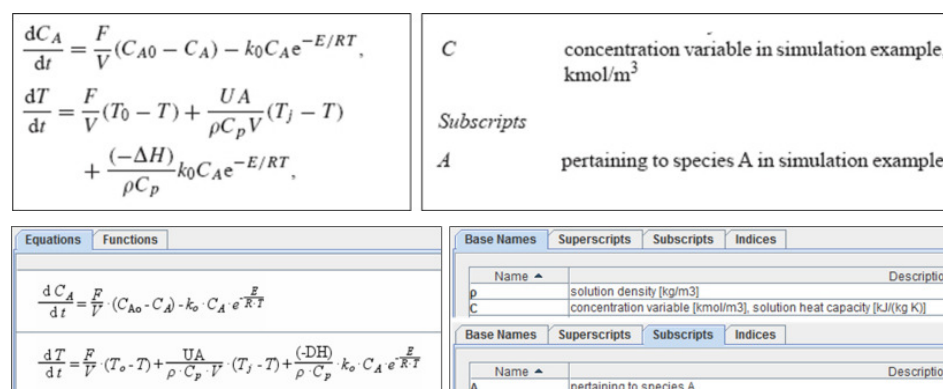
Figure 2. Equations and notation of the example from [3]. Top: as they are presented in the article, Bottom: after they have been transferred to MOSAIC.

In contrast to other modeling environments, a notation element is created before any equation is written. Every equation, equation system and function must use a notation. Equations are modular elements and can be connected to an equation system. Figure 2-

bottom shows the view on the equation system and on the notation within MOSAIC. It can be seen that the model in the literature is reflected directly in the modeling environment.
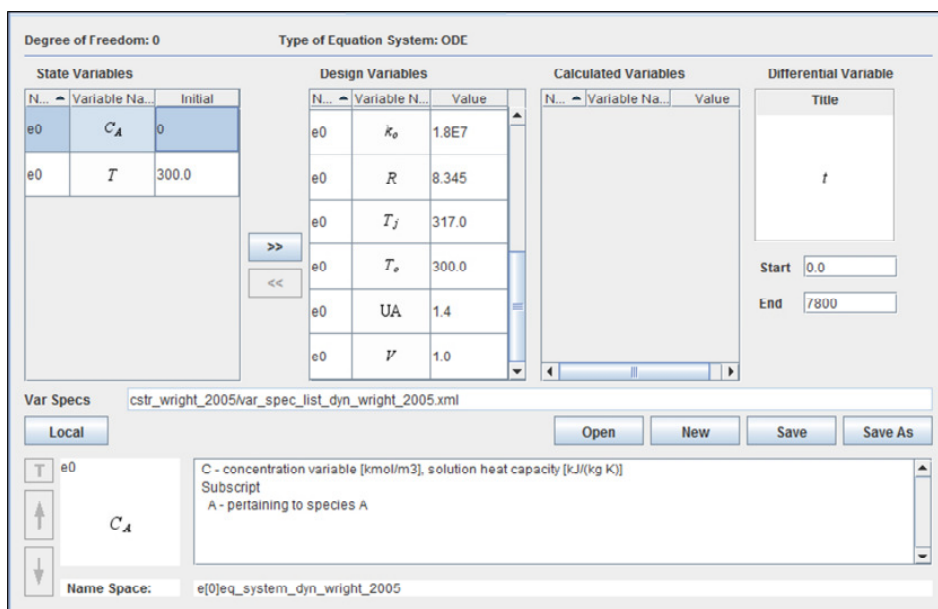
Figure 3. Analysis step of the equation system. The notation information is displayed for the selected variable.

The mathematical equations are stored in presentation MathML and can be parsed so that the mathematical content of the equation system can be analyzed. The results of such an analysis for the given example are shown in figure 3. The degree of freedom and the type of the equation system (i. e. NLE, ODE or DAE) are displayed. The variables are classified in tables and it is possible to assign values to them. Depending on the values given here, different problems are specified. Therefore it is possible to store value lists for the current set of variables independently from the equation system and from the other evaluation information. Once the variable values are specified, the information for the given problem is complete. Now code can be generated in different languages and for different purposes. For this example, C++ code has been generated that makes use of the BzzMath numerical library [4]. In this example, the results have been produced by compiling and executing the code on the modeling server. The code and the results are shown in figure 4.

2.5. Modularity and reuse

Writing and testing equations and models afford time and personal resources. It makes sense to re-use equations and equation systems that have been well established. Thus the focus is put on equations that are different from existing models, while ready-made equations can be used for standard model parts. The reuse of small model parts down to the level of equations is reflected in MOSAIC. This is done by establishing the equation as an independent model element. The model element equation system is a combination of several equation elements. Several equation system elements can be put together to create a new equation system. Furthermore it is possible to use functions and interfaces.

```
/* evaluate the function values */
f[1] = (e0_F)/(e0_V)*(e0_C_Ao-e0_C_A)-e0_k_o*e0_C_A*pow(e0_e,-(e0_E)/(e0_R*e0_T));
f[2] = (e0_F)/(e0_V)*(e0_T_o-e0_T)+(e0_UA)/(e0_greek_rho*e0_C_p*e0_V)*(e0_T_j-e0_T)+((-e0_DF
```
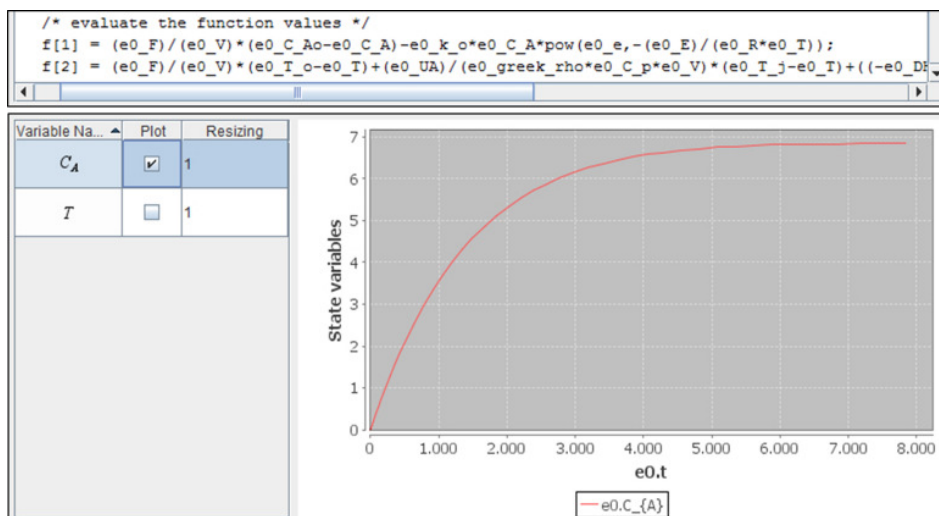
Figure 4. View of the generated code and plot of the results within the modeling environment.

3. Model elements of MOSAIC in detail

3.1. Equations

The equation is the simplest and smallest element in MOSAIC. It must use a notation, it must have a description, and it has a MathML expression containing the mathematical content. It may also use a parameter list to point out variables that should be treated as global parameters. In the current version of MOSAIC a Latex subset is used to enter the symbolic mathematical content. The Latex code is then translated to create the MathML expression.

3.2. Equation Systems and Variables

Equation systems are the central modeling object in MOSAIC. They are created by adding together equations, functions, and other equation systems. On the other hand equation systems can be evaluated by specifying the design variables and other problem specific information. Thus, they are the center piece between the nucleate pieces of information in equations and concrete modeling tasks that give you the information you need.

3.3. Functions

Functions are very useful if values have to be calculated but it is unwanted to introduce additional variables or equations to the equation system. Applications are e. g., the calculation of the enthalpy or phase equilibrium constants. Functions in MOSAIC have one output value and several input values.

3.3.1. Appliances of a Function

In general the notation of a function is independent from the notation used in the equation system. Thus, there must be a mechanism to define which of the variables in the equation system are the input values and which is the output value. In programming languages, the input values are assigned according to the order in which they appear in the function call. The modeling in MOSAIC, however, is focused on the variable names and the concept of synonyms. Thus, the function is applied in assigning the variable names of the input variables and the output variable of the function explicitly to the variable names of the corresponding variables in the equation system. The advantage of

this kind of applying the function is that the focus is really kept on the physical meaning of the variable. During this connection step, the notation information of all variable names is present in the modeling environment. Errors from handing over variables in the wrong order are thus eliminated. It is possible to provide the output variable and the input values with an index so that the corresponding appliance is automatically extended to all index values.

3.4. Connectors

Connectors are modeling objects that translate between two different notations. Suppose there is an equation system named 'I' that uses notation 'A'. To use 'I' in an equation system using notation 'B' you need to specify a connector that provides all pertinent variables in notation 'A' with a synonym complying with notation 'B'.

3.5. Interfaces

Interfaces provide a norm for the shared use of variables. In principal, interfaces represent an independent list of variable names that are either allowed or expected to be shared with another modeling element. They are used in ports and streams to further extend the reusability of existing units.

3.6. Ports and Streams

3.6.1. Units

Large equation systems may represent a unit in the physical engineering sense. It makes sense to restrict the accessibility of the variables of such equation systems and to apply standardized output interfaces in a second step. This way all incoming and outgoing material streams a process unit model could have exactly the same notation. But it would also be possible to define interfaces that represent control input and output, etc. To provide this kind of standardized access points, MOSAIC allows specifying ports. Two units that have ports can be connected by streams.

3.6.2. Ports

Ports belong to an equation system and cannot exist independently. They have a distinct name and an indication of the interface they use.

3.6.3. Streams

A stream can combine exactly two ports. It has a distinct identification number, the names of the two ports it connects and it indicates the interface it uses.

4. Connection Strategies

One special characteristic resulting from using notations is the implicit connection of variables. This approach minimizes the modeling effort and avoids the introduction of additional equations. The advantage of this approach is motivated in the following.

In traditional object-oriented modeling languages equations 'I' and 'II' might be given as

```
eq_I: 0 = a * (1 - x)
eq_II: 0 = b * (y - pow(x,2))
```

To combine both equations to a system one has to add the following information:

```
eq_I.x = eq_II.x
```

Such a proceeding, however, introduces a further equation to the system. Additional equations of that kind enlarge the equation system and have to be eliminated by

symbolical math calculations in later steps. In MOSAIC using the connection mode 'integrate', by contrast, variables with the same name are automatically matched, if they use the same notation. This allows the use of different equations using the same notation without having to specify connecting data structures for every single variable. If some equations use a different notation, however, it is necessary to establish the information about which variables of the two different notations belong together. This synonym information accounts for all the equations or equation systems that use the two notations involved. Again, matching of single variables is avoided. The effect of matching variables with the same name is unwanted in many cases. For these cases, the connection mode 'encapsulate' allows to set equal only those variables that have been assigned to one another specifically by a connector. Furthermore, it is possible to assemble equation systems via standardized interfaces, which also allows to model ports and streams. Ports can also be used to describe connections to standard interfaces. In this context, a code generator has been implemented that allows to create CHEMCAD® User Added Models in C++. These models interact with the flow sheeting software CHEMCAD® by reading out stream variables and by interacting with a custom user interface. This technique will be applied in a more general way in further steps of development. MOSAIC ports and functions will also be used to describe calls to standard interfaces, as they are defined by CAPE-OPEN, and the code generation will be extended to comply with these standards in the supported languages. Another ongoing step represents the introduction of engineering units into the modeling concept. Variable units will be specified in the MOSAIC notation element allowing the use of all connection techniques described above with automated unity conversion.

5. Conclusions

A web-based modeling environment has been presented that follows a modeling approach based on two-dimensional mathematic expressions. In this approach the notation is a separate, modular element of the model. Reuse of equations is made easier by connection procedures that take account of the notations used. The descriptive information is present throughout the entire modeling procedure.

6. Acknowledgements

References

[1] Elmqvist, S. E. Mattsson, M. Otter, ESM'98, (1998)
[2] M. Oh, C. C. Pantelides, Comp. chem. Eng., 20(1996) 611-633
[3] R. Wright, C. Kravaris, Chem. Eng. Sci., 60 (2005) 4323-4336
[4] D. Manca, G. Buzzi-Ferraris, Comp. Aided Chem. Eng., 25(2008) 587-592
[5] http://jeuclid.sourceforge.net/
[6] http://www.jfree.org/jfreechart/
[7] http://www.gnu.org/software/gsl/

Solving CAPE models from Microsoft Office applications

Karim Alloula[1], Jean-Pierre Belaud[1], Jean-Marc Le Lann[1]

1 Laboratoire de Génie Chimique (CNRS UMR 5503), INPT-ENSIACET
4, allée Emile Monso, 31432 Toulouse Cedex 04, France, Karim.Alloula@ensiacet.fr –
JeanPierre.Belaud@ensiacet.fr – JeanMarc.LeLann@ensiacet.fr

Abstract

Even if today problem solving environments may include editing capabilities, CAPE engineers keep on exchanging their simulation models and results through *Microsoft Office* documents. In order to maintain consistency between data written in textual documents and data handled by calculation systems, we suggest to make *Office* documents CAPE software frameworks. The *eXMSL for Microsoft Office* plug-in forwards equations present in documents to a calculation system which solves them. Results are available as equations which can be inserted anywhere in the document. Such an approach seems to be fruitful mainly because it is based on software components, a paradigm now widely spread inside our CAPE community.

Keywords: Framework for CAPE, component based software architecture, CAPE-OPEN numerical solvers, computer algebra

1. Introduction

The activity of CAPE engineers and researchers consists of defining problems, solving them, and exchanging those problems and their solutions. Today, presentation and exchange tasks mainly use electronic documents. Most of the documents describing the problem to be solved and its given solution are *Microsoft Office* documents. So, the CAPE engineer or researcher has two parallel activities:

1. Define and solve problems using general-purpose environments, CAPE dedicated products, mathematical libraries or in-house codes;
2. Present these problems and their solutions in *Microsoft Office* documents.

Naturally, these highly complementary tasks take place in distinct working environments: the tools are different and the documents are different. The consequences of such lack of interoperability are well known. First, the end user must have access to and become familiar with two distinct computer environments (each with their own operating systems, applications, file formats…). Second, they have to ensure consistency between the documents handled in those distinct environments during the life cycle of the case study.

Those drawbacks having been noticed, part 2 wonders how both editing and simulating tasks could take place in an integrated CAPE software environment. Our proposal is to make *Microsoft Office* documents active: CAPE models inside them serve as well for documentation as for execution purposes. Part 3 introduces the *eXMSL Microsoft Office* plug-in, a try to make office documents and CAPE solvers interact through mathematical expressions. Usage of this plug-in is illustrated in part 4, where we show how some thermodynamic model can be edited and solved from a *Microsoft Word* document. Part 5 concludes this paper, defending the interest of using office documents as the presentation layer for some of our activities.

2. Working within an integrated CAPE software environment?

Software editors have already brought solutions to the problems coming from considering edition and calculation as two parallel activities. They either increased the editing capabilities of their computing environment or increased the calculation capabilities of their text editors.

Computer algebra systems, such as *Mathematica* or *Maple*, propose to enter problem specifications and output solutions in notebooks, using both mathematical notations and rich text formatting.

CAPE simulators become more open systems. Input data, handled by a friendly graphical user interface, remain in a proprietary format. But some outputs may be provided also as office documents (HTML, PDF or *Microsoft Office* files).

Most of the time, mixing edition and calculation activities is achieved by connecting an edition solution to some calculation system. *Scientific Workplace*, a technical editor from McKichan Software, Inc. allows users, not only to compose scientific documents, but also to interact from them with a built-in computer algebra system *MuPAD*. In many domains, *Microsoft Excel* plays the role of the editing platform where users enter and receive data, whereas *MATLAB* plays the role of the calculation engine. In the CAPE community, according to [4] , *Microsoft Excel*, *MATLAB* and direct coding are often combined to perform some "custom modelling" tasks, which "are not adequately handled by flowsheet simulators alone". *ProSim Simulis Thermodynamics* [3] , a thermophysical calculation server, is one of such CAPE applications which can be driven from *Microsoft Excel*. The CAPE-OPEN standard provides specifications to such cooperations between CORBA or Microsoft software components.

Proprietary solutions, or open standards, make it possible to publish, solve and exchange the aforementioned problems. However, today, most of us keep on working with both *Microsoft Office* documents and raw input/output files from various CAPE software environments. This study opens the doors to an alternate solution: to make *Microsoft Office* documents CAPE software frameworks! *Office* applications already integrate text, equation and array editing capabilities; sophisticated graphical capabilities can be added through .NET or COM applications; the next step we suggest here is to integrate the rich solving features and legacy models provided by any today CAPE simulator**.**

3. The eXMSL for Microsoft Office plug-in

3.1. Methodology

We propose here the CAPE engineer to work almost entirely with office documents, enriched, modified and improved in only one place, and transferred to the co-workers as a single entity. These documents include the basic additional resources for:

- Extracting the mathematical model from the edited equations;
- Communicating with solvers providing the equations, and obtaining the solutions;
- Updating the document with the calculated results.

According to an equation oriented approach, the office documents include the textual and mathematical model descriptions in an XML format, some dedicated XML dialect being used for mathematical expressions. Equations can involve calls to common thermodynamic functions. Numerical and symbolic calculation tools and thermodynamic servers are accessed as software components. Closing the equation

editor generates a Content MathML 2.0 document, transmitted to *eXMSL Evaluation Server*, our in-house solver [1] , based on CAPE-OPEN numerical specification. This component implements a co-operative model combining computer algebra and numerical calculation for CAPE simulation [2] . The evaluated expression is sent back to the equation editor to be displayed in the office application.

Our objective is quite close to the one in [7] , but the means are different. Here, the edition tasks take place in an existing office application, not in a browser.

3.2. Standards and technical choices

XML has become a de-facto standard for office documents. This change brought several benefits, such as opening the door to version control of office documents at a fine-grained level [6] . Recent versions of office applications use new XML-based formats: *OpenOffice.org* uses ODF (Open Document Format), and *Microsoft Office 2007* uses OOXML (Office Open XML). Both formats have been accepted as international standards. In such a context, we decided to apply the aforementioned methodology within *Microsoft Office 2007* applications, first because of our skills, and second because *OpenOffice.org* usage is limited within the CAPE community.

Selecting OOXML, and not ODF, as the XML format for office documents, has strong consequences on how we deal with mathematical expressions. Both formats -ODF and OOXML- are mainly interested in presentation, not in semantics: 3/4 is considered as 3 over 4, ignoring that 3/4 is a rational number. Because *eXMSL Evaluation Server* works only with a semantics representation of mathematical expressions, communication between office applications and our calculation components requires at least one transform to be applied. In ODF, mathematical expressions are MathML 2.0 documents, whereas in OOXML mathematical expressions are OMML (Office MathML) elements. Because *eXMSL Evaluation Server* works only with MathML 2.0 elements, communication between *Microsoft Office 2007* applications and our calculation components requires at least one transform to be applied. A technical solution to this format conversion problem is to chain two XSL transforms: one in charge of converting OMML to Presentation MathML 2.0, and one in charge of converting Presentation MathML 2.0 to Content MathML 2.0 [5] . The reverse conversion problem is treated in a similar way.

For the time being, only *Microsoft Office Word 2007* includes the equation editor facility. *PowerPoint* will integrate this tool, as well as the OMML format, in its 2010 version. So, the methodology applied to *PowerPoint* will be the one applied to *Word*.

Within *Excel*, things are quite different. By default, CAPE models may only appear as numerical function evaluations. To deal with models made of numbers, function evaluations and symbols too, we have to answer to additional questions:

- How symbols should appear inside *Excel* cells?
- How built-in numerical functions should be overloaded by other functions which arguments may be symbolic expressions?
- Which transforms should be applied to formulas, coded as OOXML elements in *Excel*, to generate the suitable MathML documents for *eXMSL Evaluation Server*?

Some answers to these questions will be given in a later work, but here we generalize and apply some simple and powerful *Excel* features to deal with evaluation of mathematical expressions input through the equation editor. Whenever a mathematical expression starts with an equal sign, it is replaced by the result of its evaluation when closing the equation editor. This behavior follows the semantics of *Excel* formulas, and maintains consistency between *Office* applications. Whenever a mathematical

expression does not start with an equal sign, it is displayed with no change when closing the equation editor.

At a first glance, from the point of view of the office application, two classes of equations have to be distinguished: mathematical expressions which change from the equation editor to the document, and mathematical expressions which do not change from the equation editor to the document. In fact, some expressions –mainly assignments-, even not changing when closing the equation editor, may be evaluated by the calculation application. Consequently, the office application must allow the user to enter either active equations which are to be evaluated, or passive equations which are not forwarded to the calculation application. Among active equations, the subclass of formulas requires the default behavior to be changed, input expression being displayed in the equation editor and output expression being displayed in the document.

3.3. Software architecture
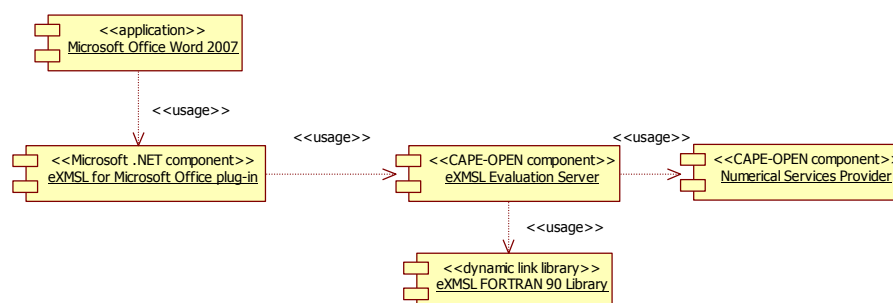
Figure 1 - eXMSL for Microsoft Office plug-in (UML 2 component diagram)

Figure 1, gives the software architecture of the edition and calculation application obtained after introducing the *eXMSL for Microsoft Office* plug-in. The *"server-side"* software components (evaluation server, numerical services provider and F90 library components) involved in evaluation of mathematical expressions have first been introduced in [1] . Between the office application and the evaluation server, the "client-side" .NET component, *eXMSL for Microsoft Office plug-in*:

- Catches some events fired by the office application (open or save a document, and open or close the equation editor) which are relevant for the calculation application;
- Applies XSL transforms to mathematical expressions coded in an XML dialect;
- Calls the evaluation server with a Content MathML 2.0 expression and gets the result as a Content MathML 2.0 expression too.

eXMSL Web Server could be inserted between the office components and the calculation components but the actual prototype does not intend to tackle secure multi-user access through the Internet.

4. Case study: active Microsoft Word documents for thermodynamics calculations

4.1. Presentation
The feasibility of such an approach has been validated on several CAPE models –non linear equations, differential algebraic equations and non-linearly constrained

optimization problems-, making use of our prototype and a market component off the shelves: *ProSim Simulis Thermodynamics*. The *eXMSL for Microsoft Office* plug-in has been coded inside *Microsoft Visual Studio .NET 2008* as a *Microsoft Word plug-in* project. This project references the *ProSim Simulis Thermodynamics* COM server and the *eXMSL Evaluation Server* .NET component. C# was used as the programming language.

To illustrate the use of the *eXMSL for Microsoft Office* plug-in, let's describe how a *Microsoft Word 2007* document can present the SRK equation of state, and apply it to a mixture for some thermodynamic calculations.

ProSim Simulis Thermodynamics is used only for getting some pure component properties, which do not depend on any variable: critical pressure, critical temperature, Pitzer acentric factor, and binary interaction parameters. All the other expressions involving symbols appear as *Microsoft Office* equations. This way, calculations are handled with a very good accuracy, because all the symbolic transformations steps, such as differentiation or integration, are applied to mathematical expressions made of numbers, symbols and arithmetic operators, without any call to a numerical evaluation routine. Of course, numerical evaluations of routines depending on symbol values could be used when modelling, for example for calculating pure component heat capacities which depend on temperature. But, in this case, an additional strategy for differentiation or integration of numerical evaluation routines would be required: provide extra code for analytical evaluation, or apply approximation schemes.

A modelling session starts when opening the *Microsoft Word 2007* application. A *Simulis Thermodynamics Calculator* component is created and displayed. Through a graphical interface, the end-user can select the components in the mixture, the thermodynamics model (SRK) and the binary interaction parameters. Each component is then identified by a unique number. When the business application is closed, calls to *Simulis Thermodynamics* routines, from the office application, use those unique identifiers.

From the point of view of the office application user, apart the initial opening of *Simulis Thermodynamics Calculator*, the *Microsoft Word 2007* interface is enriched only at one place: the Office ribbon. The original icon starting the equation editor remains the entry point for inactive equations, which will not be forwarded to the calculation application. An extra icon opens the same equation editor for typing active equations, which will be forwarded to the evaluation engine. When expressions are formulas, starting with an equal sign, the evaluated expression is displayed inline the document. Because the formula entered in the equation editor may be modified later, it is stored in the office document as an invisible OMML element.

4.2. Results and discussions

From the latter part, one can consider that the main objective of this work has been achieved. Model specification and simulation results appear in only one place, and remain consistent through all the life cycle of the office document. This comes from the tight and seamless integration of calculation facilities in an office application.

This approach brings not only more consistency, but can also make the CAPE end-user more efficient. Because the connected calculation application is a computer algebra system, some extra work, required when using a numerical problem solving environment, is omitted. Problem specifications no longer have to be adapted to the numerical solving routines [1] . Furthermore, working on models within a single environment as well for documentation and for execution purposes, can significantly augment the return on investment. Additionally, this approach should ease collaborative

work between the many engineers or scientists using *Microsoft Office* as their office applications. Learning in detail the principles and features of the *Microsoft Office* suite becomes of interest not only for editing technical document but also for modelling and simulating CAPE systems.

However, many practical difficulties raise when trying to apply this "office document centric" approach. Which format should be used for the office documents: ODF or OOXML? Which software technologies should be selected: components or Web services? And last but not least, which CAPE tasks are the more suitable do be handled only through office documents? From our short experience, we can provide a general answer to the last question. When graphical views of the processes are required, one has to work directly with the CAPE simulator interface, providing business dedicated features. When working only with numerical data of the processes, office applications, more precisely spreadsheets, provide the right user interface. When working only with the mathematical model views of the processes, office documents should be the presentation layer, provided that their equation editor is enriched with features allowing communication with the problem solving environments.

5. Conclusion

The paper deals with an integrated framework for CAPE based on *Microsoft Office* solution. Recent technologies make it possible to reconcile both the ergonomics of every day computer tools with the power of CAPE dedicated software. The *eXMSL for Microsoft Office* plug-in proves the feasibility and interest for the end user to have a restricted set of presentation applications, connected to CAPE software components dedicated to calculation. Our "office document centric" approach is particularly adapted when the end-user is working with the mathematical model views of CAPE problems.

References

[1] K. Alloula, J.-P. Belaud & J.-M. Le Lann, 2007, "Mixing computer algebra and numerical methods when solving CAPE models", *Computer-Aided Chemical Engineering*, vol. 24, pp. 135-140.

[2] K. Alloula, J.-P. Belaud, & J.-M. Le Lann, 2009, "A co-operative model combining computer algebra and numerical calculation for simulation", *Computer-Aided Chemical Engineering*, vol. 26, pp. 889-894.

[3] O. Baudouin, S. Déchelotte, P. Guittard & A. Vacher, 2008, "Simulis Thermodynamics : an open framework for users and developers", *Computer-Aided Chemical Engineering*, volume 25, pages 635-640.

[4] I.T. Cameron, G.D. Ingram, 2008, "A survey of industrial process modelling across the product and process lifecycle", *Computers & Chemical Engineering*, vol.32, pp. 420-438.

[5] W. Schweikhardt, C. Bernareggi, N. Jessel, B. Encelle, M. Gut, 2006, "LAMBDA: a European system for audio and speech access to mathematics", in *Computers Helping People with Special Needs*, Lectures Notes in Computer Science, volume 4061/2006.

[6] S. Rönnau, J. Scheffczyk & U.M. Borghoff, 2005, "Towards XML version control of Office documents", Proceedings of the 2005 *ACM symposium on Document engineering*, pp. 10-19.

[7] R. Zerry, B. Gauss, L. Urbas and G. Wozny, 2004, "Web-based object oriented modelling and simulation using MathML", *Computer-Aided Chemical Engineering*, volume 18, pages 1171-1176.

Computer-aided web-based application to modular plant design

Łukasz Hady, Günter Wozny

Berlin Institute of Technology, Chair of Process Dynamics and Operation,
Sekr. KWT-9, Str. des 17. Juni 135, D-10623 Berlin, Lukasz.Hady@TU-Berlin.DE

Abstract

In order to support the design activities as well as the reusability of available examined solutions in developing new technical projects in the field of modular plant design, a web-based application for quality- and know-how assurance, Reuse-Atlas, was developed. Using exemplary designed modules, aspects of a quality- and know-how assurance as well as the reusability of the engineering data was implemented and transferred to Reuse-Atlas. The Reuse-Atlas includes structures for the modules of different plant complexity like structural group- and plant group-modules. It allows more efficiency by tracking of undertaken decisions and assumptions during the module development and design and therefore simplicity in designing the new plants and modifying the old ones.

Keywords: computer-aided plant design, modular plant design, modularization

1. Introduction

New plant planning methods, as for example module technology, combined with computer-aided plant design offer basis for better quality assurance in the early stage of process development and plant layout. In regard to the numerous designed plants the possibility for reusing the available examined solutions in new projects has still not received enough attention. The reusability of the revised solutions and their technical know-how could lead to improvement in the efficiency of project development, project quality and consequently reduces cost. Plant engineering and construction companies or their divisions were always interested in reducing cost and therefore in using of existing solutions in developing new technical projects. However an appropriate selection of suitable documents and information could be a very difficult task in case of different customer requirements and missing "know-how warehouses". For this reason the plant layout and design will be mostly started from scratch. Within the scope of plant layout and design a complete design process based on a process simulation, cost calculation, dimensioning and layout design of process units, apparatus and machines included 2D- and 3D-plant activities will be realised. The design process is iterative and is carried out on the basis of experiences of the project engineers. In order to reduce time-to-market, a given workflow, which determines the working steps and contents, will be combined with existing engineering approaches, as for example simultaneous-, concurrent- and collaborative engineering. Despite these systematic approaches, the developed solutions are like "stand-alone" solutions. These are often only adapted to the existing project conditions without consideration of superior aspects of plant layout and design, as for example modularization. It is therefore no surprise that such solutions developed by different process engineers are unique and rather rarely usable for the new technical projects. As a consequence of this, ad-hoc solutions will be generated and the assurance of the engineering know-how will be often neglected. It comes rather seldom to the

product costing analysis and revision of the developed solutions after the technical project is finished. An Achilles' heel is searching for appropriate information. The study results of the Marketing Research Institute carried out by Vanson Bourne in May 2007, where 610 managers from European companies were questioned shown that, an employee in a typical European company wastes in average 67 minutes per day in an ineffective searching for information. In a company with 1000 employees and an average annual salary of 50.000 Euro, it could lead to annually eight million Euro expenses (Kaiser, 2007).

2. Modular Plant Design: a future plant layout and design approach?

2.1. Concept of the Engineering Reuse

The idea of module-oriented plant design is based on the definition of the process engineering units – modules, which are functionally independent (Hady et al., 2007). A modular plant consists of plant group-modules which are designed in a modular way. The plant group-modules are flexible process engineering units which are designed from smaller ones, namely from structural group-modules. Fig. 1 represents a structure of a modular plant with two approaches which can be used by modular plant engineering: Modular-Concept and Modular-Technique. The Modular-Concept characterizes the concept of the Engineering-Reuse and the Modular-Technique the concept of the Equipment Reuse. The last one is discussed within section 2.2.
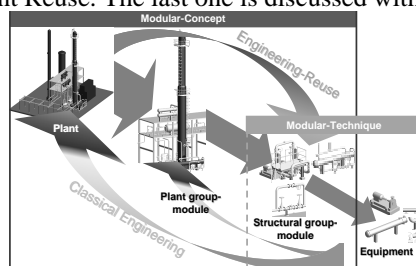
Fig. 1. Modular structure of the plant with two approaches for modular plant design: Modular-Concept and Modular-Technique

The plant group-modules are responsible for the basic operations and the structural group-modules are responsible for the basic functions as long as the whole process is covered by the plant. The configuration and layout of plant group-modules depends on the number of required basic operations. Unlike the plant group-modules, the structural group-modules are fixed process engineering units. They can only be affected with small constructional or layout changes such as near piping, order of armatures and supports. The adaptability of the structural group-modules to the conditions of new plant group-modules or whole plants during the Engineering-Reuse will be covered with a degree of flexibility around 10 to 20% for possible changes. According to this Modular Concept of Engineering Reuse (Fig. 1) various structural group-modules were developed (Hady et al., 2009a).

2.2. Concept of the Equipment Reuse

The concept of the Equipment Reuse describes the possibility for planning and designing of changeable plants, which configuration could be affected and changed due to the product or performance changes or requirements as well as by adaptation of the existing configuration from a pilot plant- or mini plant-scale to the production of the plant-scale. Using modularization at this complexity level, the changeability of a plant

should be characterized and taken into consideration within the scope of the conceptual design. In addition the standardization potentials should be developed and the suitable, standardized equipment-modules should be defined and designed. In an industrial project a mobile Absorption-module was planned and designed (Fig. 2, right hand side). This Absorption-module can be operated under changed process requirements, raw materials, operation parameters as well as production performance.

Fig. 2. Absorption-Miniplant (left) and Absorption-module (right) (Mueller et al., 2009)

The engineering activities carried out during the modular plant layout and design lead to production of a large amount of documentation, as for example P&ID´s, 2D-equipment layouts, 3D-CAD models of modules as well as of information such as design rules and planning guidelines. Due to the fact that developed modules and its documentation cover know-how principles and quality characteristics, an efficient data management of modules is very important and could be considered as a support toward successful Engineering- and Equipment Reuse.

3. Online application to support of modular plant engineering

The necessity of development of a tool to support of modular plant layout and design was stressed by Lueneburg (2003). The definition of module borders, levels of modularization as well as the development of the module libraries were the main statements which should be considered during the application of modularization (Lueneburg et al., 2003). In order to guarantee the know-how assurance as well as to increase quality of the developed modules, some additional aspects were taken into consideration during development of the Reuse Atlas. These are:

- Assurance of the modularization- and standardization rules
- Assurance of the undertaken assumptions and decisions during module design
- Assurance of the variant construction.

3.1. Concept of the Reuse-Atlas

Since the quality of the available equipment-parts and modules of a different complexity used for plant layout and design is so important, the concept for the assurance of available know-how principles and quality characteristics was developed. This concept of the documentation and management system of modules, so-called Reuse-Atlas, was discussed by Hady (Hady et al., 2009b). All of above mentioned documentation such as: 2D-equipment layouts, 3D-CAD models, know-how principles and quality characteristics and all other information, which corresponds to the developed modules, builds a main structure of the Reuse-Atlas. Moreover such a Reuse-Atlas should have:

- Easy and intuitive handling via a web browser, user friendly interface
- Possibility to be continually extended by users to cover the documentation of new structural group- or plant group-modules

- Adaptability to the requirements of design and project engineers

3.2. Reuse-Atlas: application software

For the implementation of the web-based Reuse Atlas a combination of operating system - web server - database system - scripting language for queries of the data base was required. As a platform for implementation of the Reuse-Atlas a so-called WAMP - platform was used. WAMP stands for **W**indows® - **A**pache™ - **M**ySQL® - **P**HP. The web based Reuse-Atlas was also implemented on the basis of the HTML and PHP for a web development as well as other Open-Source-Software such as MySQL® for the applied database of users, know-how principles and quality characteristics and Apache™ HTTP Server for a web server.

3.3. Structure of the Reuse-Atlas

The Reuse-Atlas enables a password protected storage, updating and management of the module documentation as well as all employees an access for reading of the stored and storage of the new data via Internet. After login, a suitable data base has to be chosen or created, if the online application is used for the first time. Storage or updating cover the modular hierarchy of the plant, such as plant group- and structural group-modules and their documentation such as e.g. P&ID´s, 2D-equipment layouts, 3D-CAD models of modules, know-how principles and quality characteristics. This hierarchical structure fulfils the Top-Down decomposition shown in Fig. 1 and has only an administrative character. After a suitable data base was chosen the available hierarchical structure of plant group- and structural group-modules is displayed. The hierarchical structure of the Deisobutanizer-plant group-module is shown in Fig. 3 (left hand side).
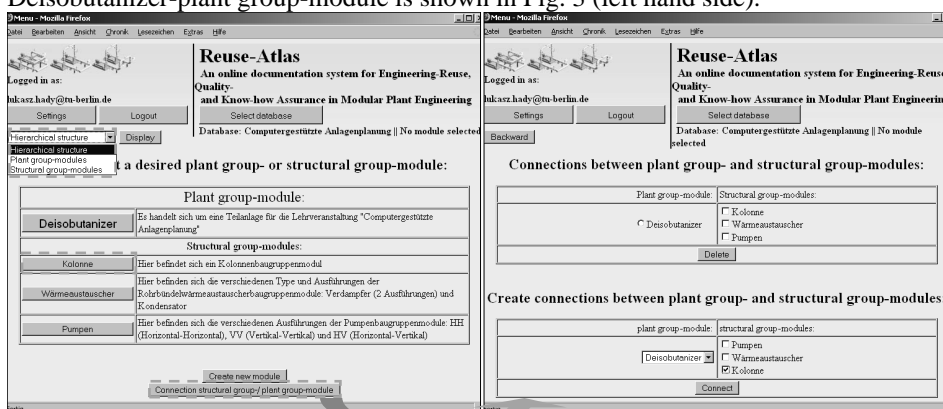
Fig. 3. Hierarchical (modular) structure (left hand side) as well as definition of connections between plant group- and structural group-modules within a chosen data base (right hand side)

For each data base as well as each modular level of a plant a short description about e.g. the scope or application can be enclosed. At the same time the stored modules can be displayed in three different views, where either plant group- or structural group-modules can be shown and with a hierarchical structure, where plant group-modules with the associated structural group-modules can be displayed. The user interface shown in Fig. 3 (right hand side) favours the suitable connection between plant group- and structural group-modules. Using it, process engineers are able not only to define some new connections but also to modify or delete the existing ones. As an outcome for using it, the redundant storage of relevant data which cover know-how principles and quality characteristics of the developed modules can be avoided. A predefined documentation structure, which is identically developed for each plant group- and structural group-

Computer-aided web-based application to modular plant design

module, gives an overview to the stored documents and information. This structure is discussed on the example of the pump structural group-module and shown in Fig. 4 and 5. The left side of the Reuse-Atlas describes the level of visualization and allows saving of information regarding the 2D- and the 3D-module layouts with characteristic 2D- and 3D-views. The buttons: H-H (horizontal-horizontal) to H-V-1 (horizontal-vertical-1) shown in Fig. 4, correspond to the pump structural group-modules with different arrangements of the pipes on the pressure- and suction side. The Reuse-Atlas gives also an access to the modules designed with commercial CAD software. Therefore the available models of modules could be easily revised and reused. The number of documents, which can be stored in the 2D- and 3D-view is unlimited and depends on the complexity of the modules. The right side of the Reuse-Atlas describes the documentation of the modules. Here the know-how principles and dimensional characteristics of the given modules could be found, edited and created as Excel-, PDF- or ASCII-files. The Reuse-Atlas gives process engineers a possibility of continual extension of the data base not only for documentation and models of the new modules but also for new sizes, arrangements and layouts within an existing module.
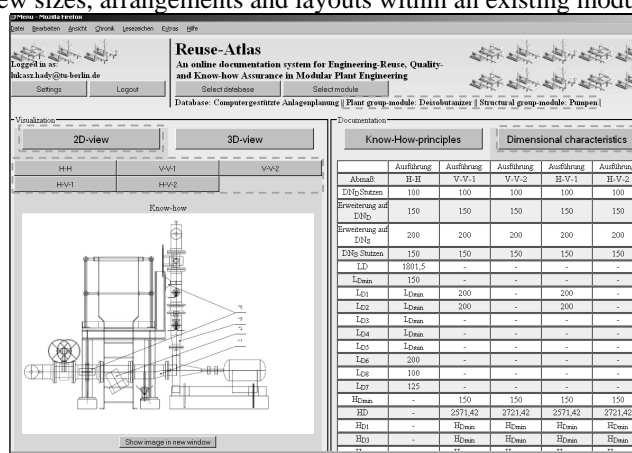
Fig. 4. 2D-view of the visualization level (left hand side) and dimensional characteristics of 2D-layouts of the documentation level (right hand side) of the pump structural group-module

The table shown in Fig. 4 covers dimensional characteristics of modules which correspond with 2D-view of the visualization level. This table can be modified and extended with additional information regarding other sizes, arrangements or dimensional characteristics of 2D-layouts. The second possibility for visualization of the stored modules is shown in Fig. 5 (3D-view), where for each module and its arrangements, movies of 3D-models of modules can be stored and played via the web browser. It gives a better and faster possibility for evaluation of the stored modules by users. Since the know-how principles cannot be always expressed with figures, movies and tables, an editable verbal description of undertaken assumptions and decisions is also available and possible within a section "know-how principles". This verbal description should favour the traceability of the undertaken assumptions and decisions. The know-how principles can be extended on new information for all sizes and arrangements of stored plant group- and structural group-module or only for the selected ones. The already stored information can be modified or, if it is necessary, deleted. In Fig. 5 (left hand side, "Add new notice" button) is shown, how some new information can be added to the existing ones. Fig. 5 (right hand side) shows an user interface used for the management of stored module documentation. This user interface makes it

possible to carry out a continual extension as well as maintenance of the Reuse-Atlas. Starting from the left side of this user interface, the module arrangements/ -layouts and sizes, stored drawings, movies and 3D-models of modules are available and displayed here.
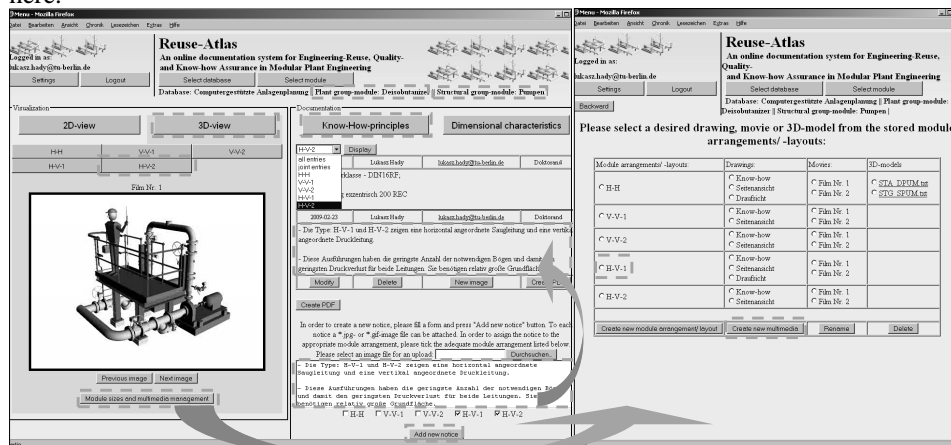
Fig. 5. 3D-view of the visualization level and know-how principles of the documentation level (left hand side) as well as management of the module documentation (right hand side)

4. Conclusion and Acknowledgement

The developed Reuse-Atlas favours the reusability of the equipment-modules and makes an internal technical know-how of a company or a division available to all employees involved in one project via Intranet or Internet. Due to easy and user friendly handling as well as extendable possibilities of this web-based application, planning works and computer-aided plant design with the Reuse-Atlas could be well supported and accelerated. Moreover, the stored modules and the know-how of modules can be reused every time when a new project is executed. As an outcome for using Reuse-Atlas, repeatable design duties could be avoided, planning works are accelerated, the know-how and the quality of the modules are assured even though the experienced engineers are no longer involved in the project. Due to the fact, that efficient Engineering Reuse and the application of the predesigned modules depend on their arrangement and configuration, the Reuse-Atlas could support the optimal development and design of the equipment-module layouts. The authors acknowledge support from the Cluster of Excellence "Unifying Concepts in Catalysis" coordinated by the Berlin Institute of Technology and funded by the German Research Foundation (DFG).

References

Ł. Hady, W. Lueneburg, M. Dyląg, G. Wozny, 2007, Modular investment cost estimate of multipurpose chemical plant, Chemical and Process Engineering, 28 (1), 17-31

Ł. Hady, G. Wozny, 2009a, Know-how and quality assurance using a web based Reuse-Atlas, Chemical Engineering Transactions, 18, 761-766, DOI: 10.3303/CET0918124

Ł. Hady, M. Dyląg, G. Wozny, 2009b, Investment cost estimation and calculation of chemical plants with classical and modular approaches, Chemical and Process Engineering, 30 (2), 319-340

J. Kaiser, 2007, Study: Waste of millions by insufficient information systems in companies, PresseBox® [BoxID 110401, Software]

W. Lueneburg, W. Zahn, 2003, Modulare Anlagentechnik, Chemie Ingenieur Technik, 75, 1022

M. Mueller, H. Thielert, G. Wozny, 2009, unpublished internal documentation

Simulation of a structured catalytic reactor for exothermic methanation reactions producing synthetic natural gas

Maria Sudiro,[a] Alberto Bertucco,[a] Gianpiero Groppi,[b] Enrico Tronconi[b]

[a] *Dipartimento di Principi e Impianti di Ingegneria Chimica "I. Sorgato" (DIPIC) Università di Padova, via Marzolo 9, 35131 Padova (Italy), maria.sudiro@unipd.it, alberto.bertucco@unipd.it*
[b] *Laboratorio di Catalisi e Processi Catalitici, Dipartimento di Energia, Politecnico di Milano, Piazza Leonardo da Vinci 32, 20133 Milano (Italy) gianpiero.groppi@polimi.it, enrico.tronconi@polimi.it*

Abstract

Aim of this work is a theoretical investigation of the catalytic methanation reactions in externally cooled tubular reactors filled with novel monolithic catalysts with high thermal conductivity.

Using the general purpose modelling tool gProms™ we have developed a steady-state, heterogeneous 1D model, representing a single, externally cooled reactor tube loaded with cylindrical honeycomb catalysts with square channels, made of conductive material. The model equations include mass and energy balances for the gas and solid phases and the momentum balance for the gas phase. Two reactions are considered: carbon monoxide and carbon dioxide methanation, whose rate equations are taken from the literature.

By reactor simulation it is shown that the problem of temperature control typical of fixed-bed methanation reactors can be overcome by the monolith reactor herein proposed. The effects of space velocity on conversion and temperature profiles are discussed, with a fixed geometrical configuration of the monolithic reactor.

Keywords: Monolith reactors, Reactor modeling, Methanation, Synthetic Natural Gas

1. Introduction

The production of synthetic natural gas (SNG) from coal or biomass is an interesting opportunity both for exploiting coal and biomass, and for replacing oil products for transportation and other uses.

Natural gas is cleaner than coal and its production has been increasing for years; however, its price has been increasing as well, so the synthesis of SNG from coal and biomass is receiving strong interest especially in USA and China (http://www.chinamining.org). SNG has many important advantages: it can be transported efficiently and cheaply using existing natural gas pipelines and distributing networks, it is an easily convertible feedstock, both in natural-gas combined-cycle power plants and in petrochemical facilities, it can count on a high social acceptance with respect to coal, and it can be stored underground, enabling efficient operation throughout the year independent of a fluctuating demand.

Unfortunately, the commercial deployment of technologies for the production of SNG is currently constrained by technical barriers, so that more research is required before extensive applications at the industrial scale. An important issue to be addressed is the

strong exothermicity of the reactions, so that conventional fixed-bed catalytic reactors cannot be safely used (Sudiro *et al*., 2009a).

The use of monolith catalyst supports offers at least two advantages with respect to conventional packed-bed reactors: pressure drops are greatly reduced (to less than 1%) and the radial heat transport can be more favorable, especially when metallic support are used. Such monoliths have in fact the advantage of favorable global heat-transfer properties resulting from heat conduction in the connected structure of their solid matrix; this problem has been extensively investigated theoretically (Groppi and Tronconi, 1996; Groppi and Tronconi, 2000). Application of monolith catalysts has already been shown as technically feasible, both by simulation in the case of the oxidation of methanol to formaldehyde and epoxidation of ethylene (Groppi and Tronconi, 2001) and experimentally at the pilot-reactor scale (Groppi *et al*., 2008; Groppi *et al*., 2000; Tronconi and Groppi, 2000) in the case of strongly exothermic gas/solid reactions, such as catalytic o-xylene selective oxidation and CO oxidation.

In this way significantly more heat can be removed and structured catalyst are eligible to be used for strongly exothermic gas/solid reactions. Favourable heat transfer properties permit reducing risk of thermal runway and catalyst deactivation. To this end, our aim is to verify the possibility of overcoming the problem of temperature control, typical of fixed-bed, for methanation reactions, if they are carried out in monolith reactors.

2. Reactor modeling

It is assumed that catalytic methanation reactions occur in an externally cooled tubular reactors filled with novel monolithic catalysts with high thermal conductivity.

2.1. Model equations

The behavior of the methanation reactor is modelled by a dynamic 1D, heterogeneous, single-tube model of an externally cooled multi-tubular fixed-bed reactor, which is loaded with honeycomb catalysts.

The model includes the specie mass balances and the energy balances for the gas and solid phases, and the momentum balance for the gas phase.

Mass balances for the gas phase are expressed by:

$$\varepsilon \frac{\partial w_{i,g}}{\partial t} = -\frac{Wt}{\rho_g} \cdot \frac{\partial w_{i,g}}{\partial z} - \frac{K_{m,i} a}{\rho_g} \cdot \left(w_{i,g} - w_{i,s} \right) \tag{1}$$

whereas those for the solid phase are:

$$0 = \sum_{j=1}^{NR} \upsilon_{i,j} \frac{\tanh \vartheta_j}{\vartheta_j} R_j PM_i \xi + K_{m,i} a \left(w_{i,g} - w_{i,s} \right) \tag{2}$$

Energy balance for the gas phase is written as:

$$\varepsilon \rho c_p \frac{\partial T_g}{\partial t} = -Wt c_p \frac{\partial T_g}{\partial z} - ha \left(T_g - T_s \right) \tag{3}$$

whereas that for the solid phase is:

Simulation of a structured catalytic reactor for exothermic methanation reactions producing synthetic natural gas

$$\rho_s c_{p,s}(1-\varepsilon)\frac{\partial T_s}{\partial t} = k_{s,ax}(1-\varepsilon)\frac{\partial^2 T_s}{\partial z^2} + \frac{ha}{1-\varepsilon}(T_g - T_s) + \sum_{j=1}^{NR}\left(-\Delta H_{R,j}\frac{\tanh \vartheta_j}{\vartheta_j}R_j\right)\xi - h_{wall}(T_s - T_{wall})\frac{4}{d} \quad (4)$$

The momentum balance is according to the following equation:

$$\left(-\frac{1}{\rho_g} + \frac{Wt}{\rho_g^2 P}\right)\frac{\partial P}{\partial z} - \frac{Wt^2}{\rho_g^2 T}\frac{\partial T_g}{\partial z} = \frac{1}{2}\frac{Wt^2}{\rho_g^2}af \quad (5)$$

Boundary conditions at reactor inlet and outlet are reported in Table 1.

Table 1. Boundary conditions

Reactor inlet (z=0)	Reactor outlet (z=L)
$w_{i,g}\big|_{z=0} = w_{i,feed}$; $T_g\big|_{z=0} = T_{feed}$; $P\big|_{z=0} = P_{feed}$	
$-k_{s,ax}\dfrac{\partial T_s}{\partial z}\bigg|_{z=0} = \sigma \cdot \varepsilon_s \cdot \left(T_g^4 - T_s^4\big|_{z=0}\right)$	$-k_{s,ax}\dfrac{\partial T_s}{\partial z}\bigg|_{z=L} = \sigma \cdot \varepsilon_s \cdot \left(T_g^4 - T_s^4\big|_{z=L}\right)$

Initial conditions are: $w_{i,g}(z,t=0) = w_{i,feed}$, $T_g(z,t=0) = T_{coolant}$, $T_s(z,t=0) = T_{coolant}$.

The following gas-solid heat and mass transfer correlations for honeycomb monoliths have been adopted to calculate Nusselt and Sherwood numbers in the square celled monoliths:

$$\text{Nu} = 3 + 6.874 \cdot \left(1000 \cdot z_{\text{Nu}}^*\right)^{-0.488} e^{-57.2 \cdot z_{\text{Nu}}^*} \quad , \text{ where } z_{\text{Nu}}^* = \frac{z}{d_{eq}\text{RePr}} \quad (6)$$

$$\text{Sh} = 3 + 6.874 \cdot \left(1000 \cdot z_{\text{Sh}}^*\right)^{-0.488} e^{-57.2 \cdot z_{\text{Sh}}^*} \quad , \text{ where } z_{\text{Sh}}^* = \frac{z}{d_{eq}\text{Re Sc}} \quad (7)$$

The following expression for calculation of friction factors in the square monolith channels was adopted in the model:

$$f\,\text{Re} = 14.227 + 17.76323 \cdot \left(1000 \cdot z_f^*\right)^{-0.4715} e^{-59.57 \cdot z_f^*} \quad , \text{ where } z_f^* = \frac{z}{d_{eq}\text{Re}} \quad (8)$$

Dimensionless numbers are defined in Table 2:

Table 2. Dimensionless numbers

$\text{Pr} = \dfrac{\mu_g c_{p,g}}{k_g}$;	$\text{Re} = \dfrac{d_{eq}Wt}{\varepsilon\mu_g}$;	$\text{Nu} = \dfrac{d_{eq}h}{k_g}$;	$\text{Sc}_i = \dfrac{\mu_g}{\rho_g D_{i,g}}$	$\text{Sh}_i = \dfrac{K_{m,i}d_{eq}}{\rho_g D_{i,g}}$

Thermal properties, such as specific heat and heat of reaction, were estimated according to the CHEMKIN correlations (Smith *et al.*, 2009), and gas properties, such as conductivity, viscosity and diffusivity, were calculated according to the correlations reported by Sudiro *et al.*, (2009b). The system of equations was solved by using the general purpose tool gProms™, adopting 200 grid points along the axial direction.

The reactor has a length of 1 m with an inner diameter of 0.254 m. The monolith pitch is set to 2 mm, values of the catalytic washcoat (ξ) and of fraction of metallic support (λ), are 0.2 and 0.25, respectively.

The wall heat transfer coefficient was set to 500 W/m^2 K, according to the results of Groppi and Tronconi, (2005) for a monolith in tight contact with the reaction tube, whereas the thermal conductivity of the monolith support was set to 200 W/m K (Al).

2.2. Kinetic scheme

The following two reactions (CO and CO_2 methanation) have been considered:

$$CO + 3H_2 \rightarrow CH_4 + H_2O \tag{9}$$

$$CO + 4H_2 \rightarrow CH_4 + 2H_2O \tag{10}$$

The CO methanation kinetic was derived from Sughrue, 1982, the one of CO_2 methanation from Weatherbee, 1982. Rate expression for CO methanation on Ni catalyst is:

$$r_{CO}\left(\frac{mol}{m^3 s}\right) = \frac{0.2 \cdot k_1 k_2 p_{H_2}}{k_1\left(1 + K_{H_2}^{0.5} p_{H_2}^{0.5} + K_{CO} p_{CO}\right)^2 + k_2\left(1 + K_{CO} p_{CO}\right)^2} \cdot \rho_{catalyst} \tag{11}$$

where k_1, k_2, k_{H2}, k_{CO} depend on the temperature according to an Arrhenius/power law expression. The rate expression for CO_2 methanation on Ni catalyst is:

$$r_{CO_2}\left(\frac{mol}{m^3 s}\right) = \frac{0.2 \cdot A \cdot p_{CO_2}^{0.5} p_{H_2}^{0.5}}{\left(1 + B\left(\frac{p_{CO_2}}{p_{H_2}}\right)^{0.5} + C\left(p_{CO_2} p_{H_2}\right)^{0.5} + \frac{p_{CO}}{D}\right)^2} \cdot \rho_{catalyst} \tag{12}$$

Constant A, B, C and D are tabulated as a function of temperature. In order to explore intensified process conditions, r_{CO} and r_{CO2} were incremented by 40% in all the simulations.

3. Results and discussion

Figure 1 shows the calculated axial profiles of gas temperature (1a) and conversion of H_2, CO and CO_2 along the reactor (1b). Coolant temperature was set to 573K, as reactions start at T>>240°C (Odermatt, 2008). A feed (60% mol H_2, 20% mol CO and 20% mol CO_2) with a H_2/CO molar ratio of 3 was assumed, and the GHSV was set to 15000 h^{-1}, as reported by Rostrup-Nielsen, 2007, which is a typical value for pilot plants. The inlet pressure was set to 6.9 bar.

(1a)　　　　　　　　　　　　　(1b)
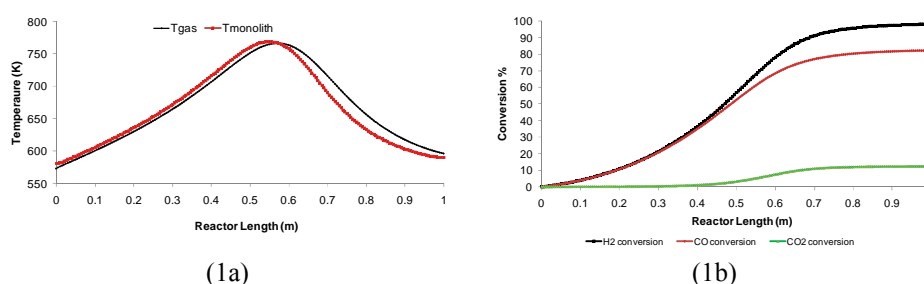
Figure 1. Axial profiles of temperatures and conversions

It is interesting to show that with only one monolithic reactor a high conversion of hydrogen and carbon monoxide can be obtained (98 and 83%, respectively), with a

Simulation of a structured catalytic reactor for exothermic methanation reactions producing synthetic natural gas

moderate and acceptable increase of temperature along the reactor, avoiding catalyst deactivation with respect to existing methanation processes. Here at least three reactors and a recycle of a part of the products are required (Sudiro *et al.*, 2009a, Topsøe, 2009): recycle is needed for controlling the temperature using products as inert and three reactors are required to achieve high conversion.

For comparison, if the same gas flow rate were sent to an equilibrium adiabatic reactor the output temperature would be higher (about 1000K), with a CO conversion of 81%, similar to that obtained with the monolith reactor, but with the consequence of a total deactivation of the catalyst.

3.1. Sensitivity analysis

The effect of a progressive increase of the flow rate value was investigated in order to improve the productivity (see Figure 2). A higher gas flow rate, and correspondingly a greater value of GHSV, results in the shift of the hot-spot towards the bed exit (in this conditions the reactions tend to the extinction). In fact, the increase of gas flow rate results in an enhanced convective heat removal, so that the temperature increment generated by the reaction heat is lower, which in turn slows down the reaction rate.

As a result, at increasing space velocity a decrease of CO conversion can be observed.

| (2a) | (2b) |
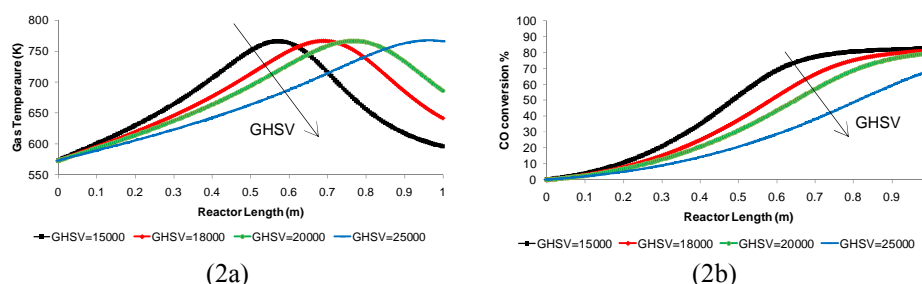
Figure 2. Gas temperature (a) and CO conversion (b) axial profiles at four different GHSV

4. Final remarks

In this work the possibility of using monolithic reactors carrying out exothermic methanation reactions from syngas was investigated by process simulation. A model of a dynamic 1D, heterogeneous, single-channel model of an externally cooled fixed-bed reactor, loaded with honeycomb catalysts, was developed and solved, using gProms™ as process simulator.

It was shown that synthetic natural gas can be produced in a single pass monolithic catalyst reactor, with acceptable conversion values and temperature hot spots compatible with the catalyst stability. This system improves the presently adopted process configurations, as it overcomes the problem of temperature control typical of fixed-bed methanation reactors.

The use of monolith reactors is also useful in view of process intensification: we have verified that the GHSV can be increased up to 20000 h^{-1}, with minimal pressure drops, increasing the cooling temperature correspondingly, without the risk of reactor runway. Further improvements might be obtained by adopting more active catalyst formulations.

Nomenclature

a = specific area (1/m), $a = 4\varepsilon / d_{eq}$

c_p = mass specific heat (J/kg K),

k_W = active washcoat conductivity (1.2 W/m K)

L = reactor length (m)

m = monolith pitch (m) (2 mm)

($c_{p,s}$=865 kJ/kg)
D = diffusivity (m^2/s)
d = monolith (or reactor) diameter (m)
d_{eq} = hydraulic diameter of the monolith

channel (m), $d_{eq} = m\sqrt{\varepsilon}$

f = friction factor
h = gas-solid heat transfer coefficient in the monolith channel (W/m^2 K)
h_{wall} = wall heat transfer coefficient
k_g = gas mixture conductivity (W/m K)
$K_{m,i}$ = gas-solid mass transfer coefficient in the monolith channel (kg/m^2 s)
$k_{s,ax}$ = effective axial conductivity (W/mK),

$$k_{s,ax} = k_s \left(\lambda + \xi \frac{k_W}{k_s} \right)$$

k_s = support conductivity (Al=200 W/m K)

Greek Letters
δ_w = thickness washcoat (m),

$$\delta_W = \left(\sqrt{\varepsilon + \xi} - \sqrt{\varepsilon} \right)$$

ε = bed void fraction (0.7)
ε_s = emissivity of solid phase (0.7 for Al)
λ = volume fraction of inert support (0.25)

Subscripts and Superscripts
ax = axial
g = gas phase
s = solid phase

Nu = Nusselt number
P = pressure (Pa); p = partial pressure
PM_i = molecular weight of the specie i (kg/kmol)
Pr = Prandtl number
Re = Reynolds number
R_j = j reaction rate (kmol/m^3 s)
R = universal gas constant (8314 J/kmol K)
Sc_i = Schmidt number
Sh_i = Sherwood number
T = temperature (K)
t = time (s)
v = gas velocity (m/s)
w_i = weight fraction of the specie i
W_t = specific mass flow rate (kg/m^2 s)
W_{tot} = mass flow rate (kg/ s)
z_f^* = axial coordinate for friction factor
z_{Nu}^* = axial coordinate for Nusselt number
z_{Sh}^* = axial coordinate for Sherwood number

μ = viscosity (kg/m s)
ξ = volume fraction of active phase (0.2)
$\Delta H_{R,j}$ = heat of reaction (J/kmol)
θ = Thiele modulus
ρ = density (kg/m^3), ρ_s=3800 kg/m^3, ρ_w=900 kg/m^3
σ = Stefan-Boltzmann constant (5.67·10^{-8} W/m^2 K^4)
υ = stoichiometric coefficient

w = active washcoat or catalytic phase
wall = reactor wall

References

Groppi G. and E. Tronconi, AIChE J., 42 (1996), 2382-2387.

Groppi G. and E. Tronconi, Chem. Eng. Sci., 55 (2000), 2161-2171.

Groppi G., G. Airoldi, C. Cristiani, E. Tronconi, Catal. Today, 60 (2000), 57-62.

Groppi G. and E. Tronconi, Catal. Today, 69 (2001), 63-73.

Groppi G. and E. Tronconi, Catal. Today, 105 (2005), 297-304.

Groppi G., E. Tronconi, F. Cruzzolin, C. Cortelli, R. Leanza, and S. Marsaud, Development and industrial pilot tests of novel conductive structured catalysts for o-xylene oxidation to phthalic anhydride. In Book of Abstracts of ISCRE20, The 20th International symposium on Chemical Reaction Engineering, 7-10 September 2008, Kyoto-Japan, pp. 150-151.

Rostrup-Nielsen J.R., K. Petersen and J. Sehested, Appl. Catal., A, 330 (2007), 134-168.

Smith, G.P., D.M. Golden, M. Frenklach, N.W. Moriarty, B. Eiteneer, M. Goldenberg, C.T. Bowman, R.K. Hanson, S. Song, W.C. Gardiner, Jr., V.V. Lissianski and Z. Qin http://www.me.berkeley.edu/gri_mech/

Sudiro M., C. Zanella, L. Bressan, M. Fontana and A. Bertucco. Synthetic natural gas (SNG) from petcoke: model development and simulation. In Proceedings of ICheaP-9, The ninth International Conference on Chemical & Process Engineering, 10-13 May 2009a, Rome-Italy, pp. 1251-1256.

Sudiro M., M. Pellizzaro, F. Bezzo and A. Bertucco, Chem. Eng. Res. Des., (2009b). In Press.

Sughrue, E. L. and C. H. Bartholomew, Appl. Catal., 2 (1982), 239-256.

Topsøe, H. (2009). http://www.topsoe.com

Tronconi E. and G. Groppi, Chem. Eng. Sci., 55 (2000), 6021-6036.

Weatherbee, G. D. and C. H. Bartholomew, J. Catal., 77 (1982), 460-472.

www.chinamining.org

Mathematical modeling and simulation of gasification processes with Carbon Capture and Storage (CCS) for energy vectors poly-generation

Victoria Maxim,[a] Calin-Cristian Cormos,[a] Ana-Maria Cormos,[a] Serban Agachi [a]

[a] Babes – Bolyai University, Faculty of Chemistry and Chemical Engineering
11 Arany Janos Street, RO-400028, Cluj – Napoca, Romania
Tel: +40264593833, Fax: +40264590818, E-mails: vgoia@chem.ubbcluj.ro;
cormos@chem.ubbcluj.ro; cani@chem.ubbcluj.ro; sagachi@staff.ubbcluj.ro

Abstract

Gasification of solid fuels is a partial oxidation process which converse the solid feedstock into syngas which can be used in a large number of applications e.g. power generation, manufacture of various chemicals and fuels (hydrogen, methanol, ammonia, fertilizers etc.). Not all of the gasification systems are suitable for energy vectors poly-generation with carbon capture and storage (CCS).

This paper is proposing to evaluate various gasification technologies by mathematical modeling and simulation methods (especially for entrained flow types as these gasifiers are more suitable for implementing carbon capture technologies). In this paper a particular accent will be put on the selection of the most promising gasifier, as not all are appropriate for a carbon capture Integrated Gasification Combined Cycle (IGCC) applied for energy vectors poly-generation (with a particular focus on hydrogen and electricity co-production case) with Carbon Capture and Storage (CCS). For the selection of the most appropriate gasifier technologies the process were mathematical modeled and simulated with process flow modeling software (e.g. ChemCAD, Aspen). In the evaluation of various gasification technologies (e.g. Shell, Siemens, GE-Texaco, Conoco-Phillips etc.) a multi-criteria analysis was performed.

Keywords: Gasification, Energy Vectors Poly-generation, Mathematical Modeling and Simulation, Multi-criteria Analysis, Carbon Capture and Storage (CCS)

1. Introduction

Coal gasification is one of the options for implementation of clean coal technologies. Gasification is the conversion of solid fuels (coal, coke, oil, tar, pitch) with air, oxygen steam or a mixture of this gases at a high temperature (above 800°C) into a gaseous product which can be used either to produce electricity either as a row material for the synthesis of chemicals or liquid fuels. In the early part of the last century the first application of fuel gas was illumination and domestic heating. Gasification of coal generates a wide range of products: power, chemicals, substitute natural gas (SNG) and transport fuels. The chemical composition of syngas varies based on many factors as: coal composition, size and rank, feeding system (dry or slurry), gasification agent used for oxidation (air or oxygen), temperature, pressure, residence time in gasifier, heating rate, gasification island configuration etc. The concept of gasification applied to electricity generation, the integrated gasification combined cycle IGCC, is very attractive for energy vectors poly-generation: electricity, hydrogen, heat and chemicals [1, 2].

The gasification reactors may be designed to gasify a wide variety of solid feedstock, either fossil fuels (e.g. coal, lignite peat etc.) or various biomass types (sawdust, agricultural wastes etc.) and solid waste (animal residue, municipal solid wastes, waste paper etc.). All coal types can be gasified, low ash content coals are preferred, but coal utilisation is regarded with concern because of bigger greenhouse gas emissions associated with it. IGCC is one of the power generation technologies having the highest potential to capture CO_2 with the lowest penalties in efficiency and cost. In an IGCC, modified for this purpose, the raw syngas (which contains mostly hydrogen and carbon monoxide) is subsequently reacted with steam in a shift converter, to maximize the hydrogen level in the syngas and to concentrate the carbon species in the form of CO_2 that can be later capture in a pre-combustion arrangement [3, 4].

The main focus of the article is to evaluate various gasification technologies by mathematical modeling and simulation methods and the selection of the most promising gasifier, as not all are appropriate for a carbon capture IGCC. The selection of the most promising gasifier investigated in the paper will be modelled and simulated using commercial process flow modelling package (ChemCAD) to produce data for the evaluation of gasification reactors. Technologies as Shell, Siemens, GE-Texaco, Conoco-Phillips will be analyzed in this paper considering coal as feedstock.

2. Gasifier options

For the gasification processes a wide range of reactors are available. For the commercial use, currently are available three types of gasifiers: moving-bed gasifiers, fluidised-bed gasifiers, and entrained-flow gasifiers [2,5]:

- Moving-bed gasifiers (also called fixed bed) are characterized by operation in a bed in which the coal moves slowly downward under gravity as it is gasified by a blast. The main drawback is that synthesis gas contains high levels of phenols, methane and tars.
- Fluidised-bed gasifiers in which the fuel and the oxidant coal particles are suspended in the gas flow. These gasifiers are suitable for reactive feedstocks such as lignite. Some of the disadvantages of these reactors are: high level of tar of the product gas, the incomplete carbon burn-out, and poor response to load changes.
- Entrained-flow gasifiers operate with feed and blast in co-current flow. They are suitable also for non-reactive feedstocks as coal. The synthesis gas contains low levels of phenols, methane and tars.

From the large range of gasifiers not all are suitable for Integrated Gasification Combined Cycle (IGCC) applied for energy vectors poly-generation with Carbon Capture and Storage. For the selection of the most appropriate gasifier several criteria must be used [6,7]:

- Oxygen purity: in conventional IGCC concept a 95% O_2 (vol.) is acceptable. The increase of the oxygen purity above 95% (e.g. 99 %) determines higher power consumption for the Air Separation Unit (ASU) with an increase of 5-10%;
- Gasifier throughputs, reliability and experience: depends of the plant size taken into evaluation (for instance in this paper in the range of 400 - 500 MW power net);
- Cold gas efficiency (CGE) and carbon conversion efficiency (CCE): is desirable that this indicators to be as high as possible on condition that hydrocarbons (mainly methane) present in syngas must be as low as possible (hydrocarbons negative influence the carbon capture plant capabilities). In case of entrained-flow gasifiers both CGE and CCE are optimum.

▪ Syngas cooling options: because of the steam requirement of the carbon monoxide shift conversion (WGS), the water quench type gasifiers are desirable. Unlike the gas quench option, the steam rising potential of the hot syngas leaving the gasifier reaction zone is severely neutralized.

▪ Influence of oxygen purity and gasifier feed system for hydrogen purification step: the influence of oxygen purity on hydrogen purification stage (done in a PSA unit) is a compromise between the need not to dilute the syngas with much nitrogen coming from the oxygen stream and decreasing the power consumption of the ASU. Although dry-feed design implies a certain syngas dilution with nitrogen, this is a preferable option against slurry-feed which imply a significant energy penalty by the water introduced with the coal slurry.

▪ Hydrogen production potential: similar to CGE is defined as the sum of carbon monoxide and hydrogen content in the syngas and it must be as high as possible, which is in the case of entrained-flow gasifiers.

▪ Downstream gas clean up issues: because of clean gas produced the entrained-flow gasifiers are the most desirable. Removing ash, hydrochloric acid, ammonia is possible using a quench system of the hot syngas

▪ Implication of gasifier reactor selection on Acid Gas Removal system, having in mind the fact that CO_2 capture process based on gas – liquid absorption is positively influenced by an increased pressure is desirable to have a high pressure.

▪ Capital cost: is a very important factor in gasifier selection and for the assessment of techno-economical indicators of the plant. However, this paper is not investigated the financial aspect of gasifier selection.

Analyzing all the above criteria the most promising reactors for energy vectors poly-generation (mainly hydrogen and electricity) with Carbon Capture and Storage are the entrained-flow gasifiers. The main characteristics of the four gasification technologies which are evaluated in this paper are presented above.

• Shell gasifier is a carbon steel vessel that contains a gasification chamber enclosed by a non-refractory membrane wall, which operates at 30-40 bar pressure, temperature range of 1500-1600°C, dry feed and one stage. Pulverised coal is stored under nitrogen,where it is pressurized and then pneumatically transported into the gasifier. The syngas is quenched with cooled recycled product gas and further cooled in a syngas cooler. Raw gas is cleaned in ceramic filters.

• Siemens gasifier is a top-fired reactor, where the reactants are introduced through a single centrally mounted burner. It is operating at similar conditions as the Shell gasifier. Unlike the Shell gasifier, which is using a gas quench, Siemens gasifier is using a water quench for cooling the syngas.

• GE-Texaco gasifier is a pressure vessel with a refractory lining, which operates at at 70-80 bar pressure, temperature range of 1300-1500°C, slurry feed and one stage. Oxygen and steam are introduced through burners at the top of the gasifier, coal is preprocessed into a slurry by fine grinding and water addition. The syngas is cooled into a water quench.

• Conoco Phillips (E-Gas) gasifier is two-stage coal-water slurry feed gasifier in a pressure shell lined with un-cooled refractory. Toward the bottom of the gasifier about 80% of the feedstock, as a coal water slurry is injected through burners. The temperature reaches 1350 - 1400°C at about 30 bar pressure. The syngas formed in the first stage flows upwards into the second stage area, where the remaining 20% of coal water slurry is injected. Hear the temperature is reduced to about 1000°C. The syngas exiting the gasifier in a fire-tube syngas cooler.

700 of 1332

3. Modeling and simulation of gasifiers for hydrogen and electricity co-production scheme

For the evaluation of various commercial gasification reactors a multi-criteria analysis will be performed considering all the criteria mentioned above. The gasification technologies presented before are analyzed in Table 1 [6,8].

Table 1. Multi-criteria analysis of coal gasifiers

Parameters	Shell	Siemens	GE Texaco	Conoco Phillips
Maximum pressure (bar)	40	40	100	40
Temperature (°C)	1400-1600	1400-1600	1200-1450	950-1150
Carbon conversion (%)	>99	>99	>98	>98
Steam/oxygen necessity	High	High	High	High
Syngas clean up issues	Low	Low	Low	Medium
H_2 production potential	High	High	High	Medium
CGE (%)	75-77	75-77	65-70	68-71
CO_2 capture capacity	High	High	High	Medium
Overall ranking	Good	OK	Bad	Bad

Considering all the above criteria when choosing a gasifier for hydrogen and electricity co-production scheme with carbon capture and storage, it appears that the most appropriate gasifiers are Shell and Siemens.

As mentioned before, the chosen gasifier is an entrained-flow type, operating at high temperatures with a high fuel conversion. From different commercial gasification technologies available on the market four are modeled and simulated using process flow modeling software (ChemCAD). As main design assumption, all gasifier concepts evaluated in the paper were considered Gibbs Free Energy Reactors (GIBS). By this model the calculations are made by the minimization of Gibbs free energy and approaching equilibrium state between reactants and products. Other gasifier design assumptions are: pressure drop 1.5 bar, pressure 40 bar (except GE Texaco 75 bar), vapour or mixed reaction phase, heat duty thermal mode (except GE Texaco which is adiabatic). The diagram for Shell gasification block is presented in Figure 1.
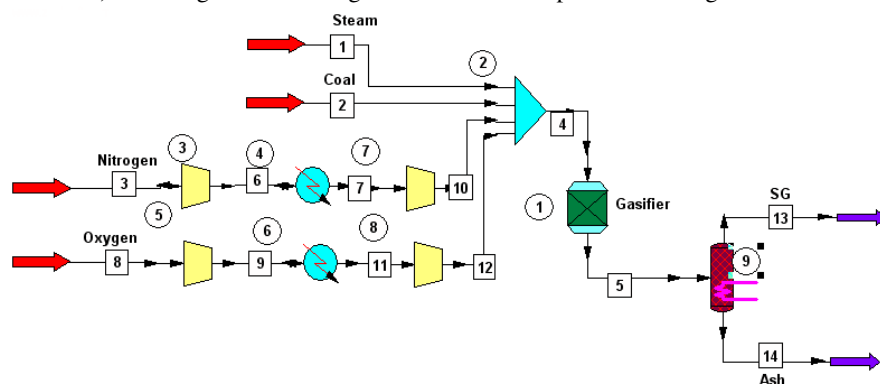
Figure 1. Shell gasification diagram

The coal is advised to have low content of ash and sulphur to reduce corrosion and SO_x emissions, but most important is that the ash has to have a relatively high melting point to prevent ash build up on the boiler heat transfer area [8]. The fuel characteristics (coal) are presented in Table 2.

Mathematical modeling and simulation of gasification processes with
Carbon Capture and Storage (CCS) for energy vectors poly-generation

Table 2. Coal characteristics

Parameter	Coal
Proximate analysis (% wt)	
Moisture	8.10
Volatile matter	28.51
Ash	14.19
Ultimate analysis (% wt dry)	
Carbon	72.04
Hydrogen	4.08
Nitrogen	1.67
Oxygen	14.17
Sulphur	0.65
Chlorine	0.02
Ash	14.19
Lower heating value - LHV (MJ/kg a.r.)	27803.29

Table 3 summarizes the syngas composition, CGE and efficiency of conversion in CO and H_2 (hydrogen production potential) for all the gasifiers that have been considered. Cold gas efficiency (CGE) and hydrogen production potential must be as high as possible on condition that hydrocarbons (mainly methane) present in syngas must be as low as possible [8].

Cold gas efficiency (CGE) shows the energy efficiency of gasification process and it is defined as follow:

$$CGE = \frac{Syngas\ thermal\ energy\ [MW]}{Feedstock\ thermal\ energy\ [MW]} * 100 \tag{1}$$

Hydrogen production potential of the gasifier gives a better idea of how much of the thermal energy of coal can be converted into hydrogen and it is calculated with the formula:

$$Hydrogen\ production\ potential =$$

$$= \frac{CO\ and\ H_2\ thermal\ energy\ [MW]}{Feedstock\ thermal\ energy\ [MW]} * 100 \tag{2}$$

Table 3. Overall gasification performance indicators

Properties	Unit	Shell	Siemens	GE Texaco	Conoco Phillips
Fuel flow	t/h	165.4	165.4	165.4	165.4
Oxygen flow	t/h	145.27	145.27	160.27	132.27
Syngas flow	t/h	314.52	314.52	382.98	354.98
H_2	% vol.	64.92	64.92	49.29	44.28
CO	% vol.	25.94	25.94	18.8	30.28
CO_2	% vol	1.2	1.2	4.8	10.04
CH_4	% vol.	0.02	0.02	0.4	1.7
H_2S	% vol.	0.19	0.19	0.1	0.17
CGE	%	79.07	79.07	73.04	80.27
$CO+H_2$ efficiency	%	78.71	78.71	71.60	74.79

As can be noticed from the Table 3 on the basis of the efficiency of conversion in H_2 + CO, meaning the hydrogen production potential, Shell and Siemens technologies are superior to the other ones. Regarding the mechanism of raw gas production these two processes are identical. The Siemens gasifier has water quench which ensures the optimal condition for shift conversion, precondition for CO_2 capture. As Table 3 shows the GE Texaco is the least appropriate for our process, because of the low CGE and hydrogen production potential. Because of the relatively high methane content in the syngas, the Conoco Phillips (E-Gas) technology has high cold gas efficiency. This will be good in a power application, but may not be the optimum choice for a synthesis gas application, in which case the (H_2 + CO) yield will provide a better guide to process selection. For gasifiers which produce a syngas with significant concentrations of methane, is difficult to capture 90% of the carbon from the coal.

4. Conclusions

The purpose of this paper is to evaluate by modeling and simulation various coal gasification technologies for hydrogen and electricity coproduction, with carbon capture and storage (CCS). The most promising gasification concepts for hydrogen and electricity co-production with carbon capture are all based on entrained-flow gasifiers
The aim was to perform a multi-criteria analysis for different gasification concepts by eliminating gasifiers which are unapropriate for this purpose (IGCC plant concept with CCS). Modelling and simulation techniques were used to evaluate four gasification technologies and the main overall performance indicators.

5. Acknowledgements

References

[1] A.G. Collot, 2006, Matching gasification technologies to coal properties, International Journal of Coal Geology, 65, 191– 212
[2] C. Higman, M. Van Der Burgt, 2008, Gasification, Elsevier Science, Second edition
[3] E. Tzimas, A. Mercier, C. Cormos, S. Peteves, 2007, Trade-off in emissions of acid gas pollutants and of carbon dioxide in fossil fuels power plants with carbon capture, Energy Policy, 35, 3991 – 3998.
[4] International Energy Agency – Greenhouse Gas Programe (GHG), Potential for improvement in gasification combined cycle power generation with CO2 capture, Report PH4/19, 2003
[5]Food and Agriculture Organization of the United Nations www.fao.org
[6] C. Cormos, F. Starr, E. Tzimas, S. Peteves, Brown A., Gasifier concepts for hydrogen and electricity co-production with CO_2 capture, Third International Conference on Clean Coal Technologies, Cagliari, Sardinia, Italy, 2007
[7]C. Cormos, 2008, Decarbonizarea combustibililor fosili solizi prin gazeificare, Cluj University Press
[8] C. Cormos, 2009, Assessment of hydrogen and electricity co-production schemes based on gasification process with carbon capture and storage, International Journal of Hydrogen Energy,34, 6065-6077
[9] Statistical Review of World Energy BP 2008, www.bp.com

Modular Simulation of a 12MW Industrial Gasifier

X. Joulia[1], P. Floquet[1], R. Sardeing[2], O. Baudoin[2], M. Vieville[3], V. Brousse[3]

(1)University of Toulouse, Laboratoire de Génie Chimique, UMR CNRS 5503, INP-ENSIACET, 4 Allée Emile Monso, BP 44362, 31432 Toulouse Cedex 4, FRANCE.
Xavier.Joulia@ensiacet.fr, Pascal.Floquet@ensiacet.fr
(2)ProSim, Stratège Bâtiment A, BP 27210, F-31672 Labège Cedex, France
(3)Europlasma, 21 rue Daugère, 33520 Bruges

Abstract

In this work, a flexible model, built from elementary modules, is developed for an industrial waste gasification process, in an industrial moving bed reactor located in Morcenx (France). This gasifier is able to treat more than 46,875 ton/year of RDF (Refuse Derived Fuel) waste for producing 12 MW. Drying, pyrolysis, combustion / gasification and plasma polishing are used to convert waste directly into a synthesis gas composed of carbon monoxide and hydrogen. This synthesis gas is then used for producing electricity via gas engine.

Keywords: Modular Simulation, Industrial Gasifier, Synthesis Gas

1. Introduction

The objective is to turn waste power potential into electricity with an environment friendly process. The gasifier is designed to treat 6.25 t/h of Refuse Derived Fuel (RDF) waste for producing 12 MW. The three main steps of waste conversion to gas and electricity are shown in figure 1.
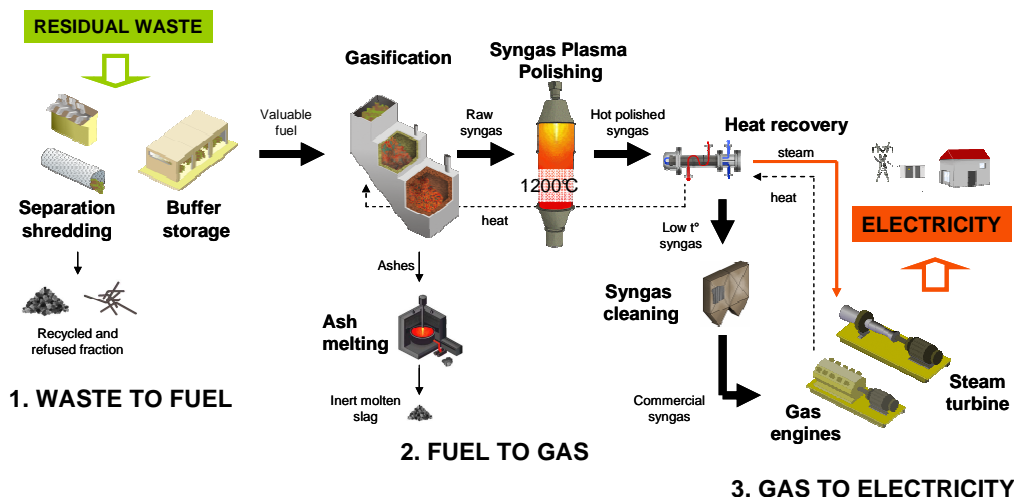
Figure 1: Main steps of the conversion of RDF waste to gas and electricity

In this paper, we focus on the second step of this scheme. Firstly the material system is briefly described. Then, the modeling approach of the three chambers of the gasifier is studied. Finally, some results on a case study, sensitivity analysis and conclusions are given.

2. Material system

The material system under consideration contains solids and gases. The main assumptions for modeling this material system are the following.

2.1. Solids

Three solids are considered in the model: the fuel part of the waste (FPW), the char and the ashes. The FPW constitutes the dry part of the waste, without ashes. It is defined from its atomic composition as an equivalent molecule: $CH_{\xi_1} O_{\xi_2} N_{\xi_3} S_{\xi_4}$. The fundamental property of RPW is its Lower Heating Value (LHV). Although some correlations have been proposed for its estimation (Niessen, 2002, Higman and Van der Burgt, 2003, Riazzi, 2005, Pröll and Hofbauer, 2008, Antonini, 2003,) it is better to determine its value from experimental data. The char is the solid residue coming out pyrolysis. Its percentage of carbon grows with the pyrolysis temperature and it is approximately 90% at 700°C (Nozahic, 2008). It is why its properties are assimilated to pure carbon. The soots are not taken into account. Finally, ashes that are the mineral part of the waste are taken into account only for mass and energy balances and are considered as chemically inert. The properties of ashes are assimilated to those of SiO_2.

2.2. Gases

The gases under consideration are the following:

- O_2, N_2 for the air feeds (drying, combustion and decarbonation sections);
- H_2O for the waste and air moisture. H_2O is also a pyrolysis product;
- H_2, CO, CO_2 that are combustion/pyrolysis/gasification products;
- CH_4 is also a pyrolysis product;
- Tar is assimilated to a mixture of two model molecules: toluene and naphthalene.

Some gaseous pollutants are also considered in this study: NO, NO_2, SO_2 and H_2S resulting from nitrogen and sulfur present in the waste.

3. Gasifier Modeling

The technology used by Europlasma for the waste gasification is confidential. It is a moving bed reactor with three chambers. For modeling purposes, the gasifier is divided in five main sections, in agreement with the gasifier structure and the elementary physico-chemical phenomena. The schematic representation of the gasifier model with all elementary components is presented on figure 2. The sections are the following:

- Drying section (chamber 1) in which waste moisture is decreasing.

- Flash pyrolysis section (chamber 2.1). This first step of the waste thermochemical transformation produces a gaseous phase, containing carbon monoxide, carbon dioxide, methane, hydrogen and water but also pollutants and tar and a solid phase, the char.

Modular Simulation of a 12 MW Industrial Gasifier

- Combustion/gasification section (chamber 2.2) of the char. Combustion is exothermic and is the energy source for balancing the endothermicity of gasification and pyrolysis. The plug-dispersion flow of the waste inside the gasifier chamber 2 is represented by a series of continuous stirred (perfect mixing) tank reactors (CSTR) 2.2.i.

- Decarbonation section (chamber 3). In this last step, the residual char is almost totally gasified by supplementary air in order to respect environmental constraint of carbon content in waste ashes: mass fraction of C less than 3 %.

- All gases produced in the three previous sections (pyrolysis, combustion / gasification and decarbonation) are then collected in the main chamber of the gasifier (chamber 2.3) where gas phase reactions, such as water gas shift, are occurring.
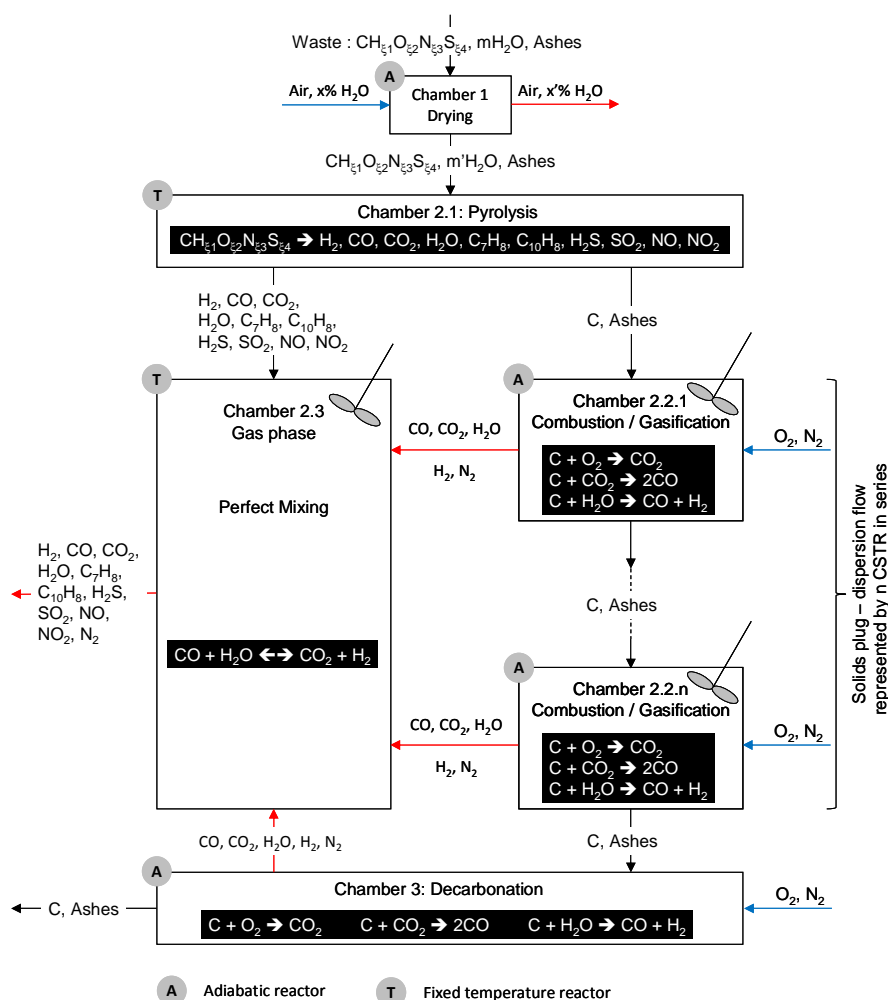
Figure 2: Schematic representation of the gasifier model

The two main assumptions of the model are the adiabaticity of the whole gasifier and the absence of cracking reactions. This last one is justified by the fact that, by using plasma technology, all organics are then transformed into CO/H_2 (see figure 1).

Each section model is built from elementary standard modules of the ProSimPlus® simulator. As illustration, figure 3 shows the simulation diagram of the chamber 2 where occur pyrolysis, combustion and gasification. The gasifier simulation diagram is obtained by aggregation of the sub-diagrams.
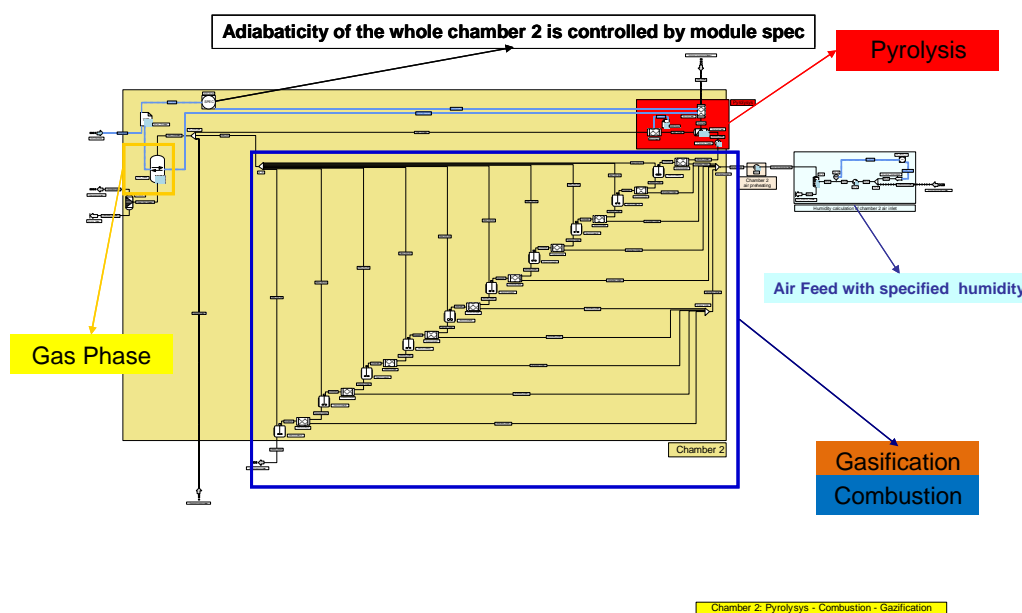
Figure 3: ProSimPlus® simulation diagram of the gasifier chamber 2

The model fundamental parameters concerns pyrolysis and are the ratio of pyrolysed carbon, τ, and the composition of the pyrolysis gas. These parameters are estimated from thermogravimetric experimental data.

4. Case study results

For the case study, the mass composition of the fuel part of the waste (FPW) is the following: C: 0.657929, H: 0.091236, O: 0.248008, N: 0.00257, S: 0,000257. Taking into account the moisture and ashes, the waste formula is:

$$CH_{1.6525}O_{0.2830}N_{0.0033}S_{0.0001}, 0.09636H_2O, 0.05625SiO_2$$
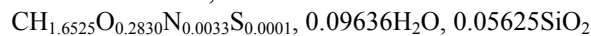
The FPW lower heating value is estimated to 27,330 kJ/kg. The waste total flowrate is 6.25t/h. The characteristics of the air feeds for drying, combustion and decarbonation are respectively:

- total flowrates: 4435, 9790 and 1000 Nm^3/h
- temperatures, after preheating: 360, 600 and 600°C

With these data, the temperature in the gasifier chamber 2, where take place pyrolysis, combustion and gasification, is 750 °C. The repartition of the head space incoming

molar flowrates are shown in figure 4. Most of gas, 64 mol%, is produced by combustion / gasification. The contribution of pyrolysis gas is 29 mol% and 7 mol% is coming from decarbonation.
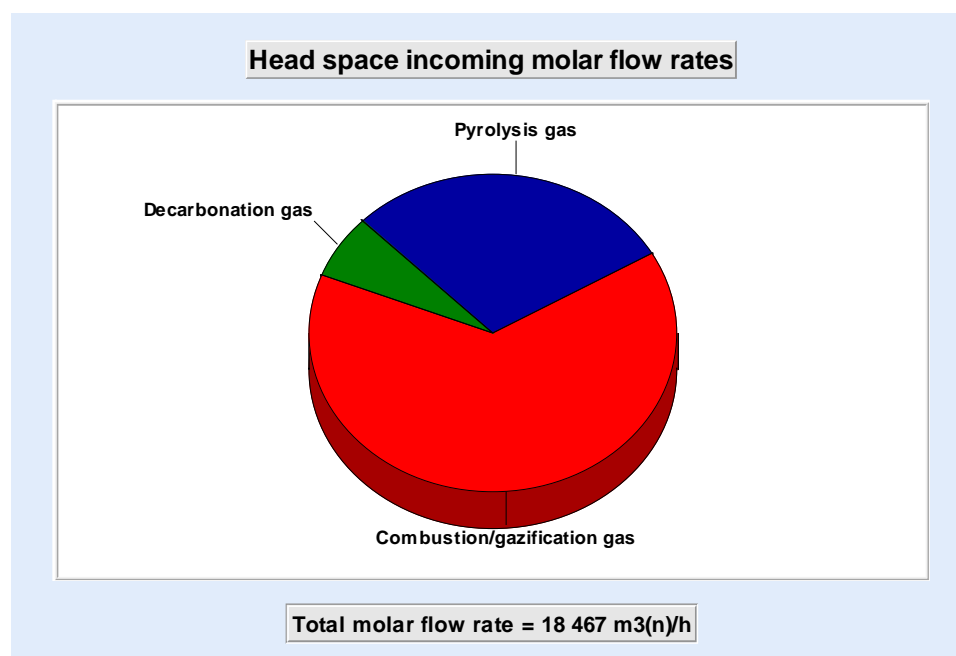
Figure 4: Head space incoming molar flowrates

Table 1 shows the results obtained by simulation. The LHV of the syngas is 5,157 kJ/Nm^3.

From sensitivity analysis, two model parameters appear essential in order to have a good representation of the gasifier operation: the waste LHV and the waste carbon conversion ratio into pyrolysis gas. The first fixed the thermal power available while the second specifies the repartition between the exothermicity and endothermicity potential. Their balancing, to ensure the gasifier adiabaticity, determines the operating temperature of the gasifier.

5. Conclusion

In this paper, we have proposed an original approach for the modeling of an industrial gasifier. The gasifier model is built up in three steps: decomposition in elementary components associated to gasifier structure and physicochemical phenomena and definition of subsystems; build up of the ProSimPlus® simulation diagrams of the subsystems from standard modules; aggregation of the sub diagrams to obtain the whole gasifier model. The first results obtained are physically correct, allowing the use of the model as decision-making tool for process design and operation. Notably a sensitivity analysis of the gasifier with respect to its operating parameters is currently underway. The final goal is to improve the efficiency of the waste conversion into electricity.

Additional experimental tests related to waste characterization and its pyrolysis are scheduled for definitively validating the proposed model.

	Wet air out of drying section	Syngas	Ashes
Total Flowrate			
Mass (kg/h)	6 173	18 743	932
Mol (Nm^3/h)	5 013	18 467	389
Mass fractions			
Char (C)	0	0	0.03
O_2	0.2154	0	0
N_2	0.7094	0.5684	0
H_2O	0.0752	0.0123	0
H_2	0	0.0182	
CO	0	0.3288	0
CO_2	0	0.0512	0
CH_4	0	0.0158	0
C_7H_8	0	0.0014	0
$C_{10}H_8$	0	0.0019	0
H_2S	0	$3.55\ 10^{-5}$	0
SO_2	0	$6.67\ 10^{-5}$	0
NO	0	$7.17\ 10^{-4}$	0
NO_2	0	$1.10\ 10^{-3}$	0
Ashes	0	0	0.97

Table 1: Simulation results

References

G. Antonini, Traitements thermiques des déchets. Processus thermochimiques. Techniques de l'Ingénieur, G2050, 2003

C. Higman, M. Van der Burgt, Gasification, Elsevier, 2003

W.R. Niessen, Combustion and Incineration Processes., *Marcel Dekker*, 3rd Edition, 2002

F. Nozahic, Production de gaz de synthèse par interactions à haute température du gaz, des goudrons et du résidu carboné issus de la pyrolyse de biomasses. Institut National Polytechnique de Toulouse, PhD Thesis, 2008

T. Pröll, H. Hofbauer, Development and Application of a Simulation Tool for Biomas, Gasification Based Processes, *International Journal of Chemical Reactor Engineering*, 6, A89, 2008

M.R. Riazzi, Characterization and Properties of Petroleum Fractions, *ASTM*, 1st Edition, 2005

Towards Integrated Product and Process Design for Biofuels

Anna Voll, Manuel Hechinger, Wolfgang Marquardt[a]

[a] Aachener Verfahrenstechnik – Process Systems Engineering, RWTH Aachen University, Templergraben 55, 52056 Aachen, Wolfgang.Marquardt@avt.rwth-aachen.de

Abstract

Sustainable processes for the conversion of whole plants into fuels are investigated in the research cluster "Tailor-Made Fuels from Biomass" at RWTH Aachen University (RWTH Aachen University, 2007). In contrast to known attempts, this project not only aims at the identification of components with promising fuel properties, but also accounts for the preservation of functional biomass structures within the biofuel production process. Accordingly, both fuel design approaches and new synthesis routes need to be explored. The solution strategy for these challenges is supported by systems engineering techniques: (i) promising target fuels are predicted by Computer-Aided Molecular Design (CAMD) and (ii) possible production alternatives are identified and classified by reaction network flux analysis (RNFA). In this contribution we introduce both methodologies and show how they can be combined to an integrated product and process design.

Keywords: biofuels, product-process design, CAMD, reaction network flux analysis

1. Motivation

Renewable raw materials are attaining increasing interest in the production of liquid transportation fuels. This change in feedstock brings along significant modifications in established processing techniques. As the physical and chemical properties of biomass strongly differ from those of crude oil, a sustainable (large-scale) production requires the development of innovative processes for the conversion, separation and purification of (by-)products, catalysts and solvents. New bio-based products should match the native molecular structures of biomass (and in particular accommodate the high O:C ratio). Consequently, product design for known applications must also be considered. Hence, a sensible approach to the transition from fossil to bio-based raw materials calls for an integrated product-process design. This idea (cf. Gani, 2009, for a review) has lately been discussed in different applications including solvent design for carbon capture (Bardow et al., 2009) or in catalyst design for upgrading biofuels (Resasco and Crossley, 2009). In this contribution, we extend the basic principles of integrated product and process design towards a sustainable production of tailored biofuels.

2. Theoretical background

In our work, two different methodologies are combined, namely reaction network flux analysis (RNFA) and product design. Their theoretical backgrounds are briefly introduced first, followed by an outline of how they can be combined to solve the integrated product and process design problem in the context of future tailor-made biofuels.

2.1. Reaction network flux analysis (RNFA)

RNFA is developed as a rapid screening method to bridge the gap between laboratory experiments and process design in the context of the transition from fossil to bio-based transportation fuels. Many alternatives need to be evaluated systematically at an early design stage to identify sustainable production routes. The backbone of the procedure derives from metabolic flux analysis (Schilling et al., 1998) which is routinely applied in systems biology. All reactions linking raw materials and products are summarized in a network, where substances and reactions are represented as nodes and arcs, respectively. Then, the mass balance for the network can be formulated as

$$A \cdot x = b, \tag{1}$$

where A is the matrix of stoichiometric coefficients. The rows of matrix A represent the substances and the columns relate to the reactions. x refers to the molar flux through the network, while b balances the product and by-product formation at each node.

Generally, the network includes more reactions than substances such that the corresponding equation system is underdetermined. Thus, optimization techniques are applied to find one or more optimal solutions for an objective function. In particular,

$$\max_{x,b} \ b_{\text{target}} \qquad s.t. \ Ax = b \qquad x,b \geq 0 \tag{2}$$

can be solved to find the reaction pathway (the combination of all required reaction steps) with the maximal yield of the target component b_{target}. In most cases more than one synthesis route is possible and all different flux scenarios need to be identified to enable a comprehensive evaluation. For this purpose the problem must be reformulated as a mixed-integer problem to be solved by adequate algorithms (e.g. Lee et al., 2000).

Subsequently the alternative solutions are evaluated with respect to multiple classification criteria. The number of reaction steps, the formation of by-products and the use of additional reactants are first indicators, which can be easily calculated as part of the analysis. Besides, the basic model can easily be expanded by additional constraints such as yield, mass or carbon efficient, toxicity, energy and material costs or even by more advanced criteria like separation effort. Promising reaction pathways as well as bottlenecks in candidate pathways can be identified from such an analysis to gain an insight in attractive production routes. More details about RNFA and an illustrative case study can be found elsewhere (Besler et al., 2009).

2.2. Property estimation from molecular structure via QSPR

Product design describes a search for a product that exhibits certain known, desired properties. Since only the property but not the product itself is known a priori, product design always requires appropriate models that link identifiers of the product to its resulting properties. In case fuel molecules are to be designed, relationships between the molecular structure of a fuel molecule and its macroscopic, thermodynamic and kinetic properties of interest have to be found. Moreover, these models not only need to supply structure-property relations for known molecules. Rather, their application in product design requires property prediction with sufficient quality for unknown molecular structures. These predictive methods can then be applied in a guided search for a suitable molecular structure by means of a Computer-Aided Molecular Design (CAMD) approach, where the search emanates from a previously defined, desired property towards a molecular structure.

Several methodologies for property modelling on a molecular basis have already been developed. Quantitative structure-property relationships (QSPR) have been successfully

used in various ways (Karelson et al., 1996). A QSPR model relates macroscopic molecular properties to the molecular structure through molecular descriptors. These descriptors capture a large variety of different attributes of a molecule which can be derived from its 3-dimensional molecular structure by means of computational chemistry (Todeschini and Consonni, 2009). The information gathered for each molecule is then used to determine those descriptors (or equivalently those molecular attributes), which have the strongest effect on the macroscopic property of interest. The resulting structure-property relations are often linear and of the form

$$\mathbf{y} = \beta_0 + \beta_1 \cdot \mathbf{x}_1 + \beta_2 \cdot \mathbf{x}_2 + ... + \beta_m \cdot \mathbf{x}_m , \tag{3}$$

where $\mathbf{y}$ represents the properties of n measured molecules in the training set (an n-dimensional property vector), the $\mathbf{x}_i$ relate to m different molecular descriptors of these n molecules, and β are model parameters to be determined via linear regression. The minimum number of required molecular descriptors for a sufficiently accurate property description is effectively determined by means of stepwise regression (Shacham and Brauner, 2003). Once this structure-property relation has been identified, it can be used to estimate the same property for new molecules of similar molecular structure. The better the new molecular structure is represented by the structures in the training set, the more reliable the resulting QSPR predicts the property of the new molecule. Since the structure of a molecule is represented by the aggregation of its molecular descriptors, structural similarity of two molecules can be measured by comparing their sets of descriptors. This is achieved by correlating the m-dimensional vectors that contain the descriptor data of both molecules (Shacham and Brauner, 2003).

2.3. Integrated Product and Process Design
The introduced strategies can be combined in two different ways to tackle the product and process design problem (Fig. 1). Either relevant fuel properties are calculated for all substances in an existing reaction network, such that they can be included as evaluation criteria in the network analysis (strategy 1). Or, a set of optimal fuel molecules is generated by the CAMD approach first and a reaction network towards these components is built next to identify and compare suitable production routes (strategy 2).

Figure 1. Integrated product and process design approach.
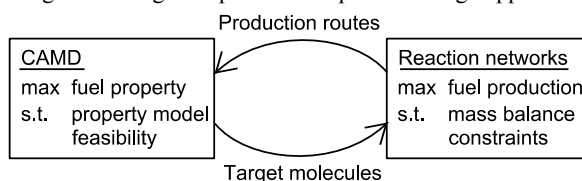
On the basis of these evaluations the trade-off between ideal properties and production effort can be discussed. The following case studies illustrate how these approaches can be used to form a profound basis towards a guided search for the tailor-made fuel.

3. Application to biofuel design
The above described methods of product and process design have been combined for the design of bio-based fuel molecules. Here, a reaction network of 64 chemical species ranging from biomass-derived intermediates to potential fuel molecules has been considered as the basis for the process design step. The quality of all molecules in terms of their suitability as a fuel has been exemplarily assessed by two properties, the

enthalpies of vaporization and combustion. Although these properties cannot be used alone for a comprehensive assessment of biofuels, they are meaningful criteria for a first assessment of the suggested method since they contribute to the integral energy balance of a combustion engine. While the enthalpy of vaporization characterizes the thermal energy that needs to be supplied to the engine to transfer the fuel from liquid to vapor state, the enthalpy of combustion refers to the chemically stored energy which becomes available during combustion. Hence, molecules with a low enthalpy of vaporization and at the same time a high enthalpy of combustion are regarded as target molecules.

3.1. Strategy 1: Evaluating reaction routes by predicted properties

Starting point in this case study is a reaction network around itaconic acid (an intermediate derived from biomass) which currently includes 64 substances and 95 reactions. Yield is introduced as a supplementary criterion which is given by

$$x_j \leq \eta_j \sum_{k,k \neq j} v_k x_k \quad \forall j. \tag{4}$$

Here, η is the molar yield of a reaction j and k is the set of fluxes entering or leaving a node. A yield of 97% is assumed to balance general losses, if no detailed information is available.

In addition, the enthalpies of combustion and evaporation are predicted for all network molecules by a QSPR approach. For this purpose, about 3200 molecular descriptors have been calculated by the Dragon software package (Talete, 2009). These calculations have not only been carried out for all the network substances, but also for about 1700 additional molecules from the DIPPR database (DIPPR, 2009), for which measurement data on enthalpies of combustion and vaporization are available. For each network substance, the 20 most similar DIPPR compounds were chosen as a training set to be used to establish a structure-property relation according to eq. (3). Since the molecules in each training set are structurally close to the respective network molecule, the resulting relations can be used to estimate the properties for the network molecules.

These properties serve as input to an analysis of the reaction network to compute two energy indicators, the specific and the net energy. The specific energy is defined as the difference between combustion and vaporization energy in absolute values. The net energy considers the production constraints in addition by multiplying the specific energy with the molar yield of the target component (converting one mol of reactant).

The ten best network molecules are chosen according to their specific energy. For each one all possible pathways are identified via a solution of the optimization problem (2). In total, 99 alternative reaction routes for the ten best fuel candidates are obtained, while in some cases more than 30 pathways are possible for the processing of a single substance.

In Fig. 2, the specific energy (gray bars) is compared to the net energy, which can be obtained from production via different reaction pathways (shaded area) for different components. All solutions of the optimization problem (2) lay within the shaded area. This area reduces to a line, if only one route is possible or if all routes have the same net energy. It is shown that the net and specific energy may strongly differ depending on the target molecule and the reaction pathway. Without focusing on any further detail, it becomes clear that the synthesis route must be considered in the selection of potential fuels. The introduction of more evaluation criteria (e.g. processing and investment costs) will further restrict the number of promising candidate fuels and reaction routes.

Towards Integrated Product and Process Design for Biofuels

Figure 2: Comparison of specific and net energy
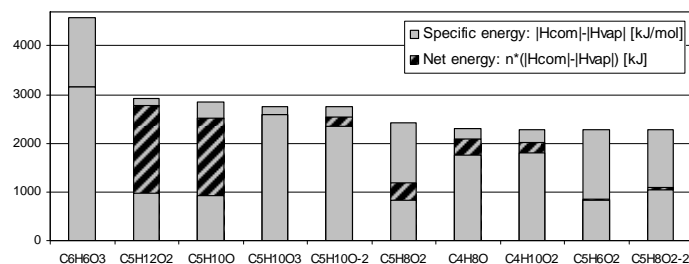
3.2. Strategy 2: From property predictions to reaction networks

Although strategy 1 in Sec. 3.1 combines the objectives of product yield and product properties, the search space for the ideal fuel molecule is restricted to the substances present in the reaction network. In order to extend the search space, a QSPR according to eq. (3) has been established by using the union $\Sigma = \cup (\sigma_i)$ of all similarity groups σ_i of the network molecules as a training set (Fig. 3). The resulting QSPRs for the 270 molecules in Σ were found to require only two descriptors each for an accurate description of both enthalpies. Since Σ contains a much larger variety of structurally different molecules than the σ_i, the resulting QSPRs describing this data can accordingly be extrapolated to a wider range of molecular structures.

Figure 3. Molecule sets and calculation scheme (left) and optimal molecule (right) for strategy 2
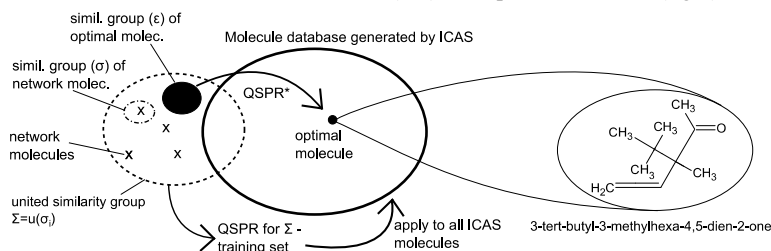
The four molecular descriptors necessary to describe the target properties of the molecules within Σ are (i) the number of carbon atoms in the molecule, (ii) the sum of conventional bond orders for the enthalpy of combustion, (iii) the first order solvation connectivity index and (iv) the number of OH-groups for the enthalpy of vaporization. Besides their physical justification, these descriptors can all be derived from the 2-dimensional graph of a molecule. They do not need any 3-dimensional or surface charge information. Accordingly, in order to extend the search space beyond the reaction network, it is reasonable to generate a large variety of molecular graphs and estimate their properties using the QSPR approaches described above (Fig. 3).

To this end, the ICAS software package (ICAS, 2009) has been used to generate about 17000 cyclic, acyclic and aromatic organic structures from a large set of molecular fragments. For all of them the four relevant descriptors have been calculated and their enthalpies of vaporization and combustion were estimated. Since the generated structures not are necessarily similar to the molecules in Σ and thus cannot be described by the identified QSPR approaches, the best molecule needs to be checked for similarity with Σ. Therefore, a similarity group ε for the optimal molecule has to be found within Σ (Fig. 3) based on which a new model QSPR* is established to predict the properties of the optimal molecule. Since now only the most similar molecules within Σ are used for

prediction, the resulting property estimate can be considered more accurate. Hence, if a similarity group ε is found and the predictions of both QSPR and QSPR* give similar property estimate, the QSPR can be considered valid for the optimal molecule.

This procedure has been performed for the molecules generated by ICAS. For the molecule exhibiting both, the best properties and a similarity group ε of 20 molecules (Fig. 3) a model QSPR* has been established. The deviations to the original QSPR models are 7.8% and 1.5% for vaporization and combustion enthalpies, respectively, which indicates the validity of the QSPR. The identified molecule is likely to be stable and can be synthesized, since a very similar structure with one instead of the three methyl groups in the middle branch has already been synthesized (Thompson, 1967).

4. Remarks and Conclusions

In this contribution we sketch how methods of CAMD and RNFA complement one another to an integrated product-process design approach: The evaluation by RNFA delivers information about the production route, while target molecules are predicted by CAMD. Knowing that the individual approaches as well as their combination still have a huge potential for improvement and extension, we consider this work as a proof of concept and a first step on the successful way towards a tailor-made fuel.

5. Acknowledgements

References

A. Bardow, K. Steur, J. Gross, 2009, A continuous targeting approach for integrated solvent and process design based on molecular thermodynamic models,

A. Besler, A. Harwardt, W. Marquardt, 2009, Reaction networks – A rapid screening method, *Proceedings of Escape 19*, Cracow, Elsevier, edited by J. Jezowski and J. Thullie, 243-248.

E.A. Brignole, M. Cismondi, 2003, Molecular Design – Generation & Test Methods. In Computer Aided Molecular Design: Theory and Practice. Amsterdam: *Elsevier*, 23.

DIPPR®801 database, 2009, Design Institute for Physical Properties, UT, USA, *http://dippr.byu.edu.*

R. Gani, 2009, Modelling for PSE and Product-Preocess Design, 10th Intl. Symposium on Process Systems Eng., Brazil.

ICAS Software package, 2009, Version 12.0, *www.capec.kt.dtu.dk.*

M. Karelson, V.S. Lobanov, A. R. Katritzky, (1996). Quantum-chemical descriptors in QSAR/QSPR studies, *Chemical Reviews*, 96, 1027.

S. Lee, C. Phalakornkule, M. Domach and I. Grossmann, 2000, Recursive MILP model for finding all the alternate optima in LP models for metabolic networks. *Comp. Chem. Eng.*, 24, 711-716.

D.E. Resasco, S. Crossley, 2009, Molecular Engineering Approach in the Selection of Catalytic Strategies for Upgrading of Biofuels, AIChE J., 55, 1082-1089.

RWTH Aachen University, 2007, Tailor-Made Fuels from Biomass – Excellence RWTH Aachen University establishes a Fuel Design Center*, NatureJobs.*

C. Schilling and B. Palsson, 1998, The underlying pathway structure of biochemical reaction networks. *Proc. Natl. Acad. Sci.*, 95, 4193-4198.

M. Shacham and N. Brauner, 2003, The SROV program for data analysis and regression model identification, *Comput. Chem. Eng.*, 27, 701.

Talete srl (2009), DRAGON for Windows. Version 1.4, *www.talete.mi.it.*

B. Thompson, 1967, US 3337634 (A).

Todeschini, R., Consonni, V. (2009). *Molecular Descriptors for Chemoinformatics*. Weinheim: *Wiley.*

Modelling of the solar sludge drying process Solia™

Nicolas Roux[a], Daniel Jung[a], Jérôme Pannejon[b], Cyrille Lemoine[a]

[a]*Veolia Environnement – Water Research Centre, Chemin de la Digue – BP76, 78603 Maisons Laffitte Cedex, France, nicolas.roux@veolia.com*
[b]*Veolia Water – Technical Department, Immeuble Giovanni Battista B, 1, rue Giovanni Battista Pirelli, 94417 Saint Maurice, France, jerome.pannejon@veoliaeau.fr*

Abstract

Greenhouses of solar sludge drying are developed as an economical alternative to the classic thermal dryers. In this study, two complementary models are used to observe and predict the operation of Solia™ units at different time and space scales. A 3D model informs on solar aspects and internal profiles of the multi-physical phenomena. A 0D model integrates drying kinetics and a new approach in stratified windrow in order to forecast the drying cycle efficiency on a large time scale. These tools are confronted to experimental data acquired on a Solia™ unit located in Fonsorbes (France).

Keywords: Solia™, solar drying, greenhouse, urban residual sludge, modelling

1. Introduction

In urban wastewater treatment plants, the majority of the waste products are represented by sludge. Agricultural spreading, incineration or landfill disposal are the different ways to evacuate this waste production. Using any of these solutions, sludge must be transformed and a major stake remains the volume reduction by water removal. In this drying step, solar drying greenhouses become an interesting economical alternative to the thermal dryers.

Solia[TM] is a solar sludge drying unit marketed by Veolia Water Solutions & Technologies adapted for small and medium-sized wastewater plants. In the greenhouse, the drying procedure consists of sludge spreading and turning into windrow, and the drying is increased with air renewal and convective exchanges. The process to be efficient requires knowledge of drying rate as a function of the greenhouses environment and their operation. The modelling is a useful tool to understand the drying mechanisms and to optimize the design and building of a Solia™ unit.

The objective of this study is to develop a predictive model of solar drying efficiency in greenhouse according to external conditions and operation.

2. Problem Statement, background

In a Solia™ unit, sludge is shaped into windrows in order to offer a most optimized transfer area with the forced air flux. This high forced air renewal allows the water evaporation from the sludge, and the transport of the water out of the unit. To avoid heterogeneous thermal and water concentration fields of air and sludge, two means operate. Firstly some fans placed inside the unit ensure homogeneous water concentration and sufficient convective transfers with the sludge. Secondly a robot named SoliaMix™ regularly adds some fresh sludge within windrows and mixes them to homogenize dryness and temperature fields of these last (Fig. 1).
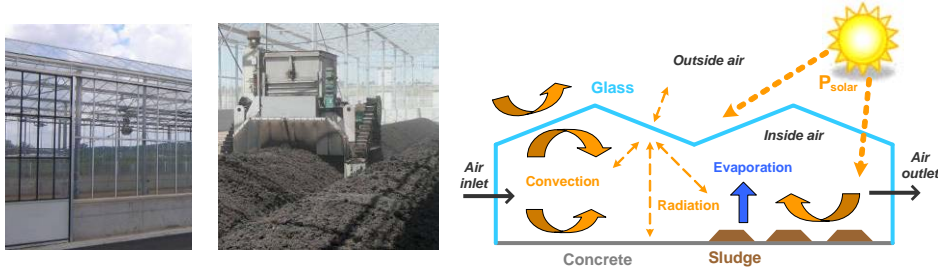
Figure 1. Solia™ unit and SoliaMix™ robot   Figure 2. Transfer phenomena in a Solia™ unit

Major operating costs of Solia™ are composed of energy consumption (extraction…). Optimized design and operation could improve both efficiency and costs of the process.

The model has to take into account all the coupled transfers occurred in the greenhouse. By specifying only the external conditions, localization and the material properties, the model should predict the thermal fields of sludge, air and boundary walls according to radiation and convection, as well as water quantity evaporated from the sludge (Fig. 2).

3. Materials & Methods

3.1. Solar greenhouse

A greenhouse based in Fonsorbes (France – 31) has been chosen as experimental case for this study. This unit is made up of two drying bays and is about 50 meters long, 4 meters high and 15 meters wide. According to the design geometry and the current operating conditions, sludge is spread only in the first bay of the greenhouse. Some sensors have been implemented to validate the numerical model: hygrometers, pyranometers, temperature probes, anemometers, air flowmeter...

3.2. Mathematical theory

Some previous works are focused on the modelling of drying kinetics but a few of them are applied and validated on a real solar greenhouse. *Jung et al. (2009)* have developed a model focused on thermal exchanges and they concluded that sludge temperature is highly sensitive to model parameters [1]. The radiative heat quantity exchanged between two walls, required by boundary walls thermal balances, is modelled by [5]:

$$R_{1-2} = \frac{\sigma\left(T_1^4 - T_2^4\right)}{\dfrac{1-\varepsilon_1}{A_1\varepsilon_1} + \dfrac{1}{A_1 F_{12}} + \dfrac{1-\varepsilon_2}{A_2\varepsilon_2}} \qquad \text{(Eqn 1)}$$

So, these boundary walls balances take into account solar effects, internal and external convections and infrared radiations. The differential equation of greenhouse roof is:

$$\rho_{V\sup}V_{V\sup}C_{pV\sup}\frac{dT_{V\sup}}{dt} = ...$$
$$...\left[P_{V\sup}a_{V\sup} + h_{V\sup}(T_a - T_{V\sup}) + h_{ext}(T_{ext} - T_{V\sup})\right]A_{V\sup} + \sum_{\substack{j \\ walls \\ \neq V\sup}} R_{j-V\sup} \qquad \text{(Eqn 2)}$$

Thermal differential equations on sludge and internal air are similar to the last one by adding evaporation effect and also, for the second balance, air renewal. For the sludge:

$$\rho_{Bo}V_{Bo}C_{pBo}\frac{dT_{Bo}}{dt} = \left[P_{Bo}a_{Bo}\tau_V + h_{Bo}(T_a - T_{Bo})\right]A_{Bo} + \sum_{\substack{j \\ walls \\ \neq Bo}} R_{j-Bo} + M_s Lv\frac{dX}{dt} \qquad \text{(Eqn 3)}$$

Modelling of the solar sludge drying process SOLIA™

Jung et al. (2009) analyse the overestimation of evaporation rate due to the lack of drying kinetics in their thermal model, which also induces some limitation on the sludge temperature prediction [1]. In order to predict more accurately this temperature, *Slim (2007)* and *Hamadou (2007)* introduced two different approaches integrating drying kinetics [2, 7]. The first one is a reduction of the mass transfer coefficient which is defined during the first drying rate by a Chilton-Colburn analogy. Their coefficient reduction is an identified function from experimental data of lab-scale, maybe difficult to extrapolate for another kind of sludge. The second approach is based on the drying theory including the drying characteristic curve which is applied on a thin sludge bed (30-40cm of thickness). As Solia™ is working until 1 meter windrow thickness, our model introduces a stratified windrow of two layers. The Hamadou's approach is implemented; the characteristic curve is useful to describe drying kinetics from initial and equilibrium water concentration in every type of air conditions.

Our works highlight and improve the method to model the drying rate and mass transfers within the windrow. The drying mechanism occurs in the thin upper layer and the windrow core is only subject to thermal exchange by conduction.

The maximum drying rate is theoretically determined by a thermal balance on sludge which is supposed by an analogy of heat and mass transfers through a thin totally wet layer. During this stage, even isenthalpic, water is evacuated at constant rate i.e. first drying rate and the sludge reaches an equilibrium temperature given by the wet temperature. By neglecting the sludge inertia during this stage, maximum drying rate can be estimated by a thermal balance on sludge:

$$\left(-\frac{dX}{dt}\right)_1 = \frac{1}{M_s Lv}\left[A_{Bo}\left[P_{Bo}a_{Bo}\tau_V + h_{Bo}(T_a - T_h)\right] + \sum_{\substack{j \\ walls \\ \neq Bo}} R_{j-Bo}\right] \qquad \text{(Eqn 4)}$$

Thus the maximum drying rate is reduced by a polynomial function $f(X_r)$ based on sludge desorption theory and experimental data (*Kechaou, 2000* and *Kouhila, 2001*) in order to obtain the drying rate of thin layer [3 – 4]:

$$\left(-\frac{dX}{dt}\right) = \left(\frac{dX}{dt}\right)_1 f(X_r) \qquad \text{(Eqn 5)}$$

The equilibrium water concentration X_{eq} is deduced from desorption isotherm using either Oswin's model [6] or an identified polynomial function. X_r is the reduced water concentration defined from equilibrium X_{eq} and critical water concentration X_{cr} by:

$$X_r = \frac{X - X_{eq}}{X_{cr} - X_{eq}} \qquad \text{(Eqn 6)}$$

This work is also dedicated to overcome the difficulty to manage layer transfers during iterations of numerical solving. The exchange depth of sludge takes place in balances and plays a significant effect on the model prediction. As the thin layer dries between two iterations therefore its volume is reduced. A mass mixing law is applied on layers after numerical resolution of ordinary differential equations in order to uniform layers dryness according to the transfer by sludge exchange depth and windrow turning.

3.3. Solar influence
All thermal fields in the greenhouse are directly (glasses, sludge…) or indirectly (air) drained by the solar influence, and so this last should be accurately represented to obtain

a representative model. Firstly, the model needs to be fed with solar radiation as a time function according to the greenhouse location. Secondly, the three dimensional geometry should be taken into account for different reasons. The glass radiation properties depend on the solar angle of incidence, and so the sun position should be integrated for each surface thermal balance. The 3D discretization is also required to evaluate solar radiation on interior surfaces, which is impacted by exterior surfaces properties and localization.

3.4. Modeling tools

Two approaches will be used to reach the previous objectives. A Computational Fluid Dynamics (CFD) code Ansys® - Fluent®, allows the three dimensional discretization and the solar influence on the different surfaces/walls of the greenhouse. A specific internal tool in the code, called "Solar Load Model", is based on the ASHRAE Handbook and gives an answer to the solar influence issue [8]. Moreover, the CFD model takes into account internal hydraulics, all thermal exchanges and water evaporation but on a short time period because a drying representative time scale would be computationally too expensive. This way, water evaporation is set in this first approach on first drying rate condition.

Then, a second model is developed on Matlab® Simulink® in order to overtake this last limitation. This zero-dimensional model would be set up thanks to the solar influence calculated by the Ansys® - Fluent® code. Thermal exchanges are identical with the 3D model, but the large time-scale resolution allows using drying kinetics as introduced previously, therefore a drying cycle simulation is achievable.

4. Results & Discussions

4.1. Solar dependence and prediction

The "Solar Load Model" gives very similar solar radiation with experimental measurements in Fonsorbes (Fig. 3).
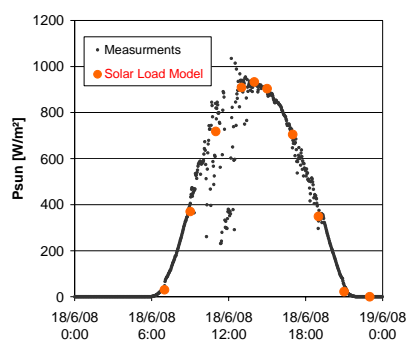
Figure 3. Comparison of horizontal solar flux [W/m²]
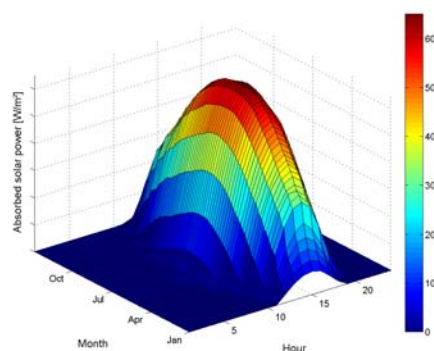
Figure 4. Absorbed solar flux [W/m²] of sludge as a function of date and hour in Fonsorbes

This model allows obtaining this radiation as a function of time and greenhouse location, and it is computed for different date or latitude. For each boundary wall, the solar flux multiplied by its absorption coefficient is stored and can be displayed (Fig. 4).

This data is locally used by the 3D model in order to parameter the solar influence in the thermal balances. Moreover, it creates a database which can feed the 0D model with solar profiles all year long depending on walls incidence angle and greenhouse location.

Modelling of the solar sludge drying process SOLIA™

4.2. Hydraulic results

The 3D model needs a numerically greenhouse construction and space discretization (Fig. 5). This model allows solving heat balances and internal hydraulics induced by fans and the extraction with the previous solar considerations. Hydraulic results underline a progressive augmentation of air temperature along the greenhouse. The internal fans induce a good homogenization of temperature without natural convection. The mass evaporation is assumed to be fixed at a maximum drying rate in this model without consequences on global results because of short time solving. The water evaporation profiles underline the internal air volume impacted by this mass transfer.
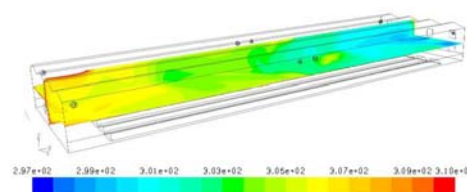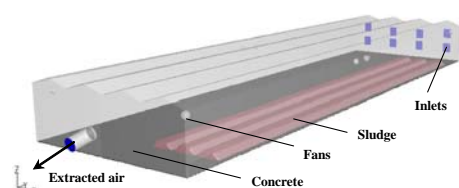
Figure 5. Geometry of Solia™ unit (Fonsorbes)   Figure 6. Example of temperature field [K]

This 3D model allows understanding internal mechanisms and is useful to optimize geometry or position dependence. Therefore, prediction of one drying cycle is required.

4.3. Drying cycle scale with 0D model

The systemic approach allows drying efficiency evaluation by predicting all temperature profiles which are coupled to sludge drying behavior. Therefore all boundary surface temperatures are computed with a 0D model (glass walls and roof, concrete and sludge) from input data like solar radiation and external moisture and temperature. The internal air moisture is also computed; consequently the drying performance can be translated to an extracted water quantity. The windrow height and dryness are obtained by computing our drying model according to simulated states and process operation.

The 0D model should firstly be confronted on experimental data acquired on Fonsorbes unit before its use to predict Solia™ efficiency on one drying cycle. Some simulation results are displayed on the figure 7 to illustrate the windrow volume reduction with time and the sludge dryness evolution. This validation has been achieved on one specific batch drying cycle i.e. 43 days to dry 51m³ of initial sludge volume from 37% to 89% dryness.
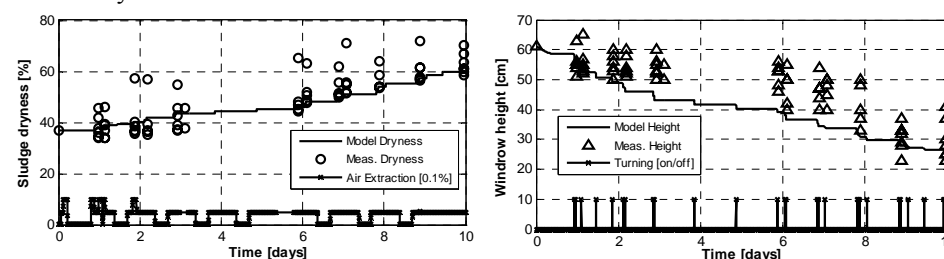
Figure 7. Example of windrow height and dryness evolution on 10 days

The results are in agreement with experimental ones despite the high variance of experimental measurements in terms of height and dryness. In the same manner, the dynamics of temperatures (sludge, internal air and walls) and air moisture are also coherent.

However, a parametric study on the exchange depth and absorption coefficients would improve the model predictions.

5. Conclusion

The complementarities of these two models lead to a good comprehension of the unit operation at different time and space scales. The 3D one is used to predict solar influence according to the greenhouse location, geometry and position, and can create a database useful for both models. Moreover, internal hydraulics and operation can be observed in accordance to operating conditions. The 0D model is focused on large time-scale in order to reach a representative drying cycle, thanks to kinetics implementation and our stratified model. The results of this model are quite close to the experimental ones; however a numerically sensibility analysis and a parametric identification with more experimental data are required. Then Solia™ units design and operation could be optimized with these tools according to its location.

Notations

Parameters				*Greeks letters*		
a	Solar absorption coefficient	[-]		ε	Infrared emissivity	[-]
A	Area	[m²]		ρ	Density	[kg.m⁻³]
Cp	Specific heat	[J.kg⁻¹.K⁻¹]		τ	Solar transmission coefficient	[-]
f	Identified function	[-]		σ	Stefan-Bolztmann's constant	[W.m⁻².K⁻⁴]
F	Vision factor between walls	[-]				
h	Convective exchange coefficient	[W.m⁻².K⁻¹]		*Indices*		
Lv	Latent heat of water vaporisation	[J.kg⁻¹]		l	First or maximum drying rate	
M	Mass	[kg]		a	Air	
P	Solar radiation (normal to wall)	[W.m⁻²]		Bo	Sludge	
R	Radiative heat	[W]		cr	Critical	
t	Time	[s]		ext	External	
T	Temperature	[K]		eq	Equilibrium	
V	Volume	[m³]		h	Wet	
X	Sludge water concentration	[kg water.kg⁻¹ dry solids]		r	Reduced	
				v	Vapour	
Abbreviations				V	Glass	
0D	Zero-dimensional			$Vsup$	Glass roof	
3D	Three-dimensional					

References

[1] D. Jung, N. Roux, C. Lemoine, J. Pannejon, Modélisation du procédé Solia™, STIC & Environnement Symposium, Calais, 15p, 2009.
[2] H. Hamadou, Modélisation du séchage solaire sous serre des boues de stations d'épuration urbaines, Ph.D. thesis, Université Louis Pasteur Strasbourg I, 2007.
[3] N. Kechaou, Etude théorique et expérimentale du processus de séchage de produits agroalimentaires, Thesis, Faculté des Sciences, Tunis, 2000.
[4] M. Kouhila, Etude expérimentale et théorique des cinétiques de séchage convectif partiellement solaire des plantes médicinales et aromatiques de la région de Marrakech, Thesis, Université Cadi Ayyad, Marrakech, Maroc, 2001.
[5] M.F. Modest, Radiative heat transfer, Academic Press-Elsevier Science, 2003, 822p.
[6] G.R. Oswin, The kinetics of package life, Int. Chem. Ind., 65, 419, 1946.
[7] R. Slim, Etude et conception d'un procédé de séchage combiné de boues de station d'épuration par énergie solaire et pompe à chaleur, Ph.D. thesis, Mines Paris, 2007.
[8] ASHRAE Handbook of Fundamentals, Eq. 20 and Table 7, Chap. 30, 2001.

Modeling and Experimental Validation of a PEM Fuel Cell System

Chrysovalantou Ziogou[a,b], Spyros Voutetakis[a], Simira Papadopoulou[a,c], Michael C. Georgiadis[b]

[a]*Chemical Process Engineering Research Institute (C.P.E.R.I.), Centre for Research and Technology Hellas (CE.R.T.H.), P.O. Box 60361, 57001 Thermi-Thessaloniki, Greece, ziochr@cperi.certh.gr,paris@cperi.certh.gr*
[b]*Department of Engineering Informatics, University of Western Macedonia, Vermiou and Lygeris str., Kozani, 50100, Greece, mgeorg@uowm.gr*
[c]*Department of Automation, Alexander Technological Educational Institute of Thessaloniki, P.O. Box 141, 54700 Thessaloniki, Greece, shmira@teithe.gr*

Abstract

This work presents a detailed dynamic model and a model validation study using real data from a Hydrogen Fuel Cell Testing Unit (HFCTU). A parameter estimation technique is employed for the determination of key model parameters and the validation of the overall system behavior is carried out by comparing experimental and simulation results. Data illustrate the transient response of the system during load changes. The model is oriented towards process optimization and control and relies on mass balances and electrochemical equations implemented in the gPROMS[TM] software environment.

Keywords: PEM Fuel Cell; parameter estimation, dynamic modelling

1. Introduction

Fuel Cells (FC) systems are a potentially good clean energy conversion technology and they can be used in a wide variety of power generation applications. They are categorized mainly on the type of electrolyte used, operating conditions or fuel. The Polymer Electrolyte Membrane fuel cells (PEMFC) are currently considered a good candidate for ground vehicle applications and small portable devices as they have high power density, fast start-up time as well as long cell and stack life. The critical operating parameters are mainly the air and hydrogen feed, flow and pressure regulation, and heat and water management. This work is focused on the validation of a model that incorporates features for these parameters, against a real test bed system. The following section presents a dynamic fuel cell model while the subsequent section introduces a model validation procedure including a comparison of model predictions against experimental results.

2. Modeling

The proposed model rely on first-principle equations combined with equations having experimentally defined parameters thus resulting in a semi-empirical system. The model accounts for mass dynamics in five control volumes: the gas flow channels, the gas diffusion layers and the membrane, as well as thermal dynamics. The mathematical model equations that describe the operation of the FC consist of the voltage-current characteristics and a relationship for the consumption of the reactants as a function of the current drawn from the fuel cell.

2.1. General analysis and modelling assumptions

In order to simplify the modeling and reduce the computation time the following assumptions are drawn. The gases are ideal and uniformly distributed inside anode and cathode. The stack is fed with humidified hydrogen and air because the use of humidified fuel and air improves the efficiency of the FC, thus the model have to take this into account. The temperature is constant and uniform for each experiment. The gas channels along the electrodes have a fixed volume with small lengths, so that it is necessary only to define one single pressure value in their interior. Regarding the operation of the system, during experiments the produced water is continuously removed from the cathode flow and also the condensed water on the anode is dragged by flow of the unreactted hydrogen. The modeling of the gas diffusion layers as well as the modeling of the membrane rely on the same equations as presented in del Real (2007). The physical parameters of these equations are adjusted according to the experimental measurements taken from the real system.

2.2. Electrochemical Equations and Voltage Calculation

Typical characteristics of FC are normally given in the form of polarization curve, which is a plot of cell voltage versus cell current density. To determine the voltage-current relationship of the cell, the cell voltage has to be defined as the difference between an ideal Nernst voltage and a number of voltage losses as it is described in the current section. The main losses are categorized as activation, ohmic and concentration losses. The equation that combines these irreversibilities expresses the actual cell voltage:

$$V_{cell} = E_{nernst} - V_{act} - V_{ohm} - V_{conc} \tag{1}$$

The above equation is able to predict the voltage output of PEM fuel cells of various configurations. Depending on the amount of current drawn the fuel cell generates the output voltage according to (1). The electric power delivered by the system equals the product of the stack voltage Vcell and the current drawn I:

$$P = I \cdot V_{cell} \tag{2}$$

This description for the activation overvoltage takes into account the concentration of oxygen at the catalyst layer (Pathapati et al. 2005)

$$V_{act} = \xi_1 + \xi_2 T + \xi_3 T_{st} \ln(I) + \xi_4 T_{st} \ln(c_{O_2}) \tag{3}$$

At a later stage, as current density rises, ohmic losses (Vohm) prevail. They are derived from membrane resistance to transfer protons and from electrical resistance of the electrodes to transfer electrons.

$$V_{ohm} = \left(\xi_5 + \xi_6 T + \xi_7 I \right) I \tag{4}$$

Finally the mass transport or concentration losses result from the change in concentration of the reactants at the surface of the electrodes as the fuel is used (Larminie J., 2003):

$$V_{conc} = \xi_8 \exp(\xi_9 I) \tag{5}$$

where ξ represents experimentally defined parametric coefficients whose vary can vary from stack to stack.

2.3. Mass Balance Equations

The model equations consist of the standard material balance of each component. Every individual gas follows the ideal gas equation. Therefore mass is described through partial pressures of each gas in the material balances. Applying mass balance to the cathode channel volume, assessing the inlet and outlet flows of the channel and the exchange flow between it and the gas diffusion layer, the following equations are derived:

$$\frac{dm_{O2,cach}}{dt} = \dot{m}_{O2,cach,in} - \dot{m}_{O2,cach,out} - \dot{m}_{O2,caGDL2cach} \tag{6}$$

$$\frac{dm_{N2,cach}}{dt} = \dot{m}_{N2,cach,in} - \dot{m}_{N2,cach,out} \tag{7}$$

$$\frac{dm_{v,cach}}{dt} = \dot{m}_{v,cach,in} - \dot{m}_{v,cach,out} + \dot{m}_{v,caGDL2cach} + \dot{m}_{evap,cach} \tag{8}$$

$$\dot{m}_{l,cach,in} - \dot{m}_{l,cach,out} - \dot{m}_{evap,cach} = 0 \tag{9}$$

$$m_{ma,cach} = m_{O2,cach} + m_{N2,cach} + m_{v,cach} \tag{10}$$

$$\dot{m}_{O2,caGDL2cach} = \frac{N_{fc}M_{O2}I}{4F} \tag{11}$$

The amount of water vapor in the fuel and air is calculated from the value of relative humidity ($\varphi_{[an,ca]ch,in}$), the saturation pressure (p_{sat}) and the temperature ($T_{[an,ca]ac,in}$) :

$$w_{v,[an,ca]ch,in} = \frac{M_{H2O}}{M_{air}} \frac{\phi_{[an,ca]ch,in}p_{sat}\left(T_{[an,ca]ch,in}\right)}{p_{[an,ca]ch,in} - \phi_{[an,ca]ch,in}p_{sat}\left(T_{[an,ca]ch,in}\right)} \tag{12}$$

The equations that give the amount of each species going into the channels which will be useful for the determination of the pressure inside the channels are presented at Pukrushpan et al. (2005). In order to describe the evaporation and condensation dynamics inside the channel, the proposed equations refer to those used by Golbert & Lewin (2007). In conjunction with these equations the inlet mass flow rate of the nitrogen, oxygen and vapor going into the anode channel can be determined:

$$\dot{m}_{k,cach,in} = w_{k,cach,in} \frac{1}{1 + w_{v,cach,in}} \dot{m}_{cach,in} \qquad {}_{,k=[O_2,N_2,v]} \tag{13}$$

At the cathode the liquid water condensed is dragged by the air, so the outlet flows are:

$$\dot{m}_{cach,out} = K_{cach,out}\left(p_{cach} - p_{out}\right) \tag{14}$$

$$\dot{m}_{k,cach,out} = \frac{m_{k,cach}}{m_{ma,cach}} \dot{m}_{cach,out} \quad {}_{,k=[O_2,N_2,v]} \qquad (15)$$

The equations that describe the anode channel are analogous to the ones that describe the cathode flow channel.

3. Experimental Validation

In order to assess the validity of the developed model a real Polymer Electrolyte Membrane (PEM) fuel cell system has been used to generate experimental data under various conditions. The effect of operational conditions such as temperature, pressure and humidity in the performance of the system was investigated, as the underlying operating conditions significantly affect profitability, effectiveness and safety aspects. Thus the validation procedure relies on experiments under steady state and dynamic conditions. The optimally defined values of the parameters were then used to validate the model under other experiments at different operational conditions.

3.1. Experimental Setup

The fuel cell unit is integrated with several auxiliary components to form a complete fuel cell system. The setup is comprised of a PEM Fuel Cell working at a constant pressure and a Power Conversion Device capable of controlling the current drawn from the FC. All experiments were conducted on a Fuel Cell Testing Unit (FCTU) composed of a humidification system, two mass flows for the regulation of the hydrogen and the air flow and two PID controllers for the anode and cathode pressure regulation. Also the temperature control subsystem includes an air cooling system and a heat up system. A simplified process and instrumentation diagram of the unit is presented in Fig 1.
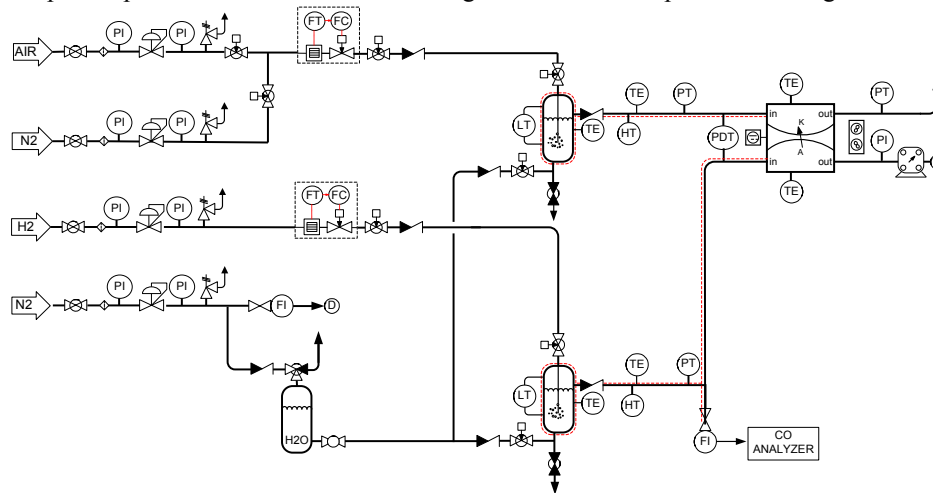
Figure 1 Fuel Cell Testing Unit

3.2. Experimental Procedure

The experimental procedure consists of reading the dynamic response of the cell voltage and cell power after the occurrence of small changes in the load demand. This experimental procedure intended to derive the fuel cell polarization curve and acquire

Modeling and Experimental Validation of a PEM Fuel Cell System

data from the overall operational range. The overall experiment is conducted using constant pressure and temperature and fixed intervals for the load demand. More specifically   the load varies from 0A to 20A with a step change of 2A every two minutes. Two sets of experiments were used for the generation of data for the parameter estimation procedure. The varying condition between the set of experiments was the pressure, 0Barg and 1Barg. Regarding the operational settings of the system: a) the FC temperature was stable at 65°C and the humidification temperature was 75°C, b) the air flow was 2000cc/m and the hydrogen flow 500cc/m. At each condition four identical experiments were conducted.

3.3. Parameter Estimation

The developed model has been implemented in the gPROMS modelling environment. Simulation runs indicated the sensitivity of the system concerning the most critical parameters to be selected for the estimation. Once the model is constructed, estimation is performed to define a set of selected parameters in activation and ohmic losses. A nonlinear regression technique with a constant variance model defining a maximum likelihood estimation problem was employed to determine the optimum values for the selected parameters, including the parametric coefficients in activation losses (ξ_1) and in ohmic losses (ξ_7). The characteristic cell voltage and the applied current density were measured through an on-line supervisory control and data acquisition system. The bounds for the parameters and the estimated values are presented in Table 1.

Table 1. Parameters for estimation

Parameter	Est. Value	Up Bound	Low Bound	Std. Dev	95% Conf Int
ξ_1	1.3205	1.4	0.954	$1.04 \cdot 10^{-3}$	$2.04 \cdot 10^{-3}$
ξ_7	$7.85 \cdot 10^{-4}$	$4.3 \cdot 10^{-4}$	$1.1 \cdot 10^{-6}$	$4.12 \cdot 10^{-6}$	$8.1 \cdot 10^{-6}$

Model predictions are in a very good agreement with the experimental data as indicated in. Fig 2. As it can be observed both experimental and simulation results show that a pressure increase raises cell voltage and consequently the power output.
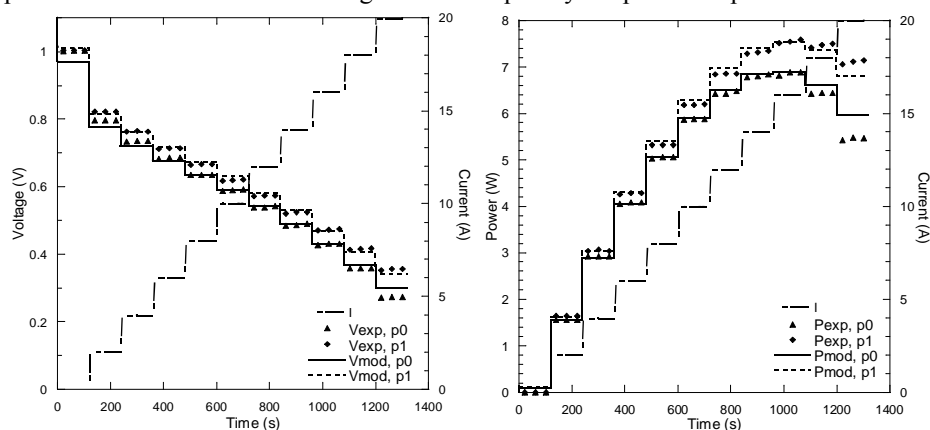
Figure 2 Voltage and power output to current changes (0Barg, 1Barg)

The validated model demonstrated an excellent behaviour both at steady and transient conditions and therefore it can be used both in system startup and during variable load changes.

3.4. Model Validation

A new set of experiments were conducted to assess the accuracy of the validated model. The varying condition was pressure of 05.Barg and 1.5Barg. Fig 3 that the model predictions are in good agreement with the experimental results. It is clear that under different pressure conditions the model response is similar to the experimental behaviour for the whole range of current variation.
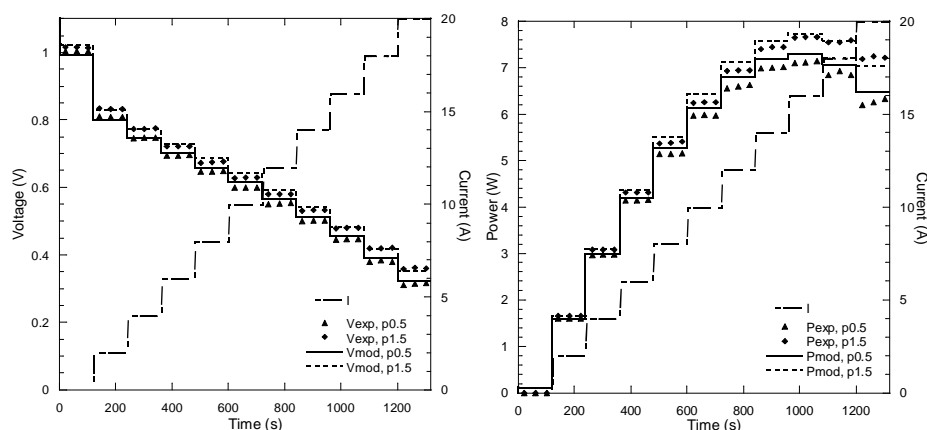
Figure 3 Voltage and power output to load changes using the validated model (0.5Barg, 1.5Barg)

4. Results and Future Work

A dynamic model of a PEM fuel Cell system has been presented followed by a validation procedure relying on the optimal determination of selected model parameters.. The predictions of the model are in good agreement with experimental data under various operating conditions. Therefore the validated model can provide the basis for deriving flexible design options and real-time control policies which are the subject of future work.

References

del Real, A.J., Arce, A., Bordons, C. (2007) Development and experimental validation of a PEM fuel cell dynamic model. Journal of power sources, 173 (1). 310-324.

Golbert, J. & Lewin, D.R. (2007) Model-based control of fuel cells: (2) optimal efficiency. Journal of power sources, 173 (1). 298-309.

Larminie J, Dicks A. (2003), Fuel Cell Systems Explained, 2nd Edition, John Wiley & Sons Ltd

Pathapati, P.R., Xue, X. & Tang, J. (2005) A new dynamic model for predicting transient phenomena in a PEM fuel cell system. Renewable energy, 30, (1). 1-22.

Pukrushpan, J.T., Stephanopoulou, A.G. & Peng, H. (2005) Control of Fuel Cell Power Systems. Principles, Modelling, Analysis and Feedback Design. Grimble, M.J. et al (ed.), London Springer.

Inventory Management MILP Modeling for Tank Farm Systems

Susana Relvas,[a] Ana Paula F.D. Barbosa-Póvoa,[a] Henrique A. Matos[b]

[a]*CEG-IST, UTL, Av. Rovisco Pais, 1049-001, Lisbon, Portugal, susanaicr@ist.utl.pt, apovoa@ist.utl.pt*
[b]*CPQ, IST, UTL, Av. Rovisco Pais, 1049-001, Lisbon, Portugal,henrimatos@ist.utl.pt*

Abstract

Process industries face complex problems when dealing with the optimization of production, control and operational tasks. While production and control are widely covered, the side operations related with logistics, supply and demand are usually underestimated, since they are not core activities. However, their contribution to business efficiency and margins maximization is crucial, because they guarantee process inputs and outputs at the desired quantities, timings and quality. One of these activities relies on inventory management. This paper presents MILP model that can be integrated with other process models, which represents a flexible storage tank farm. Under this general tank view, different logistic structures can be modeled: process feeding tanks, intermediate tanks or a final product tanks. The proposed MILP model is tested for two example problems.

Keywords: Tank, Inventory Management, MILP, flexible model

1. Introduction

Supply chain inventory management is a critical issue to address in order to enhance service level, competitive advantage and cost reduction, being a common source of mismanagement (Lee and Billington, 1992 and You and Grossmann, 2008). Given the benefits related with inventory management, process industries should give more focus to logistic issues when designing and planning their supply chain.

Inventory management policies are strongly developed for discrete products (e.g. components, parts) either accounting for the ordering or for inventory management. In the first case there are policies from traditional approaches as the Economic Order Quantity to tailor-made methodologies. Inventory management is based on warehouse optimization approaches that take into account flows and cost minimization. On the other hand, for continuous quantities stored in tank farms the problem is rather complex to represent, model and optimize. This is due to the high dependency that exists between inventory, process, product specificities and requirements, which may lead to complex models to be coupled to the complete process optimization models.

The published literature on inventory management for process industries appears on two directions: i) storage policies applied to operations scheduling models, such as zero wait storage, unlimited intermediate storage, non-intermediated storage, finite intermediate storage (as revised in Mendez et al., 2006) or using simple capacity limitations (taking as example supply chain planning models) and ii) operation specific models, such as the ones proposed for the petroleum industry (Relvas et al. 2006). The present work develops a tank farm scheduling model, where inventory is managed under common constraints related with process industries.

2. Problem Statement

The main system dealt in this work is a common inventory tank farm that can be located either in a chemical plant or other supply chain node, such as distribution centers or ports. For the purpose of the paper, we will consider that the tank farm is connected to a supply system that provides inputs and to a demand system where outputs are supplied, as presented in Figure 1. A tank farm requires some operations' management related to tank service, storage capacity, settling periods, quality tasks and fulfilling demands. The inputs required are product quantities and timings of arrival while outputs are the product demand and timing.
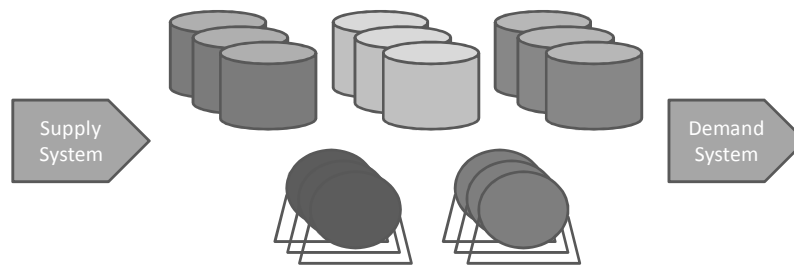
Figure 1 – Multiproduct tank farm system

The tank farm is connected through internal lines to the supply and demand systems. It will be considered that only one product is received at a given instant. At any point this assumption can be levered. On the other hand, the demand system has an independent delivery schedule per product. It is considered that each tank as a fixed service, i.e., the allocation of products to tanks is a problem data, which is a common procedure for chemical products. The tanks can store either liquid products or gases. For this end, the operating volume of each tank is assumed as its operating capacity.

For this system, it is required to determine the scheduling of storage activities at the tank farm that minimizes product mixing with origin in different receiving batches. Other objectives can be implemented for specific problem systems.

3. Inventory Tank Farm MILP Modeling

At the tank level, some usual procedures imply that each tank, at any instant, can only have a maximum of one active connection, i.e., inputs from the supply system and outputs for the demand system. For this reason, it is defined as rotation scheme the procedure that permits knowing, during the current time horizon and at any time point, the schedule of tanks' states, for each tank of each product. The states are defined from the tank operational cycle, which is represented in Figure 2. Each tank must accomplish this cycle in normal operation at the tank farm. It starts to be filled up until the total storage capacity is reached. At this point the settling period is carried out, to meet quality standards. Whenever the settling period is completed, the product stored is now available for clients' satisfaction, until the tank is completely empty.
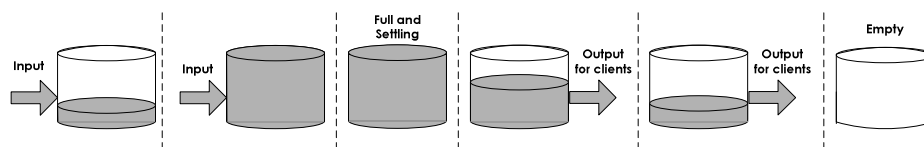
Figure 2 - Individual tank's operational cycle

The modeling challenges consist on the representation of the dynamic states (being filled up or being discharged), the boundary states detection (completely full or completely empty) and the model flexibility.
The proposed model formulation is described as follows.

Sets: $i \in I$ set of arriving batches of products, $t \in T$ set of tanks, $p \in P$ set of products and $s \in S$ set of tank states (ss = supplying state, rs = receiving state, fs = full state and es = empty state).
Note: time points are referred to the arrival of batch i to the tank farm, thus set I can either represent batches or time points. Both descriptions will be used throughout the model details.

Parameters: $hmax$ maximum time horizon extent, δ small value, C_i time scale referred to the arrival of batch i, $D_{i,p}$ volume of arriving batch i of product p, $MD_{p,i}$ market demand of product p at time point i, $Tset^0_{p,t}$ initial settling time for product p in storage tank t, $TCap_{t,p}$ tank capacity of tank t of product p, $ID^0_{p,t}$ initial inventory at tank t of product p, $ts^0_{p,t,s}$ initial state s of tank t of product p, $Trep_{p,t}$ minimum settling period of product p at tank t.

Positive variables: $Int_{p,t,i}$ volume of product p assigned to tank t at time interval i, $Outt_{p,t,i}$ volume that is supplied to fulfill market demand of product p from tank t at time interval i, $ID_{p,t,i}$ inventory level of product p at tank t at time interval i, $Tset_{p,t,i}$ settling time of product p stored at tank t at time interval i.

Binary variables: $ts_{p,t,s,i} = 1$ if tank t of product p is at state s at time interval i, $ba_{p,t,i} = 1$ if batch i of product p is assigned to tank t, $mdf_{p,t,i} = 1$ if tank t supplies the market of product p at time i.

Model Constraints:

$$ts_{p,t,s,i} + ts_{p,t,s',i} = 1, \quad \forall p,t,i, s = ss, s' = rs \tag{1}$$

$$ts_{p,t,s,i} + ts_{p,t,s',i} \leq 1, \quad \forall p,t,i, s = es, s' = fs \tag{2}$$

$$ts_{p,t,s,i-1} + ts_{p,t,s',i} \leq 1 + ts_{p,t,s'',i-1}, \quad \forall p,t,i > 1, s = rs, s' = ss, s'' = fs \tag{3}$$

$$ts^0_{p,t,s} + ts_{p,t,s',i} \leq 1 + ts^0_{p,t,s''}, \quad \forall p,t,i = 1, s = rs, s' = ss, s'' = fs \tag{4}$$

$$ts_{p,t,s,i-1} + ts_{p,t,s',i} \leq 1 + ts_{p,t,s'',i-1}, \quad \forall p,t,i > 1, s = ss, s' = rs, s'' = es \tag{5}$$

$$ts^0_{p,t,s} + ts_{p,t,s',i} \leq 1 + ts^0_{p,t,s''}, \quad \forall p,t,i = 1, s = ss, s' = rs, s'' = es \tag{6}$$

Equations (1) and (2) establish the occurrence of simultaneous tank states whereas equations (3) to (6) ensure how tanks can transit between states.

$$D_{p,i} = \sum_t Int_{p,t,i}, \quad \forall p,t \tag{7}$$

$$Int_{p,t,i} \leq ba_{p,t,i} \times Tcap_{p,t}, \quad \forall p,t,i \tag{8}$$

$$\sum_t Outt_{p,t,i} = MD_{p,i}, \quad \forall p,i \tag{9}$$

$$Outt_{p,t,i} \leq mdf_{p,t,i} \times Tcap_{p,t}, \quad \forall p,t,i \tag{10}$$

Equations (7) and (8) determine batch assignments to tanks, regarding the product. Equation (9) ensures outputs to fulfill market demands whereas equation (10) captures whether tank *t* supplies the market with product *p* at time *i*.

$$Int_{p,t,i} \leq ts_{p,t,s,i} \times Tcap_{p,t}, \quad \forall p,t,i, s = rs \tag{11}$$

$$Outt_{p,t,i} \leq ts_{p,t,s,i} \times Tcap_{p,t}, \quad \forall p,t,i, s = ss \tag{12}$$

Equations (11) and (12) ensure the relation between tank state and input reception and output supply.

$$ID_{p,t,i} = ID_{p,t,i-1} + Int_{p,t,i} - Outt_{p,t,i}, \quad \forall p,t,i > 1 \tag{13}$$

$$ID_{p,t,i} = ID_{p,t}^0 + Int_{p,t,i} - Outt_{p,t,i}, \quad \forall p,t,i = 1 \tag{14}$$

$$ID_{p,t,i} \leq Tcap_{p,t}, \quad \forall p,t,i \tag{15}$$

Equations (13) and (14) calculate the current inventory level at each tank whereas equation (15) limits inventory to maximum tank capacity.

$$ID_{p,t,i} \geq ts_{p,t,s,i} \times (Tcap_{p,t} - \delta), \quad \forall p,t,i, s = fs \tag{16}$$

$$ID_{p,t,i} \leq Tcap_{p,t} - \delta \times (1 - ts_{p,t,s,i}), \quad \forall p,t,i, s = fs \tag{17}$$

$$ID_{p,t,i} \leq \delta + Tcap_{p,t} \times (1 - ts_{p,t,s,i}), \quad \forall p,t,i, s = es \tag{18}$$

$$ID_{p,t,i} \geq \delta - Tcap_{p,t} \times ts_{p,t,s,i}, \quad \forall p,t,i, s = es \tag{19}$$

Equations (16) to (19) identify inventory level values that allow changing to full and empty tank states. For these equations a small difference from the inventory boundaries is used.

$$Tset_{p,t,i} \leq ts_{p,t,s,i} \times hmax, \quad \forall p,t,i, s = fs \tag{20}$$

$$Tset_{p,t,i} \geq Tset_{p,t,i-1} + C_i - C_{i-1} - (1 - ts_{p,t,s,i}) \times hmax, \quad \forall p,t,i > 1, s = fs \tag{21}$$

$$Tset_{p,t,i} \geq Tset_{p,t}^0 + C_i - (1 - ts_{p,t,s,i}) \times hmax, \quad \forall p,t,i = 1, s = fs \tag{22}$$

$$Tset_{p,t,i} \leq Tset_{p,t,i-1} + C_i - C_{i-1}, \quad \forall p,t,i > 1 \tag{23}$$

$$Tset_{p,t,i} \leq Tset_{p,t}^0 + C_i, \quad \forall p,t,i = 1 \tag{24}$$

$$Tset_{p,t,i} \geq Trep_{p,t} \times (ts_{p,t,s,i-1} + ts_{p,t,s',i} - 1), \quad \forall p,t,i > 1, s = fs, s' = ss \tag{25}$$

$$Tset_{p,t}^0 \geq Trep_{p,t} \times (ts_{p,t,s}^0 + ts_{p,t,s',i} - 1), \quad \forall p,t,i = 1, s = fs, s' = ss \tag{26}$$

$$Tset_{p,t,i-1} \leq Trep_{p,t} + ts_{p,t,s,i} \times \text{hmax}, \quad \forall p,t,i>1, s=ss \tag{27}$$

$$Tset^{0}_{p,t,s} \leq Trep_{p,t} + ts_{p,t,s,i} \times \text{hmax}, \quad \forall p,t,i=1, s=ss \tag{28}$$

Equation (20) ensures that the settling time is accounted when the tank is full. Equations (21) to (24) update the settling period for each tank at each time interval. Equations (25) to (28) ensure that whenever the minimum settling period is achieved, the tank state can change from full to supply.

$$\min\left[\left(\sum_{p}\sum_{t}\sum_{i} ba_{p,t,i}\right)-|I|\right]+\left[\left(\sum_{p}\sum_{t}\sum_{i} mdf_{p,t,i}\right)-|P|\times|I|\right] \tag{29}$$

Equation (29) represents the objective function which intends to minimize the difference between total number of batch assignments to tanks and the cardinality of set *I*, so as to avoid that one batch is assigned to several tanks and, therefore, that each tanks receives several batches before settling. The second term minimizes the difference between number of supplies from different tanks and total number of market demands. This objective function is relevant when product quality traceability is important.

4. Implementation and Results

Two scenarios (S1 and S2) are proposed to evaluate model performance and model size. S1 represents a motivation problem with 3 products, 3 tanks per product and 10 arriving batches. S2 represents a larger problem and accounts for 5 products, 25 tanks distributed by all the products and 25 arriving batches. S1 addresses a time horizon of 10 days whereas S2 has 25 days. The model was implemented in GAMS 22.8, CPLEX 11.1 on an Intel Pentium Core 2 Duo P9400, 4 GB Ram. The objective is to obtain either the optimal solution or a solution in 600 s (plus 120 s of solution polishing). Table 1 summarizes the model statistics and performance.

Table 1. Model Performance

	S1	S2
Time (s)	2.598	720.854
Objective function (OF)	3	12
OF Term 1	2	2
OF Term 2	1	10
Relative gap (%)	0.0	97.52
Iterations	14879	1991790
Nodes	495	41633
Continuous Variables	361	2501
Integer variables	540	3750
Constraints	2311	15876

From table 1 it can de seen how model size increases with the size of the scenario. Additionally, in S1 only two batch splits occurred whereas for a larger problem as in S2, two batch splits were necessary to accommodate all the incoming products. Regarding tank supply, in S1 there was only one market demand fulfillment that required two tanks whereas in S2 this situation occurred ten times. However, as it can be seen, there is space for improvement in the model performance. Despite locating solutions in an early stage of the tree search, the lower bound improvement is slow. The relaxed solution was 0.0 and after a high number of iterations and node exploration, the lower bound improved to 0.2976.

Figure 3 represents the inventory profile of tank 1 of product P1 for example S2. Table 2 summarizes the monitoring of this tank cycle, which has a capacity of 20000 m^3 and a minimum settling period of 24 h. The dynamic and boundary states are matched against the inventory profile evolution, so as to verify the effectiveness of the model. The boundary states only occur when the inventory is as its limits and the settling period is referred to the full state.
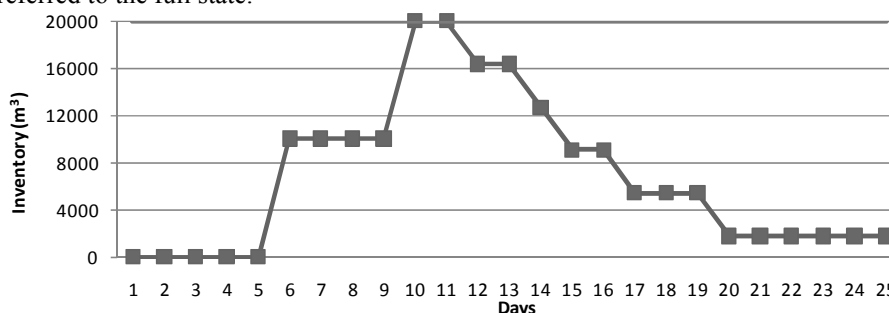
Figure 3 – Inventory profile for tank 1 of product P1, S2

Table 2. Tank cycle monitoring: S2, P1, tank 1

Day	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21	22	23	24	25
Boundary	ES	ES	ES	ES	ES					FS	FS														
Dynamic	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	RS	SS	SS	SS	SS	SS	SS	SS	SS	SS	SS	SS	SS	SS	SS
Settling										0	24														
Input						10000				10000															
Inventory	0	0	0	0	0	10000	10000	10000	10000	20000	20000	16350	16350	12700	9050	9050	5400	5400	5400	1750	1750	1750	1750	1750	1750
Output												3650		3650	3650		3650			3650					

5. Conclusions and Future Work

In this paper it was presented a flexible MILP model that can be coupled to scheduling and/or planning models for process industries, where liquid or gas products must be stored and managed under the requirements of a precedent supply system and a demand system ahead. The model accounts for tank farm constraints such as settling periods but can be further extended to account for other tank farm requirements that may describe other type of systems.

As future work, the authors propose to integrate the model with real world scheduling models for further testing. The development of the model should account for the option of having different time scales for the supply and demand systems. The formulation can also be revised so as to become tighter. Research on alternative system types requiring other constraints is also an objective of the authors.

6. Acknowledgements

References

H.L. Lee; C. Billington, 1992, *MIT Sloan Management Review*, v.33, 3 (Spring), 65-73;

C.A. Mendez; J. Cerda; I.E. Grossmann; I. Harjunkoski; M. Fahl, 2006, *Computers and Chemical Engineering*, v.30, issue 6-7, 913-946;

S. Relvas; H.A. Matos; A.P.F.D. Barbosa-Póvoa; J. Fialho; A.S. Pinheiro, 2006, *Ind. Eng. Chem. Res.*, v.45 (23), 7841-7855.

F. You; I.E. Grossmann, 2008, *Ind. Eng. Chem. Res.*, v.47, 7802–7817.

Simulation of ethanol production from sugarcane in Brazil: economic study of an autonomous distillery

Marina O.S. Dias[a,b], Marcelo P. Cunha[a], Charles D.F. Jesus[a], Mirna I.G. Scandiffio[a], Carlos E.V. Rossell[a,b], Rubens Maciel Filho[b], Antonio Bonomi[a]

[a] *CTBE – Bioethanol Science and Technology National Laboratory, PO Box 6170 – CEP 13083-970, Campinas – SP, Brazil, marina.dias@bioetanol.org.br*
[b] *School of Chemical Engineering, University of Campinas, PO Box 6066 – CEP 13083-970, Campinas – SP, Brazil*

Abstract

Simulation of the production of ethanol from sugarcane in an autonomous distillery was carried out using software SuperPro Designer and electronic spreadsheet. Analysis of the ethanol production costs was performed for different production scenarios, considering improvements on the energy production from sugarcane bagasse and the selling of surplus electricity. It was verified that selling of surplus electricity positively influences the ethanol production costs.
Keywords: ethanol, simulation, sugarcane, economic evaluation.

1. Introduction

Brazil produces bioethanol from sugarcane on a large scale basis since the 1970s (Bake et al., 2009); increase on the demand for the biofuel as a substitute or complement of gasoline has motivated the search for more efficient means of production. As a consequence, the evaluation of conventional ethanol production and the identification of critical process parameters are required. Brazil and the United States, which produces ethanol from corn, are the largest ethanol producers in the world (Balat et al., 2008); however, net energy of ethanol production from sugarcane is more positive than that from corn (Leite et al., 2009): sugarcane bagasse, one of the main by-products of sugarcane processing, is used as fuel in cogeneration systems, which provide steam and electric energy to supply the bioethanol production process. Thus, an autonomous distillery may produce electric energy to sell to the grid, if there is a surplus produced during cogeneration. For the past few years, growing interest on production of electricity in ethanol production plants has been observed, which may improve revenues and competitiveness of sugarcane ethanol.

In this work, simulations of ethanol production from sugarcane were carried out using SuperPro Designer 7.5 (Intelligen, Inc) in order to evaluate production costs within the industrial site. A "standard" autonomous distillery is considered, in which 500 tons of sugarcane (TC) per hour are processed, producing 1000 m³/day of anhydrous bioethanol, during 180 days (sugarcane harvest season). Data used to simulate the unit operations were obtained from industrial sites and from literature.

2. Ethanol production process from sugarcane

In an autonomous distillery, all the sugarcane processed is used to provide sugars for fermentation, from which ethanol is produced.

The main steps required for the ethanol production process from sugarcane in an autonomous distillery are illustrated in Fig. 1.
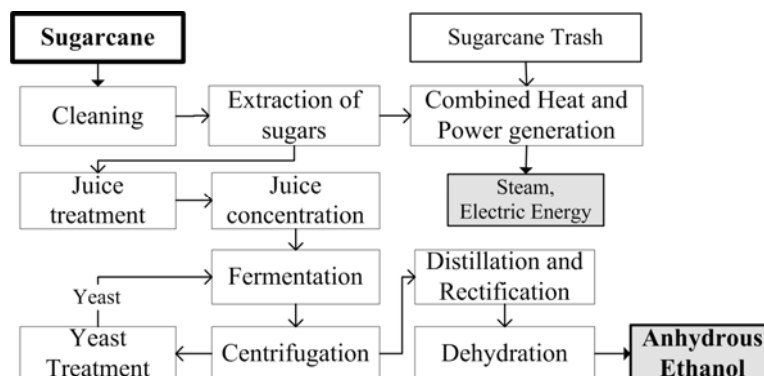
Figure 1. Simplified block flow diagram of the anhydrous bioethanol production process from sugarcane in an autonomous distillery.

3. Process simulation procedure

Operating and process parameters of the autonomous distillery were obtained in the literature and from operating industries. A "standard" plant is considered, with unit operations typical of those found in the Brazilian bioethanol industry. 500 TC per hour are processed for production of around 1000 m³/day of anhydrous bioethanol (99.3 wt % ethanol).

3.1. Simulation

Firstly, a mass balance of the process was carried out using a spreadsheet, in which usual industrial parameters (efficiency of unit operations, amount of raw materials and yields) were employed. Then, simulation was carried out using software SuperPro Designer 7.5 from Intelligen, Inc, considering the unit operations indicated in Fig. 1 except for the operations related to the cogeneration system. In these operations, where steam and electric energy are produced, calculation was performed using a conventional spreadsheet, since the simulation software does not have the procedures required to perform its simulation at the moment. A simplified flowsheet of the simulation of bioethanol production process is shown in Fig. 2.

Several hypothetic components (sugarcane bagasse constituents, sugarcane impurities, etc.) were inserted into the simulator database, in order to represent the bioethanol production process more accurately. Energy demand of unit operations that are not available in the simulator database, such as extraction of sugars in the mills, azeotropic distillation and adsorption onto molecular sieves (processes used on ethanol dehydration), was obtained in the literature (Ensinas et al., 2007; Andrietta, 2009).

The main process parameters are listed in Table 1.

3.2. Dehydration processes

Two dehydration processes were considered: azeotropic distillation using cyclohexane as entrainer, which is the most common and most energy-intensive process used in ethanol dehydration, and adsorption onto molecular sieves, which presents the lowest energy consumption among the commercial dehydration processes available, but requires a larger investment.

Simulation of ethanol production from sugarcane in Brazil: economic study of an autonomous distillery
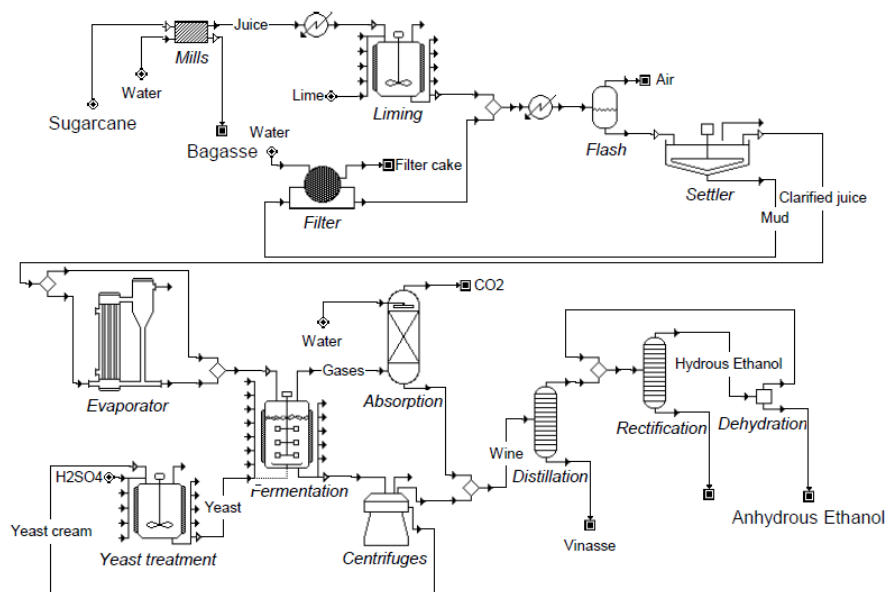
Figure 2. Simplified flowsheet of the simulation of bioethanol production process carried out using SuperPro Designer.

Table 1. Main parameters adopted in the simulation of the "standard" plant.

Parameter	Value	Unit
Sugarcane crushing rate	500	TC/h
Days of operation	180	days/year
Fibre on sugarcane	12	wt%
Sugars on sugarcane	14	wt%
Amount of sugarcane trash produced in the fields	140	kg/TC
Dirt removal on sugarcane cleaning	90	%
Sugar losses on sugarcane cleaning	0.8	kg/TC
Sugars recovery on the mills	96	%
Sugarcane bagasse water content	50	wt%
Recovery of sugars on juice treatment	99.5	%
Fermentation yield	90	%
Ethanol recovery on distillation and dehydration	99.7	%

3.3. Cogeneration system

Conventional plants are equipped with boilers for the production of 22 bar steam, in which sugarcane bagasse, produced in the mills, is used as fuel. The steam produced in the boilers is used to produce electricity in steam turbines and as thermal energy for the process, besides being used in mechanical drivers in the sugarcane preparation and juice extraction systems.

A 90 bar cogeneration system with back pressure and condensing steam turbines for production of steam and electric energy is considered as well, in which surplus electricity is sold to the grid. The amount of sugarcane bagasse available for cogeneration and the process steam demand are obtained in the simulation and used to determine the parameters of the coproduction of heat and power (CHP) plant; as a

result, the amount of surplus electric energy available for sale is determined. The main parameters of the CHP system are shown in Table 2.

Table 2. Main parameters considered in the simulation of the CHP system (Dias et al., 2009; Ensinas, 2008; Seabra, 2008).

Parameter	Value	Unit
22 bar – boiler thermal efficiency[*]	75	%
90 bar – boiler thermal efficiency[*]	86	%
High pressure steam turbines isentropic efficiency	72	%
Intermediate pressure steam turbines isentropic efficiency	81	%
Condensation turbine isentropic efficiency	70	%
Mechanical drivers – turbine isentropic efficiency	55	%
Generator efficiency	96	%
Sugarcane bagasse LHV (50 wt% water)	7565	kJ/kg
Sugarcane trash LHV (15 wt% water)	12960	kJ/kg
Electric power demand of the distillery	12	kWh/TC
Mechanical power demand – cane preparation and juice extraction	16	kWh/TC
Electric power demand of the distillery – electric drivers	18	kWh/TC
Outlet pressure of high pressure steam turbine	22	bar
1st extraction pressure	6	bar
2nd extraction pressure	2.5	bar
Process steam pressure	2.5	bar

[*] Low Heating Value (LHV) base

Besides the use of more efficient boilers (90 bar), two other process improvements were considered: the use of electric drives for mills and other equipments, replacing the mechanical drivers, and the use of sugarcane trash as fuel in boilers. Sugarcane trash is composed by leaves and tops, and nowadays is burned before harvest or left in the field, but a fraction of the trash generated may be recovered and used as a fuel in the plant. In all the cases where excess steam is produced, it is condensed on condensing steam turbines, increasing the amount of electricity produced. In this work, 50 % of trash is used as fuel for the production of steam and electricity; the remaining fraction is left in the fields in order to provide control of weeds and diseases (Hassuani et al., 2005).

4. Simulation results and discussion

Different configurations were analysed, combining the options shown in Table 3.

Table 3. Parameters considered in the studied scenarios.

Parameter	I	II	III	IV
1st generation anhydrous ethanol production	X	X	X	X
22 bar boilers	X			
Dehydration by azeotropic distillation	X			
Dehydration by adsorption onto molecular sieves		X	X	X
Sell of surplus bagasse	X			
90 bar boilers		X	X	X
Burning of surplus bagasse		X	X	X
Sell of surplus electricity		X	X	X
Electrification of drives			X	X
50 % of trash used				X

Simulation of ethanol production from sugarcane in Brazil: economic study of an autonomous distillery

Case I presents the typical traditional autonomous distillery; cases II through IV present increasing levels of technologic improvements considered in this work, which influence the production of electricity. Simulation results for each of the studied scenarios are displayed in Table 4.

Table 4. Simulation results for each of the studied scenarios.

Parameter	I	II	III	IV
Anhydrous ethanol production – L/t sugarcane	83.3	83.3	83.3	83.3
Surplus bagasse – kg/t sugarcane	16.6	0	0	0
Surplus electricity sold – kWh/t sugarcane	0	68.2	73.7	154.9

5. Economic evaluation and discussion

On Table 5 the basic parameters for the economic analysis are displayed.

Table 5. Basic parameters used in the economic analysis[1].

Parameter	**Value**
Project lifetime	25 years
Salvage value of equipment	-
Construction and start-up	2 years
Sugarcane price[2]	US$ 16.58/t
Trash price (15% moisture)[3]	US$ 12.21/t
Depreciation (linear)	10 years
Tax rate (income and social contributions)	34.0%
Ethanol (producer price)[4]	US$ 0.40/L
Bioelectricity (producer price)[3]	US$ 67.05/MWh
Bagasse (producer price)[5]	US$ 16.58/t

[1] Considered the exchange rate US$ 1.00 = R$ 2.088 (average of the past 12 months)
[2] Average of the last 12 months (UDOP, 2009)
[3] Seabra, 2008
[4] Average of the last 12 months (CEPEA, 2009)
[5] Considered equal to the sugarcane price

In order to evaluate ethanol production costs, equipment costs were evaluated for each of the studied scenarios based on data provided by the industry. Firstly, economic analysis of each of the studied scenarios was carried out using the parameters presented in Table 5. Production costs were then calculated reducing ethanol, surplus bagasse and electricity prices (amount paid to the producer) simultaneously, at the same proportion, until real profits reached zero (i.e, internal rate of return per year equal to zero). Table 6 presents the investment required on equipments for each scenario and the calculated ethanol production costs.

Table 6. Investment and ethanol production costs on each scenario.

Parameter	**I**	**II**	**III**	**IV**
Equipment costs – 10^6 US$	144	185	185	205
Ethanol production costs – US$/L	0.313	0.289	0.286	0.265

Since sugarcane costs are equivalent to roughly 60 % of the final production costs of ethanol, a sensitivity analysis was carried out in order to evaluate sugarcane costs on the final ethanol cost; results are presented in Fig. 3 for the four studied scenarios.
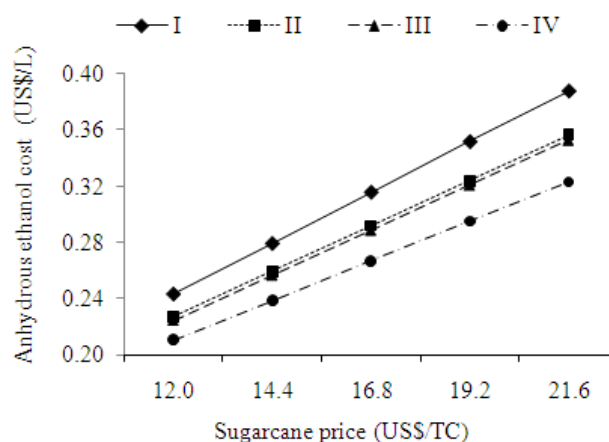
Figure 3. Evaluation of anhydrous ethanol production costs for different sugarcane prices.

6. Conclusions

In this work simulation of a typical autonomous distillery for anhydrous bioethanol production from sugarcane was carried out. Selling of surplus electricity (cases II through IV) can improve the profitability of ethanol production from sugarcane, since it leads to a reduction on ethanol production costs. The use of sugarcane trash as a fuel on cogeneration systems can significantly reduce ethanol production costs, since large amounts of electricity are available for sale.

7. Acknowledgements

References

S.R. Andrietta, Optimal Industrial Fermentation, In: BIOEN Workshop on Process for ethanol production, FAPESP, 2009. Available online at http://www.fapesp.br/eventos/2009/09/10_bioen/Silvio_Roberto.pdf, retrieved on sept 15, 2009 (in Portuguese)

M. Balat, H. Balat, C. Öz, Prog. in En. and Comb.Sci., 34 (2008) 551-573

J. Bake, M. Junginger, A. Faaij, T. Poot, A. Walter, Biom. Bioen., 33 (2009) 644-658

CEPEA - Center for Advanced Studies on Applied Economics, 2009, Available online at www.cepea.usp.br, retrieved on nov 15, 2009

M.O.S. Dias, A.V. Ensinas, S.A. Nebra, R. Maciel Filho, C.E.V. Rossell, M.R.W. Maciel, Chem. Eng. Res. Des., 87 (2009) 1206-1216

A.V. Ensinas, S.A. Nebra, M.A. Lozano, L.M. Serra, En. Conv. Manag., 48 (2007) 2978–2987

A.V. Ensinas, PhD Thesis (Mechanical Engineering), School of Mechanical Engineering, State University of Campinas (in Portuguese), 2008

R. Leite, M. Leal, L. Cortez, W. Griffin, M. Scandiffio, Energy, 34 (2009) 655-661

S.J. Hassuani, M.R.L.V. Leal, I.C. Macedo (eds.), Biomass Power Generation—Sugarcane Bagasse and Trash. Piracicaba: PNUD-CTC, 2005

J.E.A. Seabra, PhD Thesis (Energetic Systems Planning), School of Mechanical Engineering, State University of Campinas (in Portuguese), 2008

UDOP – Union of Biofuel Producers, Sugarcane prices, 2009. Available online at www.udop.com.br/index.php?item=cana. Retrieved on nov 15, 2009

A hybrid neural approach to model batch fermentation of dairy industry wastes

Alessandra Saraceno, Sascha Sansonetti, Stefano Curcio, Vincenza Calabrò, Gabriele Iorio

Department of Enginering Modeling-University of Calabria, via P. Bucci-cubo 42/A, Arcavacata di Rende 87036, Italy, alessandra.saraceno@unical.it

Abstract

In this work, the fermentation of "Ricotta cheese whey" for the production of ethanol was simulated by means of a Hybrid Neural Model (HNM), obtained by coupling neural network approach to mass balance equations describing the time evolution of lactose (substrate), ethanol (product) and biomass concentrations. The realized HNM was compared with a pure neural network model (NM) and the advantages gained from the hybrid approach were emphasized. The experimental data, necessary to develop the model, were collected during batch fermentation runs. For all the proposed networks, the inputs were chosen as the operating variables exhibiting the highest influence on the reaction rate. The simulation results showed that the HNM was capable of an accurate representation of system behavior by predicting biomass, lactose and ethanol concentration profiles with an average error percentage lower than 10%. Moreover, especially if compared with the NM, the HNM showed good forecasting capability even with fermentation run never seen during the training phase.

Keywords: grey-box models, artificial neural networks, batch fermentation, modeling

1. Introduction

Modeling of biotechnological processes represents a key issue to achieve proper design and control aimed at process optimization (Lubbert and Jorgensen, 2001). Biochemical reactions actually involve many parallel-serial reaction steps and depend on several transport phenomena that may limit the observed reaction rates. Usually, the fundamental approach, largely considered as the most rigorous, cannot be applied in reaction kinetic modeling due to inherent non-linearity, lack of information, experimental inaccuracy, deviations from ideal conditions (Feyo de Azevedo et al., 1997). On the contrary, black-box models, suggested by the empirical approach, suffer from a restricted validity domain depending on the range of data collected during the experiments. A reasonable trade-off between theoretical and empirical approach is represented by hybrid modeling, leading to a "grey-box" model capable of good performance in terms of data interpolation and extrapolation. The main advantage of hybrid modeling regards the possibility of describing some well-assessed phenomena by means of a fundamental theoretical approach, leaving the analysis of other aspects, very difficult to interpret and describe in a traditional way, to rather simple "cause-effect" models. Among these, Artificial Neural Networks (ANNs) were successfully used in bioreactor modeling (James at al., 2002; Laursen et al., 2007; Simutis et al., 1995) ANNs consist of interconnected computational elements called neurons or nodes: each

neuron receives input signals from the related units, elaborates these stimuli by an activation or transfer function and generates an output signal that can be transferred to other neurons. ANN can be used both in a pure empirical model structure and in a hybrid model formulation. When ANNs are utilized in a hybrid model, a very important step is identifying the respective domain of theoretical and empirical parts in such a way as to decide which aspects of the process are to be described by a black-box model.

Ricotta cheese whey (RCW),"scotta", is a dairy industry waste characterized by a lactose content of about 45-50 g/l and a very low amount of proteins ranging from 0.22 – 0.24 % in weight. BOD reduction of "scotta" can be achieved, for instance, by alcoholic fermentation of lactose carried out by *Kluyveromyces Marxianus*. Even if several authors dealt with biotechnological utilization of cheese whey (Gonzàles Siso, 1996; Ozmihci and Kargi, 2006, 2007), as far as our knowledge is concerned, a single work exists about "Ricotta Cheese whey"(Sansonetti et al., 2009).

The aim of this paper was to propose a model that could describe batch fermentation of lactose in ethanol investigating two different strategies. At first, the concentration profiles of the reacting species were described by a pure neural model; afterwards, with the aim of introducing theoretical information (mass balance) into the model, a multiple hybrid neural model was realized.

2. Development of the models

The proposed models, i.e. NM and HNM, are aimed at the description of the batch fermentation of RCW predicting biomass, lactose (substrate) and ethanol (product) concentration profiles. A preliminary phase of the model development was performed. It consisted on the identification of the operating variables that, among all the parameters that could affect reaction progress, exhibited the highest influence on process performance. These variables, i.e.: temperature (T), pH (pH), reactor stirring rate (rpm), initial lactose concentration (C^0_{lat}) and reaction time (t), represented the inputs for both the pure neural and the hybrid neural model.

2.1 Neural model structure

A multiple neural model was proposed to predict concentration profiles. It consisted of three neural networks, i.e. NM1, NM2 and NM3, respectively capable of predicting biomass, lactose and ethanol concentration profiles as outputs. The structure of the three neural networks was very similar: they were characterized by the same number of input nodes, corresponding to the pre-identified input variables and by a single output node that had to produce, as output, the corresponding concentration profile. Between input and output nodes two hidden layers were put and the number of nodes in each layer was determined according to an iterative procedure.

2.2 Hybrid model structure

The realized hybrid neural model consisted of three models characterized by a common structure. A mass balance equation was written for each of the compounds and an ANN was set up to evaluate the kinetic parameters actually necessary to solve the balance equations. Some logic conditions, expressing the fulfilment of physical constraints, were also introduced into the model.

The three mass balance equations, written with reference to biomass, lactose and ethanol assumed the following form:

$$
\begin{cases}
\dfrac{dX}{dt} = \mu X \\[2mm]
-\dfrac{dS}{dt} = qX \\[2mm]
\dfrac{dP}{dt} = \mu_P X
\end{cases}
\tag{1}
$$

where *X, S, P* were, respectively, biomass, lactose and ethanol concentration profiles, μ was biomass specific growth rate, *q* was the lactose consumption rate function, μ_p was the product growth rate and *t* was the reaction time. Eqs. (1) were approximated by Euler's discretization and the discretized form was used recursively to determine the biomass, lactose and ethanol concentration values at the time instant *t+Δt,* knowing all the information concerning time instant *t*. The values of kinetic parameters at time instant *t,* strictly necessary to solve the discretized form of the equations, were provided by three neural networks named HNM1, HNM2 and HNM3, respectively predicting μ, *q* and μ_p. The three neural networks were set up according to an iterative procedure. The kinetic parameters, evaluated using HNM1, HNM2 and HNM3, were subsequently processed by some logic conditions that verified the agreement between the net outputs and some well-assessed process information. The logic condition assumed the following form:

$$
\begin{cases}
if \ \ \mu \ < 0 \Rightarrow \mu \ = 0 \\
if \ \ q \ < 0 \Rightarrow q \ = 0 \\
if \ \ \mu_p < 0 \Rightarrow \mu_p = 0
\end{cases}
\tag{2}
$$

Eqs. (2) faced the fact that, during the fermentation runs, biomass and ethanol concentrations always grew, whereas lactose concentration always decreased.

3. Materials and methods

3.1. Experimental design

The experimental data necessary to develop the models were collected from a set of anaerobic fermentations performed on RCW. According to the factorial design method (Box et al., 1978), the following variables and operating condition were chosen: 1) Temperature in the range 32- 40 °C; 2) pH in the range 4-6; 3) Stirring rate in the range 100-300 rpm; 4) The lactose concentration was varied between 50-90 g/l since this is the expected range in which the process could be operated. The lowest value of the range was set to 50 g/l due to the typical lactose concentration of RCW, i.e. 45-50 g/l. The highest value of the range, i.e. 90 g/l, derived from the consideration that it could be useful to concentrate RCW before the fermentation process so to achieve a more concentrated ethanol solution flowing out the reactor. Nevertheless this value cannot be increased as much as one likes since a limit imposed by the osmotic pressure at the cell wall does actually exist. The value of 90 g/l value was chosen according to several preliminary experiments (not reported in the paper), which showed the technical and economical feasibility of RCW pre-concentration (up to 90 g/l) by nanofiltration and, therefore, by a well-assessed procedure.

A total of 16 batch runs, having a duration of 18 hours with a sampling time of one hour, were therefore performed and a total of 912 experimental points, expressing the

time evolutions of biomass, lactose and ethanol concentration, were collected during fermentation.

3.2. Neural network development

To realize both the NMs and the HNMs, six different neural networks were developed following a common iterative procedure implemented according to Matlab Neural Network Toolbox Ver. 4.0.1. For the neural model only the concentration values relative to the batch fermentation runs were sufficient to train the networks. For the hybrid model development it was necessary to determine the experimental values of μ, q, μ_p by interpolating the collected experimental data by cubic splines. Subsequently, the experimental points relative to 15 of the 16 available runs were used to train/test the developed networks: the data points were randomly split into two groups, reserving 2/3 of data (570 points) to the training phase and remaining 1/3 (285 points) to test the neural networks. Finally, the six realized neural models were validated using two complete experiments never exploited during both the previous training/test phases. The neuron transfer function was always the hyperbolic tangent, except for the output layer where a linear transfer function was used. To develop the ANNs, an iterative procedure was implemented. The performance index, chosen to evaluate simulation results reliability, was the percentage error between each predicted ($C_{p,k}$) and measured ($C_{m,k}$) concentration value: in a whole batch run the average value of the percentage error had to be lower than 10%. The iterative procedure consisted in placing a layer with a single neuron between the input and output layers, thus verifying the fulfillment of convergence criterion; if at the end of the training procedure the performance objective was not achieved, a neuron was added into the hidden layer, repeating the procedure until the performance objective was achieved. If the neurons of the 1st hidden layer reached the number of 10, an additional hidden layer with a single neuron was put into the network and, the procedure described for the first hidden layer iterated.

4. Results and discussion

On the basis of the previous discussion, a multiple neural model and a multiple hybrid neural model were realized. Table 1 summarizes the architecture of all the developed networks.

Table 1. ANNs architecture

Neural Network	Number of neurons 1st hidden layer	Number of neurons 2nd hidden layer	Number of neurons output layer
NM1	8	5	1
NM2	10	10	1
NM3	10	4	1
HNM1	6	5	1
HNM2	10	3	1
HNM3	10	2	1

A comparison between models predictions and experimental data is reported in Fig.1. Fig. 1 reports the simulation results relative to a batch run used during training/test phase of the models. A remarkable agreement between model predictions and the corresponding experimental data was found, both with the pure neural and the hybrid-neural approach so that an average percentage error much lower than 10% was found.

A hybrid neural approach to model batch fermentation of dairy industry

Figure 1. Predictions during training/test. Run conditions: T=32°C, pH=6, rpm=100, C^0_{lat}=90 g/l
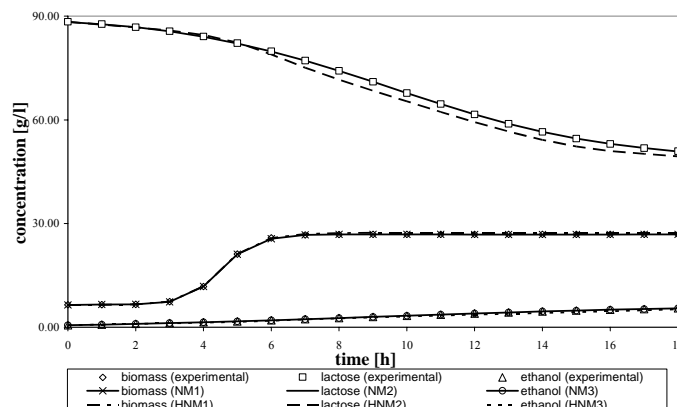
Moreover, it should be observed that even if models performance is very similar, NMs were more accurate in predicting the system dynamics with the experimental point reserved to training/test phase. This result emphasizes how neural networks, being data-based models, are able to perform very well on training data. Indeed, the training phase of a neural network consists in structuring it so as to predict the training points extremely well. Fig. 2. shows two different validation tests aimed at comparing hybrid neural model and pure neural model predictions under a combination of operating conditions never exploited during the training and the test phases, even if included in the same range chosen to perform the experimental design. From Fig. 2 it is possible to appreciate how NMs and HNMs validation results were quite different. The hybrid model was capable of giving good simulation results even during validation: the model performance, measured in terms of ε%, is comparable to that obtained with the experimental data used to perform the training and the test phases.

Figure 2. Models validation. Run conditions: a) T=32°C, pH=4, rpm=100, C^0_{lat}=50 g/l; b) T=37°C, pH=5, rpm=300, C^0_{lat}=50 g/l
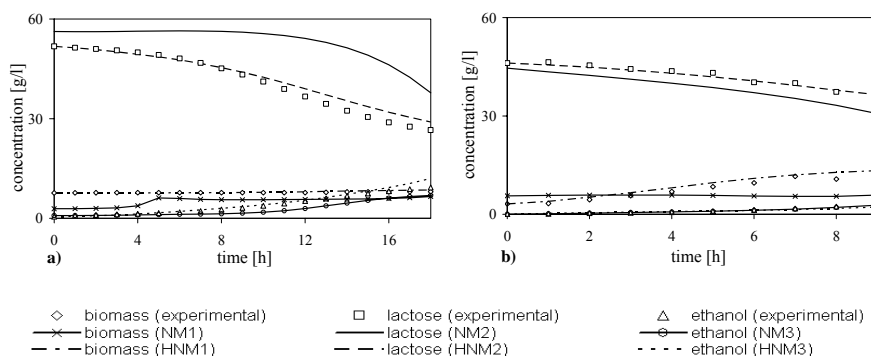
This result shows how hybrid models do not just learn in order to recognize the training points; instead, the models are able to predict the system behaviour even when they are operated in conditions unexploited during training phase. On the contrary, neural

models, during the validation, give worse performance that those obtained during the training phase. Indeed, even if the neural models are able to predict smoothly the system behaviour, the error percentage is significantly higher than 10% thus confirming how pure black-box models are very reliable into their definition domain whereas they become less accurate when used in interpolation and extrapolation. Therefore, the results obtained in this paper, indicate how the introduction of theoretical information into black-box models can lead to a more robust modeling approach.

5. Conclusions

The purpose of this work was to make a comparison between a pure neural and a hybrid neural approach in modeling the batch fermentation of RCW. The hybrid modeling approach showed better forecasting capability than the pure neural model, being capable to predict the system behavior even in operating conditions never exploited during the training phase. On the other hand, NMs did not show the same reliability during the model validation thus stating how the introduction of simple theoretical information into black-box models can lead to a strong improvement in model performance. As a matter of fact, HNMs turned out to be a very efficient tool for the simulation of the RCW fermentation process and, more generally, it could be used in modeling several complex biotechnological processes. Overcoming the difficulties to describe very complex reaction mechanism, HNMs are able to exploit all the available theoretical knowledge all the same. This modeling effort can lead to more robust models than empirical models and this feature make the hybrid neural approach a promising alternative in process engineering because most of the advanced control and optimization strategies are based on mathematical models.

References

G. Box, W. Hunter and S.Hunter , 1978, Statistics for Experimenters, An Introduction to Design, Data Analysis and Model Building, John Wiley and Sons, New York.

S. Feyo de Azevedo, B. Dahm and F.R. Oliveira, 1997, Hybrid modelling of Biochemical Processes: A comparison with the conventional approach., Comput. Chem. Eng., 21,751.

M. I. Gonzàles Siso, 1996, The biotechnological utilization of cheese whey: a review, Bioresour. Technol., 57, 1.

S. James, R. Legge and H.Budman, 2002, Comparative study of black-box and hybrid estimation methods in fed-batch fermentation, J. Process Control, 12, 113.

S. O.Laursen, ,D. Webb and W. F. Ramirez, 2007, Dynamic HNM of an industrial fed-batch fermentation process to produce foreign protein, Comput. Chem. Eng., 31, 163.

A. Lubbert and S. B. Jorgensen, 2001, Bioreactor performance: a more scientific approach for practice, J. Biotechnol., 85, 187.

S. Ozmihci and F. Kargi, 2006,Utilization of cheese whey powder for ethanol fermentation: effects of operating parameters, Enzyme and Microbial Technol., 38, 711.

S. Ozmihci and F. Kargi, 2007, Kinetics of batch ethanol fermentation of (CWP) solution as a function of substrate and yeast concentrations, Bioresour. Technol., 98, 2978.

S. Sansonetti, S. Curcio, V. Calabrò and G. Iorio, 2009, Bio-ethanol production by fermentation of Ricotta Cheese Whey as an effective alternative non-vegetable source, Biomass and Bioenerg., 12, 1692.

R. Simutis, M. Dors and A. Lubbert, 1995, Bioprocess optimization and control: Application of hybrid modelling, J. Biotechnol., 42, 285.

Numerical Solution of Batch Crystallization Models

Qamar S.[a,b], Seidel-Morgenstern A.[a]

[a] *Max-Planck Institute for Dynamics of Complex Technical System, Sandtor Str. 1, 39106 Magdeburg, Germany, seidel-morgenstern@mpi-magdeburg.mpg.de*
[b] *COMSATS Institute of Information Technology , Park Road Chak Shahzad , Islamabad, Pakistan, qamar@ mpi-magdeburg.mpg.de*

Abstract

An efficient and accurate numerical technique is introduced for the simulation of a batch crystallizer equipped with a fines dissolution unit incorporating a time-delay. The dissolution of small crystals (fines dissolution) improves the product quality and facilitates the downstream processes. The proposed method follows two steps. In the first step, a coupled system of ordinary differential equations (ODEs) for moments and solute mass is numerically solved in the time domain of interest, giving the discrete values of growth and nucleation rates. In the second step, theses discrete values are used along with the initial crystal size distribution (CSD) to construct the final CSD. The method of characteristics and Duhamel's principle are employed for deriving an expression for CSD from the given population balance model (PBM). An alternative quadrature method of moments (QMOM) is introduced for approximating integrals in the ODE system of moments and mass balance. In this technique, orthogonal polynomials, obtained from the lower order moments, are used to find the quadrature abscissas and weights. The numerical results of our scheme are validated against the results of high resolution finite volume scheme results. Our scheme was found to be efficient, accurate, and free from numerical dissipation and dispersion.

Keywords: Batch crystallization, nucleation and growth, fines dissolution, time-delay, method of characteristics, Duhamel's principle.

1. Introduction

Crystallization from solution is an important industrial operation due to marketing of large number of materials as crystalline particles. It is therefore one of the most widely separating and purifying technique in chemical, pharmaceutical, semiconductor, and food industries. Population balance equations (PBEs) are widely used for modelling crystallization processes. Several numerical methods have been developed for solving PBEs, see for example Ramkrishna (2000) and Qamar et al. (2009). In Qamar et al. (2009), a batch crystallization models without fines dissolution was considered.

In this article, the numerical method in Qamar et al. (2009) is extended for solving a batch crystallization model with fines dissolution. The main ingredients of the scheme are the initial data for CSD and solute mass, a coupled system of ODEs for moments and solute mass, as well as an expression for CSD obtained by utilizing the method of characteristics and Duhamel's principle. To overcome the closure problem, a Gaussian quadrature method based on orthogonal is employed for approximating integrals appearing in the moment-system of ODEs polynomials (e.g. in Qamar et al. (2009). In this manuscript, a third order orthogonal polynomial, needing first six moments, is used. A comprehensive discussion on other QMOM methods and their limitations can be found in Grosch et al. (2007). The current QMOM was found robust and the problems

mentioned by Grosch et al. (2007) were not observed. The proposed method is efficient, accurate, and easy to implement compared to other numerical methods available in the literature. Further details about this method can be found in Qamar et al. (2009). For validation, the numerical results of our scheme are compared with the finite volume scheme of Koren (1993).

2. Mathematical Model

In one-dimension, the crystals size is represented by a characteristic length x. The CSD, i.e. $n(t,x)$, represents the number of crystals per crystal length. A balance for the number of crystals in an infinitesimal interval of crystal length leads to the following PBE (e.g. Randolph and Larson (1988))

$$\frac{\partial n(t,x)}{\partial t} = -\frac{\partial[G(t,x)n(t,x)]}{\partial x} - \frac{\dot{V}}{V_c}h(x)n(t,x) + B_0(t)\delta(x-x_0), \quad n(0,x)=n_0(x), \tag{1}$$

where, $(t,x) \in \Re_+^2$ and $\Re_+ := (0,\infty)$. Here, $n_0(x)$ denotes the CSD of seed crystals, $G(t,x)$ is the size-dependent crystal growth rate, $B_0(t)$ is the nucleation rate at minimum crystal size x_0 and δ represents Dirac delta distribution. Moreover, V_c is the volume of crystallizer, $\dot{V}$ is volumetric flow rate from the crystallizer to the dissolution unit, and the death function $h(x)$ describes the dissolution yield of small particles below some critical size. The ith moment of CSD is defined as

$$\mu_i(t) = \int_0^\infty x^i n(t,x)dx, \quad i = 0,1,2,\cdots. \tag{2}$$

A balance for the liquid phase yields an ODE for the solute mass

$$\frac{dm(t)}{dt} = \dot{m}_{in}(t) - \dot{m}_{out}(t) - 3\rho_c k_v \int_0^\infty x^2 G(t,x)n(t,x)dx, \quad m(0) = m_0, \tag{3}$$

where, ρ_c is crystals density, and k_v is a volume shape factor defined such that the volume of a crystal is $k_v x^3$. In equation (3), $\dot{m}_{out}$ is the mass flux in the liquid phase which is being taken out from the crystallizer to the dissolution pipe, and $\dot{m}_{in}$ is the incoming mass flux in the crystallizer from the dissolution unit. They are defined as

$$\dot{m}_{out}(t) = w(t)\rho_{solu}\dot{V},$$
$$\dot{m}_{in}(t) = \dot{m}_{out}(t-t_p) + \frac{k_v\rho_c\dot{V}}{V_c}\int_0^\infty x^3 h(x)n(t-t_p,x)dx - k_v\rho_c x_0^3 B_0(t-t_p), \tag{4}$$

where, $w(t) = m(t)/(m(t)+m_{solv})$ represents the mass fraction. Here, ρ_{solv} is the density of the solution and m_{solv} is the mass of the solvent. Moreover, $t_p \geq 0$ represents residence time in the dissolution unit. It is defined as $t_p = V_p/\dot{V}$, where V_p denotes the volume of pipe. A size dependent growth rate can be defined as

$$G(t,x) = k_g(S(t)-1)^g(1+\alpha_1 x)^{\alpha_2} := g_1(t)g_2(x), \tag{5}$$

where k_g is the growth rate constant. The exponent g denotes the growth order, and α_1 and α_2 are constants. Moreover, $S(t) = w(t)/w_{eq}$ is the supersaturation of the dissolved component which is decreasing with time. Here, w_{eq} is the equilibrium mass fraction. The nucleation rate is defined as

$$B_0(t) = k_b(S(t)-1)^b \mu_3(t), \tag{6}$$

where, k_b is nucleation rate constant and the exponent b gives the nucleation order. Both growth and nucleation are of empirical nature. The above model reduces to the case of fines dissolution without time-delay when $t_p = 0$ and to without fines dissolution case when the second last term on the right-hand side of (1) and the first two terms on the right-hand side of (3) are zero. Then equation (4) is not required.

3. Numerical Technique

Here, we give a brief overview of our numerical scheme, named as constructed scheme. A complete derivation of the scheme can be found in the recent article by Qamar et al. (2009). The method has following two steps:

Step 1: In this step of the method, we solve the moment-system of ODEs, obtained by using the PBE (1), equation (2), and the mass balance equation (3)

$$\frac{d\mu_i(t)}{dt} = i\int_0^\infty x^{i-1}G(t,x)n(t,x)dx - \frac{\dot{V}}{V_c}\int_0^\infty x^i h(x)n(t,x)dx + x_0^i B_0(t), \quad i = 0,1,2,\cdots,5,$$

$$\frac{dm(t)}{dt} = \dot{m}_{in}(t) - \dot{m}_{out}(t) - 3\rho_c k_v \int_0^\infty x^2 G(t,x)n(t,x)dx. \tag{7}$$

In the case of time-delay, the conditions in the pipe are taken equal to the initial conditions in the crystallizer for $t < t_p$ in equation (4). The above equation can not be solved by a standard quadratur method due to the closure problem, therefore the quadrature method of moments (QMOM) is needed. By using QMOM, we obtain

$$\frac{d\mu_i(t)}{dt} = i\sum_{k=1}^N w_k x_k^{i-1}G(t,x_k) - \frac{\dot{V}}{V_p}\sum_{k=1}^N w_k x_k^i h(x_k) + x_0^i B_0(t),$$

$$\frac{dm(t)}{dt} = \dot{m}_{in}(t) - \dot{m}_{out}(t) - 3\rho_c k_v \sum_{k=1}^N w_k x_k^2 G(t,x_k), \quad i = 0,1,2,\cdots,5. \tag{8}$$

where, x_k are the abscissas and w_k are the quadrature weights. A third order orthogonal polynomial is given as (see Qamar et al. (2009)):

$$p_3(x) = x^3 + \frac{(\mu_1\mu_2\mu_4 - \mu_0\mu_3\mu_4 + \mu_0\mu_2\mu_5 + \mu_1\mu_3^2 - \mu_1^2\mu_5 - \mu_2^2\mu_3)x^2}{\mu_2^3 - \mu_0\mu_2\mu_4 - 2\mu_1\mu_2\mu_3 - \mu_0\mu_3^2 - \mu_1^2\mu_4}$$

$$+ \frac{(\mu_1\mu_2\mu_5 - \mu_0\mu_3\mu_5 - \mu_1\mu_3\mu_4 + \mu_0\mu_4^2 - \mu_2^2\mu_4 + \mu_2\mu_3^2)x}{\mu_2^3 - \mu_0\mu_2\mu_4 - 2\mu_1\mu_2\mu_3 - \mu_0\mu_3^2 - \mu_1^2\mu_4}$$

$$+ \frac{(2\mu_2\mu_3\mu_4 + \mu_1\mu_3\mu_5 - \mu_1\mu_4^2 - \mu_2^2\mu_5 - \mu_2\mu_3^3)x}{\mu_2^3 - \mu_0\mu_2\mu_4 - 2\mu_1\mu_2\mu_3 - \mu_0\mu_3^2 - \mu_1^2\mu_4}. \tag{9}$$

The roots of above polynomial are the abscissas x_k of Gaussian quadrature in (8). After having the abscissas, the next step is to calculate weights w_k. According to Press et al. (2007) the weights are given as

$$w_k = \frac{\langle p_2 | p_2 \rangle}{p_2(x_k)p'(x_k)}, \qquad \langle p_2 | p_2 \rangle = \int_0^\infty n(t,x)p_2^2 dx, \quad k = 1,2,3, \tag{10}$$

where, the second order orthogonal polynomial $p_2(x)$ is defined as

$$p_2(x) = \frac{x^2(\mu_0\mu_2 - \mu_1^2) + x(\mu_1\mu_2 - \mu_0\mu_3) + \mu_1\mu_3 - \mu_2^2}{\mu_0\mu_2 - \mu_1^2}. \tag{11}$$

Step 2: After solving the ODE system (8), the discrete values of six moments and the corresponding growth and nucleation rates are available in the time domain of interest. The next step is to get an expression for the CSD. This can be obtained by solving the PBE (1) with the method of characteristics and Duhamel's principle (see e.g. Qamar et al. (2009)). Let $\xi(t,x)$ be a location where the characteristic through (t,x) intersects the x-axis in the backward direction. Then, equation (1) gives (e.g. Qamar et al. (2009))

$$n(t,x) = n_H(t,x) + \begin{cases} \dfrac{g_2(x_0)B_0(\tau)}{G(\tau,x)} \exp\left(-\dfrac{\dot{V}}{V_c}\int_\tau^t h(\xi(t-s,x))ds\right), & \xi(t,x) < 0, \\ 0, & \text{otherwise} \end{cases} \tag{12}$$

where,

$$n_H(t,x) = \frac{G(t,\xi(t,x))n_0(\xi(t,x))}{G(t,x)} \exp\left(-\frac{\dot{V}}{V_c}\int_0^t h(\xi(t-s,x))ds\right). \tag{13}$$

The analytic expression for $\xi(t,x)$ is given as (e.g. Qamar et al. (2009))

$$\xi(t,x) = \begin{cases} \dfrac{1}{\alpha_1}\left[-1 + \varphi(x)e^{-\alpha_1\int_0^t g_1(\tau)d\tau}\right], & \alpha_2 = 1, \quad \alpha_1 \neq 0, \\ \dfrac{1}{\alpha_1}\left[-1 + \left([\varphi(x)]^{\alpha_3} - \alpha_1\alpha_3\int_0^t g_1(\tau)d\tau\right)^{\frac{1}{\alpha_3}}\right], & \alpha_2 \neq 1, \quad \alpha_1 \neq 0, \\ x - \int_0^t g_1(\tau)d\tau, & \alpha_1 = 0, \end{cases} \tag{14}$$

where $\varphi(x) = (1+\alpha_1 x)$ and $\alpha_3 = 1-\alpha_2$. Note that $\xi(0,x) = x$ and $\xi(t,x) < x$ for $t > 0$ because the characteristics moves from left to right in the $x-t$ plane. The last step is to find τ needed for any $\xi(t,x) < 0$ by finding the root of

$$Q(\tau) = \xi(t-\tau,x) - x_0, \qquad Q'(\tau) = \frac{\partial \xi(t-\tau,x)}{\partial \tau}. \tag{15}$$

Thus, for a given t and $\xi(t,x) < 0$, one can find τ by using Newton's formula

$$\tau^{k+1} = \tau^k - \frac{Q^k(\tau)}{Q'^k(\tau)},$$

where, k represents the iteration steps. Note that, $G(t,x)$ and $B_0(t)$ are only available at discrete points in the time domain. Therefore, a linear interpolation is used for calculating their values at any $\tau \in [0,t]$.

4. Numerical Test Problem

In order to validate our numerical schemes for the given model, we consider the following numerical test problem. The initial data are given as

$$n(0,x) = \frac{m_{seeds}}{k_v\rho_c\mu_3(0)\sqrt{2\pi}\sigma} e^{\frac{(x-\bar{x})^2}{2\sigma^2}}, \qquad m(0) = 0.009915. \tag{16}$$

Here, $\rho_c = 1250\,kg/m^3$, $k_v = 0.029$, $m_{seeds} = 0.0025\,kg$, $\bar{x} = 0.0014\,m$, $\sigma = 0.00032\,m$. The minimum and maximum crystal sizes considered are $x_0 = 0$ and $x_{max} = 0.005\,m$,

respectively. The interval $[x_0, x_{max}]$ is subdivided into 300 grid points and the final simulation time is 800 minutes. In the case of size-dependent growth rate we have chosen $\alpha_1 = 200$ m^{-1} in equation (5), while $\alpha_2 = 1$ in all cases. For size-independent growth rate, $\alpha_1 = 0$. Moreover, $k_g = 1.37 \times 10^{-5}$ *m/min*, $g = 0.73$, $k_b = 3.42 \times 10^7$ *(m³min)⁻¹*, $b = 2.35$, $w_{eq} = 0.091$, $m_{solv} = 0.802$ *kg*, $\rho_{solu} = 10^3$ *kg/m³*, $V_c = 10^{-3}$ *m³*, $\dot{V} = 2 \times 10^{-5}$ *m³*, $V_p = 2.4 \times 10^{-4}$ *m³*, and $T = 33\,^oC$. The death function $h(x)$ is considered as

$$h(x) = \frac{1}{\sqrt{2\pi}\sigma_1} e^{-\left(\frac{x}{\alpha\sigma_1}\right)^2}, \qquad \sigma_1 = \frac{1}{0.6\sqrt{2\pi}}, \qquad \alpha = 1.1547 \times 10^3. \tag{17}$$

In Figure 1, the final CSDs for size-independent and size-dependent growth rates are given. The CSDs of our scheme are also compared with those obtained from the high resolution finite volume scheme (FVS) of Koren (1993). In the figure, symbols denote the results of our scheme (constructed) and lines are used for FVS results. The fines dissolution unit dissolves small crystals below some critical size and therefore reduces the number of small crystals in the crystallizer as depicted in the figure. The dissolution of small crystals enhances the solute mass in the solution and therefore could increase the growth rate of the seed crystals. However, in the case of fines dissolution without time-delay, the dissolution of a large number of small nuclei produces negligible effect on the growth rate of seeds crystals. In this case, small nuclei get no time to grow and instantaneously dissolve back in the solution as soon as they achieve a stable size. On the other hand, the fines dissolution with time-delay allows the small nuclei to grow in the crystallizer and the more concentrated solution from the dissolution unit comes back to the crystallizer after certain time-delay. Hence, the concentration of the solution increases in the crystallizer. In other words, the seeds crystals and those introduced by nucleation grow at a faster rate. The numerical results of finite volume scheme and our scheme are almost overlapping. However, our scheme has better resolved the sharp discontinuities. Moreover, Figure 1 also shows plots of the normalized moments. One can see that, the total number (μ_0), size (μ_1) and surface area (μ2) of the crystals reduces, but the total volume (μ_3) increases with fines dissolution. Table 1, gives the errors in mass balances and CPU time of our scheme. It is clear from the table that our scheme preserves mass balance.

Table 1. Errors in mass balances for the size-independent growth rate

Description	Absolute error	Relative error	CPU time (sec)
Without fines dissolution	1.39×10^{-7}	1.37×10^{-6}	11.6
Fines dissolution without delay	1.39×10^{-7}	1.37×10^{-6}	11.8
Fines dissolution with delay	5.8×10^{-4}	4.16×10^{-3}	15.8

5. Conclusions and Remarks

In this manuscript an efficient and accurate numerical technique is derived for solving one-dimensional batch crystallization model with fines dissolution and time-delay. The main ingredients of the scheme are the initial CSD and solute mass, a coupled ODE system of moments and solute mass, as well as an expression for the CSD obtained by employing the method of characteristics and Duhamel's principle. A Gaussian quadrature method based on orthogonal polynomials was used to close the moment-system. The method was found to be efficient, accurate, and free from numerical

dissipation and dispersion. Moreover, no grid refinement is needed for improving the quality of the solution. This property of the numerical method is fruitful in the multi-dimensional case or in situations where more computations are involved. The significant effects of fine dissolution and delay on CSD are illustrated in a case study.
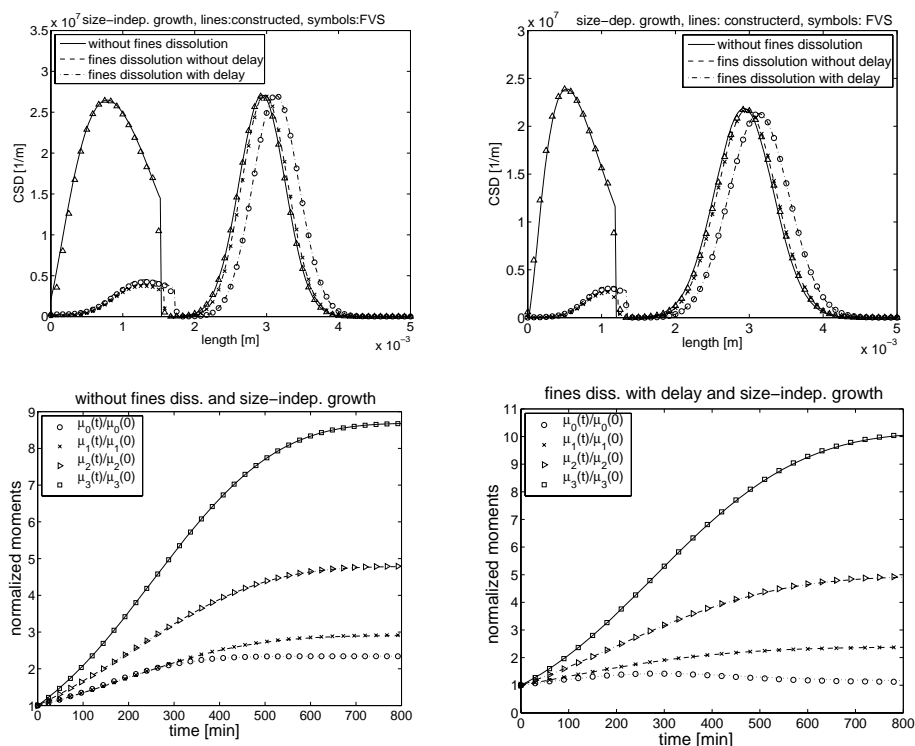
Figure 1: A comparison of results, Lines: our scheme (constructed), symbols: FVS.

6. Acknowledgements

References

R. Grosch, H. Briesen, W. Marquardt, M. Wulkow, (2007), Generalization and numerical investigation of QMOM, AIChE J. 53, 207-227.

H.M. Hulburt, S. Katz, (1964), Some problems in particle technology, Chem. Eng. Sci. 19, 555-574.

B. Koren, (1993), A robust upwind discretization method for advection, diffusion and source terms. In: C.B. Vreugdenhil and B. Koren, editors, Numerical Methods for Advection-Diffusion Problems, Volum 45 of Notes on Numerical Fluid Mechanics, Vieweg Verlag, Braunschweig, 117-138.

S. Qamar, S. Mukhtar, A. Seidel-Morgenstern, M.P. Elsner, (2009), An efficient numerical technique for solving one-dimensional batch crystallization models with size-dependent growth rates, Chem. Eng. Sci. 64, 3659-3667.

D. Ramkrishna, (2000), Population balances: Theory and applications to particulate systems in engineering, Academic Press.

D. Randolph, M.A. Larson, (1988), Theory of particulate processes, second edition, Academic Press.

Optimal control solutions for crystal shape manipulation

Naim Bajcinca[a], Vinicius de Oliveira[a], Christian Borchert[a], Jörg Raisch[a,b]
Kai Sundmacher[a,c]

[a]*Max-Planck Institute for Dynamics of Complex Technical Systems, Sandtorstr.1, 39106 Magdeburg, Germany*
[b]*Technische Universität Berlin, Einsteinufer 17, 10857 Berlin, Germany*
[c]*Otto-von-Guericke Universität Magdeburg, Universitätsplatz 2, 39106 Magdeburg, Germany*

Abstract

Optimal control algorithms and strategies for crystal shape manipulation are explored in this paper. We focus on minimum time trajectories with the temperature as the control input. A challenge results from the inherent constraints in the particle growth vector, which is confined to lie within a cone in the model state-space. As a consequence, switching trajectories with a number of subsequent growth and dissolution phases may be indispensable. Such a strategy employs the unequal growth and dissolution rates in order to achieve crystal morphologies which do not result directly from a pure growth process only. For instantaneous switching between subsequent phases we need to extend the temperature solution profiles to piecewise-continuous. By a suitable specification of the switching manifolds, we are then able to avoid unnecessary long paths in the state-space. Diverse solution approaches to the optimal control problem are suggested using the minimum principle and efficient numerical techniques which utilize the invertibility property of the particle growth dynamics. In particular, we prove that minimum-time scenarios are composed of constant supersaturation trajectory sections.

Keywords: Crystal shape manipulation, optimal control, dynamic inversion, batch processes, multidimensional crystallization

1. Introduction & Motivation

Crystal shape is critical for sophisticated particle products in numerous industries. Properties of dispersed phase products are strongly linked to their shape. For instance, the surface structure and binding energies, and thus reactivity, varies with crystallographic orientation [10]. From the engineering point of view, manipulation of the crystal morphology is therefore essential. While much of the effort in the past has gone into the control and optimization of 1-D crystal-size-distributions [2,4], research on multidimensional crystallization processes has been scarce despite of its evident importance in practice. The limitations in monitoring of crystal shape has been recognized as the major bottleneck [1]. As a consequence, traditional approaches in crystal shape manipulation, as applied in industrial processes like crystallization or precipitation, shifted towards purely chemical techniques. For instance, utilization of additives (see [9]) for blocking or promoting of certain crystal faces has been widely used, despite the fact that, thereby, chemical impurities may arise. However, recent progress in image processing techniques for particle shape monitoring has been an impetus for intensification of theoretical research efforts in modeling and control of multidimensional crystallization processes. In this paper, optimization algorithms for crystal shape manipulation by temperature control only are proposed, by making use of the measurement of concentration and online monitoring of the crystal particle shape.
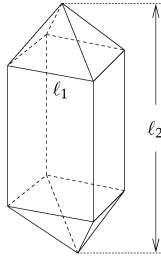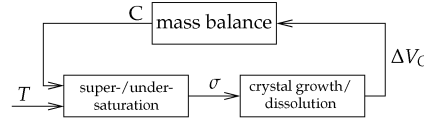
Figure 1: KDP   Figure 2: Batch feedback loop   Figure 3: Reachability regions
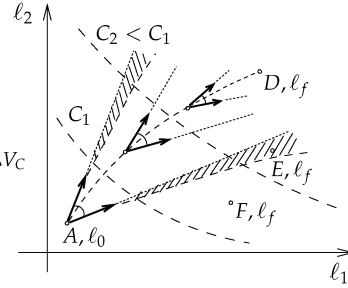
2. Particle shape dynamics

Crystal particle shape dynamics is modeled as a negative feedback structure, as shown in Figure 2. Any temperature (T) change affects supersaturation σ, which in turn forces the crystal to grow or dissolve in order to retain the mass balance equilibrium. The corresponding equations are standard, *e.g.*, see [7],

$$\text{particle growth: } \mathrm{d}\ell_i/\mathrm{d}t = k_{g_i}\sigma^{g_i} \triangleq G_i, \text{ with } \sigma \geq 0 \text{ and } i = 1,2,\cdots,n \tag{1}$$

$$\text{particle dissolution: } \mathrm{d}\ell_i/\mathrm{d}t = -k_{d_i}(-\sigma)^{d_i} \triangleq G_i, \text{ with } \sigma \leq 0 \text{ and } i = 1,2,\cdots,n \tag{2}$$

$$\text{supersaturation: } \sigma = C/C_{\text{sat}} - 1, \ C_{\text{sat}} = a_0 + a_1 T + a_2 T^2, \ a_i > 0, \ i = 0,1,2 \tag{3}$$

$$\text{mass balance: } C = C_0 - \rho_C/V_t M \cdot \Delta V_C, \ V_C = V_C(\ell_1,\ell_2,\cdots,\ell_n) \tag{4}$$

where C stands for the concentration, C_{sat} for saturation concentration, $\Delta V_C = V_C - V_{C,0}$ for crystal volume change, $V_{C,0}$ for the initial crystal volume, ρ_C for crystal mass density, M for solvent molar mass, and V_t for the solvent volume. As case study we consider the two-dimensional potassium dihydrogen phosphate (KH_2PO_4, abbr. KDP) particle, see Figure 1. The numerical data are taken from [5].

The model setting as suggested in (1)-(4) is simplifying for a number of reasons. The growth rate coefficients in different dimensions are assumed to be independent in (1), though experimental reports on coupled growth rate kinetics exist. Nucleation is neglected, but this simplification will not contest our key results. Furthermore, we consider here a single particle only, not a population distribution.

Figure 3 provides a qualitative reachability analysis for trajectories starting at an arbitrary point A in the ℓ-space ($\mathbb{R}^n_+$) for $n = 2$. The growth rate vector $G = \mathrm{d}\ell/\mathrm{d}t$, as defined in (1)-(4), is constrained to lie within a cone parameterized by the supersaturation level σ lying between a minimal value $\sigma_{\min}$ corresponding due to the maximal temperature $T_{\max}$, and a maximal value $\sigma_{\max}$ corresponding to the minimal temperature $T_{\min}$. Obviously, as the particle grows, the supersaturation σ decreases, while the growth cone narrows and rotates clock-wise (for $g_2 > g_1$). In Figure 3, a reachable (dashed) region below the cone with the apex at A appears, while an upper (dashed) piece is lost. This delineated reachability picture is further affirmed by the decrease in the concentration due to the nucleation phenomenon. In this paper, we primarily address (but not confine to) the *reachable* targets within the constraint cone at the starting point A (*e.g.*, D). All such points are connected by a straight line segment. Such trajectories are particularly important in this study.

For $\Delta V_c \rho_C \ll V_t M$ the concentration C remains nearly constant, $C \approx C_0$, and the mass balance feedback decouples in Figure 2. In the next section, this particularly simple situation is examined to gain insight into the optimal solutions in the general case.
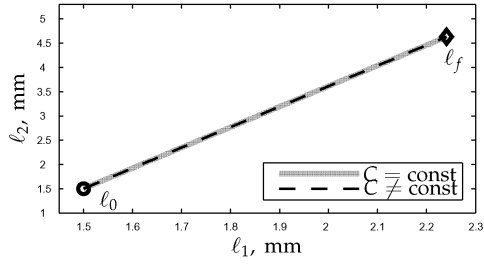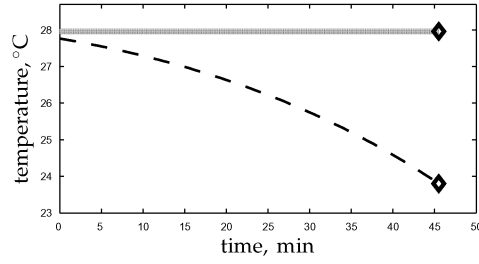
Figure 4: Path profile   Figure 5: Temperature profile

3. Optimal control algorithms

3.1. A single growth trajectory

Consider a two-dimensional particle and let $C = C_0 = const$. We construct first the minimum-time trajectory starting from $\ell_0 = \ell(0)$ for a reachable target $\ell_f = \ell(t_f)$ within its growth constraint cone. According to the Pontryagin's minimum principle (see *e.g.*, [8]), we seek for the optimal inputs $u = u^*$ which minimize the Hamiltonian H defined by

$$H(u,p) = 1 + p^T G(u), \tag{5}$$

where $p = [p_1, p_2]^T$ is the costate vector, and $G = [G_1, G_2]^T$. Note that instead of T, we can equivalently set $u = \sigma = C_0/C_{\text{sat}} - 1$ as the independent input. The costates evolve as $\dot{p}_1 = -\partial H/\partial \ell_1 = 0$ and $\dot{p}_2 = -\partial H/\partial \ell_2 = 0$, i.e. they are constant, $p_1 = p_1^*$ and $p_2 = p_2^*$. Hence, H is a function of σ only, and, the optimal $u^* = \sigma^*$ is obtained from $\partial H/\partial u = 0$

$$\sigma^* = \left(-p_1^* k_{g_1} g_1 / p_2^* k_{g_2} g_2\right)^{\frac{1}{g_2 - g_1}}. \tag{6}$$

The costates $p_1 = p_1^*$ and $p_2 = p_2^*$, as well as, $t_f = t_f^*$, are obtained as solutions to the boundary value problem

$$\ell_f = G(\sigma^*) \cdot t_f^* + \ell_0, \quad H(\sigma^*, p^*) = 0. \tag{7}$$

To summarize, the solution to the minimum-time problem is a linear trajectory driven by a *constant supersaturation* $\sigma^* = const$. Note that σ^* computed from (6) must be feasible ($\sigma_{\min} < \sigma^* < \sigma_{\max}$), since there is a unique line segment which connects ℓ_0 and ℓ_f. The corresponding optimal temperature $T^* = const$ is also constant and it can be analytically computed from (6) and (3).

The equations (6) and (7) hold even if the assumption $C = const$ is dropped. Indeed, we only need to readjust the temperature profile to fulfill the mass balance equation (4) by solving the equation

$$C_{\text{sat}}(T^*) = C(\ell)/(1 + \sigma^*), \tag{8}$$

where σ^* is from (6) and $\ell = \ell(t)$. Numerical results for our case-study are shown in Figure 4 and 5. The plots in gray refer to the simplified case $C = C_0 = const$, and the dashed ones to the complete model. Notice that, in both cases, the optimal trajectory is linear, indicating a constant supersaturation level. The concentration $C = C(\ell)$ is now effected by the mass balance, but the temperature decreases in order to keep the supersaturation at the constant level σ^*. The two optimal solutions provide the identical optimized value for the process duration $t_f^* \approx 45$ [min], see Figure 5.

The extension of the minimum principle approach to the case with an n-dimensional particle is rather straightforward. The boundary value problem (7) provides now $n + 1$ conditions for the n unknown costates p_i, $i = 1, \cdots, n$ and the time duration t_f. To keep things simple, again (as in the two-dimensional case), any two state equations from (7) are sufficient to solve for the optimal σ^* and t_f^*. The remaining $n - 2$ states affect the temperature profile via (8).

3.2. Switching trajectories

Target points, such as F in Figure 3, are not reachable from A without further ado. One is forced to switch at some point from the growth to a dissolution mode, see [3]. In this section, we provide conditions for minimum-time trajectories comprising a fixed number of switching phases. As switching manifolds we specify the manifolds in ℓ−space corresponding to a constant concentration or (according to (4)), equivalently, to a constant particle volume

$$\text{growth to dissolution: } m_u(\ell) = V_C(\ell) - V_u = 0, \quad V_u \text{ fixed} \tag{9}$$

$$\text{dissolution to growth: } m_l(\ell) = V_C(\ell) - V_l = 0, \quad V_l \text{ fixed} \tag{10}$$

For simplicity, consider a single-switching trajectory. Let t_b be the switching time instant from a growth to a dissolution mode. Then the Hamiltonian function reads

$$H = 1 + p_+^T G_+ \triangleq H_+, \ 0 < t < t_b, \text{ and } H = 1 + p_-^T G_- \triangleq H_-, \ t_b < t < t_f. \tag{11}$$

where the subscript "$+$" refers to the growth and "$-$" to the dissolution mode. Using the arguments of the previous section we know that the costates p_+ and p_- must remain constant within a switching mode. In accordance with the minimum principle, the boundary value problem now extends to

$$\text{growth mode: } \ell_b = G_+ \cdot t_b + \ell_0, \quad m_u(l_b) = 0 \tag{12}$$

$$\text{switching conditions: } H_+(t_b) = H_-(t_b), \quad p_+ = p_- + \lambda \cdot \partial m_u / \partial \ell |_{\ell_b} \tag{13}$$

$$\text{dissolution mode: } \ell_f = G_- \cdot (t_f - t_b) + \ell_b, \quad H_-(t_f) = 0, \tag{14}$$

where we require continuity for the Hamiltonian at the switching time instant, and introduce a Lagrange-multiplier λ for the switching manifold $m_u(\ell) = 0$ at the switching point ℓ_b.

The above system of nonlinear algebraic equations provides $3n + 3$ conditions on ℓ_b, p_+, p_-, t_b, t_f and λ, i.e. on $3n + 3$ unknowns. Notice that, due to the continuity of the Hamiltonian, the costates must jump from p_+ to p_- at the switching moment. Again, the minimum-time single-switching trajectory consists of two linear trajectory sections with *constant supersaturation levels*. At the switching point the optimal value of supersaturation jumps from $\sigma_+^* > 0$ to $\sigma_-^* < 0$, which satisfy equations of the form (6). In other words, the optimal control policy imposes a *discontinuity* in the optimal temperature profile.
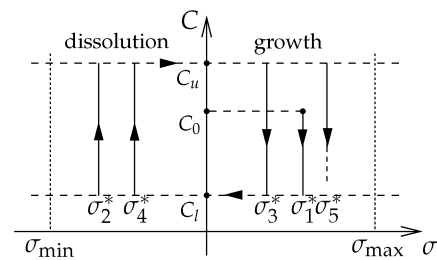
Figure 6: Optimal supersaturation levels

The optimal switching conditions (12)-(14) are easily generalized for a multiple switching trajectory. Figure 6 illustrates a switching control policy between optimal supersaturation levels in the $(\sigma - C)$-plane. The number of unknown variables and conditions of the boundary value problem increases by $2n + 2$ with each additional switching. In total, a trajectory including η switching modes, involves a system of $(2\eta + 1)(n + 1)$ nonlinear algebraic equations.

Optimal control solutions for crystal shape manipulation
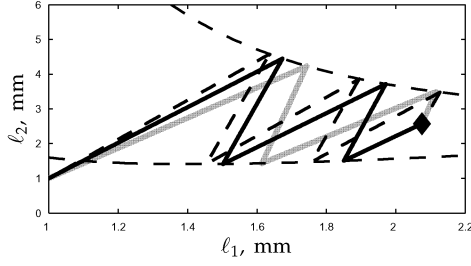
Figure 7: Path profile
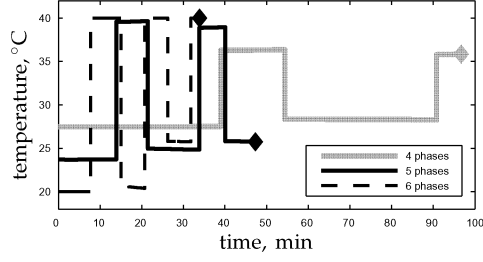
Figure 8: Temperature profile

3.3. Numerical optimization algorithms

For a trajectory involving η switchings we expect from an optimization algorithm to decide on $\eta + 1$ constant optimal supersaturation levels σ_i^*, $i = 1, \cdots, \eta + 1$, see Figure 6. Using the facts from Section 3.2, we formulate the following nonlinear programming problem in terms of the decision variables σ_i and $\ell_1^{(i)}$

$$\min \sum_{i=1}^{\eta+1} \frac{\Delta \ell_1^{(i)}}{G_1(\sigma_i)}$$
$$\text{s.t. } m_u(\ell^{(i)}) = 0, 0 \le \sigma_i \le \sigma_{\max}, i \in \mathbb{N}_o, m_l(\ell^{(i)}) = 0, \sigma_{\min} \le \sigma_i \le 0, i \in \mathbb{N}_e \quad (15)$$
$$\text{where } \Delta \ell_1^{(i)} = \ell_1^{(i)} - \ell_1^{(i-1)}, \ \ell^{(i)} = \frac{G(\sigma_i)}{G_1(\sigma_i)} \Delta \ell_1^{(i)} + \ell^{(i-1)}, \ \ell^{(0)} = \ell_0, \ \ell^{(\eta+1)} = \ell_f,$$

where the symbol $\ell^{(i)}$ stands for the ith switching point on the manifolds $m_u(\ell)$ (if i is even) or $m_l(\ell)$ (if i is odd). Note that the ith summing term in the cost function represents the time length of the ith switching phase. Without loss of generality, we start here with a growth phase. The decision variables $\ell_1^{(i)}$ and σ_i count up to $2\eta + 1$. The number of variables which the optimizer in (15) has to decide on reduces to $1/(n+1)^{\text{th}}$ as compared to the boundary value problem in the previous section.

Obviously, the problem formulation (15) applies for trajectories consisting of linear sections only. For "non-linearly" reachable targets, such as E in the dashed region in Figure 3, the profiles provided by (6) are not feasible. Therefore, we propose a numerical algorithm, which makes use of the invertible structure of the model (1)-(4). Indeed, given the functions $\ell_1 = \ell_1(t)$ and $G_1 = d\ell_1/dt$, all the system variables, *i.e.* $\ell(t) = [\ell_1(t), \ell_2(t), \cdots, \ell_n(t)]^T$, $\sigma(t)$ and $T(t)$, are uniquely determined as follows

$$\sigma(t) = (G_1/k_{g1})^{\frac{1}{g_1}}, \ \ell_i(t) = \int k_{gi} \sigma^{g_i} dt, \ C_{\text{sat}}(T) = C(\ell)/(1+\sigma). \quad (16)$$

Now, introduce a parametrization $\ell_1(t) = \ell_{1,0} + \alpha^T \psi(t)$, where α^T is a parameter vector and $\psi(t)$ includes functions from a convenient basis (*e.g.*, B-splines). It is important to note that $\ell_i(t)$ in (16) are integrable in closed form, i.e. the expression $\ell(t) = \ell(\alpha, \psi(t))$ is analytically available. The dynamic optimization problem reduces again to a nonlinear program with α as the decision variable

$$\min t_f(\alpha^*) \ \text{s.t.} \ \ell_0 + \ell(\alpha^*, \psi(t_f)) = \ell_f, \ k_{g1} \sigma_{\min}^{g1} \le G_1(\alpha^*, \psi(t)) \le k_{g1} \sigma_{\max}^{g1}. \quad (17)$$

The notation $t_f(\alpha^*)$ indicates that t_f can be computed from $\ell_{1,0} + \alpha^{*T} \psi(t_f) = \ell_{1,f}$. The algorithm (17) is easily extended to switching trajectories. A switching time instant t_b can be computed by solving algebraic equations of the form $m_{u/l}(\ell(\alpha, \psi(t_b))) = 0$, which apply when a trajectory section hits onto a switching manifold.

A slightly different, but a more direct trajectory parametrization, uses the fact that the whole information on a growth trajectory $\ell = \ell(t)$ is contained in its projection into any two-dimensional (*e.g.*, $(\ell_1, \ell_2)-$) plane. Indeed, then

$$\sigma(\ell_1) = \left(\frac{k_{g1}}{k_{g2}} \frac{\mathrm{d}\ell_2}{\mathrm{d}\ell_1} \right)^{\frac{1}{g_2 - g_1}}, \ t_f = \int_{\ell_{1,0}}^{\ell_{1,f}} \frac{\mathrm{d}\ell_1}{k_{g1}\sigma^{g_1}}, \ \ell_i = \int \frac{k_{gi}}{k_{g1}} \sigma^{g_i - g_1} \mathrm{d}\ell_1, \ C_{\mathrm{sat}}(T) = C(\ell)/(1+\sigma).$$

$$(18)$$

One can introduce a parametrization $\ell_2 = \ell_{2,0} + \alpha^T \psi(\ell_1)$ in terms of ℓ_1 and proceed analogously to the formulation in (17). While, typically, the order of the functions $\psi(\ell_1)$ is low, the expressions $\ell_i = \ell_i(\alpha, \psi(\ell_1))$ and $t_f = t_f(\alpha, \psi(\ell_{1,f}))$, $i \geq 3$, must be computed numerically.

Figures 7 and 8 present the numerical results for the case-study with different number of switching phases. Note that the optimal t_f^* reduces as the number of switchings increases. For the indicated switching manifolds in the figure, the optimization problem is not feasible for less than four and more than six switching phases. Clearly, optimal profiles tend to a bang-bang profile with an increasing number of switchings.

4. Conclusions

Different optimization strategies for multidimensional crystal shape manipulation are proposed in this article. Optimal scenarios with subsequent growth and dissolution phases have been investigated. The class of switching trajectories driven by optimal constant supersaturation levels emerge as particularly important since they produce minimum-time trajectories and are easy to implement in practice by standard constant supersaturation control. In addition, we use the invertibility property of the crystal growth dynamics to provide two efficient numerical algorithms for the construction of nonlinear minimum-time trajectories. Such algorithms are good candidates for extension to the case of crystal population distributions. Moreover, the design of the switching manifolds could be naturally embraced in a more general optimization setting.

References

[1] D. Patience and J. Rawlings (2001): *Particle-shape monitoring and control in crystallization processes*, AIChE J., 47(9):2125-2130.

[2] R. Braatz (2002): *Advanced control of crystallization processes*, An. Rev. in Control, 26 , 87-99.

[3] R. C. Snyder, S. Studener, and M. F. Doherty (2007): *Manipulation of crystal chape by cycles of growth and dissolution*, AIChE-J., 53 (6), 1510-1517.

[4] U. Vollmer and J. Raisch (2003): *Control of batch cooling crystallisers based on orbital flatness*, Int. J. Control(1635–1643).

[5] D. L. Ma, D. K. Tafti, R. D. Braatz (2002): *High-resolution simulation of multidimensional crystal growth*, Ind. Eng. Chem. Res., 41, 6217-6223.

[6] C. Borchert, D. Ramkrishna and K. Sundmacher (2009): *Model based prediction of crystal shape distributions*, In ESCAPE19

[7] A.D. Randolph and M.A. Larson (1998): *Theory of particulate processes*, Academic Press.

[8] D.E. Kirk: *Optimal control theory: An introduction*, Dover Publications, 1998.

[9] L. Weissbuch, et al. (1995): Acta Cryst. B, 51, 115-148.

[10] H.G. Yang, et al. (2008): Nature, 453, 638-641.

CALS-model of innovative technology for plasmachemical synthesis of nanopowders

A. Bessarabov,[a] A. Kvasyuk,[a] M. Ivanov,[b] N. Menshutina[c]

[a]The State Scientific-Research Institute of Chemical Reagents and High Purity
Chemical Substances (IREA); Bogorodsky Val, 3, 107076, Moscow, Russia; E-mail:
bessarabov@irea.org.ru
[b]LOHR PLASMA, 29 rue du 14 Juillet, 67980 Hangenbieten, France; E-mail:
otrok@live.ru
[c]Mendeleev University of Chemical Technology of Russia (MUCTR); Miusskaya square,
9, 125047, Moscow, Russia; E-mail: chemcom@muctr.edu.ru

Abstract

Information model of plasmochemical processes for synthesis of nanodispersed
materials was developed on the basis of the CALS concept for the example of ultrapure
compounds of tin, iron, silicon, titan, tungsten.

Keywords: CALS, plasmachemistry, nanopowders, oxides, thermodynamic modelling.

1. Introduction

In the last years, a special interest was generated by some nanopowders which are used
in production of ceramic products, such as: tungsten carbide, tantalum, niobium,
hafnium, molybdenum, silicon; silicon nitride, titanium nitride, aluminium nitride;
silicon oxide, ferric oxide, aluminium oxide, zirconium oxide, titanium oxide and tin
oxide. In our works special attention has been given to producing of oxide nanopowders
of high purity.

2. CALS-project of universal plasmachemical apparatus

To obtain a nanodisperse oxide of high purity the universal plasmachemical apparatus
[1] was developed, which allows applying not only initial hardphase product by means
of powder feeder, but liquid-phase reagents with the help of special sprayer.
Universality of the plant allows obtaining nanodisperse oxide of metals of 2nd, 3rd and
4th group of the periodic system. Depending on amount of the parent material, plasma-
formation gas flow and power insertion it is possible to obtain nanopowders of different
dimension series.
Development of the plasmachemical process was carried out in the context of the most
current and perspective system of computer support – CALS-technology (Continuous
Acquisition and Life cycle Support). Within design CALS-project a typical scheme was
created (protocol of application) - «Initial data for designing» (fig. 1).
Design electronic description according to STEP standard (fig. 1) contains the structure
and variants of item configuration, geometrical models and drawings, properties and
features of components. At the element of this scheme universal plasmachemical
apparatus is shown allows to transfer to reactor (Fig. 1-a) not only the initial solid
product by means of powder feeder, but the liquid reagents (chlorides and alchoxides)
with a special sprayer (Fig. 1-b).
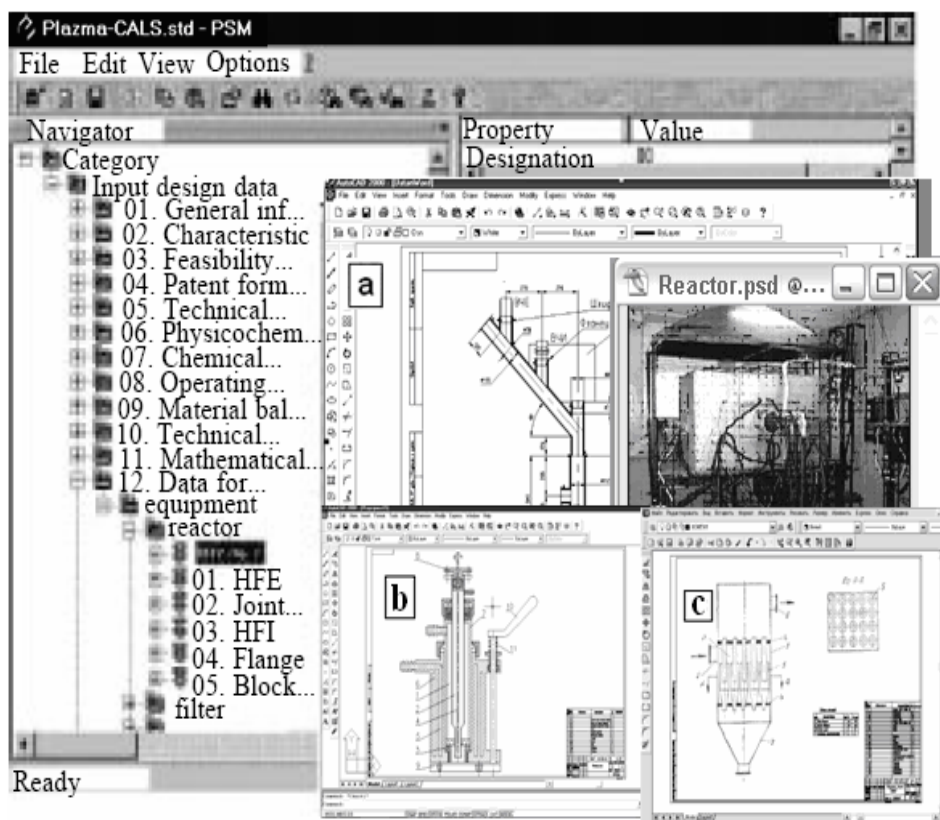
Fig. 1. Element of design CALS-project of plasmachemical apparatus for synthesis of nanomaterials (a – reactor, b – sprayer, c – filter).

For this CALS-project of apparatus (Fig. 1) includes metering device for the transfer of initial materials powders, pulverizer for transfer of plasma-creating gas, filter for the product recovery (Fig. 1-c) and plasma torch. Apparatus universality allows obtaining nanodisperse compounds of tin, iron, silicon, titan, tungsten on it.

3. Regulation of nanopowders dispersity

To the work and CALS-project researches, dealing with influence for nanodispersity of two parametrical complexes: aggregate condition of initial substance; ratio of speed pressures of plasma stream (PS) and stream of input gas (SIG) were included. Research of influence of aggregate condition was carried out for plasmachemical synthesis of nanopowders of silicon oxide (required granulated content: $d = 10$ nm). To the proper subcategory of information CALS-project the table of obtained results (Fig. 2-a) was included. It is shown that for obtaining of required granulated content when using of initial substance (tetraethoxysilane – TEOS) ratio of SIG/PS is enough to 1. When input through sprayer of liquid TEOS ratio of SIG/PS equal to 12 is required. When input through feeder of quartz powder ($d_0 = 10$ mcm) for obtaining of nanodisperse silicon oxide (10 nm) high ratio of SIG/PS equal to 50 is required.

In CALS-project researches of influence for dispersity of final product of ratio of SIG/PS (Fig. 2-b) are shown. Quartz powder was used as initial product ($d_0 = 10$ mcm). Ratio of SIG/PS (β) varied from 20 till 50.
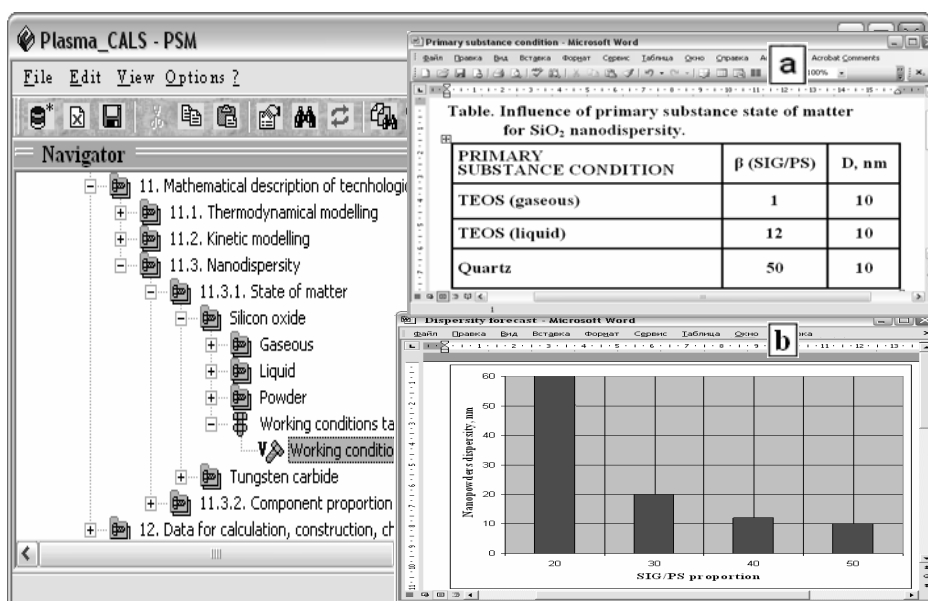
Fig. 2. Element of CALS-project «Modeling of nanopowders dispersity» (influence on dispersity: a – aggregate condition; b - ratio of SIG/PS).

As the result there were nanopowders with the diameter from 60 till 10 nm. This relation is approximated by the next exponential relation: $\ln(d) = a_0 + a_1\beta$. Linear equation of correlation of ratio of SIG/PS and input power (W) also included to the model: $W = W_0 + b_1\beta$. Application in the CALS-project of methods of computer modeling and forecasting allow to create optimum flexible structure of high technology plasmachemical production and to provide full post sell support, including the documentation in electronic form.

4. Thermodynamic modeling

Section no. 11 (mathematical description of the process) of the CALS project (Fig. 1) presents the results obtained in simulating the plasmachemical synthesis. Analysis of chemical and heat-and-mass exchange processes at elevated temperatures leads to severe difficulties already in the stage of formulation of the simulation problem. It is appropriate to use thermodynamic simulation methods as a first approximation. These methods presume that, in the processes under consideration, the working body forms a conditionally closed, isolated system in which a local thermodynamic equilibrium (LTE) is attained. In this approximation, the state of the system is only determined by the content of chemical elements in the system and by values of two parameters of state. The use of the thermodynamic equilibrium approximation is justified by the high concentration of energy in the volumes under consideration and by the resulting high rates of conversion processes, which instantaneously bring the system in the LTE state.

A calculation of equilibrium for isolated multicomponent thermodynamic systems can be reduced to a problem of determination of a state with the minimum entropy. Therefore, to compose the sought-for system of equations, it is necessary to find an analytical relationship between the entropy of a unit mass of the working body and the thermodynamic parameters determining its composition, properties, and existence conditions.
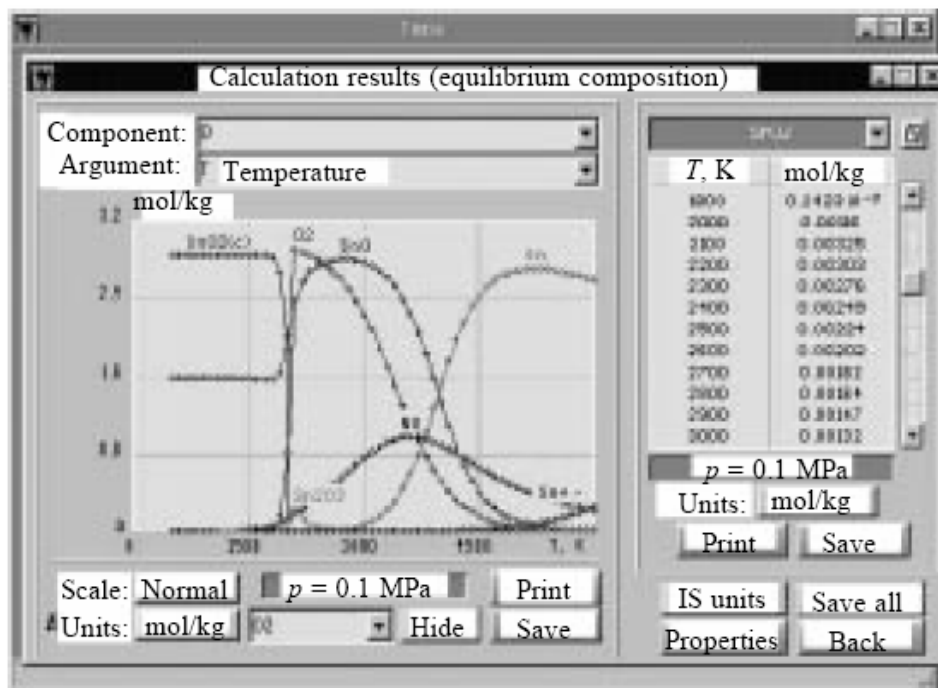
Fig. 3. Screen form of the results of thermodynamic simulation of the plasmachemical synthesis.

A thermodynamic calculation of the equilibrium states of the system is performed in a wide range of basic parameters of the plasmachemical process: starting component ratios, temperatures, and pressures (Fig. 3). A thermodynamic simulation makes it possible to choose the synthesis conditions, analyze the environmental safety of the production process, and assess the mechanism of thermal dissociation of the starting compounds.

5. Computer quality management

The system of a computer quality management is developed for a choice and the analysis of initial reagents and target products of plasmachemical synthesis (Fig. 4). The system has hierarchical structure of databases. Three basic information categories are allocated: «Analyzed substance»; «Analysis procedure» and «Output documentation» [2]. The developed information structure allows choosing optimum methods of the analytical control for as much as possible exact definition of qualitative characteristics of analyzed products.

On the basis of information model the program complex of the CALS-project of analytical monitoring is developed. The program interface is performed taking into account an optimality of work of the user. Special procedures and the screen forms including a complex of modern elements of representation of the information and interaction with the user are developed for each stage of functioning of system.

For assortment of initial reagents considered by us and target products of plasmachemical synthesis the following inorganic clusters are entered into the first category (Fig. 4): «alchoxides» (tetraethoxysilane, tetrabuthoxytitan); «oxides» (oxides of silicon, titan, tin, iron, and tungsten); «salts» (tungsten carbide).

Fig. 4. Element of CALS-project CQM-system in category «Analyzed substance» (a – SiO$_2$ high purity «HP 7-5»; b – SiO$_2$ «HP 12-4»; c – tetraethoxysilane «HP 11-5»).

The chosen structure of classification of high purity nanomaterials corresponds to the applied All-Russian qualifier of standards (ARQS), a part of Uniform system of classification and coding of the technical and economic and social information (USCC) of the Russian Federation. The qualifier is harmonized with the International qualifier of standards (IQS) and the Interstate qualifier of standards.

6. Conclusions

Implementation of CALS technologies and modern computerized quality management systems in development of science-intensive plasmachemical processes raises the productivity and diminishes the expenditure of time and material resources. This is achieved by simplified access to information, reorganization of the working activities (without change of the tasks to be accomplished), computerization of the working environment, and significant improvement of relationships between partner designers.

7. Acknowledgements

References

[1] A.M. Bessarabov, A.N. Ponomarenko, M.Ya. Ivanov, 2007, Information CALS-technologies (ISO-10303 STEP) in development of plasmachemical processes of obtaining of ultradisperse oxides of special pure, Journal of applied chemistry., Vol. 80, No. 1, p. 15-19.
[2] A.M. Bessarabov, O.A. Zhdanovich, A.M. Yaroshenko, G.E. Zaikov, 2007, Development of an analytical quality control system of high-purity chemical substances on the CALS concept basis, Oxidation Communications, Vol. 30, No. 1, p. 206-214.

Simulation and Experimental Evaluation of Seed and Supersaturation Control Design Approaches for Crystallisation Processes

Erum Aamir, Zoltan K. Nagy*, Christopher D. Rielly

Department of Chemical Engineering, Loughborough University, Loughborough, LE11 3TU, United Kingdom, z.k.nagy@lboro.ac.uk.

Abstract

The paper presents a methodology for the systematic design of operating recipes for batch cooling crystallisation systems with the aim to produce a desired target crystal size distribution. The population balance model is solved using the method of characteristics under the assumptions of a constant supersaturation and growth dominated process, yielding a simplified analytical expression for the size distribution. A method is proposed for designing the seed distribution, which can be used in conjunction with the supersaturation set-point design to shape the product crystal size distribution (CSD). The approach designs the seed as a mixture of crystals obtained from standard sieve analysis. A second approach is also proposed which uses dynamic seed addition during the batch to control the final size distribution.

Keywords: Seed recipe, supersaturation control, size-dependent growth, dynamic seeding.

1. Introduction

Batch cooling crystallisation is an industrially important unit operation due to its ability to provide high purity separation. Recently, the production of crystalline compounds with consistent properties throughout batches has become very important, especially in the pharmaceutical industries with high value-added products and strong regulatory constraints. Most of the product qualities are directly related to the crystal size distribution [1-2]. The main difficulty in batch crystallization is to accomplish a uniform and reproducible CSD [3]. One way to enhance the control of CSD is to use supersaturation control (SSC), which drives the process within the metastable zone to avoid nucleation [4]. Although this approach has proved to produce high quality crystals, the set-point operating profiles for the supersaturation controller are usually chosen arbitrarily, or by trial-and-error experimentation [4] and generally without taking into account the characteristics of the seed. Recently it has been recognised that for the optimisation of process conditions for crystallisation, the most suitable candidates as manipulated variables can be the temperature trajectories and seed characteristics [5].

Seeding has been known for a long time as an effective technique to stabilize batch crystallization processes and is often treated as an art [6-8] rather than science. In addition, quantitative information on seeding is limited. Generally there is a lack of systematic methodologies related to the amount and size of seeds that should be added into a crystallizer to obtain a product of desired size distribution. The paper presents a novel approach for the systematic design of the seed recipe for supersaturation controlled crystallization processes so that a target CSD with a desired shape is

obtained. In the case of seeded batch cooling crystallization processes controlled at constant supersaturation, the main governing phenomenon is growth. For these systems, an analytical solution of the population balance equation can be obtained, which gives the entire CSD at any moment of the batch. A design parameter, as a function of the batch time and supersaturation can be used to characterise and design supersaturation controlled processes. The approach is extended by designing the optimal seed distribution required to achieve the desired shape of the final CSD, which due to the characteristics of the growth kinetics may not be possible to achieve by optimizing the constant supersaturation only. Additionally a novel methodology is proposed which designs dynamic seed addition profiles required for a desired target final CSD.

2. Systematic design of supersaturation controlled crystallization processes based on explicit solution of the PBE

Considering a single growth direction with one characteristic length L, and a well-mixed crystalliser with supersaturation control and growth as the only dominating phenomena, the population balance equation (PBE) has the form,

$$\frac{\partial f_n(L,t)}{\partial t} + \frac{\partial(G(S,L;\theta_g)f_n(L,t))}{\partial L} = 0, \tag{1}$$

where $f_n(L,t)$ is the crystal size distribution expressed as a number density function (number of crystals per unit mass of slurry), t is time, $G(S,L;\theta_g)$ is the rate of crystal growth, $S = (C - C_{sat})$ is the supersaturation, C is the solute concentration, $C_{sat} = C_{sat}(T)$ is the saturation concentration with T being the temperature and θ_g is a vector of growth kinetic parameters, respectively. The generic PBE (1) can be reduced to a system of ODEs by applying the method of characteristics (MOCH) [8]. Seed is added to suppress nucleation and in the case of supersaturation controlled crystallisation the process will be dominated by growth. For the generic case of size dependent growth, the kinetics are assumed to be given by,

$$G = k_g S^g (1 + \gamma L)^p, \tag{2}$$

where $\theta_g = [k_g, g, \gamma, p]$ is the growth parameter vector. Applying the MOCH, reduces eq. (1) to the following system of two ODEs:

$$dL / dt = k_g S^g (1 + \gamma L)^p, \tag{3}$$

$$df_n(L,t) / dt = -k_g S^g \gamma p (1 + \gamma L)^{p-1} f_n(L,t). \tag{4}$$

In the case of well-controlled constant supersaturation, which follows the desired set-point value, S_{sp} the system (3)-(4) can be solved analytically [8] with the solution (for $p \neq 1$ and $\gamma \neq 0$) given by,

$$L = (((1 + \gamma L_0)^{1-p} + k_g S^g t \gamma (1-p))^{\frac{1}{1-p}} - 1) / \gamma \tag{5}$$

$$f_n(L) = f_{n,0}(L_0) \left(1 + (k_g S^g t \gamma (1-p) / (1 + \gamma L_0)^{1-p})\right)^{\frac{p}{p-1}} \tag{6}$$

Equations (5)-(6) can be used to compute the dynamic evolution of the CSD for a generic growth dominated process by discretizing the initial (seed) distribution $f_{n,0}(L_0) = f_{seed}(L_0)$ for different values of L_0, and computing the corresponding L and f_n values at various times. The CSD given by the system (5)-(6) is determined by the

product between S^g and t, and hence a design parameter (ϕ) can be defined as, $\phi = S^g t$. The optimal supersaturation control design parameter (ϕ^*) can be determined by minimizing the difference between the discretized target CSD and the predicted CSD obtained from the analytical estimator (5)-(6),

$$\min_{\phi} \left\{ \sum_{i=1}^{N_d} (f_{v,i} - \tilde{f}_{v,i})^2 \right\}, \qquad s.t : 0 \leq \phi \leq \phi_{\max}; C(t_{batch}) \leq C_{f,\max}. \tag{7}$$

where N_d is the number of discretizations, $\tilde{f}_{v,i}$ is the discretized target CSD (volume particle density function), $C(t_{batch})$ is the solute concentration at the end of the batch and $C_{f,\max}$ is the maximum acceptable concentration at the end of the batch to achieve the required yield. Once the optimal ϕ^* is determined the batch time or the supersaturation can be fixed and the other parameter can be calculated. The corresponding temperature profiles can be calculated in each case, following the direct design concept [9], which will all yield the same target CSD used in the design.

The approach has been applied to a model system of potassium dichromate in water for which the size-dependent growth parameters have been reported in the literature [9]. Figure 1 (a) shows the results of temperature profiles corresponding to design parameter $\phi = 0.206$ min, obtained for different supersaturation set-points and batch times in the case of a fixed target CSD (Figure 1 (c)). The corresponding concentration profiles are shown in Figure 1 (b). The proposed simplified model-based direct-design approach provides a systematic methodology to operate crystallisation processes at a constant supersaturation by controlling a temperature trajectory throughout the batch.
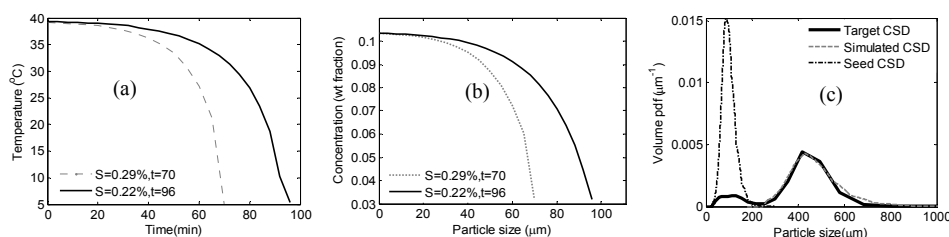
Figure 1: a) Temperature profiles b) concentration profiles obtained at different set-point supersaturation, S_{sp} and batch time t_{batch}, corresponding to the same design parameter $\phi = 0.206$ min, optimized to achieve the target and seed CSDs shown in (c).

3. Seed recipe design

In addition to the supersaturation and batch time, the seed recipes may also be optimised to obtain a desired target CSD, which may not be achievable by optimising the parameter ϕ alone. Experimentally, seed can be obtained by mixing different amounts of seeds with different size distributions [9]. In practice most often the mean and standard deviations characterizing the seeds in particular size ranges is fixed depending on the mesh sizes used for seed sieving. The optimization was solved considering only the amounts of seeds used in the various size ranges as the decision variables, and is formulated as follows:

$$\min_{\beta} \left\{ \sum_{i=1}^{N_d} (f_{v,i} - \tilde{f}_{v,i})^2 \right\}, \qquad s.t: 0 \leq \beta \leq \beta_{\max}; C(t_{batch}) \leq C_{f,\max}. \tag{8}$$

$$f_{n,seed} = \sum_{i=1}^{N_G} w_i N_i (L_{m,i}, \sigma_i), \tag{9}$$

where $\beta = [w_1, w_2, ..., w_N]$, w_i the weights, $L_{m,i}$ the mean sizes (μm) and σ_i (μm) the standard deviations of the respective sieves. The means and standard deviations are calculated for the consecutive sieves as follows:

$$L_{m,i} = (\mathcal{L}_i + \mathcal{L}_{i+1})/2, \qquad \sigma_i = (\mathcal{L}_{i+1} - \mathcal{L}_i)/2, \tag{10}$$

where $\mathcal{L}_i$, $i = 0, 1, ..., N_G$, are the standard sieve sizes, with N_G being equal to the total number of selected sieves. These values provide approximately 2σ overlap between the distributions of seeds from adjacent sieve ranges, and correspond to the experimental observations of the sieve analyses of the studied compound. Figure 2 shows the results of the seed design, for arbitrary bimodal and tri-modal target distributions respectively, using the optimal $\phi^* = 0.1981$ obtained by solving an optimization problem of the form of eq. (7).

Table 1: Optimised seed parameters for the target CSDs.

Target CSD	N_G	$L_{m,i}$	σ_i	w_i
Bimodal	2	152.5	27.5	0.222
		215.0	35.0	0.779
Tri-modal	3	100.0	25.0	0.752
		152.5	27.5	0.226
		196.0	16.0	0.023
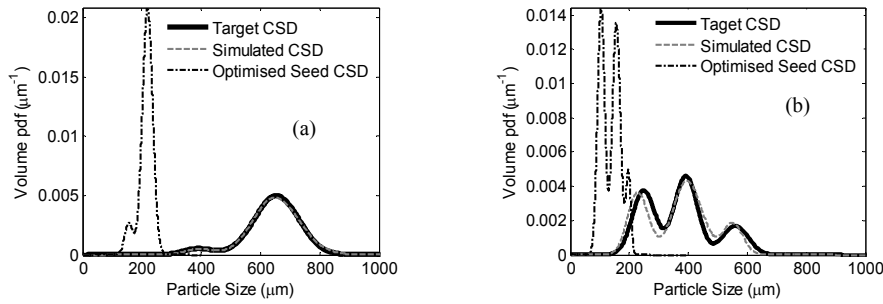
Figure 2: Optimized seed for a) bimodal target CSD and b) tri-modal target CSD.

Table 1 contains details of the optimised seeds for the two target distributions. Figure 2 (a) indicates that the resulting CSD is in good agreement with the desired bimodal CSD. The sieve sizes used were 125, 180 and 250 μm and the optimal seed distribution is a sum of two Gaussians. If a more complex, tri-modal target CSD is used, as shown on Figure 2 (b) and the optimization is solved with seeds from sieve sizes 75, 125, 180 and 212 μm, the optimal seed distribution consists of a sum of three Gaussians. However the agreement with the target distribution is inferior, due to the limitation in the sieve sizes.

4. Shaping of the CSD via optimal dynamic seed addition

Shape the final CSD by designing the initial size distribution as a mixture of seed with different size distributions can be difficult on industrial scale. Similar results can be achieved if mono-modal seed is introduced in the crystallizer during the crystallisation process. The only requirement is that the seed CSD must be narrower than the target distribution; the narrower the seed distribution, the closer the final CSD will be to the target CSD. According to the dynamic seeding approach the batch time is discretised in

N intervals and the amount of seed added to the system at the start of each time interval is determined by solving the following optimisation problem:

$$\min_{w_1,...,w_N} \left\{ \sum_{i=1}^{N_d} (f_{v,i} - \tilde{f}_{v,i})^2 \right\} \tag{11}$$

where w_i, $i = 1,2,...,N$, are the amounts of seed in weight %. Figure 3 (a) shows the result of the optimised dynamic seed addition profile in the case of a tri-modal distribution, which cannot be captured precisely using the initial seed recipe design with seeds in size ranges available from standard sieve analysis (Figure 2(b)). In this case the seed distribution was Gaussian with a mean size of 60 μm and standard deviation of 5 μm. The simulated and target distribution are in very good agreement, indicating that the dynamic seed addition approach can provide multi-modal target distributions using mono-modal seed only. It can be seen that in the case of the tri-modal target distribution seed addition was required during the first 90 min of the batch to achieve the desired shape of the target distribution. The narrower the seed, the smaller the error is in achieving different target distributions. Figure 3 (c) shows the results for the same tri-modal target distribution using a broader seed size distribution, with a mean size of 120 μm and a standard deviation of 35 μm, indicating that if the seed is too large and/or the distribution too broad the target CSD may not be achievable.
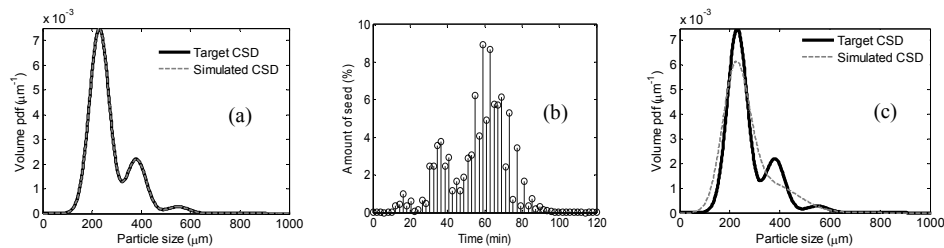
Figure 3: a) Results of the dynamic seed addition for a tri-modal distribution (the two curves overlap); b) dynamic seed addition profile (seed with narrow distribution); c) tri-modal target distribution with broad seed distribution.

5. Experimental evaluation of crystallization using mixture of seeds

In experiments, potassium dichromate ($K_2Cr_2O_7$) (>99.95% purity, Fisher BioReagents) solution was prepared, corresponding to a solubility of 20 g of potassium dichromate per 100 g of water at 30 °C. After the solution was equilibrated at 40 °C for 15 min, the temperature of the solution was lowered to 29 °C and after 10 minutes 1.2 g of mixture of sieved seed was added having equal masses of seed in the size ranges 40-63 μm and 90-106 μm; the temperature was reduced from 29°C to 20°C following a cubic profile. FBRM and ATR-UV/Vis spectrometer measurements showed constant values after the addition of seed, which indicated no dissolution occurred at 29°C and the solution was supersaturated. The concentration from the ATR-UV/Vis was obtained by developing a calibration model of the form:

$$C = a_0 + a_1 A_1 + a_2 A_2 + a_3 T \tag{12}$$

where A_1 and A_2 are the absorbance values at the two wavelengths 270.15 nm and 377.89 nm respectively, C is the concentration (g/ml) and T is the temperature (°C). The parameters were obtained using least-squares minimization. The values obtained

with their uncertainty bounds are: $a_0 = -0.0086 \pm 0.0002$, $a_0 = -0.6737 \pm 0.0025$, $a_0 = 1.7332 \pm 0.0013$ and $a_0 = 0.0004 \pm 0.0001$. Figure 4 (a), indicates that the simple form of eq. (12) provides a good calibration model. For simulation purposes the kinetic parameters used for potassium dichromate are: $k_g = 9.565 \; (\mu m \, s^{-1})$ $g = 0.8$, $p = 1.235$ and $\gamma = 0.0075 \; \mu m^{-1}$, which were obtained by using the model identification procedure described in [9], based on experimental results with mono-modal seed. Figure 4 (b) shows the comparison between the simulated and experimental CSDs, when the bimodal seed was used. The simulated and the experimental data are in good agreement, however the distribution sizes are slightly over predicted using the model. Figure 4 (c) shows the seed and product crystals obtained at the end of the experiment. The microscopic images also indicate that the final distribution of crystals is a mixture of two different sizes. The experimental results indicate that a desired multi-modal distribution may be possible to achieve by designing an appropriate seed blend from various fractions of sieved seeds.
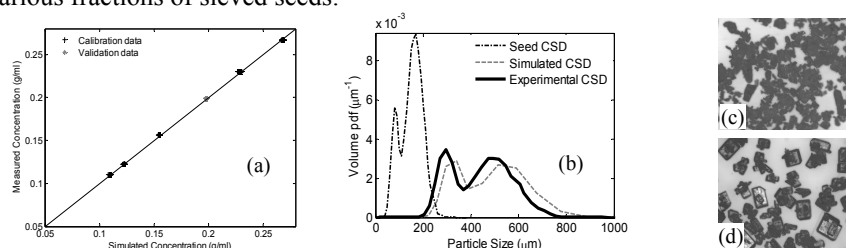
Figure 4: a) Estimation and validation of calibration parameters using measured and simulated concentration b) amount of seed (seed has a narrow distribution) c) microscopic image of the seed distribution and d) microscopic image of the target distribution.

6. Conclusions

The paper describes an approach for optimal seed recipe design for supersaturation controlled crystallisation processes, by determining the amounts in which seeds obtained from standard sieve analysis must be combined to achieve a desired shape of the product CSD. Another approach using dynamic seed addition to shape the CSD is also presented. The methodology of seed recipe design has been tested experimentally for potassium dichromate system. Microscopic image gives a good qualitative validation of the simulation results.

References

[1] M.J. Hounslow, G.K. Reynolds, AIChE J., 52 (2006) 2507.
[2] C. Wibow, K.M. Ng, AIChE J., 47 (2001) 2474.
[3] R.D. Braatz, 2002, AIChE J., 26 (2002) 87.
[4] M. Fujiwara, Z.K. Nagy, J.W. Chew, R.D. Braatz, J. Process Control, 15(2005) 493.
[5] S.M. Nowee, A.Abbas, J.A. Romagnoli, Chem. Eng. Sci., 46 (2007) 1096-1106.
[6] N. Kubota, N. Doki, M. Yokota, A. Sato, Powder Technol., 121 (2001) 31.
[7] N. Doki, N. Kubota, A. Sato, M. Yokota, Chem. Eng. J, 81 (2001) 313.
[8] E. Aamir, Z.K. Nagy, C.D. Rielly, T. Kleinert, B. Judat, Ind. Eng. Chem. Res., 48 (2009) 8575.
[9] E. Aamir, Z.K. Nagy, C.D. Rielly, Proceedings of 16th International Workshop on Industrial crystallisation, Finland, (2008) 61.

Development of a global mathematical model for reactive extrusion processes in corotating twin-screw extruders

Marie-A. de Ville d'Avray,[a,b] Arsène Isambert,[a] Stéphane Brochot[b]

[a]*Laboratoire de Génie des Procédés et Matériaux, Ecole Centrale Paris, Grande Voie des Vignes, 92295, Châtenay-Malabry, France, ma.devilledavray@caspeo.net*
[b]*CASPEO,3 avenue Claude Guillemin, BP 36009, 45060, Orléans Cedex 2, France, s.brochot@caspeo.net*

Abstract

Reactive extrusion involves complex interactions between operating parameters, flow conditions, material rheological behavior and reaction kinetics. Although reactive extrusion modelling has interested many authors, it still remains a challenge. We propose here a steady-state reactive extrusion model combining chemical engineering methods and simplified fluid mechanics laws. This steady-state model was derived from the dynamic model proposed by Choulak (2004). A rheo-kinetic model for a biopolymer oxidation process induced by coupled thermo mechanical and chemical effects was developed and integrated into the twin-screw extrusion model. This modelling approach enables to provide a predictive model involving very rapid calculation. The reactive extrusion model was then integrated into a static process simulator. The simulations reproduce available experimental data with a satisfying accuracy.

Keywords: reactive extrusion, modelling, simulation

1. Introduction

Reactive extrusion, which consists in using an extruder as a chemical continuous reactor to transform highly viscous materials, involves complex interactions between operating parameters, flow conditions, material rheological behavior and reaction kinetics. Moreover, reactive extrusion is generally conducted in corotating twin-screw extruders which are characterized by a modular screw profile including a succession of different screw elements with various shapes and functions. This modular structure complicates process understanding, design and control and thus reinforces the need to resort to process modelling and simulation. Although many authors have worked on reactive extrusion modelling, this still remains a challenge. Two categories of models are reported in literature (Vergnes and Berzin, 2006). The first group of models is based on chemical engineering methods and aims at representing the extruder as a series of ideal chemical reactors. Such models offer a global description of the system and provide rapid calculations, but they are generally limited to a narrow range of operational conditions and cannot be transposed from a screw configuration to another. And yet, the screw profile has a great impact on the efficiency of a desired reaction. The second modelling approach consists in solving the classical equations of fluid mechanics. These models are more accurate and predictive, but involve more calculation and are based on the assumption of a homogeneous reactive medium. Choulak et al. (2004) proposed a hybrid dynamic model combining the advantages of both methods, i.e. rapid calculation and improved flexibility and accuracy. The extruder is modeled as a series of perfectly

stirred tank reactors with possible backflows. Material flow rates between reactors are calculated using simplified one-dimensional fluid mechanics equations. The only adjustable parameter is the number of reactors, which has to be determined from an experimental residence time distribution (RTD) measurement.

The dynamic reactive extrusion model proposed by Choulak et al. (2004) was developed for process control. However, as far as process development and off-line optimization are concerned, steady-state models are more appropriate. Considering the advantages of the approach adopted by Choulak et al. (2004) in terms of rapidity of calculation and flexibility, we intended to derive a steady-state model from it. The model was integrated into the static process simulator USIM^{TM} PAC (Brochot et al., 2002).

A few modifications were brought to the initial model. Although the model developed by Choulak et al. (2004) enables to describe the main effects of operating parameters on the residence time distribution, the mean residence times are slightly underestimated by the model. This is notably due to a simplification of flow description in kneading blocks. We therefore modified the way of modelling material flows in kneading blocks, by integrating the experimentally observed upstream/downstream flow subdivision in such elements. The relative value of upstream and downstream flow components depends on the staggering angle between successive kneading discs. We also added a solid transport model for the first screw elements, and we developed rheo-kinetic model for a biopolymer oxidation induced by coupled thermo mechanical and chemical effects.

2. Steady-state mathematical model for reactive extrusion

Reactive extrusion is generally conducted in corotating twin-screw extruders. Material conveying is provided by the relative movement of the screws which act as a positive displacement pump, so that screw channels do not need to be totally filled to ensure material conveying. Moreover, corotating twin-screw extruders are characterized by a modular screw profile. The screws are composed of a succession of screw elements of different shapes and functions. Some of these elements are referred to as "restrictive elements", because their rotation creates a resistance to material flow. To cross such elements, the processed material has to be subjected to a rise in pressure. The extruder can be divided into three functional sections: the solid transport zone, the melt zone and the die. The desired reaction is assumed to occur in the melt zone.

2.1. Flow modeling

The melt zone is modeled by a series of n perfectly stirred tank reactors. Material flows between reactors are represented as the combination of two components: "shear flows" created by the relative movement of the screws, and "pressure flows" induced by axial pressure gradients in the neighborhood of restrictive elements. The expressions of the different types of flows are reported in Table 1, in which the coefficients K_d, K_r, K_p and K_{fp} are functions of geometrical parameters (Choulak et al., 2004). The coefficients x_{av} and x_{am} are used to take into account the lowered conveying capacity of kneading blocks compared to classical screw elements and the possible flow subdivision into downstream and upstream components respectively in kneading blocks (Carneiro et al., 1999). In classical direct screw elements, $x_{av}=1$ and $x_{am}=0$.

Development of a global mathematical model for reactive extrusion processes in corotating twin-screw extruders

Table 1. Expressions of material volumetric flow rates between reactors

Screw element represented by reactor i	Flow direction	Flow type	Flow rate expression
Direct screw or kneading block ($1 \leq i < n$)	Downstream	Shear	$F_{d,i}^{av} = x_{av,i} K_{d,i} N y_i V_i$
		Pressure	$F_{r,i}^{av} = x_{am,i} \dfrac{K_{r,i}}{\eta} (P_i - P_{i+1})$
	Upstream	Shear	$F_{d,i}^{am} = x_{am,i} K_{d,i} N y_i V_i$
		Pressure	$F_{r,i}^{am} = x_{av,i} \dfrac{K_{r,i}}{\eta} (P_i - P_{i-1})$
Reverse screw ($1 < i < n$)	Downstream	Pressure	$F_{r,i} = \dfrac{K_{r,i}}{\eta} (P_i - P_{i+1})$
	Upstream	Shear	$F_{d,i} = K_{d,i} N y_i V_i$
		Pressure	$F_{p,i} = \dfrac{K_{p,i}}{\eta} (P_i - P_{i-1})$
Last before die ($i=n$)	Downstream	Pressure (die flow)	$F_{fp} = \dfrac{K_{fp}}{\eta} (P_n - P_0)$
	Upstream	Shear	$F_{d,n}^{am} = x_{am,n} K_{d,n} N y_n V_n$
		Pressure	$F_{r,n}^{am} = x_{av,n} \dfrac{K_{r,n}}{\eta} (P_n - P_{n-1})$

N=screw rotation speed (rpm), y_i= filling ratio of reactor i, V_i=free volume of reactor i (m^3), η=viscosity (Pa.s), P_i=material pressure in reactor i (Pa)
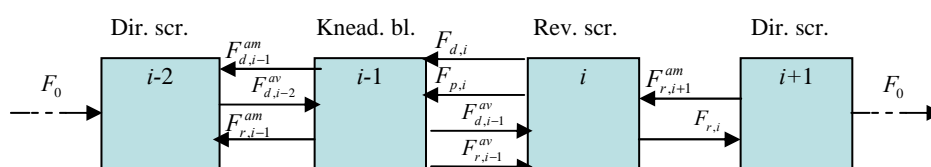
Fig. 1. Flow modeling in the melt zone

The calculation of reactors filling ratio, material pressure and flow rates is based on the two following assumptions:

1. At the steady-state, the net downstream flow rates between successive reactors are equal to the feed flow rate entering the melt zone (F_0).
2. A reactor is partially filled if and only if its internal pressure is equal to atmospheric pressure (P_0).

Let us consider the restrictive screw element represented by reactor i on Fig. 1. At the steady-state, the net flow rate leaving reactor i is equal to F_0 (Eq. (1)).

$$F_{r,i} - F_{r,i+1}^{am} = \frac{K_{r,i}}{\eta}(P_i - P_{i+1}) - \frac{K_{r,i+1}}{\eta}(P_{i+1} - P_i) = F_0 \qquad (1)$$

Hence,

$$P_i - P_{i+1} = \frac{F_0}{\dfrac{K_{r,i}}{\eta} + \dfrac{K_{r,i+1}}{\eta}} > 0 \tag{2}$$

As $P_i > P_{i+1}$, reactor i is totally filled. Assuming that following restrictive elements are partially filled (assuption based on experimental observations), P_{i+1} is equal to P_0 and we have:

$$P_i = P_0 + \frac{F_0}{\dfrac{K_{r,i}}{\eta} + \dfrac{K_{r,i+1}}{\eta}} \tag{3}$$

To calculate the filling ratio of reactor i-1, we first assume that reactor i-1 is partially filled. Its pressure is thus equal to P_0. At the steady-state, we have:

$$F_{d,i-1}^{av} + F_{r,i-1}^{av} - F_{d,i} - F_{p,i} = F_0 \tag{4}$$

Replacing the flow rates by their expressions we obtain:

$$y_{i-1} = \frac{F_0 + K_{d,i}NV_i + \left(P_i - P_0\right)\left(\dfrac{K_{p,i}}{\eta} + x_{am,i-1}\dfrac{K_{r,i-1}}{\eta}\right)}{x_{av,i-1}K_{d,i-1}NV_{i-1}} \tag{5}$$

If y_{i-1} is found to be greater than or equal to 1, this means that the assumption about the filling ratio of reactor i-1 was false. In this case, reactor i-1 is totally filled and its internal pressure is calculated so:

$$P_{i-1} = P_i + \frac{F_0 - x_{av,i-1}K_{d,i-1}NV_{i-1} + K_{d,i}NV_i}{x_{am,i-1}\dfrac{K_{r,i-1}}{\eta} + \dfrac{K_{p,i}}{\eta}} \tag{6}$$

The filling ratio and pressure of the previous reactors are calculated in the same way until the previous restrictive element (or the first element) is reached. The same reasoning is used for all the restrictive elements of the screw profile.

2.2. Modeling thermal, rheological and kinetic phenomena

The thermal, rheological and kinetic phenomena involved are coupled.

The rheo-kinetic model is here dedicated to a biopolymer oxidation reaction, which induces a decrease in the biopolymer mean molecular weight and an increase in carbonyl and carboxyl groups content. All these transformations are the result of combined effects of a chemical reaction (oxidant consumption) and thermo-mechanical constraints.

The oxidative depolymerisation induced by the oxidant (X) is described as a random scission of a polymer chain whose molar weight is x into two polymers chains P_1 and P_2 whose molar weight are respectively x' and x-x' (Eq.(7)).

$$P(x) + X \xrightarrow{\ k_1\ } P_1(x') + P_2(x - x') \tag{7}$$

However some of the oxidant molecules react with the polymer without inducing chain scission, contributing only to the increase in carbonyl and carboxyl content (Eq.(8)).

Development of a global mathematical model for reactive extrusion processes in corotating twin-screw extruders

$$P(x) + X \xrightarrow{k_2} P(x) \tag{8}$$

The polymer chains can also be degraded under thermo-mechanical effects (Eq.(9))

$$P(x) \xrightarrow{k_3} P_1(x') + P_2(x - x') \tag{9}$$

where k_3 is a function of temperature and shear rate.

The polymer number-average mass is calculated by applying the moment operation to the following population balance equation:

$$\frac{\partial p_i(x)}{\partial t} = 0 = -k_1 c_{ox,i} p_i(x) + 2\int_x^\infty k_1 c_{ox,i} \Omega(x,x') p(x') dx' - k_3(x) p_i(x)$$

$$+ 2\int_x^\infty k_3(x')\Omega(x,x')p(x')dx' + \frac{1}{y_i V_i}\left[Q_{out,i-1}p_{i-1}(x) + Q_{in,i}p_{i+1}(x) - (Q_{out,i} + Q_{in,i-1})p_i(x)\right] \tag{10}$$

where $p_i(x)dx$ is the molar concentration of polymer in the range $(x,x+dx)$, $c_{ox,i}$ is the oxidant content, $\Omega(x,x')$ is the stoechiometric kernel (McCoy and Madras, 1997) and $Q_{out,i}$ and $Q_{in,i}$ (m^3.s^{-1}) are the total volumetric flow rates oriented from reactor i to reactor $i+1$ and from reactor $i+1$ to reactor i respectively.

The carbonyl and carboxyl content are estimated from the quantity of oxidant consumed, which is obtained by performing a material balance in oxidant in a reactor i. The evolution of material temperature along the screw is calculated by an energy balance including convective heat transfer, conduction and viscous heat dissipation. A mean viscosity value is calculated as a function of mean molacular weight, temperature and shear rate in each reactor.

3. Model validation

3.1. Material and methods

The model was coded in FORTRAN and integrated into the simulation platform USIM PAC. The number of reactor to use in the model was determined by minimization of the error between an experimental and a simulated RTD. The number of reactors was set to 30 for all the simulations. Simulation results were compared to experimental data collected from biopolymer oxidation trials. Two series of data were used. The first one deals with the effect of the flow rate and screw rotation speed on the RTD. The second one is a study on the effect of the flow rate and initial oxidant content on the final functional groups content.

3.2. Results and discussion

The effects of flow rate and rotation speed on the RTD observed experimentally are well reproduced by the model (Fig. 2). The simulated final functional groups content as a function of initial oxidant content and feed flow rate also fit well with experimental data (Fig. 3). The number of experimental data used to adjust the kinetic parameters was significantly lower than the total number of extrusion experiments conducted, which enables to verify the predictive character of the model.

4. Conclusions and future works

The steady-state reactive extrusion model proposed here has been shown to reliably reproduce the available experimental results. Although further experimental validation should be carried out, the model already seems very promising. Its flexible structure enables to envisage its use in various applications. Work is being carried out currently

to simulate a crosslinking reaction. Moreover, model availability in the simulation tool USIM[TM] PAC makes it possible to simulate whole processes including upstream and downstream unit operations.
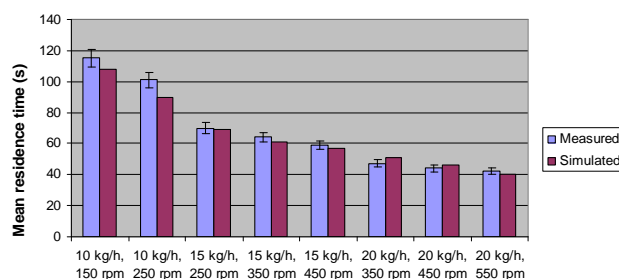
Fig. 2. Comparison of measured and simulated mean residence times
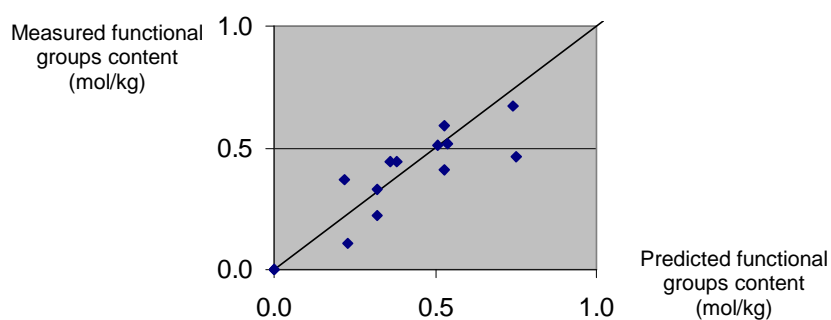
Fig. 3. Comparison of measured and simulated final functional groups content

5. Acknowledgements

References

S. Brochot, J.Villeneuve, J.C. Guillaneau, M.V. Durance and F. Bourgeois, 2002, USIM PAC 3: Design and Optimization of Mineral Processing Plants from Crushing to Refining. Mineral Processing Plant Design, Practice and Control, Ed. A.L. Mular, D.N. Halbe & D.J. Barratt, SME, pp 479-494.

O. S. Carneiro, G. Caldeira and J.A. Covas, 1999, Flow patterns in twin-screw extruders, Journal of Materials Processing Technology, 92-93, pp 309-315

S. Choulak, F. Couenne, Y. Le Gorrec, C. Jallut, P. Cassagnau, A. Michel, 2004, Industrial & Engineering Chemistry Research, 43, 23, pp 7373-7382

B.J. McCoy, G. Madras, 1997, Degradation kinetics of polymers in solution: dynamics of molecular weight distributions, AIChE J., 43, pp 802-810

B. Vergnes and F. Berzin, 2006, Modelling of reactive systems in twin-screw extrusion: challenges and applications, C.R. Chimie, 9, pp 1409-1418

A mathematical model of a fluidized bed combustor coupled with a Stirling engine

Francesco Miccio

Istituto Ricerche sulla Combustione CNR, P.le Tecchio 80, 80125 Napoli, Italy, miccio@irc.cnr.it

Abstract

The paper reports on a mathematical model of a fluidized bed combustor coupled with a Stirling engine for co-generation purposes. It consists of three fundamental blocks describing i) the heat transfer, ii) the fluidized bed combustor, and iii) the Stirling engine. The model produces as relevant outputs the bed temperature and the efficiency of the Stirling engine, at changing the operating conditions and geometrical parameters. The model results indicates that the proposed "integrated system" is of interest for micro-scale co-generation from low grade fuels.

Keywords: fluidized bed, Stirling engine, biomass, co-generation, heat transfer

1. Introduction

The biomass fuels have intrinsic CO_2 neutral impact on the environment, independently of the process adopted for energy production. The combustion is the most suitable option for its simplicity of implementation, in particular at small scale. The Stirling engine (Walker, 1980) has been conceived and developed since the 19th century, although having never reached a high level of penetration in the energy production systems. Main advantages are the smoothness, the flexibility toward the external heat source and the high thermodynamic efficiency. Coupling a burner with Stirling engines is an interesting option for micro-scale co-generation (10-50 kW$_{th}$, 1-10 kW$_e$). Unfortunately, dirtying/fouling problems arising from deposition of unburned species and ashes on the heat exchanger are likely to occur, in particular when low grade fuels are considered. As consequence, the system reliability is affected and the performances worsened.

The fluidized bed (FB) combustion boasts many advantages, as the flexibility toward the fuel characteristics and the good thermal properties (CTI, 1993). Adopting a fluidized bed combustor can contribute to overcome the problem of heat exchanger fouling as well as to assure very high heat transfer coefficients with the Stirling engine. Also in presence of complex tube arrays, the abrasion exerted by the bed material contributes to keep clean the external surface of the tubes and to remove residues locked in the interstices.

The present research provides a contribution in modeling an integrated system that is basically composed of a fluidized bed combustor and a Stirling engine for co-generation purposes.

2. Modeling

The mathematical model of the fluidized bed allows to compute the efficiency of biomass fuel combustion in the bed, as well as the heat transfer rate between the bed and a tube exchanger of a Stirling engine, considering both external and internal resistances. The Stirling engine is modeled adopting a lumped approach, according to the constitutive equations available in the literature that provides the engine efficiency and power output at changing upper/lower temperatures and angular speed. The model predicts the dynamic response of the system. The efficiency in energy conversion can be estimated at changing of the operating conditions for different schemes of system implementation and geometrical parameters.

The numerical model consists of three fundamental blocks describing i) the heat transfer, ii) the fluidized bed combustor, and iii) the Stirling engine.

2.1. Heat transfer

Following the chapter 11 of "Fluidization Engineering" [Kunii and Levenspiel, 1991], the general equation for computing the heat transfer coefficient h in a bubbling fluidized bed is given by Eq. 1.

$$h = \delta_w \left(h_r + h_g \right) + \left[\cfrac{1 - \delta_w}{\cfrac{1}{h_r + \cfrac{2k_{ew}}{d_p} + \alpha_w c_{pg} \rho_g u_0} + \cfrac{1}{h_p}} \right] \tag{1}$$

The heat transfer rate from the bed to a generic immersed surface S_w is calculated via Eq. 2.

$$Q = h S_w \left(T - T_w \right) \tag{2}$$

The explanation of symbols and parameters of Eq. 1 can be found in the cited source [Kunii and Levenspiel, 1991]. The whole procedure for the calculation of the heat transfer coefficient has been implemented in a MS-Excel worksheet. Figure 2 shows the results in terms of h as a function of the bed particle size d_p, at three bed temperature (973, 1073, 1173 K). The results of Fig. 1 indicate that the bed particle size exerts a larger influence on the heat transfer, small particles being recommended.

A mathematical model of a fluidized bed combustor coupled with a Stirling engine
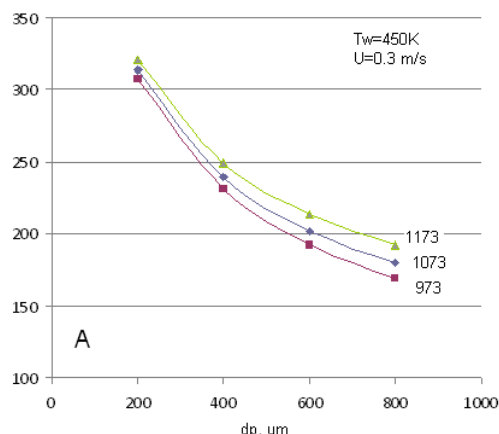
Fig. 1. Heat transfer coefficient in fluidized bed as function of the particle size and temperature

2.2. Fluidized bed combustion

A small-scale fluidized bed combustor (300 mm ID) is considered as combustion chamber for a solid biomass fuel, that is directly fed into the fluidized bed at a rate depending of the assigned excess air ratio e. The fluidization velocity U is also given in excess with respect to the minimum fluidization velocity U_{mf}. Due to the fast release of volatile matters and imperfect mixing not full combustion is achieved in the bed. Therefore, an overall combustion efficiency in the bed η_c is introduced, which mainly depends on the fluidization velocity and bed temperature. According to studies on the combustion of high-volatiles fuels [Scala and Salatino, 2002], a reasonable expression for η_c is given by Eq. 3.

$$\eta_c = 1 - c_1 \frac{U}{U_{mf}} T \exp\left(-\frac{T}{c_2}\right) \tag{3}$$

Figure 2 shows the asymptotic increasing dependence of the efficiency with the temperature as well as its decay with the fluidization ratio U/U_{mf}, as predicted by Eq. 3.
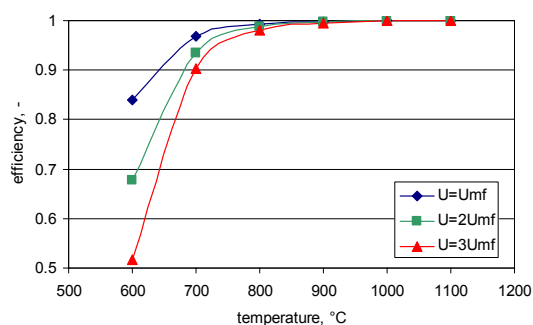
Fig. 2. Combustion efficiency in the bed as function of the temperature and fluidization velocity

Adopting a lumped approach, the unsteady energy balance for the fluidized bed is expressed by Eq. 4,

$$m_b c_{pb} \frac{dT}{dt} = AU\rho_g c_{pg}(T_0 - T) + S_l K_t(T_0 - T) + hS_{SE}(T_{SE} - T) - m_f \Delta H_c \tag{4}$$

where the terms at the right side, listed in the order as they appear, are: the enthalpy of the fluidizing gas flow-rate $AU\rho_g$, the heat dispersed at the lateral wall S_l, the heat exchanged with the Stirling engine throughout the surface S_{SE}, and the enthalpy of combustion.

2.3. Stirling engine

The displacement and the internal pressure of the Stirling engine are assigned (97cm^3 and 10Mpa). The tube exchanger of the Stirling engine is at direct contact with the fluidized bed, its surface being assigned (0.543m^2). The performance of the Stirling engine is predicted by using the software tool "StirlingPro.exe" developed by F. Normani (2009). The software accounts for different engine types and sizes, allowing for the calculation of the engine power, efficiency, and torque for assigned geometrical parameters and physical properties (e.g. pressure, upper temperature, lower temperature, fluid density/viscosity). The SW program has been executed more times in order to produce engine data-sheets as those reported in Fig. 3, showing the dependence of the mechanical power W_{SE} (A) and efficiency η_{SE} (B) on the angular speed ω for different upper temperatures. By interpolation of computed data, the 3rd order fitting equations (Eqs. 5 and 6) are obtained.

$$W_{SE} = (b_{3.1} + b_{3.2}T)\omega^3 + (b_{2.1} + b_{2.2}T)\omega^2 + (b_{1.1} + b_{1.2}T)\omega + b_{0.1} + b_{0.2}T \tag{5}$$

$$\eta_{SE} = (d_{3.1} + d_{3.2}T)\omega^3 + (d_{2.1} + d_{2.2}T)\omega^2 + (d_{1.1} + d_{1.2}T)\omega + d_{0.1} + d_{0.2}T \tag{6}$$
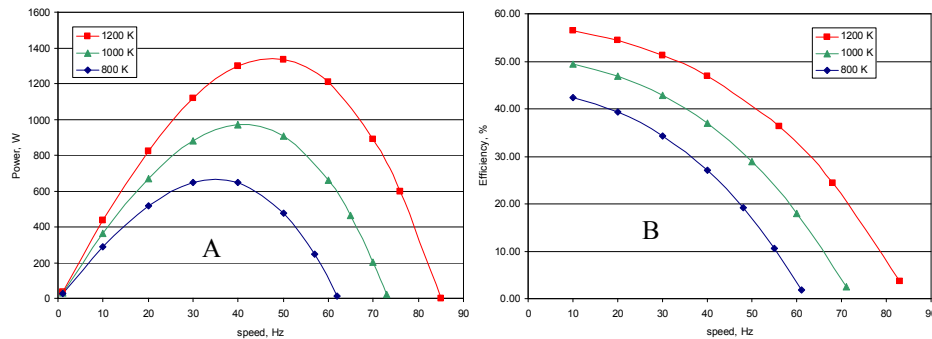
Fig. 3. Mechanical power (A) and efficiency (B) of a Stirling engine as function of the angular speed and upper temperature (Helium, P=10Mpa, T_{low}=25°C)

2.4. Computation technique

The ordinary differential equation (Eq. 4) is numerically integrated over the time for an assigned initial value of the bed temperature. For each time step, an iterative procedure is adopted for equating the heat transfer rate to the SE calculated via Eq. 2 to the ratio W_{SE}/η_{SE}, both terms being dependent on T_{SE}.

A mathematical model of a fluidized bed combustor coupled with a Stirling engine

3. Results

Figure 4 shows the time dependent profiles of the bed temperature, upper temperature in the Stirling engine and combustion efficiency as evaluated by the present model. The bed temperature increases steadily in the left side of the diagram thanks to the excess of the combustion enthalpy and approaches the asymptotic final value in the right side of the diagram. The upper temperature T_{SE} closely follows T_{bed}.
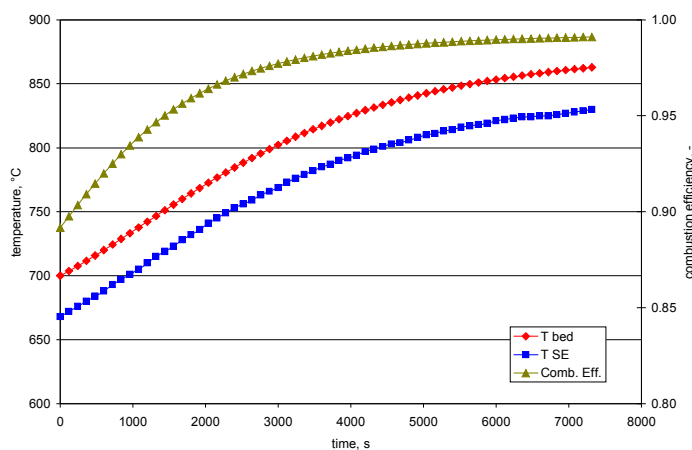
Fig. 4. Bed temperature, upper temperature of the Stirling engine and combustion efficiency as function of the time for a standard set of operating variables (U=0.33 m/s, e=2.0, ω=50Hz)
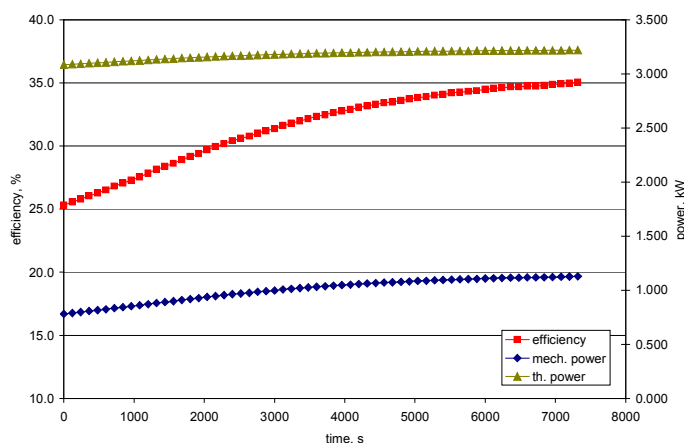
Fig. 5. Mechanical efficiency, thermal power and mechanical power of the Stirling engine as function of the time for a standard set of operating variables (U=0.33 m/s, e=2.0, ω=50Hz)

Figure 5 shows the time dependent profiles of the mechanical efficiency, thermal power and mechanical power of the Stirling engine. As consequence of the bed temperature increase (Fig. 4), both η_{SE} and W_{SE} increases, approaching asymptotic values as the in

temperature stabilizes. The thermal power exchanged between the bed and the Stirling engine is rather constant over the time.

Table I. Model results

			1	2	3	4	5
I	Room temperature	°C	25	25	25	25	25
I	Combustor size	m	0.3	0.3	0.3	0.3	0.3
I	Minimum fluidization velocity	m/s	0.1	0.1	0.1	0.1	0.1
I	Air flow rate	m3/h(25°C)	26	26	26	26	30
I	Air temperature	°C	25	25	200	25	25
I	Stoichiometric coefficient	kg/kg	6.24	6.24	6.24	6.24	6.24
I	Excess air ratio		2	1.8	2	2	2
O	Fuel flow rate	kg/h	2.46	2.73	2.46	2.46	2.84
I	Fuel heating value	KJ/kg	17000	17000	17000	17000	17000
O	Nominal power	kW	11.61	12.90	11.61	11.61	13.39
O	Bed temperature	°C	870	1000	1037	984	934
O	Fluidization velocity	m/s	0.392	0.436	0.449	0.431	0.477
I	Bed height	m	0.3	0.3	0.3	0.3	0.3
O	Bed mass	kg	27.6	27.6	27.6	27.6	27.6
O	Combustion efficiency		0.992	0.999	0.999	0.999	0.997
I	Surface of SE exchanger	m2	0.543	0.543	0.543	0.543	0.543
I	SE angular speed	Hz	50	50	50	30	50
O	SE upper temperature	°C	837	966	1004	961	901
O	SE efficiency	%	35.47	43.21	45.50	52.79	39.31
O	SE thermal power	kW	3.22	3.29	3.30	2.19	3.26
O	SE mech. Power	kW	1.14	1.42	1.50	1.16	1.28

Table I reports the results of calculations carried out at steady state. Selected input variables of the model are marked by I in the first column, whereas O is used for outputs. The changes of the input variable with respect to the case 1 are marked by yellow. The model congruently predicts the effect of relevant input variables: the bed temperature increases as the excess air decreases (case 2), the air preheating occurs (case 3), the SE angular speed is lowered (case 4), and the air flow rate increases (case 5). Increasing T_{bed} is beneficial for the SE efficiency; the latter is much higher at $\omega=30$ Hz, by virtue of the decreasing dependence reported in Fig. 3B. The mechanical power, in the range $1.1 - 1.5$ kW, is maximized at the highest bed temperature (case 3).

4. Conclusions

An integrated model of fluidized bed combustion coupled with a Stirling engine has been developed. The model provides as relevant outputs the bed temperature and the efficiency of the Stirling engine, at changing the operating conditions and geometrical parameters.

Realistic values of the mechanical power generated by the engine are in the range 1.0-2.0 kW for a 20 kW$_{th}$ FB combustor. These results indicates that the proposed "integrated system" is of interest for micro-scale co-generation from low grade fuels.

References

F. Walker, Stirling Engines, Oxford University Press, 1980

1. CTI Comitato Termotecnico Italiano: La combustione in letto fluido: Principi e tecnologie. CTI, Milano, 1993

D. Kunii, O. Levenspiel, Fludization Engineering. Butterworth-Heinemann, Boston, 1991

F. Scala, P. Salatino, Modelling fluidized bed combustion of high-volatile solid fuels, Chem. Engng Sci., 57 1175, 2002

F. Normani, Stirling Engine Design Program, http://newenergydirection.com/blog/2009/03/stirling-engine-design-program, 2009

Oxidative Coupling of Methane: Reactor Performance and Operating Conditions

Stanislav Jašo,[a] Hamid Reza Godini,[a] Harvey Arellano-Garcia,[a] Günter Wozny [a]

[a]Belin Institute of Technology, Strasse des 17. Juni 135, 10618 Berlin ,
stanislavjaso@mailbox.tu-berlin.de

Abstract

In this work, the achievable performance in case of desired-product selectivity, yield and conversion are evaluated systematically for different reactors in order to find the optimum range of operation for the OCM process. This approach is applied to a nonisothermal plug flow reactor and a nonisothermal porous packed bed membrane reactor using different types of catalysts in the wide range of operating conditions. Moreover, a fluidized bed reactor is also considered. The results show that tracking the optimum area of operation has a monotonic direction under some range of operating conditions, whereas it reflects a qualitative trade-offs under some other ranges of operating conditions. For all investigated reactor concepts the likelihood of optimal operating conditions are found, and the best corresponding performance for all of them are reported.

Keywords: Oxidative coupling of methane, reactor design, membrane reactor, fluidized bed reactor

1. Introduction

The production of C_2 hydrocarbons (i.e. ethylene + ethane) through the oxidative coupling of methane (OCM) has been prohibited from commercial practice due to low overall yield. In particular, today's catalysts exhibit either high selectivity (>70%) coupled with low conversion (<5%) or high conversion (>75%) with low selectivity (<15%) [1]. This implies that an optimum ratio of conversion - selectivity exist, in order to achieve the highest possible yield of C_2 hydrocarbons.

Many different reactor concepts were proposed for the oxidative coupling of methane, for instance : counter current simulated moving bed reactor [2], solid oxide fuel cell reactor [3], catalytic dense membrane reactor [3,4], fluidized bed reactor [5,6], porous membrane rector [7,8], fixed bed reactor [5]. All of those reactor concepts are having their advantages and drawbacks. However, only the last three reactor types have the potential to be exploited industrially. Therefore only those three reactor types will be investigated in this study.

Fixed bed reactor represents a state of the art in the industry, and has to be examined in detail. Several studies on this reactor type were done, and the biggest drawback are severe hot spots, formed as a consequence of the poor heat removal from the highly exothermal reaction. Feed dilution is necessary if any application is expected.

Fluidized bed reactor has been investigated only briefly [5,6,9], and it showed similar performance like the fixed bed reactor. The biggest advantage of the fluidized bed reactor is the isothermal operation and a possibility to operate using undiluted feeds [6,9]. This option is very attractive for industry (even in the case of lower selectivity) because costly separation of nitrogen and methane is prevented, and the downstreaming process is simplified.

Membrane reactor concept allows fine distribution of one reactant into the reactor, using a porous membrane as a gas distributor. Fine oxygen distribution has been proven to give enchased selectivity even with significant methane conversion. However, feed dilution and reactor size are main drawbacks of this reactor type. Feed dilution is necessary because of the safety reasons as well as for low heat transfer rates. Reaction rates are approximately one order of magnitude smaller compared to fixed and fluidized bed reactor, therefore much bigger reactor is needed to achieve same conversion level.

In this work three reactor concepts are compared using a same graphical approach for obtaining information about their optimal performance. Moreover same kinetic model [10] for catalyst La_2O_3/CaO is used in order to give comprehensive comparison of these reactor concepts.

2. Modeling

2.1. Fixed bed reactor

Fixed bed reactor was simulated using a standard pseudo homogeneous plug flow reactor model :

$$\frac{dF_i}{dl} = r_{c,i} \, A \tag{1}$$

Where F_i (mol/s) is the molar flow-rate for component i and A (m^2) is the cross sectional area of the tubular reactor. Both isothermal and nonisothermal simulations were performed with a typical 2:1 methane to oxygen feed ratio and a reactor of 7 mm diameter, with a flow rate of 10 cm$_n^3$/s. Additionally, nonisothermal simulation was done adding a differential reactor heat balance into the system of equations.

2.2. Membrane reactor

Membrane reactor model is based on pseudo-homogeneous one dimensional flow. No radial and axial diffusion was taken into account. This was possible only due to the small reactor diameter of 7 mm used in the simulation. The reactor model equations are:

$$\frac{dF_i^t}{dl} = r_{c,i} \, A_t - N_i \, \pi \, d^t \qquad (Tube \ \ side) \tag{2}$$

$$\frac{dF_i^s}{dl} = N_i \, \pi \, d^t + r_{g,i} \, A_s \qquad (Shell \ \ side)$$

Where F_i (mol/s) is the molar flow-rate for component i in the tube and the shell side of the membrane reactor, and A (m^2) is the cross sectional area of the tubular reactor. N_i represents the mass transport of the components through the membrane. Reaction occurs in both tube and shell side of the reactor. More details of this model can be found in [11]. In case of membrane reactor, heat balance equations similar to Eq. (2) were added in order to simulate nonisothermal reactor behavior similarly like for the plug flow reactor.

2.3. Fluidized bed reactor

Fluidized bed reactor was simulated using a modified two phase model suggested by Werther et al.[12] :

$$\left[u - u_{mf} \, (1 - \varepsilon_b) \right] \frac{dC_{b,i}}{dh} + k_{g,i} \, a \, (C_{b,i} - C_{e,i}) = \varepsilon_b \, r \, (C_{b,i}, T, \varepsilon_{bb}) \tag{3}$$

$$\left[u_{mf} \, (1 - \varepsilon_b) \right] \frac{dC_{e,i}}{dh} = k_{g,i} \, a \, (C_{b,i} - C_{b,i}) + (1 - \varepsilon_b) \, r \, (C_{e,i}, T, \varepsilon_e)$$

Oxidative Coupling of Methane :Reactor Performance and Operating Conditions

Here C_i (mol/m^3) represents the concentration of component *i* in bubbles and emulsion, ε represents porosity, $k_{g,i}$ is the mass transfer between bubbles and emulsion and r_i is reaction rate. Diameter of the simulated fluidized bed reactor is 40 mm, for the Geldart A group of particles of 110 µm diameter. The range of operation of the fluidized bed reactor lies between fluidization numbers of 3-15, therefore in the bubbling regime. More details about the reactor model applied in this study can be found in [12]

3. Results and discussion

Fixed bed reactor has been simulated for the isothermal and nonisothermal case in order to investigate the achievable performance of this reactor type. Only two variables are influencing the performance of this reactor type: inlet composition which can be represented as inlet ratio of methane to oxygen, and temperature of the feed. For the isothermal case temperature in the reactor was kept constant and equal to the inlet temperature. For the nonisothermal simulation this temperature was only an inlet boundary condition, and the heat balance determined temperature at other points in the reactor. The simulation results for the isothermal case are represented in the figure 1.
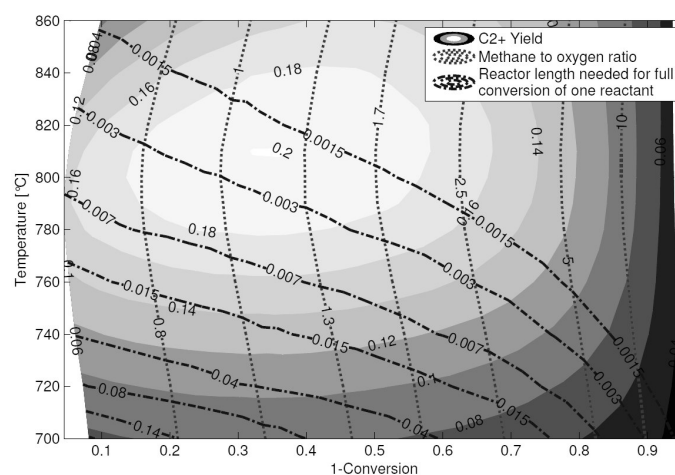
Fig. 1 Performance of the isothermal fixed bed reactor

As can be seen from the Fig.1, there is a range of optimal operating conditions for this reactor type. These operating conditions are temperature of 810°C and inlet methane to oxygen ratio of 1.2. The maximum yield of higher hydrocarbons obtained by simulation accounts for 20% as reported by experiments as well [5].

Nonisothermal simulation shows the necessity to use diluted feeds in order to efficiently remove heat from the reactor. Wall heat transfer coefficient was calculated by correlation of Li and Finlayson (1977), (260 W/mK) with environment temperature of fixed to 600 K. When environment temperature was kept same like the inlet temperature (like for the oven), temperature runaway was inevitable for any applied condition.

It can be concluded that significant dilution of the catalyst bed and dilution with nitrogen is required in order to avoid hot-spot formation in the fixed bed reactor. On the other hand where dilution of the feed or the catalyst surpasses certain critical value, hot spots do not appear and temperature in the reactor drops monotonously. Here we can emphasize that controlling the temperature in the fixed bed is almost an impossible task. Reason for this is very fast and exothermal reaction, which is self-accelerating when heat balance is disturbed. Consequently hot spots of more than 300K are observed.
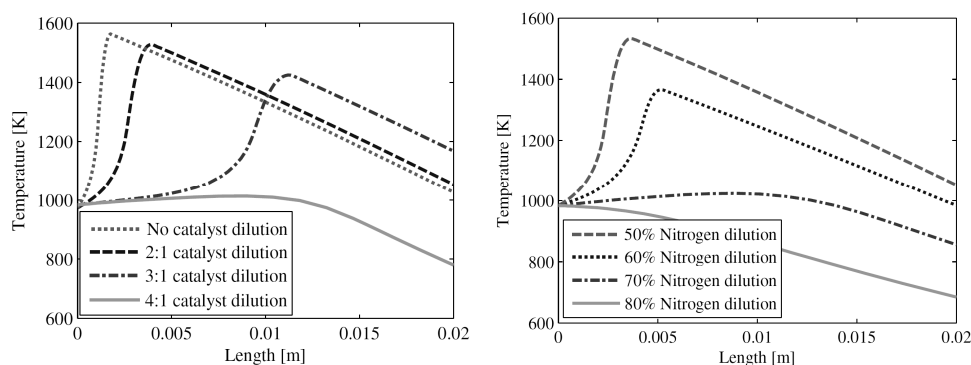
Fig. 2 Influence of catalyst dilution (left) and dilution by nitrogen (right) on the hot spot formation in the fixed bed reactor

Similar investigations were conducted for the membrane reactor. Several variables such as flow of the feed in the tube side and in the shell side were manipulated, as well as oxygen concentration in the shell, temperature and the membrane thickness. We concluded that the thickest membrane investigated (0.1 mm) provided the best overall reactor performance. Moreover, equal flow rate of the tube and the shell stream allowed higher yield in comparison to other investigated flow rates. Influence of the shell oxygen fraction and the inlet temperature on the yield can be seen on the figure 3.
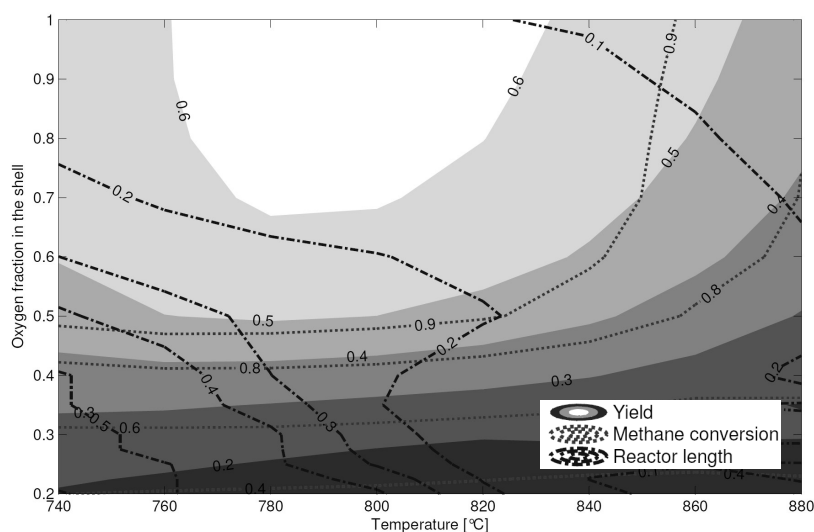
Fig. 3 Influence of the temperature and the oxygen ratio on the performance of the porous membrane reactor

Porous membrane reactor gave remarkably high yields of over 50%. It should be noted that there are no experimental investigations for use of undiluted feeds and such high membrane thickness to testify whether simulated results are reasonable or not. Highest yield achieved so far in a porous membrane reactor is 27.5 % obtained by Lu et al [8].

Nonisothermal simulation for the porous membrane reactor showed some disadvantages for membrane reactor application. In most cases temperature runaway was observed, similarly like for the fixed bed reactor. However, because of the slower reaction rate in the membrane tube, it was possible to maintain hot spot formation below desirable level

by manipulating the feed dilution and membrane thickness (fig. 4). During this investigation, tube and shell flow rate were kept constant – 4 cm_n^3/s with a stoichiometrical ratio of methane/oxygen and inlet temperature of 750°C.

It can be noticed that membrane reactor has to be operated with significant feed dilution and with a thick membrane in order to provide isothermal operation. This leads to a significant increase in reactor size required for necessary conversion level.

— ∎ — ∎ Tube temperature (50% Dilution)
∎ ∎ ∎ ∎ Shell temperature (50% Dilution)
— — — Tube temperature (75% Dilution)
∙∙∙∙∙∙ Shell temperature (75% Dilution)
------- Tube temperature (83% Dilution)
∙∙∙∙∙∙∙ Shell temperature (83% Dilution)

-∙------ Tube temperature (d = 0.15 mm)
∙∙∙∙∙∙∙ Shell temperature (d = 0.15 mm)
—∙—∙— Tube temperature (d = 0.10 mm)
∙∙∙∙∙∙ Shell temperature (d = 0.10 mm)
∎ ∎ ∎ Tube temperature (d = 0.05 mm)
∎ ∎ ∎ ∎ Shell temperature (d = 0.05 mm)
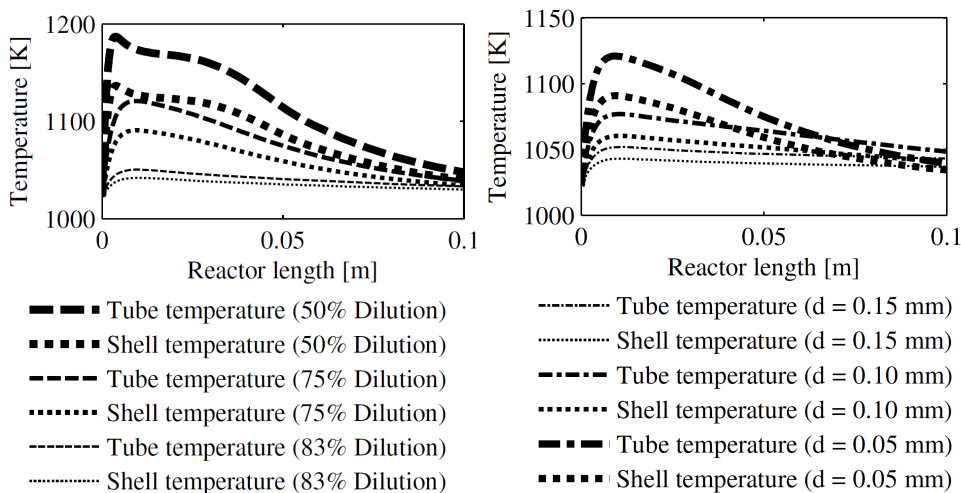
Fig. 4 Nonisothermal simulation of the membrane reactor : influence of the feed dilution (left, $\delta = 0.1$ mm) and membrane thickness (right, 75% dilution) on the hot spot formation

Finally, fluidized bed reactor was investigated under different operating conditions. Manipulation of the reaction temperature, feed composition and the fluidization velocity showed influence of these variables on the reactor performance. Increase in fluidization velocity showed monotonous increase in yield and selectivity, most probably because "reaction front" was moved away from the distributor zone to the bulk, and allowed bubble mass transfer to play an important role. For further simulation only the highest flow rate of 15 u_{mf} was applied. Results are shown on the figure 5.
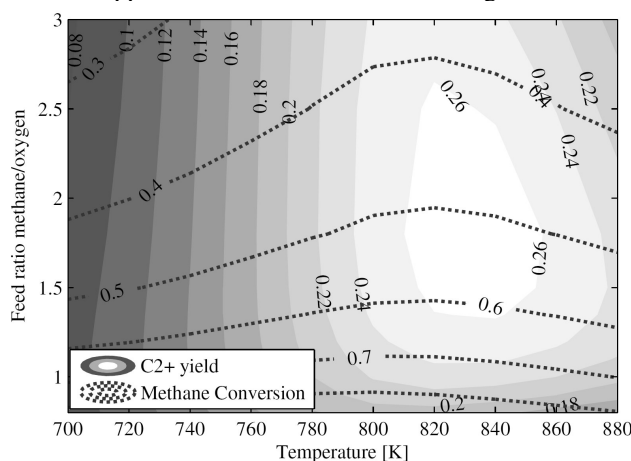
Fig. 5 Influence of the reaction temperature and feed composition on the performance of the fluidized bed reactor

The highest yield in the fluidized bed of 26% is in reasonable agreement with experiments [6,9]. Nonisothrermal simulation for the fluidized bed was not applied, because most of the authors of experimental studies on the OCM in the fluidized bed reactor have reported isothermal behavior, even when they used undiluted feed [6,9]. We can therefore adopt the fact that the fluidized bed reactor provides the possibility to operate under isothermal operation under all simulated conditions.

4. Conclusions

A model-based performance analysis of fixed bed, fluidized bed, and porous membrane reactor were conducted. Simulation results undoubtedly show that fixed bed reactor can not be used industrially, even when high heat transfer rates are applied. This reactor type has no control over the reaction heat in case of small dilution, and behaves almost ideally adiabatically. C_{2+} yields of 20% are not satisfactory for an industrial application. In contrast to this, the membrane reactor offers the possibility to increase the yield by fine oxygen distribution through the membrane. However, application of undiluted feeds and thin membranes result in hot spots of over 100 K. Even though a yield of over 50% yield is achievable for such conditions, these are only theoretical results. For an isothermal operation thick membranes and over 80% feed dilution are necessary. The fluidized bed reactor shows an improved performance in comparison to the fixed bed reactor. Although yields of 26% are still below industrial requirements, the use of undiluted feeds and an isothermal operation are the main advantages of this reactor type. A further economical analysis of the membrane reactor and/or the fluidized bed reactor including the downstreaming process is however necessary in order to answer definitively the question which is the most suitable OCM reactor concept. For this purpose, a mini-plant is currently being built at the Berlin Institute of Technology.

5. Acknowledgements

References

[1]   G.Hutchings, in E.E. Wolf (Ed.), Methane Conversion by Oxidative Processes : Fundamental and Engineering Aspects, Van Nostrand Reinhold, New York, 1992, pp. 200.

[2]   A.L. Tonkovich, R.W. Carr, R. Aris, R, Science 262 (1993) 221.

[3]   W. Kiatkittipong, T. Tagawa, S. Goto,S. Assabumrungrat, K. Silpasup, P. Praserthdam, Chem. Eng. J. , 115 (2005) 63

[4]   S. Bhatia, C.Y. Thien, A. R. Mohamed, Chem. Eng. J. 148 (2009) 525

[5]   L. Mlezko, M. Baerns, Fuel Processing Technology , 42 (1995) 217.

[6]   L. Mlezko, U. Pannek, Ind. Eng. Chem. Res. 35 (1996), 54-61

[7]   Y.K. Kao, L. Lei, Y.S. Lin, Catal. Today, 82 (2003) 255

[8]   Y. Lu, A.G. Dixon, W.R. Moser, Y.H. Ma, Chem. Eng, Sci. ,55 (2000) 4901.

[9]   K. T. Do, J. H. Edwards, R. J. Tyler, Can. J. Chem. Eng. 73 (1995) 327

[10] Z. Stansch, L. Mleczko, M. Baerns, Ind. Eng. Chem. Res. 36 (1997) 2568.

[11] H. R. Godini, H. Arellano-Garcia, M. Omidkhah, and G. Wozny, proceedings of The 19th European Symposium on Computer Aided Process Engineering (ESCAPE19), Cracow, Poland, 1, pp. (2009) 123

[12] Werther, J, Ind. Eng. Chem. Res. 43 (2004), 5593

Ultrafiltration of Surfactant Micelles: Cross-flow Experiments and Flux Modelling

Anja Drews [a,b], Harvey Arellano-Garcia [c], Reinhard Schomäcker [d], Matthias Kraume [b], Günter Wozny [c]

[a] *Chemical Engineering in Life Sci. Eng., HTW Berlin, Engineering II, Wilhelminenhofstr. 75a, D-12459 Berlin, Germany*
[b] *Chair of Process Engineering, Technische Universität Berlin, MA 5-7, Str. des 17. Juni 135, D-10623 Berlin, Germany*
[c] *Chair of Process Dynamics and Operation, Technische Universität Berlin, MA 5-7, Str. des 17. Juni 135, D-10623 Berlin, Germany*
[d] *Dept. of Chemistry, Technische Universität Berlin, TC3, Str. des 17. Juni 135, 10623 Berlin, Germany*

Abstract

Due to the high cost of catalysts containing optically active ligands and rare metals, their retention in continuous processes is extremely valuable. The vast majority of enantioselective homogeneous hydrogenations take place in organic solvents, but many can also be carried out in aqueous solvents if catalyst and substrate are "solubilised" by means of surfactant micelles. Micelles can then be separated from the emulsion using an ultrafiltration (UF) membrane. However, since MEUF has so far mainly been investigated for application in water treatment and lab experiments have been carried out in dead-end stirred cells, a screening for the most suited membrane under conditions that allow scale-up is still pending. In this work, we investigate the Micellar-Enhanced Ultrafiltration (MEUF) of surfactants solutions (SDS and Triton X-100) under technically more realistic cross-flow (CF) and steady state conditions. Moreover, for an efficient process design, a novel practical and reliable flux model based on the resistance-in-series model is proposed.

Keywords: Ultrafiltration, micelles, modelling, process intensification

1. Introduction

The use of catalysts in the production of fine chemicals and pharmaceuticals increases every day. Catalysts with optically active ligands result in a better selectivity in most of this type of reactions [Schwarze et al. 2008]. However, their price is very high. Therefore, catalyst recovery plays an important role so as to attain a useful and economically viable process. Entrapping the catalysts in micelles serves two purposes: Organic solvents can be replaced to a large extent by aqueous solutions because catalysts can be "solubilised" in them, and the micelle shell enlarges their size, so from that point of view, homogeneous catalysts are "heterogenised" and can be separated more easily. Membrane technology is gaining importance for a wide range of applications because of its low energy requirements and its good results in separation and selectivity in comparison with other unit operations. Micellar-enhanced ultrafiltration (MEUF) has been proposed for the recuperation and reuse of those valuable compounds. Micelles can be separated from the emulsion using an

ultrafiltration (UF) membrane (micelles diameters $\approx$ 1-100 nm) [Schwarze et al. 2008]. Most studies on suited membranes and conditions have so far been carried out using stirred – or sometimes even unstirred [Purkait et al. 2004] – dead-end cells and/or small feed solution volumes which resulted in a change of feed concentration over time, despite the fact that hydrodynamic conditions and unsteady states are well-known to affect permeability and retention. In addition, investigations have so far mainly focused on water treatment applications, e.g. to remove heavy metals [Li et al. 2006]. The aim of this study is thus to investigate MEUF under technically more realistic cross-flow (CF) and steady state conditions in order to enable representative membrane screening and to provide a framework for operating conditions optimisation.

2. Experimental Procedure

UF of two surfactants solutions (anionic: SDS; nonionic: Triton X-100) was carried out in a test cell using two membranes (C005 (regenerated Cellulose with a nominal MWCO of 5kD) and P010 (PES with a nominal MWCO of 10kD)) (see Fig. 1). Transmembrane pressure (TMP), temperature, cross-flow velocity (CFV) and feed concentration were varied (see Table 1). Feed volume was 10L to enable almost constant feed concentrations despite permeate withdrawal and sampling. Permeate flux development was determined by an electronic balance, and regular feed and permeate samples were taken and analysed for their surfactant concentration (SDS: conductivity, Triton X-100: spectrometry) to enable calculation of micelle retention or rejection.

High micellar retentions were achieved (>96% for C005 and >87% for P010) which were influenced especially by the transmembrane pressure and cross-flow velocity. Interestingly, different or even opposing effects of CFV, temperature and transmembrane pressure (TMP) on rejection were found for different membrane/surfactant combinations. Filtration of Triton X-100 through C005 was the optimum combination in terms of micelles rejection (>99%). In the presence of an organic solvent (dodecene), rejections were found to be in the same range.

Membrane surface area: 44cm^2
Channel height: 2.5 mm
Channel width: 15 mm
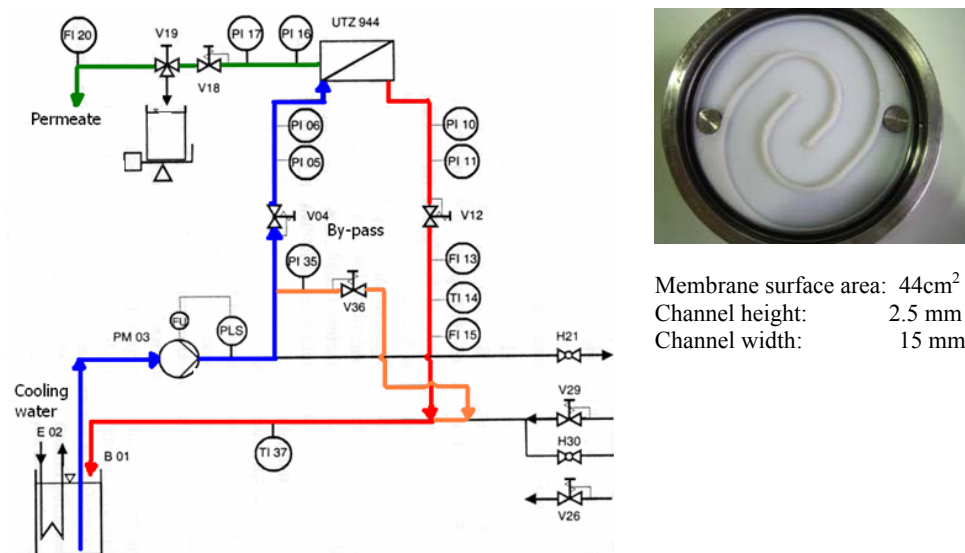
Fig. 1. Flowsheet of the cross-flow ultrafiltration and meander channels of the test cell.

Ultrafiltration of Surfactant Micelles: Cross-flow Experiments and Flux Modelling

Table 1: Range of experimental conditions

variable	symbol	experimental range	unit
Feed concentration	c_F	10…150	g/L
CF velocity	v	0.5…3	m/s
Temperature	T	20…30	°C
Transmembrane pressure	TMP	2…10	bar

Fig. 2 shows that in contrast to literature [Zaghbani et al. 2007] where CF filtrations were carried out but with only 200mL feed volume, i.e. even any permeate withdrawal or sampling would have increased the feed concentration, steady states are quickly achieved. This means that steady state models can be used to predict flux and rejection.
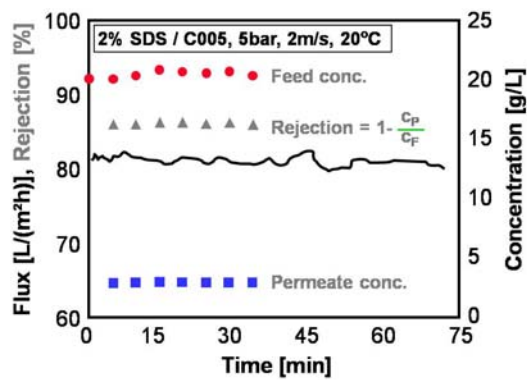
Fig. 2. Typical flux and rejection trajectories.

3. Modeling Approach

The developed model is based on the resistance-in series-model. The model assumes a constant overall hydraulic resistance to flow including the effect of surfactant fouling and native membrane resistance.

$$J = \frac{TMP}{\eta_P (R_T)} \tag{1}$$

The total resistance is the sum of the resistance due to the membrane (R_m) and due to the solute (R_s):

$$J = \frac{TMP}{\eta_P (R_m + R_s)} \tag{2}$$

The solute resistance can be divided into two parts: R_f, which is the fouling layer resistance due to specific solute-membrane interactions (adsorption), and R_p, the resistance due to concentration polarisation, which is a result of the applied pressure.

$$J = \frac{TMP}{\eta_P (R_m + R_f + R_p)} \tag{3}$$

$$J = \frac{TMP}{\eta_P \left(R_m + R_f + b \cdot TMP\right)} \tag{4}$$

First of all, the intrinsic membrane resistance must be determined from water flux experiments. Usually, to determine R_m, it is sufficient to plot the permeate flux against the applied pressure, and the slope, according to Eq. 5 is the inverse of the viscosity and the membrane resistance.

$$J = \frac{TMP}{\eta_P \left(R_m\right)} \tag{5}$$

Sometimes, as a result of membrane compaction the plot does not show a linear correlation, and in that case, the membrane hydraulic resistance increases with pressure as follows:

$$R_m = R_{mC} + a \cdot TMP \tag{6}$$

In order to obtain R_{mC} and a, R_m calculated with Eq. (5) is plotted against the applied pressure: the origin ordinate is R_{mC} and the slope is a. Once the values of the membrane resistance are obtained, the value of R_S is calculated based on Eq. 2. By plotting the values of R_S against pressure, we obtain the parameter b from the slope, and R_f from the interception. Both R_f and R_p were found to be influenced by cross-flow velocity v, especially R_p. The expression that best fits the experimental data for the parameter b is the following (see Fig. 3):

$$b = \frac{e}{v^\beta} \tag{7}$$
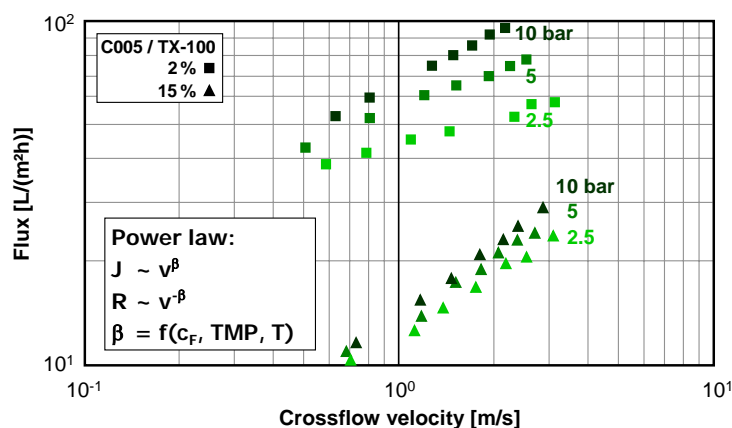
Fig. 3: Flux depency on crossflow velocity at different pressures

The resistance due to substrate fouling is also found to be influenced by cross-flow velocity. The simplest expression that relates both parameters, which agrees with experimental data, is a linear equation:

$$R_f = c \cdot v + d \tag{8}$$

Including both relationships in the model, the following expression is obtained:

$$J = \frac{TMP}{\eta_P (R_m + R_s)} = \frac{TMP}{\eta_P (R_m + R_f + R_P)} = \frac{TMP}{\eta_P \left(R_m + c \cdot v + d + \dfrac{e}{v^\beta} \cdot TMP \right)} \tag{9}$$

When the membrane resistance is not constant, the model is expressed by the equation below:

$$J = \frac{TMP}{\eta_P \left(R_{mC} + a \cdot TMP + c \cdot v + d + \dfrac{e}{v^\beta} \cdot TMP \right)} \tag{10}$$

When only one CFV is used, eq. (4) – in the case of compressible membranes together with eq. (6) – can be used by fitting the 3 (compressible: 4) parameters to experimental data (model 1). Eq. (10), on the other hand, contains more fitting parameters (6) but is also more generally valid since the same parameters can be used for each cross-flow velocity. The derived model is also a base model that can be extended with other terms and parameters, e.g. to include the influence of temperature and bulk concentration.

4. Results

In contrast to other studies [Markels et al. 1995] it was found that no osmotic pressure model is needed to describe the phenomena. Starting from a model which requires one set of fitted parameters per operating condition (M1, eqs. (4) and (6)), extensions were made as described above to account for physical phenomena such as CF velocity (M2). As can be seen in Fig. 4, both models fit experimental data well. The extended model, however, has many advantages over M1: only one parameter set is needed for all systems, and only two experiments at two different CF velocities are necessary. This is useful in plant design, e.g., to optimise CF velocity, i.e. achieving the desired permeate flux at minimal energy consumption. M2 also allows obtaining the different contributions to resistance, and to study their variation with TMP and CF velocity.
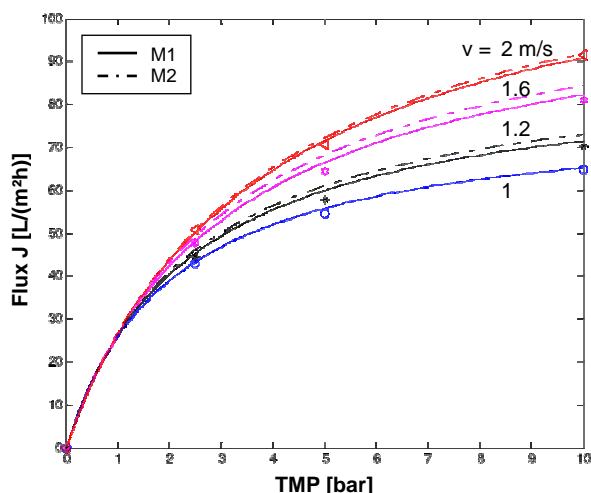
Fig. 4: Model comparison for 2 w-% Triton X-100 at T = 20°C

It was found that while for P010 the adsorptive fouling resistance R_f was always positive, it was typically negative for C005 (except for dilute SDS at low CFV). In other words, wettability of C005 was actually improved in the presence of surfactants.

5. Conclusions

Cross-flow ultrafiltration of surfactants solutions in a test-cell installation has been carried out in order to investigate the behaviour of the permeate flux under different operating conditions and to analyse the rejection of micelles and their affecting factors. Due to technically realistic cross-flow conditions and larger feed volumes, steady states were achieved more quickly then so far reported in literature. Ultrafiltration of surfactant solutions resulted in high rejections and high permeate flux, especially when working at higher pressures and higher cross-flow velocities. A temperature increase also has a positive effect on permeate flux.

A new model approach has been formulated with the integration of cross-flow velocity in the resistance-in-series model, which especially through concentration polarisation has a large influence on permeate flux. In this model, only one expression is needed for every system, and with only two experiments at two different cross-flow velocities the system can be characterised. This is useful in order to design a plant or to determine the optimal cross-flow velocity so as to achieve a desired permeate flux with minimal energy consumption. The proposed resistance-in-series model describes the ultrafiltration of surfactant solutions well, with no need of including the osmotic pressure model.

6. Acknowledgements

References

M. Schwarze, A. Rost, T. Weigel, R. Schomäcker, Chem. Eng. and Processing 2008, 48, 356-363.
J. H. Markels, S. Lynn, C. J. Radke, AIChE Journal 1995, 41 (9), 2058-2066.
N. Zaghbani, A. Hafiane, M. Dhahbi, Sep. Purif. Technol. 2007, 55, 117-124.
M.K. Purkait, S. DasGupta, S. De, J. Coll. Interface Sci. 2004, 270, 496-506.
C.-W. Li, C.-K. Liu, W.-S. Yen, Chemosphere 2006, 63, 353-358.

Modelling and simulation of extensional-flow units in emulsion formation

Cristhian Almeida-Rivera[a], Peter Bongers[a,b]

[a]*Structured Materials and Process Science Department, Unilever Research Vlaardingen, Olivier van Noortlaan 120, PO Box 114, 3130 AC Vlaardingen, The Netherlands, cristhian.almeida-rivera@unilever.com, peter.bongers@unilever.com*
[b]*Chemical Engineering and Chemistry, Eindhoven University of Technology, POBox 513, 5600 MB Eindhoven, The Netherlands, p.m.m.bongers@tue.nl*

Abstract

In this contribution we studied the emulsification process carried out in an extensional-flow unit. By means of rigorous population and momentum balances we captured the phenomenological description of the first principles occurring in such unit.
The strong feature of our model approach resides in the fully mechanistic description of the governing phenomena. Namely, a population balance equation was formulated and solved to account for the disappearance and appearance of droplets at each size class. Moreover, coalescence mechanism was included to account for the instability of newly created droplets. Additionally, our model estimated the prevailing break-up mechanism at each size class as a function of droplet diameter, the acting forces on the droplet and the exposure time in the high energy zone. We validated the accuracy of the results obtained from our equation-based model with experimental data obtained at pilot-plant scale. The results obtained by simulation showed that at a given set of operational conditions and pre-emulsion properties the product obtained was within the desired and narrow specifications space. As a concluding remark of this study we suggest further exploring the design and development of extensional-flow units for the creation of structured emulsions.

Keywords: emulsification, extensional flow, oil-in-water emulsion, population balance, momentum balance

1. Introduction

Every day we are in direct contact with emulsions and dispersions. Many food products do not form homogeneous mixtures but are composed of at least one component in the form of solids or droplets. This microstructure is responsible for key sensory attributes in food products, such as mouth-feel, colour, aroma and spreadability. Reaching a desired microstructure involves not only the close marriage between product composition and processing, but also the development of a process synthesis methodologies embracing both (see e.g. [1]). A model-based approach is a fundamental building block of such design methodologies.

Current technologies for the production of emulsion-type foods include, among others, static mixers [2], stirred tanks [3], rotator-stator devices [4,5], and high-pressure and ultrasonic homogenisers [6]. The governing droplet break-up mechanism in each device is highly dependent on the physical properties of the system (e.g. viscosities of continuous and dispersed phases), and on the flow conditions. The type of flow profile

that the droplets experience depends on the mechanical configuration and operational regime of the unit and can be characterized by the flow parameter $\alpha \in$ [-1,1] (Fig. 1-left). For instance, stirred tanks are basically featuring simple shear in the laminar regime, whereas turbulent inertial and cavitation mechanisms are responsible for droplet break-up in high-pressure homogenisers. Despite the sound expertise generated over the last decades regarding droplet disruption and coalescence mechanisms [3, 7, 8], the effort has been exclusively channelled to those units where either single shear flow (α=0) or rotational flow (-1<α<0) is dominating. Contrary to that research focus, both single shear flow and rotational flow are rarely the dominant droplet break-up mechanisms in commercial emulsion technologies. As a result, little (or rather non-existing) attention has been paid to explore units based on, for instance, extensional flow (0<α≤1). In this regard, this contribution intends to broaden the spectrum of current emulsification devices to this flow pattern.

2. Modelling framework

We address here the modelling and simulation aspects of a unit characterised by the extensional flow of material. In this type of flow regime, normally referred to as 'shear-free' flow, a preferred molecular orientation occurs in the direction of the flow field. Moreover, it is characterised by the absence of competing forces to cause rotation, resulting in a maximum stretching of molecules and large resistance to deformation [9]. As turbulent flow is responsible for droplet break-up in high stress emulsification units, we focus here on this flow regime. Note that from an industrial perspective, production of emulsions in the laminar regime ($Re_p < 10^{-2}$-10^{-6} [2]) is a highly energy-demanding operation for systems where $\eta_{disperse}/\eta_{continuous} < 0.1$ and even impossible for systems where $\eta_{disperse}/\eta_{continuous} > 4$ (Fig. 1-left). Under single shear conditions, the droplets are not able to deform as fast as the flow regime induces deformation [10].

From a mechanical point of view, the unit is materialised, for instance, in a converging element of the nozzle-type (Fig. 1-right).
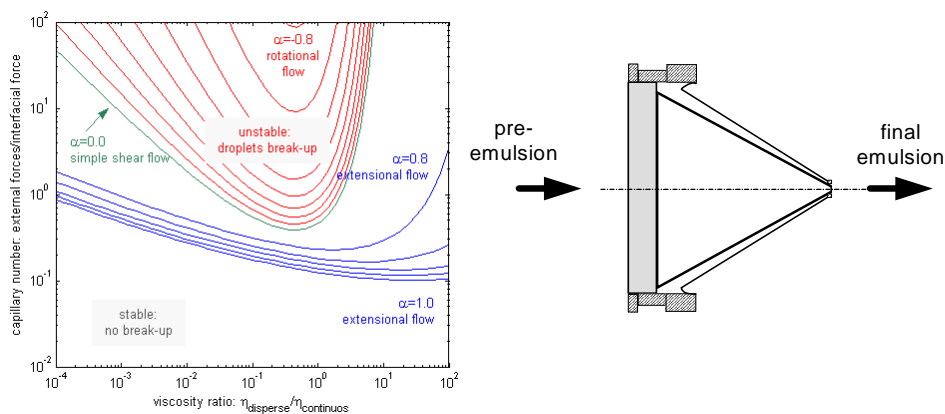
Figure 1. <u>Left</u>: Critical capillary number for droplets break-up in single shear, extensional flow and only rotational flow; α represents the flow type; <u>right</u>: nozzle-like unit featuring extensional flow

Modelling and simulation of extensional flow-units in emulsion formation

The dynamic modelling of the production process of an oil-in-water structured emulsion involves the description of all first principle phenomena together with a reliable estimation of relevant physical properties.

2.1. Population balance

Population balances (PB) are performed to account for the disappearance and appearance of droplets/particles at each size class in a size distribution. A relevant aspect of these balances is the estimation of the prevailing mechanism (break-up or coalescence) at each size class as a function of droplet diameter, acting forces on the droplet and time domain. Knowing the governing mechanism at a given size class allows us to calculate the rate of production and breakage of droplets. Thus, depending on its droplet size, the stress exerted on it and the its residence time at a given size class, a droplet is disrupted into a given number of daughter droplets, coalesce with a colliding droplet or remains unchanged in size.

The PB equations for a size class of mean volume v is given by the contributions of birth and death terms,

$$\frac{dn(v,\mathbf{z},t)}{dt} = \phi_{\text{in},\mathbf{z}}(V,t) - \phi_{\text{out},\mathbf{z}}(V,t) + \int_{v'>v}^{\infty} N(v',\mathbf{z})S(v',\mathbf{z})B(v|v',\mathbf{z})n(v',\mathbf{z},t)dv' - S(v,\mathbf{z})n(v,\mathbf{z},t) \tag{1}$$

where $n(v,\mathbf{z},t)dv$ represents the number of droplets per unit volume of the dispersion at time t and location $\mathbf{z}$ with volumes between v and $v+dv$, $\phi(v,t)$ is the convective flow of droplets of volume v, $N(v',\mathbf{z})$ is the number of daughter droplets at location $\mathbf{z}$ produced by breakage from parent droplets of volume v'; $B(v|v',\mathbf{z})$ is the droplet size distribution of daughter droplets of volume v at location $\mathbf{z}$ and produced by breakage of parent droplets of volume v'; $S(v,\mathbf{z})$ is the break-up frequency of droplets of volume v at location $\mathbf{z}$ and $n(v,\mathbf{z},t)$ is the number density of droplets of volume v at time t and location $\mathbf{z}$. The initial condition of the differential system is $n(v,\mathbf{z},0)$ and represents the number density of the coarse emulsion.

An analytical solution for the number-based PB (Eq. (1)) is provided by [11]. Their approach relies on simplifications of the breakage frequency and experimental correlations. In this contribution, the PB is solved numerically by discretisation of the size classes in the birth term (third term at right-hand-side of balance equation).

Droplet break-up occurs when the external stress exerted on the droplet surface exceeds the form-maintaining interfacial force locally for a sufficiently long time,

$$\sigma(\mathbf{z}) > P_L(\mathbf{z}) \tag{2}$$

and

$$t_r(v,\mathbf{z}) > t_b(v,\mathbf{z}) \tag{3}$$

where σ is the deformation stress; P_L is the capillary pressure (or Laplace pressure) and t_b, and t_r are the break-up and residence times, respectively.

The ratio between the external forces exerted on each droplet and the counteracting interfacial force that is responsible for keeping the droplet in a spherical shape defines a key dimensionless variable in droplet break-up. This variable –termed capillary number for viscous break-up mechanism (Fig. 1-left)- enables the prediction whether or not will occur under the operating conditions. In the turbulent flow regime the break-up of droplets is caused by two size-dependent mechanisms,

- **inertial break-up** mechanism, where the size of a droplet is larger than the length of Kolmogorov energy dissipating eddies; the pressure fluctuations are much larger than the pressure due to interfacial forces,
- **viscous break-up** mechanism, where the size of a droplet is smaller than the length of Kolmogorov energy dissipating eddies; the viscous forces are larger than the interfacial forces.

As stated in Eq. (3) for both mechanisms, the residence time of a droplet in the volume element where break-up occurs should be longer than the time required for break-up (t_b).

Understanding **droplet coalescence** in emulsions has been a hot research topic for the last decades [12]. Droplet coalescence occurs under the conditions that (*i*) droplet surfaces are not sufficiently stabilised by emulsifier agent and (*ii*) the contact time between adjacent droplets is long enough for the phase between them to be effectively drained. The coalescence mechanism involves the following events [3]: two adjacent droplets collide with a frequency given by their size and energy input. According to the coalescence efficiency both droplets will either coalesce into a larger droplet or separate. The coalescence efficiency has been found to be a function of the contact time between droplets, draining time of the phase between them and the interface characteristics between phases [6, 13]). Combining the collision frequency and coalescence efficiency results in the estimation of the coalescence frequency. Thus,

$$\Gamma = \lambda_c \varphi_c \tag{4}$$

where Γ is the coalescence efficiency; φ_c is the collision efficiency and λ_c is the coalescence frequency.

The functions to estimate the collision efficiency and the coalescence frequency have been reported to be dependent on the following variables [12]:

$$\varphi_c = f_\varphi(t_r, t_{mbc}) \tag{5}$$

$$\lambda_c = f_\lambda(\lambda_\eta, \Omega, d) \tag{6}$$

where t_{mbc} is the time between collisions, Ω is the capillary number, λ_η is the ratio between viscosities of the disperse and continuous phases and d is the mean droplet size.

2.2. Momentum balance

We describe the flow behaviour of the oil-in-water emulsion system with the Hershel-Bulkley (HB) model [9]. According to that model, the oil-in-water emulsion has a yield

stress (σ_0) with power-law dependence on shear rate after yield (γ). The model expression can be written in the form given by [14],

$$\sigma(v,\mathbf{z}) = \begin{cases} \sigma_0 + K\gamma(v,\mathbf{z})^n, & \sigma(v,\mathbf{z}) \geq \sigma_0 \\ 0, & \sigma(v,\mathbf{z}) < \sigma_0 \end{cases} \tag{7}$$

where n is the flow-behaviour index and K is the consistency factor.

Our hypothesis is that for a nozzle-like unit the overall pressure drop, ΔP_{in}, is composed of contributions originated from shear stress and a dominating extensional force flow [9]. Namely,

$$\Delta P_{in}(\mathbf{z}) = \Delta P_{in,\text{shear stress}}(\mathbf{z}) + \Delta P_{in,\text{extensional force}}(\mathbf{z}) \tag{8}$$

Descriptive models for break-up and coalescence mechanisms are embedded within a detailed droplet population and momentum balances. The unit under consideration consists of a conical section, through which the material flows and simultaneously pressure drop and extensional stress are built-up. The geometry of the unit resembles a nozzle with a converging angle close to 90° (see Fig. 1- right). The Sauter diameter (d_{32}) of our product target ranges 4-8μm. The initial pre-emulsion is represented by a log-normal distribution of droplets of a diameter of 70-75μm.

3. Results and discussion

The results obtained by simulation show that at a given set of operational conditions (e.g. flowrate) and pre-emulsion properties (e.g. initial Sauter diameter, $d_{32\text{-init}}$, and oil content, $x_{oil} > 0.74$), the diameter of the finished product is smaller than the targeted value.
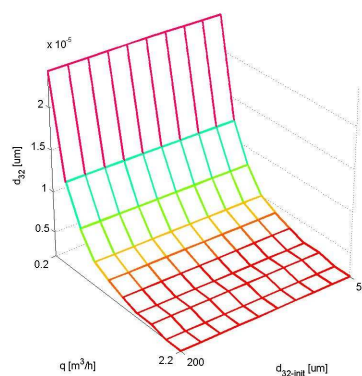
Figure 2. Overall variation of d_{32} with volumetric flow and initial Sauter diameter for $x_{oil} > 0.74$.

In terms of energy density, this fact implies that the efficiency obtained by means of extensional flow is higher than that of shear flow. Moreover, a sensitivity analysis suggests that the breaking mechanism changes with feed flowrate. For various oil contents it is noticed that large throughputs result in finer dispersions. Additionally, coarser pre-emulsions lead to coarser products regardless the oil level. The simulated data and experimental data obtained at pilot plant scale show a remarkable agreement (deviation <3%) and within the expected experimental error.

As depicted in Fig. 3-left, the dominating contribution to pressure drop comes from the extensional flow along the unit. This fact confirms our initial hypothesis regarding the nature of the comminution mechanism. Moreover, due to the short mean time between collisions (for x_{oil}>0.74) and large deforming stresses we found that break-up dominates over coalescence mechanism.
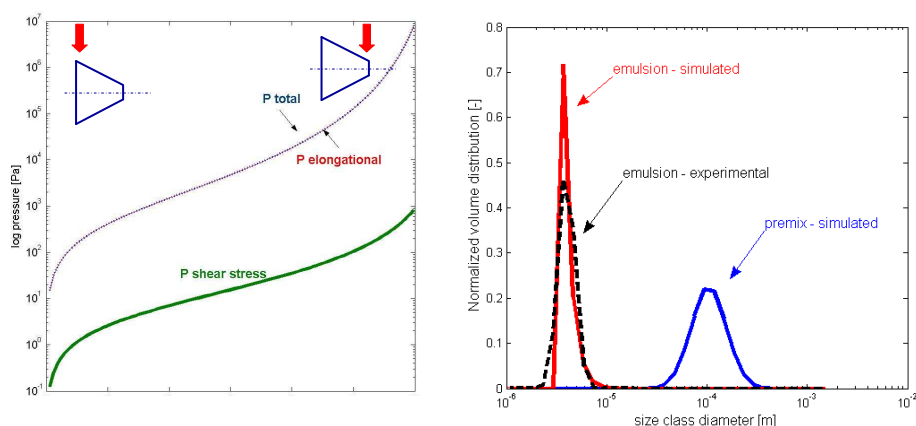
Figure 3. <u>Left</u>: pressure drop profile in the nozzle-type emulsification device; x-axis denotes axial location along the nozzle unit; <u>right</u>: simulated and experimental droplet size distributions. System features: q=1.8-2.0m^3/h, d$_{32-init}$=70-75μm, x_{oil}>0.74.

4. Conclusions and future work

Based on the development of a rigorous model for the nozzle-type unit and on experimental data we demonstrated the capability of such geometry when it comes to emulsification efficiency. Thus, the mean Sauter diameter obtained by simulations and experimental work (<4μm; see Fig. 3-right) suggests the possibility of reducing the disperse phase fraction without compromising the final structure and stability of the product. This last topic is worthwhile further exploring.

References

[1]   P. Bongers and C. Almeida-Rivera, Comp.Aided Proc. Eng., 20 (2009), 231.
[2]   H.P. Grace, Chem.Eng.Commun., 14 (1982) 225.
[3]   A.W. Nienow, Adv.Colloid Interface Sci., 18 (2004) 95.
[4]   J. Wieringa, F. Dieren, J. Janssen and W.G.M. Agterof, Trans.I.Chem.E. 69(1996) 554.
[5]   R.V. Calabrese, M.K. Francis, V.P. Mishra and S. Phongikaroon, Tenth European Conference on Mixing (H.E.A. van den Akker and J.J. Derksen, Editors), (2000) 149.
[6]   H. Karbstein, PhD thesis, University Karlsruhe, Germany, 1994.
[7]   W.J. Tjaberinga, A. Boon and A.K. Chesters, Chem.Eng.Sci., 48 (1993) 285.
[8]   A. Saboni, S. Alexandrova, C. Gourdon and A.K., Chesters, Chem.Eng.J., 88 (2002) 127.
[9]   J. Steffe, Rheological methods in Food process engineering, Freeman Press, USA, 1996
[10]   P. Walstra, Chem.Eng.Sci., 48 (1993) 333.
[11]   N.K. Nere and D. Ramkrishna, Ind.Eng.Chem.Res., 44 (2005) 1187.
[12]   A.K. Chesters, Trans.I.Chem.E., 69 (1991) 239.
[13]   H. Karbstein and H. Schubert Chem.Eng.Proc., 34 (1995) 205.
[14]   J.D. Sherwood and D. Durban, J. Non-Newtonian Fluid Mech., 77 (1998) 115.

Simulation of commercial dimethyl ether production plant

Ik Hyun Kim[a], Seunghyok Kim[a], Wonjun Cho[b], En Sup Yoon[a]

[a] *School of chemical and biological Engineering, Seoul National University, San 56-1, Sillinm-dong Gwanak-Gu, Seoul 151-744, Korea*
[b]*DME Project, R&D Division, Korea Gas Corporation, Incheon, Korea*

Abstract

The process of Dimethyl ether(DME) production consists of the four parts which are syngas synthesis from natural gas, absorbing CO_2 from syngas, DME synthesis reactor and DME separation/purification. KOGAS has developed a process in which syngas is produced from natural gas and converted to DME using a single reactor[1]. For the construction of commercial scale DME plant, the modeling of one-step DME synthesis reactor is required prior to beginning the construction. Since then, the simulation of DME production process should represent actual operation data of pilot or demo scale plant. The simulations of reactor had been conducted using a one-dimensional steady-state model of a shell-and-tube type fixed-bed reactor[2]. Using the result of a reactor analysis, we have conducted simulations of all processes using steady-state models in Aspen Plus[®]. The simulation of process in this paper reflects the result of a reactor simulation and the real operation data of demo scale DME plant. And the simulation results are satisfied with the requirements for the basic design and engineering of commercial DME plant construction.

Keywords: dimethyl ether, single step process, simulation, Aspen Plus

1. Introduction

Dimethyl ether(DME, CH_3OCH_3) is the simplest ether and is considered a leading alternative to petroleum-base fuels and liquefied natural gas. Its physical properties are similar to liquefied petroleum gas(LPG) and can be stored and delivered using existing infrastructures with minor modifications. And DME is considered as a substitute for diesel fuel because it has a cetane number of between 55 and 60.

Korea Gas Corporation(KOGAS) has developed a process in which syngas is produced from natural gas in proprietary auto-thermal reformer(Tri-Reformer) and then converted to DME in a single reactor. The 3,000 metric ton per day KOGAS DME plant will be built close to a remote natural gas field and will convert natural gas to dimethyl ether.

For the basic design of commercial DME production plant, a commercial steady state process simulator, Aspen Plus, has been used to model the KOGAS DME process and a number of cases were run and reviewed with KOGAS.

The previous simulation of process reflects only conceptual design of pilot and demo scale DME plants. In this paper, we have conducted simulations improved to reflect the data of experimental vapor-liquid equilibrium in the literature[3-7] and the test operation data of the KOGAS' demo scale plant. This work will provide theoretical supports for the basic engineering of commercial plant construction.

2. Process Description

The process producing dimethyl ether will have four main sections: Reforming, CO2 Absorption and Recycle, DME Synthesis, and DME Separation and Purification. Within each of these sections various technologies might be used and design decisions made.
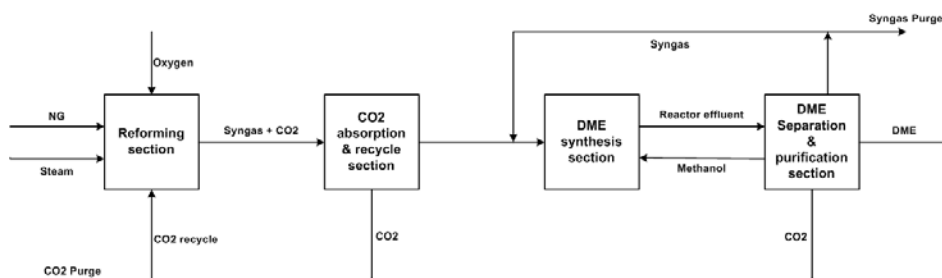
Figure 1. Simple block diagram of dimethyl ether production

In this process, natural gas and CO_2 are combined with a recycle carbon dioxide stream and heated in a fired heater. The recycle CO_2 comes from two locations: the first point is after the clean-up of the syngas and the second point is after the DME synthesis reaction. In the same fired heater steam is heated in a separate convection coil. The natural gas and recycled CO_2 stream combines with the steam and flows to the burner of the Tri-Reformer. Oxygen is heated in a separate heater that uses high pressure steam for the heating function. The temperatures are maintained such that reaction occurs instantaneously and a significant flame front is established.

The syngas exits the Tri-Reformer at temperatures of around 1000°C and pressures of about 3000 kPa. This hot gas is cooled down and compressed to 6000kPa and routed through the CO_2 Absorber which is designed the extract the CO_2 out of this stream using cold methanol.

The decarboxylated syngas along with unconverted reaction gas is reheated up to temperatures of around 230°C and directed into the DME synthesis reactors. The gases leaving the DME synthesis reactors are cooled down in a series of sequential heat exchangers. The final heat exchanger is designed to cool the syngas down to about -40°C. Most of the DME is condensed along with all of the methanol and water exiting the DME Reactor. The uncondensed gases, primarily comprising CO and hydrogen are recycled back to the DME synthesis reactors. A small amount of purge is maintained to control the inert levels in the recycle loop.

The liquid DME/CO_2/methanol/water stream is then depressurized down to 3500 kPa and flows into the CO_2 Column. All the CO_2 is rejected in the overheads. The bottoms are depressurized down to 1800 kPa and directed into a DME Column. The product DME is recovered in the overheads and directed to storage tanks.

The bottoms are a small stream of methanol with water. This is combined with a portion of methanol solvent from the CO_2 stripper and directed into the Methanol Dehydration Columns where the methanol is recovered. A portion of recovered methanol is recycled back to the CO_2 absorber and the rest is directed into the Methanol Dehydration Reactor for converting methanol into DME.

A large portion of CO_2 stream from CO_2 absorption and DME purification section is recycled back to the Reforming section and the rest is vented.

Simulation of commercial dimethyl ether production plant

3. Simulation Description

KOGAS has constructed and operated demo scale plant since 2008. The primal simulation studies of the pilot or demo scale plant have been accomplished in terms of conceptual design of DME processes. So they still have limits for the basic design of commercial scale DME plant. This means that the primal simulation does not represent sufficiently detailed utilities and process considerations like steam generation rate in the fired heater, significant heat exchangers, pressure drop, reflecting demo plant experiences of the key units, and so on.

Consequently, the purpose of simulation in this paper is the basic engineering of commercial DME plant. It should contain significant changes driven by the new process knowledge gained during the operation of the KOGAS DME Demo plant.

The first, physical property model is improved by using Aspen Plus[®], SRK(Soave-Redlich-Kwong equation of state) for vapor phase system and RKSMHV2(Redlich-Kwong-Kwong equation of state with modified Huron-Vidal mixing rules) for two phase vapor-liquid system. The Mathias-Copeman and UNIFAC molecular R_v and Q parameters from the article are used where available[3]. For components which are not covered by this literature, we regressed data from another literatures[4-7].

The second, we conduct the modification of DME synthesis reactor. The model of DME reactor represented by a single yield reactor block(RYIELD) in Aspen Plus[®] with a FORTRAN user subroutine that determines the reactor output composition based on KOGAS' specified CO conversion and product composition profile. It can accept and calculate five values which are one conversion and four selectivities, which is different from traditional definition of selectivity. It can also reflect the considered inert.

The third, we conduct the modification of Tri-Reformer model changing from Gibbs reactor block(RGIBBS) to yield reactor block(RYIELD). It can reflect the experimental data of demo scale Tri-Reformer operation more precisely.

The fourth, we conduct the adding heat exchangers, heaters, and coolers for realistic heat integration strategy. This represents good engineering practice, and makes it possible to properly calculate the operation costs. It can also represent the generation rate of high pressure, medium pressure, and superheated high pressure steam so that we can decide the specifications of fired heater.

Finally, the flowsheet layout and simulation convergence is improved for good engineering practice and leading to accurate results.

4. Simulation Result

The following Aspen Plus Modeling areas are being considered:
- Component List
- Physical Properties
- Flowsheet Layout and Unit Operation Blocks
- Convergence
- Reactors

4.1. Overall Design and Documentation Issues

There are significant design differences between the former case and the improved case as indicated in the table below:

Table 1. Overall design and documentation issues

Design Element	Former Case	Improved Case	Concern
Columns	5	5	Considerable impact on capital & operating costs
Process-Process Heat Exchangers	0	8	Increases process efficiency
Air Coolers	0	3	Decreases cooling water consumption
Compressors	1	2	Feasible utility costs

4.2. Component List
The component list is the same as that used former case:

- H_2, O_2, N_2, H_2O, CH_4, CO, CO_2, C_2H_6, C_3H_8, n-C_4, i-C_4, i-C_5, Methanol, DME

Our understanding is that most catalytic reactors producing methanol or DME also produced small amounts of higher molecular weight alcohols and ethers. But there is no document that KOGAS have measured these compounds during the Demo Plant operation before.

4.3. Physical Properties
The improved model uses the same physical properties as the former case for the vapor-only situation but modifies three different physical property packages for vapor-liquid situations.

Table 2. Physical properties

Situation	Former Case	Improved Case	Concern
Vapor-Only Flash	SRK (A+ default)	SRK (A+ default)	None
Vapor-Liquid Flash	SRK, NRTL, RKSMHV2 (A+ default)	RKSMHV2 (Literature parameters)	Validated with the literature data

4.4. Unit Operation Blocks and Flowsheet Layout
The number of unit operation block in the improved model flowsheet is significant different from the former case as shown in the table below. This is the result of design changes, the addition of all process heat integration, and the addition of a number of blocks used to characterize real process constitution.

Table 3. Unit operation blocks

CASE	RadFrac	Heater	Compr	MCompr	FLASH2	FSPLIT	Pump
Former	5	15	1	0	4	3	3
Improved	5	44	1	1	6	4	5

CASE	Mixer	RGIBBS	RYIELD	RSTOIC	SEP	VALVE	(Total)
Former	7	2	1	1	1	2	(45)
Improved	7	2	2	1	0	0	(78)

The lack of heat integration is unrealistic, does not represent good engineering practice, and makes it impossible to properly calculate the operation cost. So the ten heat exchangers and the three waste heat boilers are added in the improved case, it will impact the reduction of the steam generation and uses on the all process.

Simulation of commercial dimethyl ether production plant

4.5. Convergence

The former simulation model contains three TRANSFER blocks. This has been for easy convergence when the model calculations of the various case studies or analysis are performed.

But this use of TRANSFER blocks is unnecessary in basic engineering process and is not good simulation practice since they can lead to inaccurate results. In essence, it means that the model recycle streams are not fully connected and may not be properly converged. In addition there will be no warning from Aspen Plus that the results are inaccurate since the TRANFER blocks are outside of the normal mass balance checks.

This is modified by creating a Broyden convergence block and making a partial revision of convergence options in Aspen Plus®.

4.6. Reactors

The former and improved models contain the reactors shown in below. And all of reactors reflect new process knowledge from Demo Plant operation.

Table 4. Reactor type

Reactor	Former Case	Improved Case	Concern
Sulfur Guard Bed	none	RGIBBS	Added for representing capital cost and pressure drop
Pre-Reformer	RGIBBS	RGIBBS	
Tri-Reformer	RGIBBS	RYIELD	Modified for representing exactly real experiment data
DME Reactor	RYIELD (with Fortran subroutine)	RYIELD (with Fortran subroutine)	
Methanol Dehydration Reactor	RSTOIC	RSTOIC	

The Sulfur Guard Bed that would be required in an actual plant is included in this case. This is a relatively minor improvement and means that the pressure drop of this equipment is taken into account.

The Tri-Reformer using an RGIBBS as does in former case does not reflect actual operation condition or operation data of reaction. In the improved case, the Tri-Reformer satisfies the operation condition and represents the actual Demo plant reformer data more exactly.

The former and improved model represent the DME reactor using the RYIELD/Fortran subroutine method developed during the feasibility study. So the reaction calculation in the DME synthesis reactor is same as before.

5. Conclusion

The process simulation of the commercial DME plant, which has 3,000 metric ton per day capacity, has been completed. The former study on the DME synthesis reactor modeling provides the important specifications of overall process, and can help to reflect the real operation data more precisely. A commercial steady-state chemical process simulator, Aspen Plus®, is used to simulate the KOGAS DME plant in order to ensure the accuracy and reliability. Aspen Plus® includes the largest database of

physical properties and provides easy application of modifying its property models from literatures. So we modify the built-in model of SRK and RKSMHV2 so that simulation results can be entirely consistent with the experimental vapor-liquid equilibrium data of the literatures. It is also conducted to modify essential unit operation blocks in Aspen Plus$^{®}$' in order to reflect the real data and behavior of DME demo plant. The improvement of convergence strategy and flowsheet layout has been conducted to make good simulation practices and lead to accurate results. Based on the result of this simulation, basic design and engineering will be in progress.

6. Acknowledgements

References

[1] Seung-Ho Lee, Wonjun Cho, Taekyoung Song and Young-Jin Ra, 2009, Scale up study of DME direct synthesis technology, World Gas Conf., [Pap.], 24th, WOC1

[2] Daesung Song, Wonjun Cho, Gibaek Lee, Dal Keun Park, and En Sup Yoon, 2008, Numerical Analysis of a Pilot-scale Fixed-Bed Reactor for Dimethyl Ether(DME) Synthesis, Ind. Eng. Chem. Res, 47(13), 4553-4559

[3] Soren Dahl, Aage Fredenslund, and Peter Rasmussen, 1991, The MHV2 Model: A UNIFAC Based Equation of State Model for Prediction of Gas Solubility and Vapor-Liquid Equilibria at Low and High Pressure, Ind. Eng. Chem. Res., 30, 1936-1945

[4] Takashi Katayama, Kazunri Ohgaki, Goro Maekawa, Motojiro Goto, and Tamon Nagano, 1975, Isothermal vapor-liquid equilibria of acetone-carbon dioxide and methanol-carbon dioxide systems at high pressure, Journal of Chemical Engineering of Japan, Vol. 8, No. 2

[5] Torben Laursen and Simon Ivar Anderson, 2002, High-Pressure Vapor-Liquid Equilibrium for Nitrogen+Methanol, J. Chem. Eng. Data, 47, 1173-1174

[6] Elaine Chang, Jorge C.G. Calado, and Willian B. Streett, 1982, Vapor-Liquid Equilibrium in the System Dimethyl Ether/Methanol from 0 to 180 C and at Pressure to 6.7 MPa, J. Chem. Eng. Data, 27, 293-298

[7] E. Brunner, W. Hultenschmidt, and G. Schlichtharle, 1987, Fluid mixtures at high pressures; ⅠⅤ.Isothermal phase equilibria in binary mixtures consisting of (methanol+hydrogen or nitrogen or methane or carbon monoxide or carbon dioxide), J. Chem. Thermodynamics, 19, 273-291

On the Prediction and Shaping of the PSD in Crystallization Operations

Massimiliano Grosso[a], Omar Galan[b], Roberto Baratti[a], Jose A. Romagnoli[b]

[a] *Dipartimento di Ingegneria Chimica e Materiali, Università degli Studi di Cagliari, Piazza d'Armi, Cagliari, I-09123, Italy, E-mail:grosso;baratti@dicm.unica.it*
[b] *Department of Chemical Engineering, Louisiana State University, E-mail: jose@lsu.edu*

Abstract

The time behavior of the crystal size distribution in anti-solvent based crystal growth processes is investigated via statistical tools. The data are provided by the NaCl-water-ethanol anti-solvent crystallization system in a bench-scale fed-batch crystallization unit. It was found that the crystal population can be reasonably described in terms of a Gaussian or a Gamma distribution depending on the operating conditions.

Keywords: Anti-solvent crystallization, image data acquisition, maximum likelihood estimation, particle size distribution

1. Introduction

In crystallization processes, and particle synthesis in general, control of the size distribution is a crucial point of interest. Recently, a novel approach (Grosso et al., 2009; Galan et al. 2009) for the modeling, prediction and control of the particle size distribution (PSD) was introduced. The crystal growth is described as a stochastic process also driven by a deterministic growth term. Thus, a stochastic mathematical model (i.e. a Fokker-Planck equation) for the probability density distribution provides the evolution of the PSD at any instant of time. The approach was tested on experiments obtained from the crystallization of sodium chloride in water using ethanol as anti-solvent, performed in an experimental bench-scale semi-batch crystallizer. In particular it was found that modeling the diffusive term in the FPE is a crucial task that can strongly affect the final shape of the PSD, as already reported in the literature (Fa, 2005; Matsoukas and Lin, 2006). The previous considerations strongly motivate the study of the experimental PSD shape, since such analysis could give useful insights on the subsequent growth modeling.

The present work tackles this problem. Samples obtained from an experimental setup are analyzed through visual examination under microscope in order to obtain the particle size distribution. The procedure is repeated at different anti-solvent flow rates and times. A first analysis is provided to characterize and classify the particle size dispersion and a number of possible statistical distributions are compared. In addition, the distribution parameters are estimated.

2. Experimental setup

The NaCl-Water-Ethanol system is used in this study as anti-solvent crystallization system. The experimental apparatus is made up of one liter glass, cylindrical crystallizer submerged in a temperature controlled bath. The anti-solvent addition is carried out by a slave peristaltic pump.

At the start-up condition, the crystallizer is loaded with an aqueous solution of *NaCl* deionised water. The sodium chloride and ethanol are regent grade with purities og *99.9*% and *190* proof respectively. The temperature is controlled at *20*°C. Then ethanol was added to the aqueous NaCl solution using a peristaltic pump. Three different anti-solvent (ethanol) feeding conditions, at a feed-rates of 0.7 *mL/min* (hereafter referred as low feed rate), 1.5 *mL/min* (intermediate feed rate) and 3.0 mL/min (high feed rate), were performed. A detailed description of the experimental set-up and the experiment is reported in the literature. Along the operation, 5 *ml* samples were taken in an infrequent way. The sample was filtered over filter paper ($\geq$ 3.5 μm) and then dried in an oven at *50*°C for 24 h. The crystal size distribution is determined by visual inspection of images taken using a digital camera mounted in a stereo-microscope at 25× magnification. The captured images are then processed by means of sizing computer software (Amscope©). Image analysis was conducted on the process samples during the semi-batch operation. In Figure 1 typical images are shown, illustrating the growth of the particles, for the run at the calculated flow-rate policy. Similar images were obtained for the other conditions and times (not shown here).
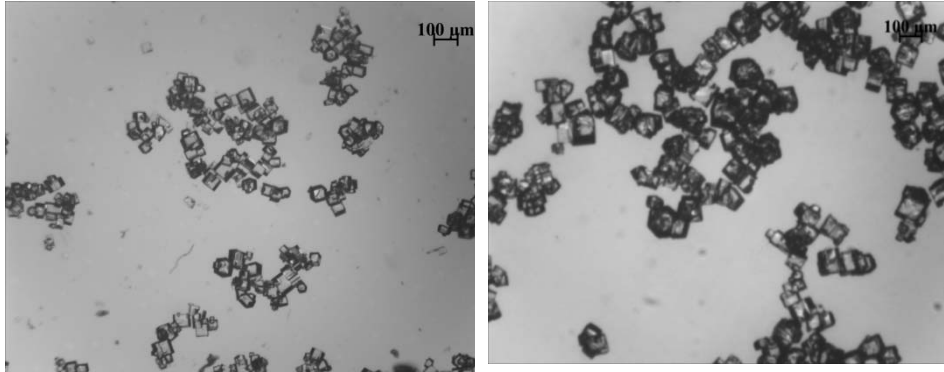
Figure 1: *Typical image acquired with the digital camera. for antisolvent feed flow of u_0=0.5 mL/min for t=0.17h and t = 5.0 h.*

3. Analysis of the data

The population of the measured crystals size was fitted to most common statistical distributions in order to establish which was the most proper for their description. In this work we concentrate on the following statistical distributions:

Normal (N) distribution

$$f(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma} exp\left(-\frac{1}{2}\frac{(y-\mu)^2}{\sigma^2}\right) \quad y \in \mathbb{R}, \; \boldsymbol{\theta} = [\mu,\sigma] \in \mathbb{R}^2 \tag{1}$$

Lognormal (LN) distribution

$$f(y|\mu,\sigma) = \frac{1}{\sqrt{2\pi}\sigma y} exp\left(-\frac{1}{2}\frac{(ln(y)-\mu)^2}{\sigma^2}\right) \quad y \in \mathbb{R}^+ \; \boldsymbol{\theta} = [\mu,\sigma] \in \mathbb{R}^2 \tag{2}$$

Weibull (W) distribution

$$f(y|\alpha,\beta) = \frac{\alpha}{\beta}\left(\frac{y}{\beta}\right)^{\alpha-1} exp\left(-\frac{y}{\beta}\right) \quad y \in \mathbb{R}^+ \; \boldsymbol{\theta} = [\alpha,\beta] \in \mathbb{R}^2 \tag{3}$$

On the prediction and Shaping of the PSD in Crystallization Operations

Gamma (G) distribution

$$f(y|\alpha,\beta) = \frac{1}{\beta^\alpha \Gamma(\alpha)} y^{\alpha-1} exp\left(-\frac{y}{\beta}\right) \quad y \in \mathbb{R}^+ \quad \boldsymbol{\theta} = [\alpha,\beta] \in \mathbb{R}^2 \tag{4}$$

The parameters in the models are computed by means of the Maximum Likelihood Estimation procedure. In order to establish whether the fitted distributions model our data we use standard goodness of fit tests (see e.g. D'Agostino and Stephens, 1986). In particular, we have chosen two consolidated statistical tests: the Kolmogorov-Smirnov (KS) and the Anderson-Darling (AD) test.

The Kolmogorov-Smirnov statistic D_{KS} for a given cumulative distribution $F(y)$ to be tested, is based on the empirical cumulative distribution function (ECDF) evaluated by experiments. Thus the statistics is defined as

$$D_{KS} = max_{1 \le i \le n}\left(F(y_i) - \frac{i-1}{n}, \frac{i}{n} - F(y_i)\right) \tag{6}$$

Where i is the number of points less than y_i (and the y_i are ordered from smallest to largest value). The null hypothesis to be investigated that the sample comes from $F(x)$ is rejected at a significance level a if the test statistic is greater than the critical value obtained from a table.

The Anderson-Darling test compares the fit of an observed cumulative distribution function to an expected distribution function. This test gives more weight to the tails than the KS test (D'Agostino and Stephens, 1986).

The *AD* statistics A^2 is defined as:

$$A^2 = -n - \frac{1}{n}\sum_{i=1}^{n}\left((2i-1)\left(lnF(y_i) + ln(1 - F(y_{n+1-i}))\right)\right) \tag{8}$$

The AD is a modification of the KS test and gives more weight to the tails than the KS test.

Both the statistics here introduced can give a measure of the *closeness* of the data to the given distribution: roughly speaking the smaller the statistic, more related to the distribution are the data.

4. Results and discussion

In the following, for sake of space, the detail of the analysis will not be presented for all the flow-rates. Figure 2 shows the AD statistics with respect to the time, for the data related to $u = 0.7$ *ml/min* (low flow rate), with the Normal, Lognormal, Weibull and Gamma distributions. The Lognormal distribution (up-triangles) gives the poorer performance whereas the Gamma and Normal distribution behave better: the Gamma distribution should be preferable during the initial transient while the Normal (and/or the Weibull) fits properly at larger times.
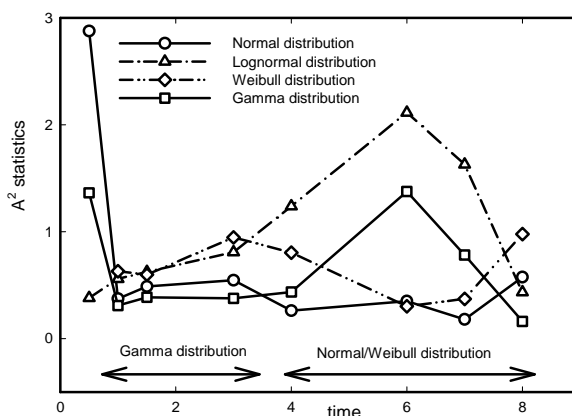
Figure 2: The AD statistics compute for the data provided at low flow rate (u = 0.7 ml/min) for the Normal distribution (solid line), the Lognormal distribution (dashed line) and the Gamma distribution (dot-dashed line), the Weibull distribution (dot-dot-dashed line)

Figure 3 reports the KS statistics with respect to the time, for the data related to *u* = 1.5 *ml/min* (intermediate flow rate), with the Normal, Lognormal and Gamma distributions. Similar results are obtained by considering the AD statistics. Again, it is found that the scenario changes with time: as reported in the figure, in the transient after the experiment start-up, the Normal distribution appears to be the best choice for the description of the crystal sample. Conversely, as time increases, the Gamma distribution should be preferred. As a final remark, one should observe that, as already commented, the Log-normal distribution gives the poorer results, at least with respect to the other curves here considered.
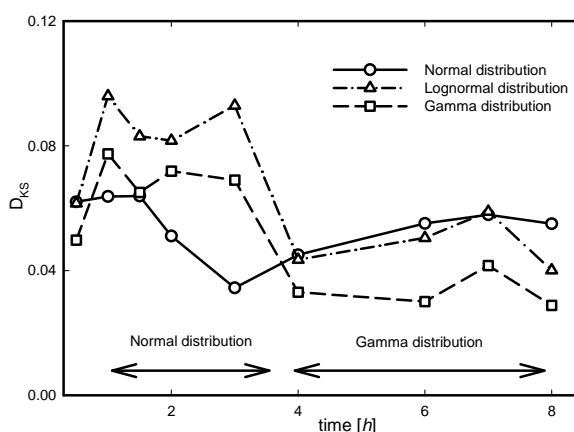
Figure 3: The KS statistic computed for the data provided at intermediate flow rate (u = 1.4 ml/min) for the Normal distribution (solid line), the Lognormal distribution (dashed line) and the Gamma distribution (dot-dashed line)

Finally, the results obtained at *u* = 3.0 *mL/min* are reported in terms of *p*-values computed with the KS statistics in Table 1. One should remind that the *p*-value represent the probability that the data are generated from the assumed distribution. Thus

p-values close to the unity indicate a good fit to the hypothesized distribution. The main features observed at lower concentration feed are confirmed: the distribution is initially close to a Gamma distribution, and then resembles more with the Weibull and with a Normal distribution.

Table 1: p-values for the KS statistics evaluated at the feed concentration u = 3.0 *mL/min*

	Normal distribution	Gamma distribution	Weibull distribution	Lognormal distribution
t = 0.17 h	**0.83093**	**0.80191**	0.55924	0.54379
t = 1.0 h	0.6234	**0.80721**	0.58606	0.58606
t = 1.5 h	**0.87795**	**0.86522**	**0.82088**	0.54993
t = 2.0 h	0.3513	0.69853	0.29788	0.42674
t = 3.0 h	**0.75387**	0.21857	0.69518	0.1005
t = 4.0 h	0.68889	0.4353	**0.76206**	0.25982
t = 5.0 h	**0.9704**	0.75298	**0.99273**	0.51641

In summary, it was observed that the best fit is addressed in most of the cases with the Gamma and the Normal distribution. This led us to the issue of addressing the characterization of the parameters of such distributions with respect to the process conditions. To this end, Figure 4 reports the estimated μ and σ parameters of the Normal distribution together with their 95% confidence intervals. It can be noticed indeed that the μ parameter increases with the time, and it decreases with the *u* feed concentration. A similar scenario is found for the standard deviation σ, meaning that a spreading of the crystals population is experienced as time increases. The only (significant) exception is observed at the highest feed concentration, where a shrinkage in the crystals population is evident for *t > 3 h*. As a final comment, for *t < 2.0 h* the μ and σ parameters do not differ significantly for the different operating conditions and the related confidence intervals partially overlap. On the other hand, a clear distinction is observed for *t > 4.0 h* and one can conclude that the operating conditions do have an impact.
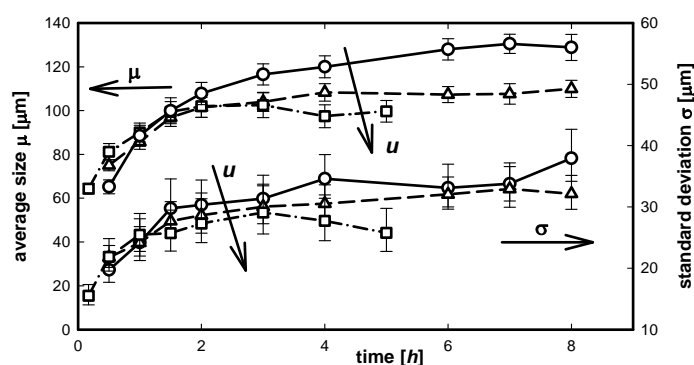
Figure 4: The parameters μ and σ of the Normal distribution evaluated at the different flow rates. Circles: u = 0.7 ml/min; Up-Triangles: u = 1.5 ml/min; Square points: u = 3.0 ml/min. Confidence intervals for the estimation are also reported.

Regarding the Gamma distribution, the estimated α (shape) and β (scale) parameters are reported in Figure 5. In more detail, Figure 5.a shows the point estimations with respect

to time, for the different process conditions. A clear trend is appreciated for both the parameters as time increases. On the other hand, no clear dependece on the *u* (antisolvent flow) can be envisaged. A partial explanation of this scenario is given in the Figure 5.b where detail of the parameter inference at *u* = 1.5 *mL/min* with the related confidence intervals are shown. It is eviden that the confidence intervals are much larger than the ones observed with the Normal distribution, thus the parameter inference appears to be more affected by the uncertainties that, in turn, cannot allow a clear distinction among the different process conditions.
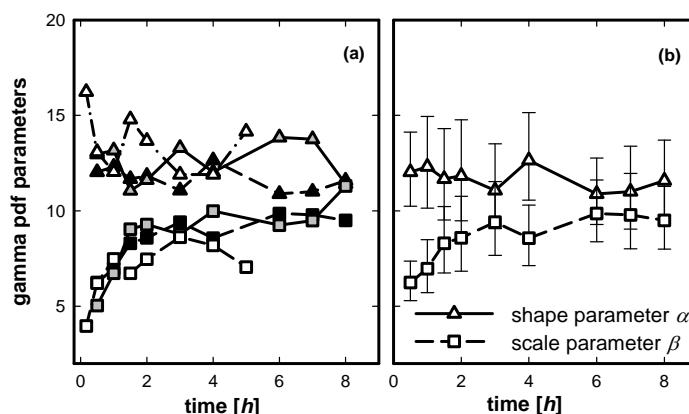
Figure 5: a) The shape parameter α (up-triangles) and scale parameter β (squares) of the Gamma distribution evaluated at the flow rate u=1.5 ml/min, together with their confidence intervals b) Point estimation of a (up-triangles) and b (squares) at the different anti-solvent feed rates. White points: u = 3.0 ml/min; gray points: u = 1.5 ml/min; white points: u = 0.7 ml/min

5. Conclusions

In this work different statistical distributions were compared in order to determine the distribution that best fits the experimental data towards the estimation of the PSD in crystallization processes. The statistical analysis gave no univocal selection of the most proper distribution for the current data and the results depend on the operating conditions. It was found that i) the Normal and the Gamma distributions gave in most of the cases the best fit to the data, ii) the Lognormal distribution seldom leads to a fair description of the crystal size sample (in that case, this occurs especially at the process start-up) iii) in some cases the Weibull distribution could also be a reasonable choice.

References

R.B. D'Agostino and M.A. Stephens, 1986, Goodness-of-fit Techniques, Marcel Dekker, New York

K.S. Fa, 2005, Exact Solution of the Fokker-Planck equation for a broad class of diffusion coefficients, Phys. Rev. E, 72, 020101(R)

O.Galan, M.Grosso, R.Baratti and J.A. Romagnoli, 2009, Stochastic Approach for the Calculation of Anti-Solvent Addition Policies in Crystallization Operations: An Application to a Bench-Scale Semi-Batch Crystallizer, submitted for pubblication on Chem. Eng. Sci.

M.Grosso, O.Galan, J.A.Romagnoli and R.Baratti, 2009, A Stochastic Approach for the Prediction of PSD in Antisolvent Mediated Crystallization, accepted for pubblication on AIChE J.

T.Matsoukas and Y.Lin, 2006, Fokker-Planck for particle growth by monomer attachment, Phys. Rev. E, 74, 031122

Systematic approach towards inspecting equipment defects in chemical plants

Seunghyok Kim, Young Hun Kim, Won So, Ik Hyun Kim, En Sup Yoon*

School of Chemical and Biological Engineering, Seoul National University, 151-744, Seoul, Republic of Korea, hyok20@pslab.snu.ac.kr

Abstract

Today, there are so many chemical plants in the world. The more chemical plants are established, the more it is important to maintain them properly. The inspection and maintenance of risk equipments in chemical plants are of fundamental importance in process safety system; however, strategic methods in maintaining equipments like vessels, pipes and valves are applied on few plants only.

This paper presents the systematic approach towards inspecting equipment (pipe lines centrally) defects in chemical plants. It is accomplished by analyzing the inspection and maintenance records in specific plants (specific data), as well as the offered standard database (generic data) first. Then, equipment that is frequently defected can be distinguished. By using the Dow Fire and Explosion Index (DF&EI), risk index of the process units is generated. In addition, risk priority of equipment is generated by applying the damage mechanism standards. As a result, the inspection priority based on risk can be achieved that is based on semi-quantitative method.

Keywords: Inspecting equipment, Dow Fire and Explosion Index

1. Introduction

The effective procedure is needed to inspect and detect equipment defects in chemical plants considering costs and time required. In other words, the regular inspections and maintenances should be based on the risk priority. But existing indexes and methodologies cannot reflect this inspection method based on risk priority. There are inspection standards for the unit equipments or facilities but there are not standards for the whole processes taking into account process risks. For estimating process risks quantitatively or semi-quantitatively, relative risk priority decision, fault tree analysis, event tree analysis can be used. And for estimating equipment risks, damage mechanism and inspection standards for equipment can be used. Nevertheless, they are restricted within the specified limits in chemical plants. So, the integrated method is required considering process and equipment risks.

In this paper, this integrated inspection algorithm will be introduced hierarchically. First, Dow Fire and Explosion Index are applied to evaluate the relative priority in unit processes. Then, in equipment level, the semi-quantitative evaluation is conducted by using damage mechanism and configuration. Finally, considering equipments data, the integrated algorithm is composed and makes the risk ranking. This ranking is reflected to priority of inspection planning and the effective diagnosis of chemical plants can be achieved.

2. Background

　　1) Relative ranking decision

　Relative ranking decision method is more practical strategy than former complex analysis methods. This strategy makes the analyzer catch the risk potential by comparing a few process characteristics. And this informs the proper location selection of processes, design, arrangements of equipments as well as relative risks. Relative ranking also point out the risk potential during the operation.

　There are several methods of relative ranking decision, for example, Dow Fire and Explosion Index (DF&EI), Mond Index, Substance Hazard Index (SHI), Material Hazard Index (MHI), Chemical Exposure Index (CEI), etc. In this paper, Dow Fire and Explosion Index is applied to evaluate relative risk of processes and unit processes in a NCC plant.

　　2) Dow Fire & Explosion Index

　The Dow Fire and Explosion Index (DF&EI) calculation is a tool to help determine the areas of greatest loss potential in a particular process. It also enables one to predict the physical damage that would occur in the event of an incident.

　The first step in making the Dow F&EI calculation requires using an efficient and logical procedure to determine which process units should be studied. A process unit is defined as any major item of process equipment. Unloading and loading facilities, storage tanks treating combustible and flammable material, reactors, distillation columns, quench vessels, storage vessels could be identified in a typical plant. The main evaluation items are material factor (MF), general process hazards factor (F_1) and special process hazards factor (F_2). The Material Factor (MF) is the basic starting value in the computation of the F&EI and other risk analysis values. The MF is a measure of the intrinsic rate of potential energy release from fire or explosion produced by combustion or chemical reaction. The MF is obtained from the flammability and instability rankings according to NFPA 704. The next is to calculate the process unit hazards factor (F_3), which is the term that is multiplied by the material factor to obtain the F&EI. The numerical value of the process unit hazards factor is determined by first determining the general process hazards factor (F_1) and special process hazards factor (F_2) listed on the F&EI form. Each item which contributes to the process hazards factors contributes to the development or escalation of an incident that could cause a fire or an explosion. The usual process unit hazards factor value is about 1~8.

$$F_3 = F_1 \times F_2 \tag{1}$$

$$F\&EI = F_3 \times MF \tag{2}$$

　The risk classification of Dow F&EI is shown in following table.

Table 1. Dow F&EI risk classification

Index	Risk	Grade
1~60	Light	E
61~96	Moderate	D
97~127	Intermediate	C
128~158	Heavy	B
159~	Severe	A

3) Damage mechanism and Configuration

To estimate damage of equipments, the damage mechanism and configuration based on API 581 is identified. Following table indicates the kind of damage mechanisms and configurations applied in this research.

Table 2. Damage mechanism and configuration

Damage mechanism	Configuration
High Temperature Hydrogen Attack	Flow Velocity Influence
Hydrogen Blistering/Embrittlement	Injection Point
High Temperature Sulfuric Corrosion	Turbulent Flow
Galvanic Corrosion	Corrosive Material Accumulation
Stress Corrosion Cracking	Phase Change

3. Algorithm and Case Study

1) Algorithm

The algorithm is suggested to evaluate risk in chemical plants. This evaluation procedure starts with the step of analysis of characteristics in target process. And then, relative risk of the target process is estimated by Dow F&EI. If there is no flammable material or MF is too low to value in the target process, Dow F&EI might be useless to estimate the risk of the process. So the process which cannot apply Dow F&EI could be classified separately.
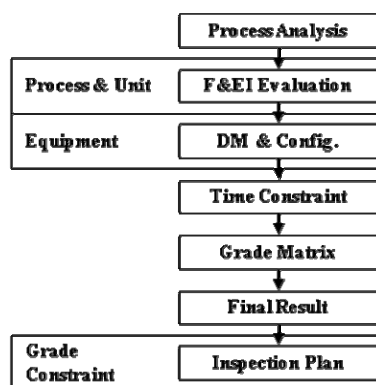
Figure 1. Suggested Algorithm

Based on relative risk decided by Dow F&EI, the damage mechanism and configuration could be judged whether they can be applied. In equipment inspection level, more detail process analysis is needed (damage mechanism and configuration). The opinion of experts and skilled engineers could contribute to increase reliability.

And since it is more effective to estimate old equipments prior to new or recent-fixed equipments, the time constraints are added in suggested algorithm.

Then the grade matrix (Figure 2) could be composed by deciding relative risk based on Dow F&EI with time constraints. The higher number means the higher priority.

Figure 2. Grade Matrix by Dow F&EI evaluation

The inspection priority is derived from the evaluation of relative risk ranking. This makes the plant managers plan the inspection schedule more efficiently like reducing time and cost.

2) Case study

NCC (Naphtha Cracking Center) plant is used for estimating the algorithm performance. In this case, a NCC plant is divided into 14 processes and 14 processes are divided into several separate unit processes. This classification is based on PFD in process and unit process level to estimate Dow F&EI and P&ID and Line Index in equipment level to analyze damage mechanism and configuration. Past inspection history and data is considered for time constraints.

4. Results and Discussions

Dow F&EI classification is conducted in the process and unit level. Table 3 is a result of top priority. (A: the highest priority, E: the lowest priority)

Table 3. Process and unit level classification

Process	F&EI	Grade	Ranking
Cracking Heater[*]	165.74	A	1
Gasoline Fractioner	107.26	C	10
Quench Tower	107.17	C	11
Charge Gas Compressor&Dryer	131.55	B	2
Methanization	96.04	D	13
Chilling Section	116.73	C	6
Demethanizer	117.98	C	5
Deethanizer	112.56	C	8
Acetylene Converter	130.68	B	3
Ethylene Fractionator	110.46	C	9
Depropanizer	97.81	C	12
MAPD Converter	120.83	C	4
Propylene Fractionator	113.11	C	7
Debutanizer	90.18	D	14

Systematic approach towards inspecting equipment defects in chemical plants

Process	Dwg(unit)	Description	F&EI	Grade
Cracking Heater	E5939-01001B	GAS TURBINE AND COMBUSTION AIR SYSTEM	113.18	C
	E8015-01001A	PYROLYSIS SECTION SRT VI LIQUID CRACKING HEATERS E-BA102～E-BA111	164.95	A
	E8015-01001AA	PYROLYSIS SECTION SRT VI LIQUID CRACKING HEATERS E-BA112～E-BA114	165.74	A
	E8015-01001B	PYROLYSIS SECTION SRT VI LIQUID CRACKING HEATERS E-BA101	165.00	A
	E20833-01001A	PYROLYSIS SECTION SRT VI LIQUID CRACKING HEATERS E-BA116	164.95	A
	E5939-01001C	FUEL GAS COMPRESSOR SYSTEM	84.96	D
Quench Tower	E8015-01001D	QUENCH TOWER	107.17	C
Charge Gas Compressor & Dryer	E8015-02001A	GASOLINE STRIPPER AND CHARGE GAS COMPRESSION STAGES 1,2 &3	116.43	C
Methanization	E8015-02001E	DRYER FEED WASH, COND, STRIPPER AND CHARGE GAS COMPRESSION STAGES 4&5	131.55	B
	A1-12655-02001CC	ACID GAS REMOVAL CHARGE GAS DRYING AND DRYER REGENERATION FACILITIES	113.78	C
	E8015-02001C	ACID GAS REMOVAL CHARGE GAS DRYING AND DRYER REGENERATION FACILITIES	112.52	C
	E8015-03001A	METHANIZATION AND HYDROGEN PURIFICATION	96.04	C
Chilling Section	E8015-03001B	CRACKED GAS CHILLING	116.73	C
Demethanizer	E8015-03001BC	PYROLYSIS SECTION SRT VI LIQUID CRACKING HEATERS	114.17	C
	E8015-03001BB	CRACKED GAS CHILLING (NEW SYSTEM)	109.52	C
	E8015-03001C	DEMETHANIZER	117.98	C
Deethanizer	A1-12655-04001AA	PARALLEL DEETHANIZER	103.05	C
Acetylene Converter	A1-12655-06501A	DEETHANIZER AND ACETYLENE HYDROGENATION	100.72	C
	E8015-04001A	DEETHANIZER AND ACETYLENE HYDROGENATION	109.31	C
	E8015-04001A	DEETHANIZER AND ACETYLENE HYDROGENATION	130.68	B

And the damage mechanism and configuration is added on the result based on Dow F&EI. The following table indicates the result in case of pipe lines.

Table 4. Inspection priority of pipe lines in a NCC plant

LINE	P&ID No.	F&EI	Damage Mechanism	Configuration	Grade
P-2002	6A	C	H2/H2S	Turbulent	5
P-30001	3007H	C	H2/BM	Turbulent	5
P-20004	2006M	C	H2	Turbulent	5
P-20005	2006M	C	H2	Turbulent	5
P-2222	6J	C	H2	Turbulent	5
P-40061	4008P	C	H2	2Phase	5
P-4061	8B	B	H2	-	4
P-3002	7A	D	H2	Turbulent	4
P-4016	8A,8B	B	-	CV/2Phase	4
P-3011	7A	D	H2	CV	4
P-4176	4008F	C	BM	CV	4
P-4051	8B	B	H2	-	4

In this Table, grade is the value presented in Figure 2. From the results, the processes treating the more flammable materials or operated in severe environment tend to have risky grades. But with the integration of the damage mechanism and configuration, H2 damage becomes the dominant component of the risk grade despite the low F&EI value.

5. Conclusion

In this paper, integration of relative risk priority decision, damage mechanism and configuration could be a method to decide the equipment inspection priority in chemical plant. This method helps the plant manager to plan the inspection schedule effectively.

References

AIChE, 1994, Dow Chemical Company Dow's Fire & Explosion hazard classification guide
CCPS, 2008, Guidelines for hazard evaluation procedure
Daniel A., 2002, Chemical process safety
API 581, 2000, Base Resource Document-Risk-Based Inspection

Ab Initio Crystal Structure Prediction for Flexible Molecules

Andrei V. Kazantsev, Panos G. Karamertzanis, Constantinos C. Pantelides, Claire S. Adjiman

Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, London SW7 2AZ, United Kingdom, c.adjiman@imperial.ac.uk

Abstract

A three-stage lattice energy minimisation methodology is reported for the identification of stable crystal structures containing flexible molecules. The accuracy of the approach is improved from stage to stage by using increasingly accurate models, first by substituting an atomic charge model with a distributed multipole model computed from the isolated-molecule charge density, and then by allowing user-defined molecular flexibility. When using the most sophisticated model, the computational cost is kept manageable by the use of local approximate models (LAMs) to estimate the conformationally dependent intramolecular energy and charge density and the reuse of quantum mechanical calculations stored in a molecule-specific databases. The applicability of the approach is illustrated by the generation of crystal structures with one molecule in the asymmetric unit for hydroquinone. The methodology combines an extensive search with accurate models for the lattice energy and can be equally applied to single and multi-component systems, such as cocrystals, salts and hydrates.

Keywords: Crystal Structure Prediction, Polymorph, Lattice Energy, Minimisation

1. Introduction

A large proportion of organic molecular solids are obtained in one or multiple stable crystalline forms (polymorphs). The knowledge of the three-dimensional atomic structure of a crystal is the basis to understanding and predicting the physical properties of the material (colour, density, solubility, dissolution rate etc.) [1]. Hence, the development of algorithms to predict the structure and the thermodynamic stability of single and multi-component crystals is of significant practical importance.

One of the most successful approaches for predicting organic crystal structures is to minimise the lattice energy, E^{latt}, of a large number of candidate structures [2]. It is assumed that the experimentally observed crystal structure is amongst the most stable packing arrangements, corresponding to low-lying minima in a computational energy landscape. Crystal structure prediction is thus a complex multidimensional optimisation problem. The decision variables are the space group, the unit cell dimension, and the position of all atoms in the unit cell. The size of the problem increases dramatically with molecular size and the number of molecules in the asymmetric unit (e.g. cocrystal, salts, solvates). Many hypothetical structures are very similar in stability [3]. Consequently, the correct ranking of structures requires an accurate representation of all the components of the lattice energy - the intermolecular energy contributions, U^{inter}, dominated by electrostatic interactions, and the energy cost associated with distorting

the molecule from its gas-phase conformation, ΔE^{intra}. The accurate calculation of these two components requires elaborate energy models based on quantum mechanical (QM) calculations [4]. This is especially demanding for flexible molecules as the conformation adopted by the molecules affects the interaction energy greatly.

In this paper, we address the challenge of predicting reliably possible crystal structures of flexible molecules without relying on any experimental information. In order to keep the computational cost manageable, a 3-stage minimisation procedure, in which increasingly accurate and complex models are used, is proposed. This is described in Section 2. The approach is applied to the crystal structure prediction of hydroquinone in Section 3.

2. Computational Methodology

Intermolecular forces are generally weak compared with the forces between covalently bonded atoms [3]. As a result, the majority of intramolecular degrees freedom ($\boldsymbol{\theta}$) are not expected to deviate from their gas-phase values (for example most bond lengths and bond angles). Consequently, molecular flexibility can often be captured by using a small set of "flexible" degrees of freedom, $\boldsymbol{\theta}^f$, such as torsions around single bonds. The intramolecular energy, ΔE^{intra}, can then be calculated as a function of the flexible degrees of freedom from QM calculations:

$$\Delta E^{\text{intra}}\left(\boldsymbol{\theta}^f\right) = \min_{\boldsymbol{\theta}^r}\left[E^{\text{intra}}\left(\boldsymbol{\theta}^r;\boldsymbol{\theta}^f\right)\right] - E^{vac} \tag{1}$$

where the constant E^{vac} is defined as the global conformational minimum for the molecule in isolation. The remaining "rigid" degrees of freedom, $\boldsymbol{\theta}^r$, are assumed to relax to their gas-phase values in response to the changes in the flexible degrees freedom, and can be obtained from the same isolated-molecule calculation:

$$\boldsymbol{\theta}^r\left(\boldsymbol{\theta}^f\right) = \arg\min_{\boldsymbol{\theta}^r}\left[E^{\text{intra}}\left(\boldsymbol{\theta}^r;\boldsymbol{\theta}^f\right)\right] \tag{2}$$

The lattice energy minimisation problem is then formulated as:

$$\min_{X,\boldsymbol{\theta}^f} E^{\text{latt}} = \Delta E^{\text{intra}}\left(\boldsymbol{\theta}^f\right) + U^{\text{inter}}\left(\mathbf{X},\boldsymbol{\theta}^f;\boldsymbol{\theta}^r\right) \tag{3}$$

where, **X** denotes the lattice variables comprising the unit cell lengths and angles and the position and orientation of each molecule in the asymmetric unit. Problem (3) is solved at each of the three stages of the proposed methodology with different models for the intra- and intermolecular energy terms, summarised in Table 1.

In the first stage, CrystalPredictor [5] is used to generate a large number of crystal structures in a set of user-defined space groups and to provide a first estimate of their relative stability. An isotropic intermolecular potential, U^{inter}, is used. At this stage the molecule is either treated as rigid or the intramolecular energy is interpolated over a set of precomputed *ab initio* conformational energies. The efficiency and parallel implementation of the algorithm make it suitable for searching lattice energy surfaces of complex multi-component crystals.

In the second stage, a large set of the most stable structures from Stage 1 are minimised [6] without further conformational relaxation, by replacing the point charges used in the first stage with a distributed multipole model, computed from the QM charge density [7], to improve the quality of the representation of the electrostatic interactions [8].

In the final stage [9], the modelling accuracy is further improved by allowing molecular flexibility and re-minimising the set of most stable structures obtained from Stage 2.

Ab Initio Crystal Structure Prediction of Flexible Molecules

The computational cost is kept manageable through the use of local approximate models (LAMs) for the conformationally-dependent intramolecular energy and charge density. These are as accurate as explicit QM calculations but carry a much smaller computational burden. QM calculations are performed for a limited number of reference conformations in the course of lattice energy minimisation.

Table 1 Summary of the multistage lattice minimisation method. Indicative CPU cost is per crystal structure of the molecule used in this study.

	Stage 1	**Stage 2**	**Stage 3**
Flexible d.o.f.	fixed or optimised	fixed	optimised
Rigid d.o.f	fixed	fixed	as function of flexible d.o.f.
Intramolecular energy	fixed/hermite polynomial on grid from QM	fixed	"on-the-fly" QM/LAM
Intermolecular electrostatic	conformationally-invariant atomic charges	multipole moments	conformationally-dependent multipole moments
Intermolecular repulsion/dispersion	isotropic exp-6	isotropic exp-6	isotropic exp-6
CPU time per minimisation	seconds	minutes	hours

The intramolecular energy is estimated using a LAM based on a quadratic Taylor expansion of the intramolecular energy in terms of the flexible degrees of freedom, $\boldsymbol{\theta}^f$. The expansion is constructed around a reference conformation derived from the intramolecular energy and its first and second derivatives with respect to the flexible degrees of freedom are computed via an isolated-molecule QM calculation.

An additional QM calculation is performed to compute a distributed multipole expansion of the isolated-molecule charge density of a reference conformation, where moments up to hexadecapole are expressed with respect to the local axis system of each atom. This LAM is then used to model the conformationally dependent intermolecular electrostatic interactions by rotating the multipole moments for small conformational changes, assuming that they remain invariant with respect to the local axis system.

The LAMs are valid within a given hyper-rectangle of conformational space around the reference conformation. The QM-computed quantities used to construct the LAMs, are stored in molecule-specific databases and, thus, can be re-used during subsequent minimisations greatly reducing the computational cost. In this work, the LAMs were recalculated after a 5° change in the value of the flexible torsions.

The crystal structure prediction algorithm is illustrated by application to the prediction of polymorph β of hydroquinone (Fig 1). The results are compared with the experimental form [10].
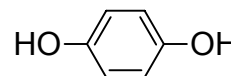
Figure 1 Molecular structure of hydroquinone.

3. Results and Discussion

In this work, all QM calculations were performed using GAUSSIAN [11] at the HF/6-31G(d,p) level of theory, except for the charge density calculations which were performed at the MP2(f,c)/6-31G(d,p) level. An empirical exp-6 potential [12] was used to model the repulsion-dispersion interactions.

3.1. Crystal structure prediction

In Stage 1, the gas-phase conformation of hydroquinone was used to generate hypothetical crystal structures from 50000 starting points in 26 common space groups ($P1$, $P\bar{1}$, $P2_1$, $P2_1/c$, $P2/c$ $P2_1/m$, $P2_12_12$, $P2_12_12_1$, $Pna2_1$, $Pca2_1$, $Pbca$, $Pbcn$, Pc, $Pnna$, $Pccn$, $Pbcm$, $Pnnm$, $Pmmn$, $Pnma$, $C2/c$, Cc, $C2/m$, $C2$, $C222_1$, Cm and $R3$), using the gas-phase conformation as rigid and generating values for the lattice variables, X, and position and orientation of molecules in the cell with a low discrepancy sequence. This led to the identification of 160 distinct structures, amongst which the expected experimental form ranked 35th from the global minimum (Fig 2a).

In Stage 2, the key approximation in the electrostatic model used in Stage 1 was removed by replacing the atomic charges with an anisotropic distributed multipole model of rank 4, computed for the gas-phase conformation from the isolated-molecule charge density. Rigid-body lattice energy minimisations were performed for all 160 structures from Stage 1, with the experimental form now ranked 10th in stability from the global minimum (Fig 2b). The extent of re-ranking indicates the strong dependence of the lattice energy on the electrostatic model used during minimisation.

a) Stage 1
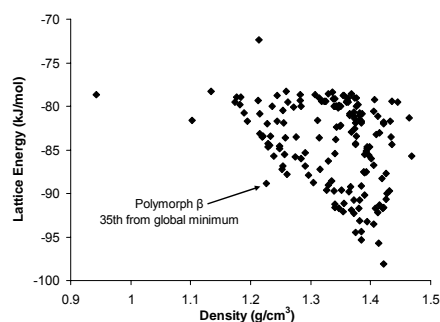
b) Stage 2
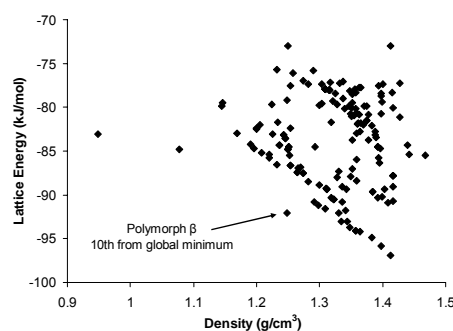
c) Stage 3
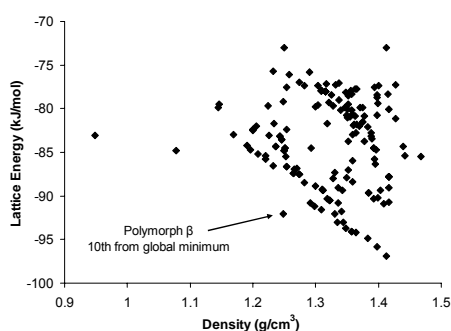
d)
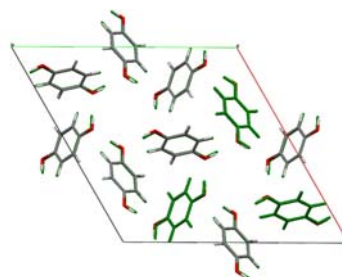
Figure 2 Lattice energy - density landscapes obtained for the three stages of the crystal structure prediction methodology (a, b and c respectively). The minimum obtained when the experimental polymorph is minimised with the same computational model is indicated on all graphs. The overlay of the experimental and Stage 3 minimised unit cell is also shown (d).

In the final stage, the 100 most stable structures from Stage 2 were re-minimised by allowing the two hydroxyl torsion angles in the molecule to change in response to the lattice forces. By lifting the rigid-body approximation, it was possible to obtain more stable (and more dense) packing arrangements (Fig 2c). The reproduction of the experimental unit cell is exceptional (Fig 2d) - the root mean square deviation of the 15-

Ab Initio Crystal Structure Prediction of Flexible Molecules

molecule coordination sphere between the predicted and minimised structure is 0.06 Å. The extent to which each structure is stabilised differs, leading to some reordering in stability: the experimental structure is ranked 9th from the global minimum. This indicates the existence of uncharacterised polymorphs, or, more likely, limitations in the repulsion-dispersion potential used.

The refinement of such a large set of structures in Stage 3 was only possible by the use of LAMs and LAM databases during lattice energy minimisation. The gain in computational performance is discussed in the following section.

3.2. Analysis of computational performance

Each minimisation in Stages 1 and 2 is relatively quick, requiring only a few CPU minutes on an Intel Xeon E5462 2.8 GHz processor. The main cost is in Stage 3, where conformationally dependent models for the intramolecular energy and charge density require multiple quantum mechanical evaluations during lattice energy minimisation. The typical cost for a single lattice energy minimisation in Stage 3 is of the order of CPU hours, without the use of LAM databases. The cumulative number of minimisation iterations as a function of the number of structures analysed for Stage 3 is shown in Fig 3a (solid line). Each of the 1400 iterations requires an evaluation of the intramolecular energy and of the charge density for the current molecular conformation, making it prohibitive to use algorithms based solely on explicit QM evaluations [13]. In the proposed algorithm, the intramolecular energy and the charge density for all conformations within a region around a reference conformation are estimated using a LAM constructed using just one QM evaluation. Since the molecular conformation does not change significantly from iteration to iteration, the same LAM can be re-used for several iterations, greatly reducing the number of QM calculations that would otherwise have to be performed. If the minimisation algorithm brings the molecular conformation outside the region of LAM validity (5° for torsions here), a new LAM is constructed through a new QM evaluation. The number of QM evaluations actually required as a function of the number of structures analysed is shown in Fig 3a (dashed line). The number of QM evaluations has decreased by a factor of 3 through the use of LAMs.
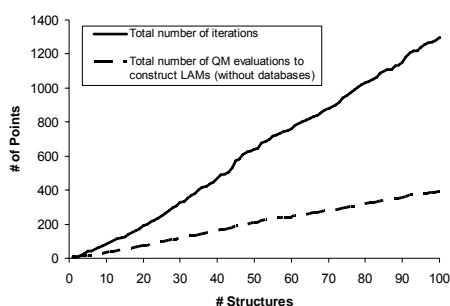
a)   b)
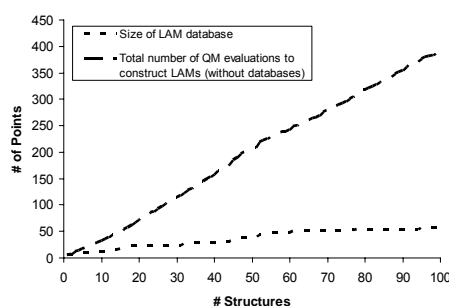
Figure 3 a) Total number of Stage 3 minimisation iterations (solid) and number of QM evaluations when LAMs are used (dashed) as a function of the number of crystal structures minimised; b) Number of QM evaluations using LAMs (dashed) and number of QM evaluations using LAMs and a database (dotted) as a function of the number of crystal structures minimised in Stage 3.

Because the molecule was treated as rigid in Stages 1 and 2, the initial conformation for all 100 structures in Stage 3 is the same. Thus, once a QM calculation is performed for the first structure and the resulting LAM stored in a database, the same LAM can be

used to initialise the minimisation of the 99 remaining structures. Re-use of additional LAMs from the database is also possible during the minimisations. The number of QM evaluations needed during the Stage 3 minimisation of 100 structures without the use of LAM databases is shown as a dashed line in Fig 3b. The actual total number of *ab initio* calculations as a function of the number of structures analysed is shown as a dotted line. For this model system, only 60 QM evaluations were performed over the 1400 iterations in Stage 3. By using LAMs and molecule-specific LAM-databases, the computational cost is kept manageable during the multi-stage crystal structure prediction methodology and a large set of hypothetical structures is assessed with the most accurate lattice energy minimisation method (Stage 3).

4. Conclusions and future work

An extensive and efficient methodology to locate thermodynamically stable crystal structures of flexible molecules using quantum mechanical, conformation-dependent atomic multipole moments and intramolecular energies has been presented. The use of local approximate models and databases allows the accurate minimisation of a large set of crystal structures with manageable computational cost. The methodology has been illustrated by identifying hydroquinone polymorph β but can easily be applied for predicting the crystal structure of multi-component systems (cocrystals, salts and solvates). Such, applications can improve our understanding of crystal formation.

The main shortcoming of the approach at present is the limited handling of molecular flexibility in Stage 1. The conformational energy is interpolated from a pre-computed QM grid covering the entire range of the flexible degrees of freedom. The calculation of such a grid becomes very expensive when more than 4 flexible degrees of freedom are considered. In principle, this limitation can be overcome by extending the use of local approximate models into the search stage. The LAM databases that will be generated in Stage 1 can then be re-used in Stage 3 keeping computational cost manageable. These ongoing developments will allow the modelling of large and flexible systems.

5. Acknowledgements

References

[1] J. Bernstein, Polymorphs in Molecular Crystals, Clarendon Press, Oxford, (2002)
[2] G. M. Day et al., Acta Crystallogr., Sect. B: Struct. Crystallogr. Cryst. Chem., 65 (2009) 107
[3] S. L. Price, Int. Rev. Phys. Chem., 27 (2008) 541
[4] B. P. van Eijck et al., J. Comput. Chem., 22 (2001) 805
[5] P. G. Karamertzanis and C. C. Pantelides, Mol. Phys., 105 (2007) 273
[6] D. J. Willock et al., J. Comput. Chem., 16 (1995) 628
[7] A. J. Stone, J. Chem. Theory Comput., 1 (2005) 1128
[8] G. M. Day et al., Cryst. Growth Des., 5 (2005) 1023
[9] C. S. Adjiman and A. Galindo (eds.), Process Systems Engineering. Volume 6: Molecular Systems Engineering, WILEY-VCH, Verlag GmbH, (in press)
[10] S. V. Lindemann et al., Cryst. Struc. Commun., 10 (1981) 1173
[11] M. J. Frisch et al., Gaussian 03, Gaussian Inc. Wallingford, CT, (2003)
[12] D. S. Coombes et al., J. Phys. Chem., 100 (1996) 7352
[13] P. G. Karamertzanis and S. L. Price, J. Chem. Theory Comput., 4 (2008) 522

Pyrolysis of Naphtha Feedstocks: Automatic Generation of Detailed Kinetics and Lumping Procedures

Mario Dente[a], Eliseo Ranzi[a*], Giulia Bozzano[a], Sauro Pierucci[a]
Florian I. Kleinendorst[b], Marco W.M. van Goethem[b]
[a]*Politecnico di Milano, CMIC Department Piazza Leonardo da Vinci, 32, 20131 Milano, Italy*
[b]*PYROTEC - TECHNIP Zoetermeer (The Netherlands)*
* *Corresponding author: Eliseo.Ranzi@polimi.it*

Abstract

This paper analyses and revises the critical steps in the computer generation of complex reaction schemes for pyrolysis reactions of naphtha feeds. Moving from the pioneering experience of the well known SPYRO kinetic model, the critical steps in the simplification and lumping procedures are explained, critically discussed and partially removed. The availability of parallel processors and always more efficient computer facilities allows enlarging the description details of the overall reacting system. With reference to typical naphtha feedstocks, different description levels of feed characterization and different simplifications of the reaction network can be used. Namely, a large number of branched paraffin isomers are used to characterize the feed and the effect of possible refinery treatments of the naphtha feed are discussed. Moreover, the complex reaction scheme of hydrocarbon pyrolysis needs to be simplified both in terms of decomposition and recombination reactions, in order to avoid an excessive increase of the overall dimension of the mechanism, without significant improvements in the obtained results. As a result of this kinetic study, the SPYRO kinetic scheme is further extended to cover new possible feeds and more complex situations. Several examples illustrate the advantages and the limits of the proposed approach

Keywords: Pyrolysis, detailed kinetics, ethylene, steam cracking

1. Introduction

Feedstocks ranging from light gases up to naphthas and gas oils are converted into valuable products at high temperatures in tubular coils in steam cracking furnaces (Dente et al. 1970; Froment, 1992). The complexity of this system is due both to the large number of elementary reactions and to the difficulty of properly characterizing the reacting mixture. The main focus of this paper is to critically review and discuss the main simplifications applied during the development and validation of these detailed kinetic schemes (Dente et al. 1992; Dente et al. 2007). Apart from the analysis of the chemistry involved in the process, it is also critical to characterize the internal composition of the naphtha feed.

2. Kinetic Mechanism

Free radical reactions dominate the thermal degradation of hydrocarbons. Only radical kinetic schemes can provide reliable descriptions of the pyrolysis process (Benson 1976; Dente et al., 2007). Radical chain propagation process is simply summarized with the following reaction classes:

2.1. - Initiation and termination reactions

$$CH_3\text{-}CH_2\text{-}CH_3 \Leftrightarrow CH_3\bullet + C_2H_5\bullet$$
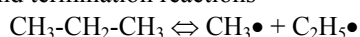
Unimolecular initiation reaction involves the breaking of a covalent C-C bond with the formation of two radicals. This initiation process is sensitive to the stability of the formed radicals. Total radical concentration in the reacting system is controlled by radical initiation and recombination reactions.

2.2. - Propagation reactions

2.2.a - H-abstraction reactions

$$R\bullet + CH_3\text{-}CH_2\text{-}CH_3 \Leftrightarrow RH + C_3H_7\bullet$$

Once initiation reactions generate the propagating radicals, the primary and main decomposition products are well explained by the H abstraction reactions and the subsequent, fast decomposition of primary radicals. Kinetic parameters of these reactions are mainly function of the H abstracting radical and of the site from which H-atom is removed.

2.2.b- β-decomposition and dehydrogenation reactions and reverse radical addition reactions.

$$\bullet CH_2\text{–}CH_2\text{-}CH_3 \Leftrightarrow CH_2\text{=}CH_2 + \bullet CH_3$$
$$CH_3\text{-}\bullet CH\text{-}CH_3 \Leftrightarrow H\bullet + CH_2\text{=}CH\text{-}CH_3$$
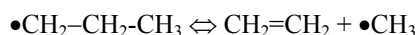
At high temperature, β-decomposition reactions constitute the prevailing fate of alkyl radicals. At high pressure and high concentration of unsaturated species, reverse radical addition reactions also become significant.

2.2.c- Alkyl radical isomerization reactions

$$CH_2\bullet\text{–}CH_2\text{-}CH_2\text{-}CH_2\text{-}CH_3 \Leftrightarrow CH_3\text{-}CH_2\text{-}CH_2\text{-}CH\bullet\text{-}CH_3$$
$$CH_2\bullet\text{–}(CH_2)_3\text{-}CH_2\text{-}CH_3 \Leftrightarrow CH_3\text{–}(CH_2)_3\text{-}CH\bullet\text{-}CH_3$$
$$CH_2\bullet\text{–}(CH_2)_4\text{-}CH_2\text{-}CH_3 \Leftrightarrow CH_3\text{–}(CH_2)_4\text{-}CH\bullet\text{-}CH_3$$
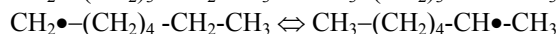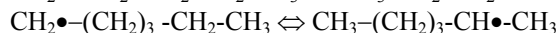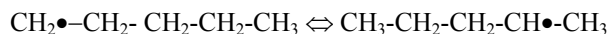
At low temperature, large radicals can isomerize through 5, 6 and 7-membered ring intermediates, before decomposition.

2.3. Concerted path molecular reactions

$$CyC_6H_{12} \Leftrightarrow CH_2\text{=}CH\text{-}CH_2\text{-}CH_2\text{-}CH_2\text{-}CH_3$$
$$CH_3\text{–}CH\text{=}CH\text{-}CH_3 \Leftrightarrow CH_2\text{=}CH\text{-}CH_2\text{-}CH_3$$
$$CH_2\text{=}CH\text{-}CH_2\text{-}CH_3 \Leftrightarrow CH_2\text{=}CH\text{-}CH\text{=}CH_2 + H_2$$
$$CH_2\text{=}CH\text{-}CH_2\text{-}CH_2\text{-}CH_3 \Leftrightarrow CH_3\text{-}CH\text{=}CH_2 + CH_2\text{=}CH_2$$
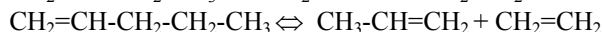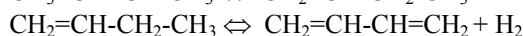
Concerted path molecular reactions, via four and six centers, can also play a significant role. Diels-Alder, cyclo-alkane and olefin isomerization as well as dehydrogenation and 'ene' decompositions are typical examples of this reaction class.

2.4. Successive condensation reactions

Once significant amounts of ethylene and propylene are formed, vinyl, allyl and benzyl radicals are present in the reacting system. Successive addition and condensation reactions of unsaturated species explain the formation of benzene and heavier species. These reactions decrease the overall ethylene selectivity. The interactions of unsaturated species are the first hierarchical step and the core of the pyrolysis mechanism. In fact,

high temperature pyrolysis of large hydrocarbons rapidly gives rise to small radicals and species, and their interactions are common ground shared by all pyrolysis systems.

Figure 1 shows the primary propagation reactions of n-decane and itself confirms the complexity of the pyrolysis mechanism. All the different H-abstracting radicals produce five isomers of n-decyl radical (nC10H21). These radicals can then isomerize and/or decompose. In the temperature range of major interest for steam cracking process, decomposition prevails at high temperatures, while isomerizations dominate at low temperatures.

3. Automatic generation of kinetic schemes and lumping procedures

The manual compilation of the whole set of reactions becomes unmanageable, particularly when the hydrocarbon molecular weight is increased. Therefore, it is convenient to classify the reactions and to apply an automatic generation of the reaction scheme on the basis of a small set of reference kinetic parameters. As an example, the resulting mechanism of n-decane pyrolysis includes 5 initiation reactions, 5 generic H-abstraction reactions and 35 primary propagation reactions of the 5 n-decyl radicals. Figure 2 shows the successive isomerization and decomposition reactions of lower alkyl radicals formed via β-decompositions. The number of elementary reactions, and reacting species, rapidly increases with the number of C-atoms, mainly in the case of cycloalkane and alkene components.
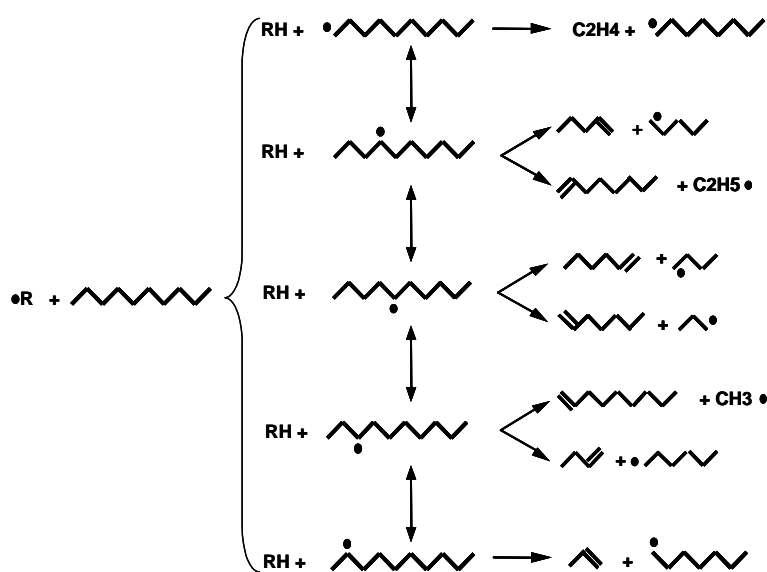
Figure 1. H abstraction reactions of n-decane.

3.1 Lumping of reactions (single event micro kinetics)

Due to the huge number of possible reactions and intermediates products, it is a need to simplify the pyrolysis mechanism. Fortunately, the monomolecular reactions of large alkyl radicals are much faster than the bimolecular ones. Thus, the μ-radical hypothesis allows assuming the pseudo steady state assumption (PSSA) for large radicals, so that the overall reaction scheme is drastically reduced (Dente and Ranzi, 1983; Van Geem et al., 2008). Alkyl radicals containing more than 4-5 C atoms are supposed to be

instantaneously transformed (via isomerisation and decomposition reactions) into their final products, without loss of model accuracy. Isomerisation and decomposition reactions largely prevail on possible H-abstraction, recombination and addition reactions. Thus large sections of the decomposition scheme are reduced to simpler equivalent or lumped reactions.
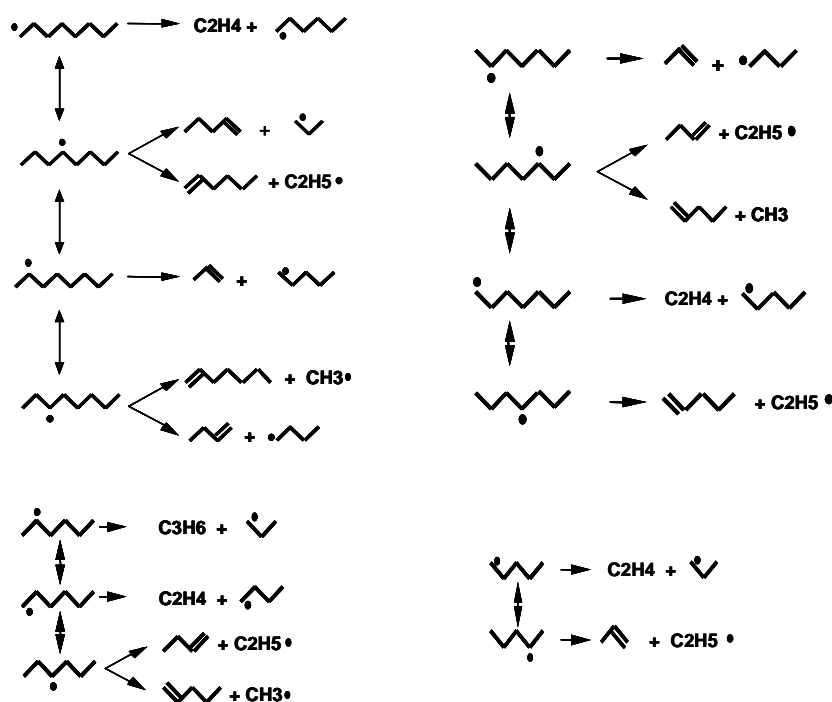
Figure 2. Successive isomerization and decomposition reactions of alkyl radicals.

The kinetic generator, MAMA program, manages these calculations and produces equivalent stoichiometries that compress and lump several reactions into a single equivalent reaction whose apparent stoichiometry is only a relatively weak function of cracking temperature. The elementary H abstraction reactions of n-decane shown in Figure 1 and 2 are thus lumped into a single equivalent reaction whose stoichiometry, evaluated at 1040 K, simply becomes:

$$R\bullet + n C_{10}H_{22} \Rightarrow RH + .0205 \; H\bullet + .0803 \; CH_3\bullet + .2593 \; C_2H_5\bullet + .406 \; nC_3H_7\bullet + .234 \; 1C_4H_9\bullet$$
$$+ .3785 \; C_2H_4 + .3127 \; C_3H_6 + .2114 \; 1\text{-}C_4H_8 + .1870 \; 1\text{-}C_5H_{10} + .1815 \; 1\text{-}C_6H_{12}$$
$$+ .1461 \; 1\text{-}C_7H_{14} + .1284 \; 1\text{-}C_8H_{16} + .0540 \; 1\text{-}C_9H_{18} + .0025 \; 1\text{-}C_{10}H_{20} + .0006 \; 2\text{-}C_5H_{10}$$
$$+ .0012 \; C_6H_{12}s + .0013 \; C_7H_{14}s + .0005 \; C_8H_{16}s + .0100 \; C_{10}H_{20}s$$
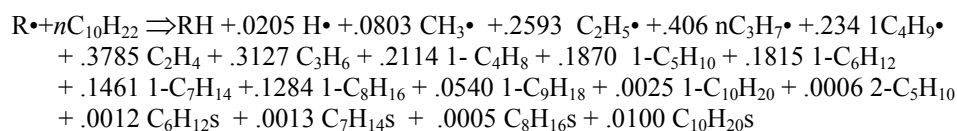
MAMA program is more than a simple mechanism generator. It may lump isomerization and decomposition reactions of large radicals, by solving the overall system and evaluating primary distribution products according to PSSA of intermediate

radicals. The net result is the generation of equivalent or lumped reactions that drastically simplify the reacting system, in terms of both species and reactions.

3.2 Naphtha Characterization and Lumping of Species

Light and heavy naphthas are mostly aliphatic but they also contain significant amounts of cyclo-alkanes and aromatics. Detailed description of naphtha feed involves several isomers, mainly in branched and cyclo-alkane fractions.

There are significant differences in the potential pyrolysis yields of major olefins from structural isomers. Table 1 shows the variability of these yields from the pyrolysis of different isomers of branched alkanes C8, at the same typical operating conditions of steam cracking process. Ethylene yields can span from less than 15% from tri-methylpentane to more than 30% from the three mono-methylheptanes..

Table 1: Yields predictions from pyrolysis of different branched C8 isomers.
(Lumping Temperature 1000 K).

(wt. %)	2methyl C7	3methyl C7	4methyl C7	23dimethyl C6	234tri-methylC5	3ethylC6
H2	0.79	0.75	0.81	0.78	0.82	0.86
CH4	12.73	14.45	13.00	16.45	18.48	14.83
C2H4	30.84	30.44	32.16	22.28	14.95	31.71
C3H6	21.44	18.88	21.61	21.22	20.98	14.94
13C4H6	4.54	5.58	5.28	6.69	7.12	7.16
1C4H8	2.38	2.81	2.01	2.18	2.15	2.92
2C4H8	0.76	1.45	0.57	2.41	3.83	0.88
iC4H8	5.56	2.20	1.88	2.04	3.63	1.31
C5-	84.75	82.46	83.97	79.09	76.00	81.69

The number of species in naphtha feeds is large and the number of elementary steps in pyrolysis mechanism amounts to more than 10^5 (Van Geem et al., 2008). In spite of the different origin of the naphtha feeds, there are evident regularities with regard to their composition. The distribution of structural isomers is largely independent of the origin of the feedstock and it is possible to describe the naphtha feed with only a limited number of pseudo-components. Mono methyl-alkanes prevail on di-methyl- and ethyl-alkanes. Tri-methyl-alkanes are less abundant and quaternary C atoms are of very limited importance. On this basis, it is possible to empirically derive lumped components with statistically defined internal isomer distribution. Thus, structural isomers of large hydrocarbons are conveniently grouped into equivalent components. One lumped component is assumed for each number of C-atoms and per class of components (branched alkanes, cyclo-alkanes and alkyl-aromatics) in the naphtha fraction. Lumped species react in a similar way and their internal distribution remains almost unchanged along the reaction process. Each pseudo-component consists of structural isomers and only one continuity equation needs to be considered.

The equivalent reactions of the different pseudo-components are obtained by averaging the equivalent reactions of the different isomers, taking all the elementary reactions of these isomers into account. As a result of the combination of PSSA of intermediate

radicals and the lumping of different isomers into the lumped pseudocomponent {isoC8}, the resulting lumped H abstraction reaction for the standard {isoC8} pseudo-component simply becomes:

R• + {isoC8} =   RH + .0167 H + .1842 CH3• + .218 C2H5• + .1835 nC3H7•+ .1296 iC3H7•
+ .0671 i-C4H9• + .1415 1C4H9• + .0592 2C4H9• + .1628 C2H4 + .2499 C3H6 + .107 1C4H8
 + .0414 2C4H8 + .1417 iC4H8 + .0704 1C5H10 + .0326 2C5H10 + .0510 2me-1C4H8
 + .0599 3me-1C4H8 + .0103 2me-2C4H8 + .009 1C6H12 + .0214 2C6H12 + .0047 3C6H12
 + .1012 me-C5H10s + .0155 1C7H14 + .1402 C7H14s + .0137 C8H16s

 .. . . …… . … . …. . .. . ……… .. . .
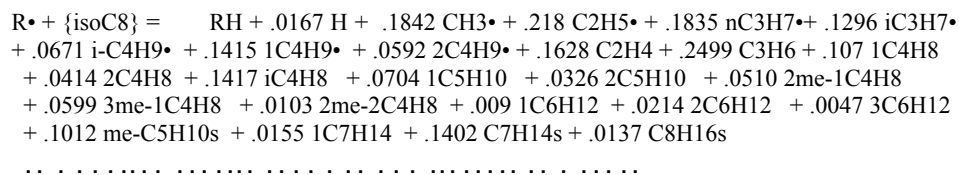
4. Adaptive kinetic schemes and flexible lumping.

The simplifications applied to the lumping of species and to reaction products, partially restrict the validity range of the overall kinetic model. It is viable to remove the above simplifications, without increasing the dimension of the overall model.

Although the stoichiometries of lumped components are temperature-dependent, they are tailored to actual cracking temperatures through proper reaction tables. Thus the kinetic model easily derives a temperature dependent stoichiometry based on three or more reference temperatures.

Furthermore, the fixed weight factors of isomers in lumped component definition makes impossible to describe the decomposition of structural isomers or particular mixtures. The correct approach is to determine the internal weight factors from the analytical composition of naphtha. In this way, new pseudo-components are defined for each feedstock. This approach has been applied and lumped components are modified according to naphtha composition, when available.

The adaptive kinetic scheme becomes more flexible to particular feedstocks and to different operating conditions. Several examples of industrial interest already shown limits and advantages of these extensions of the overall pyrolysis model.

5. Acknowledgements

References

Benson, S.W. (1976) "Thermochemical Kinetics (2nd ed)." John Wiley and Sons, New York

Dente, M., Ranzi, E., Antolini, G., Losco F. (1970) *97th EFCE Symposium*, Florence

Dente, M., Ranzi, E. (1983) Chapter 7   in 'Pyrolysis: Theory and Industrial Practice' (L.F.Albright, B.L.Crines, W.H.Corcoran Eds), Academic Press: San Diego

Dente, M., S. Pierucci, E. Ranzi, G. Bussani (1992) *Chem. Eng. Sci*. 47, 2629

M. Dente, G. Bozzano, T. Faravelli, A. Marongiu, S. Pierucci, E. Ranzi *Advances in Chemical Engineering* (G. Marin Ed.) vol. 32: 52-166, Elsevier Inc.

Froment G.F. (1992), *Chem. Eng. Sci*. 47, 2163.

Van Geem, K.M., M.F. Reyniers, G.B. Marin (2008) *Oil & Gas Science Technology* –IFP, 63 (1).

Detailed kinetics in the mathematical model of fixed bed gasifiers

R. Grana, S. Sommariva, T. Maffei, A. Cuoci, T. Faravelli, A. Frassoldati,
S. Pierucci, E. Ranzi*

Politecnico di Milano, CMIC Department Piazza Leonardo da Vinci, 32, 20131 Milano, Italy
* Corresponding author: eliseo.ranzi@polimi.it

Abstract

This work presents a general mathematical model of a fixed bed gasifier, where both the transport resistances and the chemical kinetics are important at the reactor and the particle scale. Pyrolysis and gasification of the solid fuel (coal, plastics, biomasses or wastes) is only the first step of the whole process. Successive gas phase reactions of the released species are described with a detailed kinetic scheme. Furthermore, an accurate description of heat and mass transport between gas and solid phases allows a proper characterization of combustion and gasification of the solid fuel. This work summarizes several facets of this problem with examples and validations. Examples relating to coal and biomass gasification allow evaluating the feasibility and limitations of the proposed approach.

1. Introduction

The environmental concerns towards the combustion of fossil fuels drive the interest in gasification processes of biomasses, coals, plastics, and refuse derived fuels (RDF). Gasification of solid fuels is nowadays a promising alternative to direct combustion, both electric and thermal energy are viable products, together with chemicals.

The mathematical description of such processes is rather difficult due to the complex phenomena involved, such as modelling solid devolatilization, gas phase reactions and gas-solid interactions. However, it is worth to emphasize the importance of these models, which could ease the scale-up and the optimization of the gasifier and could improve the understanding of the whole process. In this work, we proposed the methodology for solving such problems, showing the approach for the main facets involved in solid fuels gasification. Finally, an application example of a fixed bed gasifier model is provided, emphasizing the thermal features of the reactor as well as the role of feedstock characterization.

2. Devolatilization of Coals, Plastics, Biomasses and Refused Derived Fuels (RDF)

2.1. Solid fuel characterization

The different solid fuels are described with a limited number of reference compounds and for each of them a multistep kinetic scheme was developed. While plastics, such as poly-ethylene (PE), poly-propylene (PP) and poly-styrene (PS) have a very well-defined structure and composition, the available information about coals and biomasses is usually limited to the elemental composition in terms of C/H/O. Degradation of plastics

was already discussed by [Marongiu et al., 2007]. Biomass composition , if biochemical analysis is available, is simply defined in terms of humidity, ash, cellulose, hemicelluloses and lignin. If only the elemental analysis is available, then a suitable combination in terms of reference species is derived by atomic balance. [Ranzi et al., 2008]. Similarly, three reference species ($COAL_1$, $COAL_2$ and $COAL_3$) are used to describe the composition and reactivity of the different coals. $COAL_1$ ($C_{12}H_{11}$) , together with pure carbon (CHARC), is useful to describe anthracitic coals with different degree of aromaticity. $COAL_2$ ($C_{14}H_{10}O$) lies in the middle of bituminous coals, while $COAL_3$ is highly oxygenated ($C_{12}H_{12}O_5$) and is representative of lignitic coals [Sommariva et al., 2010]. In a very similar way, Refused Derived Fuels (RDF), typically with heating values of 4500-5000 kcal/kg, are described in terms of a proper combination of plastic wastes (15-30%), lignocellulosic material (30-50%), ash, and humidity. Fig. 1 shows typical compositions of different RDF in terms of C and H (wt% ash free)
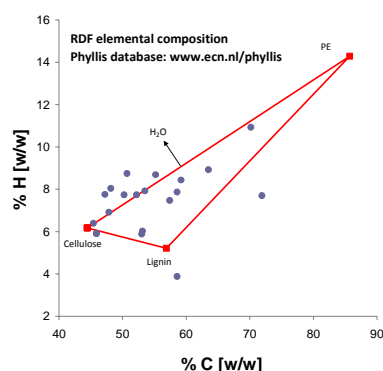
Figure 1 Typical compositions of different RDF in terms of C and H (wt% ash free)

2.2. Solid fuel devolatilization

The devolatilization of solid particles (mixture of reference compounds) is always considered as a linear combination of the pure component devolatilization. The overall kinetic scheme of RDF devolatilization is simply the combination of the multistep devolatilization models of Biomass [Ranzi et al., 2008], plastic [Marongiu et al., 2007] and coal [Sommariva et al., 2010]. The peculiarity of this approach is that all these schemes consist of a limited number of devolatilization reactions, which are able to describe not only the solid residue, but also the detailed composition of released gas and tar species. Thus, the novelty of this kinetic model, when compared with the majority of the available ones in the literature, is the effort and the challenge to define lumped stoichiometries of the devolatilization reactions with a detailed characterization of the gas and the tar fractions. Thus, Fig. 2 shows the lumped reference components of tar products released by the reference coals.

Operating conditions affect the devolatilization selectivity and yields; in particular this multistep approach allows also to describe the chemical and morphological evolution of the solid phase in terms of composition and reactivity.

Detailed kinetics in the mathematical model of fixed bed gasifiers
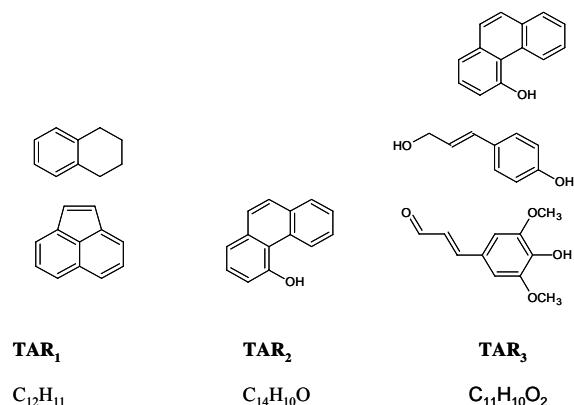
Figure 2 Lumped reference components of tar products from reference coals

3. Combustion and gasification reactions of residual char

The gasification and combustion of char, i.e. the set of heterogeneous reactions of oxygen and steam with the solid residue coming either from coal, biomass or plastics, are responsible for the autothermic behaviour of the whole gasification process. These reactions and related kinetic parameters are summarized in Table 2 [Groeneveld and van Swaaij, 1980, Kashiwagi and Nambu, 1992].

Table 1: Char gasification and combustion reactions. Units are: kmol, m^3, K, kcal, s.

	k
$CHAR + O_2 \rightarrow CO_2$	$5.7 \times 10^9 \exp(-38200/RT) [O2]^{0.78}$
$CHAR + 0.5\,O_2 \rightarrow CO$	$5.7 \times 10^{11} \exp(-55000/RT) [O2]^{0.78}$
$CHAR + H_2O \rightarrow CO + H_2$	$7.9 \times 10^9 \exp(-52000/RT) [H_2O]^{0.7}$

4. Detailed kinetics of secondary gas phase reactions

The volatile components released during the pyrolysis undergo successive decomposition or combustion reactions in the surrounding gas phase. As clearly stated by Li et al 2004, a pure thermodynamic or equilibrium approach is not able to properly predict the composition of the gas mixture in a gasifier. The equilibrium approach largely over predicts the heating value of flue gases; in particular, higher amount of hydrogen is predicted, while only trace amounts of methane and hydrocarbons are obtained. On this basis, it is clear the utility to develop a detailed or semi-detailed kinetic scheme for the gas phase.

An existing kinetic scheme for pyrolysis and oxidation of hydrocarbon species, already validated in a wide range of operating conditions [Ranzi et al., 2001], has been extended to properly describe successive gas phase reactions of released species. Due to the modularity of the detailed kinetic scheme, it is only necessary to describe the primary initiation, decomposition and H abstraction reactions of the new species. The overall gas-phase kinetic model, together with thermodynamic properties, is available at the website http://www.chem.polimi.it/CRECKModeling/.

The number of new species included in the gas phase is a compromise between accuracy and computational efforts, maintaining the ability of the model to describe the

gas composition and reactivity in a wide range of operating condition: in particular tar species are grouped into pseudo-components representative of a set of similar species with similar reactivity.

As already discussed in a previous paper [Sommariva et al., 2010], the secondary gas phase reactions can also explain relevant temperature and pressure effects during coal devolatilization, not only related to the solid phase reactions.

5. Mathematical model of fixed bed gasifier

The mathematical model of the fixed bed gasifier consists of two models. The first one at the particle scale and the latter at the reactor scale. This approach is discussed in details elsewhere [Pierucci and Ranzi, 2008; Dupont et al., 2009]. The particle model provides an internal description in order to account for intraparticle heat and mass resistances. The accuracy of this description depends on the number of discretization sectors (N). This feature becomes fundamental for the analysis of the gasification of thermally thick solid fuel particles. The equation 1 summarizes the main terms of material and energy balances for the particle:

$$\begin{cases} \dfrac{dm_{j,i}}{dt} = J_{j-1,i} - J_{j,i} + R_{j,i} \\[2ex] \dfrac{d\sum\limits_{i=1}^{NCP} m_{j,i} Cp_{j,i} T_j}{dt} = JC_{j-1} + \sum\limits_{i=1}^{NCP} J_{j-1,i} h_{j-1,i} - JC_j - \sum\limits_{i=1}^{NCP} J_{j,i} h_{j,i} + HR_j \end{cases} \tag{1}$$

where m is the mass and j refers to the jth particle sector and i to the ith component. t is the time variable and J the mass flux (Fick and Darcy's laws contributions) and R the kinetic term. Analogously, the energy balance settle the particle temperature T and accounts for conduction contribution (JC) as well as the convective contribution (Jh). HR is the heat of reactions involved in the solid phase.

This particle model together with the kinetic models previously described is embedded in the fixed bed reactor model (equation 2). The reactor is considered as a series of elemental units (Fig. 3) which exchange mass and heat to each others. The single unit accounts for gas-solid interactions with particular attention to the inter-phase resistances. The balance equations (2) refer to a perfectly stirred reactor in dynamic conditions, where $R_{g,i}$ are the gas phase reactions and $G_{0,i}$ and G_i are the inlet and outlet gas feeds [Pierucci and Ranzi, 2008; Dupont et al., 2009].

$$\begin{cases} \dfrac{dg_i}{dt} = G_{0,i} + J_{NR,i}\eta + R_{g,i} - G_i \\[2ex] \dfrac{d\sum\limits_{i=1}^{NCP} g_i Cp_i Tg}{dt} = \sum\limits_{i=1}^{NCP} G_{0,i} hg_{0,i} + \sum\limits_{i=1}^{NCP} J_{NR,i} h_{NR,i}\eta + JC_{NR}\eta + HR_g - \sum\limits_{i=1}^{NCP} G_i hg_i \end{cases} \tag{2}$$

The term $J_{NR,i}\eta$ refers to the gas-solid mass exchange ($J_{NR,i}$) multiplied by the number η of particles in the bed. Four terms contribute to gas phase temperature (Tg) evolution: the enthalpy convective terms (Ghg) between adjacent reactor layers, the gas-solid heat exchange ($J_{NR,i}\eta$), the enthalpy diffusion term ($J_{NR,i}h_{NR,i}\eta$), and finally the contribution due to gas phase reactions (HR_g).

Detailed kinetics in the mathematical model of fixed bed gasifiers

More than one thousand balance equations are obtained, when considering ~10-15 solid species, 100 gas-phase components, 10 reactor layers and 5 discretizations of the solid particle. The numerical problem is structured in tridiagonal blocks. The resulting ODE system is solved by using BzzMath Library (www.chem.polimi.it/homes/gbuzzi/).

6. Application examples
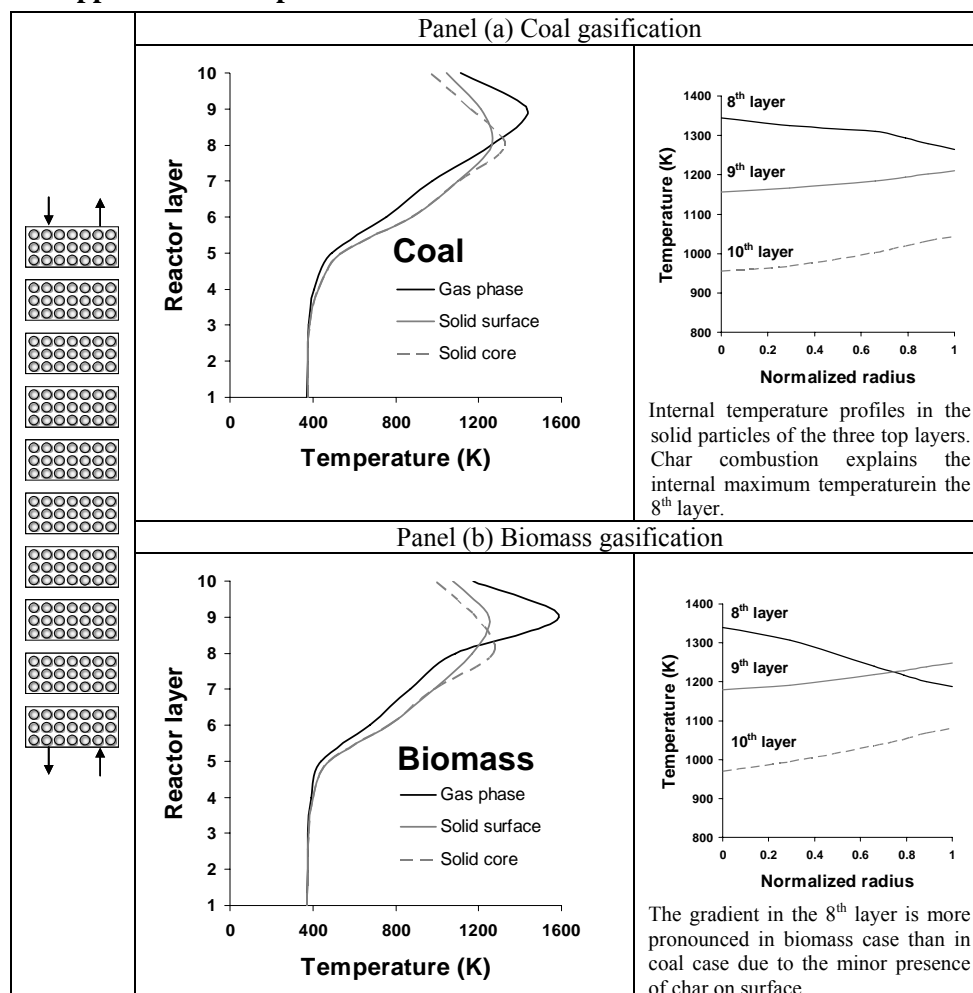
Figure 3: Predicted gas and solid temperature profiles of a counter-current fixed bed gasifier.

Fig. 3 shows the predicted gas and solid temperature profiles of a counter-current fixed bed gasifier fed with biomass (panel a) and with coal (panel b). Solid particles of 3 cm are treated with air at equivalence ratio 0.25-0.4 and with steam/fuel ratio of about 0.30 wt/wt. Typical contact times of gases are of a few seconds while several hours are necessary to treat solid particles. The model characterizes the solid and gaseous streams, giving a proper emphasis to secondary gas phase reactions of tar and gas components. Typical molar composition of flue gases is about 11-15% of H_2, 13-17% of CO, 10-12% of CO_2, 15-20 % of H_2O and ~1-3% of CH_4 and hydrocarbons.

Maximum gas phase temperature is higher when treating biomasses, due to the largest extent of volatilization and the highest oxygen availability. Large temperature gradients are predicted in the first top layers of the bed. Char combustion in the third layer from the top of the gasifier justifies the maximum internal temperature of coal particles (~1100 °C), while this effect is less evident when biomass is treated in the gasifier, due to the lower presence of residual char. Depending on the ash content in the solid fuel, the shrinking of the bed is usually rather larger than 50%.

These simulations require several hours of CPU time, due both to the stiff nature of the gas phase kinetics and to the dynamic approach to the steady solution. For these reasons a further model validation strongly demands for further simplifications in the solid and gas phase description.

7. Acknowledgements

References

A. Cuoci, T. Faravelli, A. Frassoldati, S. Granata, G. Migliavacca, E. Ranzi, S. Sommariva, 2007, A General Mathematical Model of Biomass Devolatilization, 30th Meeting of the Italian Section of the Combustion Institute.

C. Dupont, L. Chen, J. Cances, J.-M. Commandre, A. Cuoci, S. Pierucci, E. Ranzi, 2009, Biomass Pyrolysis: Kinetic Modelling and Experimental Validation Under High Temperature and Flash Heating Rate Conditions, Journal of Analytical and Applied Pyrolysis, 85, pp. 260-267.

M.J. Groeneveld and W.P.M. van Swaaij, 1980, Gasification of Char Particles with CO_2 and H_2O, Chemical Engineering Science, 35, pp 307-313.

T. Kashiwagi and H. Nambu, 1992, Global kinetic Constants for Thermal Oxidative Degradation of a Cellulosic Paper, Combustion and Flame, 88, pp. 345-368.

X.T Li, J.R. Grace, C.J. Lim, A.P. Watkinson, H.P. Chen, J.R. Kim, 2004, Biomass Gasification in a Circulating Fluidized Bed, Biomass and Bioenergy, 26, pp. 171-193.

A. Marongiu, T. Faravelli, E. Ranzi, 2007, Detailed Kinetic Model of the Thermal Degradation of Vynil Polymers, Journal of Analytical and Applied Pyrolysis, 78, pp. 343-362.

S. Pierucci, E. Ranzi, 2008, A General Mathematical Model for a Moving Bed Gasifier,Computer Aided Chemical Engineering, 25, pp. 901-906.

E. Ranzi, A. Cuoci, T. Faravelli, A. Frassoldati, G. Migliavacca, S. Pierucci, S. Sommariva, 2008, Chemical Kinetics of Biomass Pyrolysis, Energy&Fuels 2008, 22, 4292–4300.

E. Ranzi, M. Dente, A. Goldaniga, G. Bozzano, T. Faravelli, 2001, Lumping procedures in detailed kinetic modeling of gasification, pyrolysis, partial oxidation and combustion of hydrocarbon mixtures, Progress in Energy and Combustion Science, 27, pp. 99-139.

S. Sommariva, T. Maffei, G. Migliavacca, T. Faravelli, E. Ranzi, 2010, A Predictive Multi-Step Kinetic Model of Coal Devolatilization, Fuel, 89, pp. 318-328.

Physical properties predictions of indium based solid solution by artificial intelligence techniques

Liliana Bizo, Calin I. Anghel

Babes Bolyai University, Faculty of Chemistry and Chemical Engineering, Department of Chemical Engineering and Oxide Materials Science, 11 Arany Janos Street, 400028 Cluj-Napoca Romania,e-mail: lbizo@chem.ubbcluj.ro

Abstract

Transparent conducting oxides have become an important topic in the field of optoelectronics for various devices and numerous investigations have been carried out recently on these materials. Due to the significant stabilizing effect of introduction of W^{6+} cations in a M_7O_{12} structure, we are interested in the problem of formation a solid solution between $In_4Sn_3O_{12}$ and In_6WO_{12}, implementing the fully compensated cationic substitution: $3Sn^{4+} \rightarrow 2In^{3+} + W^{6+}$. Correlated predictions of properties or trends based on experimental measurements are important factors to reduce the costs of future researches in order to obtain new transparent conducting oxides. To avoid some major drawbacks of any regression techniques a potential support is the integration of artificial intelligence innovative ideas. The paper implements a novel procedure of the artificial intelligence based on support vector machine in a *minimax* approach. The procedure, involving a link between artificial intelligence and materials science, is able to reveal parameters or to predict particular variables, properties or trends. The main goal of the paper is to compare the performance of the procedure with regression techniques and to promote it as an effective technique in material science. Comparative numerical experiments with regression techniques demonstrate the capability of the proposed procedure. Presented numerical experiments reveal the real capacity of this procedure for engineers dealing with material science analyses and beyond these to many others engineering domains.

Keywords: Solid solution, Electrical properties, Optical properties, Regression techniques, Support vector machine, Minnimax approach

1. Introduction

Taking into account the existence of tungstate In_6WO_{12} previously reported [1-3], which shows the significant stabilizing effect due to the introduction of W^{6+} in a M_7O_{12} structure, we are interested in the problem of formation a new solid solution between $In_4Sn_3O_{12}$ and In_6WO_{12} implementing the fully compensated cationic substitution $3Sn^{4+} \rightarrow 2In^{3+} + W^{6+}$. Thus it is possible to be informed about the role played by a cation with empty outer electronic configuration, i.e. $6d^0$, compared to the case of d^{10} ions. In the case of introduction of W^{6+} in $In_4Sn_3O_{12}$ in the form of $In_{4+2x}Sn_{3-3x}W_xO_{12}$ solid solution, stability of the M_7O_{12} structure present throughout the entire homogeneity range is not affected by the presence of an increasing quantity of W^{6+}. But on the contrary the experimental measurements of some major properties as diffuses reflectance and electrical resistivity show a progressive degradation of electrical conductivity and optical properties. Although new scientific achievements on materials have supplied basic knowledge of the underlying phenomena, there remain many

problems where theoretical treatments are cumbersome or dismally lacking. Recent achievements in the field of artificial intelligence improve the investigations in various material science problems [4-6]. Laboratory experiments and measurements to establish properties of solid solution are cumbersome, expansive and time consuming. To avoid these drawbacks we consider opportune to integrate innovative ideas of artificial intelligence to predict particular properties or trends based on reasonable number of experimental data. In order to reduce the number of laboratory experiments, the objective of the present work is to implement a procedure named *minimax decision procedure* able to predict optical property of a new indium based solid solution using a small number of experiments. The procedure is based support vector machine in a *minimax* approach. Today we are unaware of others similar approaches related to material science. The results compared to a robust regression technique (Total Least Squares regression based on Principal Component Analysis) point out the opportunity of such assessments based on artificial intelligence in material science.

2. Problem Statement

Unlike Support Vector Machines for which the closest points to the decision boundary are most important, the *minimax* approach named *minimax probability machine* [7,8] looks at the margin between the means of classes. Thus *minimax* classification is similar to maximum margin classification with respect to the mean of the classes, where a factor that depends on the covariance matrices of each of the classes pushes the threshold towards the class with lower covariance. Moreover by kernel trick - simply mapping data into a higher dimensional feature space, *minimax probability machine* can adapt them to become a non-linear classifier. This is equivalent to a non-linear classifier into original input space. Based on the kernel formulation for *minimax* approach a regression model into feature space named as *minimax probability machine regression* [9] was built as maximising the minimum probability of future predictions being within some bound of the true regression function. The strength of this regression model comes from its ability to represent very high dimensional input space through kernel functions with great resistance to over-fitting. Our procedure named *minimax decision procedure* implements *minimax probability machine regression* into a global decision procedure. The flowchart of the proposed procedure and basic principles of *minimax decision procedure* can be found elsewhere [10,11]. Only for simplicity and to be in agreement with comparative regression technique the performance of the procedure was simply investigated based on the relative errors criterion:

$$\text{RE} = 100 \times (Y_{predicted} - Y_{test} / Y_{predicted}) \quad [\%] \tag{1}$$

To avoid the possibility of over-fitting data the whole experimental database was random divided into two sets: training (*learn* and *test*) and *validation* dataset. A model is produced using only the training data. The *validation* data are then used to check that the model behaves itself when presented with other unseen data. To ensure accuracy and stability of the procedure the *model* is generated based on data also randomly divided into a number of distinct training (*learn* and *test*) subsets. The best case over "*k*" cyclic simulations and the corresponding outputs emerged from the *minimax probability machine regression*, was defined as *the model*. Long random trials don't get improved accuracy or more reliable predictions. Based on some statements [12] the number of trials were limited to $k = 50...100$. Typically ways to choose kernels and kernel parameters are manually or determine it by cross validation. This time an

Physical properties predictions of indium based solid solution by artificial intelligence techniques

empirical but heuristic principle was applied for setting the type of the kernels and kernel parameters that give good performance. The kernel type and parameters that yields to the best performance, assessed by Eq. (1), was put aside and considered for the *model*. The implementation was developed as a user-friendly computer application in MATLAB software environment and works in random cyclic steps. The procedure was conducted in a crude manner, without outliers' detection and no features reduction.

3. Experiments and Predictions

Seven compositions (x) of $In_{4+2x}Sn_{3-3x}W_xO_{12}$ solid solution were prepared by solid state reactions from mixtures of pure In_2O_3, SnO_2 and WO_3 powder oxides in alumina crucibles heated in air. According to experimental procedure [13] pure phases corresponding to the entire range of solid solution $In_{4+2x}Sn_{3-3x}W_xO_{12}$ ($0 \leq x \leq 1$) were isolated.
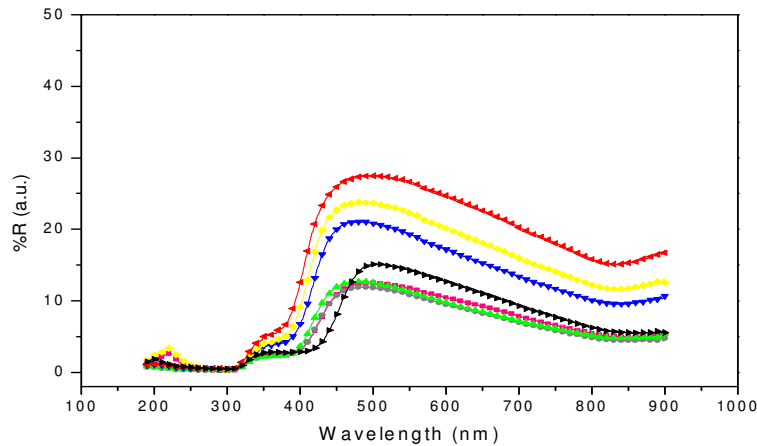
Figure 1. Measured optical reflectance in the compositions x = 0.1 (pink), x = 0.2 (grey), x = 0.3 (green), x = 0.4 (blue), x = 0.5 (yellow), x = 0.6 (red) of the solid solution $In_{4+2x}Sn_{3-3x}W_xO_{12}$ and $In_4Sn_3O_{12}$ (black).

Numerous experimental optical properties as diffuse reflectance spectra (Fig. 1) for these compositions of $In_{4+2x}Sn_{3-3x}W_xO_{12}$ solid solution registered with a double beam spectrophotometer (Cary Varian) show that the optical bandgap moves to higher energies with the introduction of W^{6+}. This increase becomes significant for values of x ≥ 0.4 (80 cm^{-1}) and it is accompanied by strong increase in the percentage of maximum reflectance. Clearly the optical characteristics of TCOs are altered from the higher contents in W^{6+} (composition x ≥ 0.4). The composition x = 0 corresponding to $In_4Sn_3O_{12}$, was also considered for comparison in every experimental measurement. The goal of the application is predictions of new diffuse reflectance values for mentioned compositions samples' of indium-based solid solution. The predictions were related to entire database of experimental conditions and not individual. Experimental parameters of interests were wavelength and composition (x) of every experimental specimen. The performance of the application was investigated based on the relative errors criterion Eq. (1). The application was conducted in comparative manner between our procedure

and mentioned regression technique. Always mentioned comparative procedures were reported to the same identical data set (training and validation). Details of database of the application, the main conditions of simulation and partial results are done in Table 1, Table 2 and Fig. 2. For simplicity we present only predictions of a single specimen of composition x = 0,4. The results for these predictions (Table 2 and Fig. 2) are at least competitive or better with those obtained by mentioned robust regression technique.

Table 1. Values of experimental measured property

Sample composition	Optical reflectance %R [a.u.]		
	Range	Mean	Standard deviation
Experimental parameter	Range of Wavelengths = 190÷900[nm]		
x1 = 0	0.46021÷15.145	7.0682	4.7589
x2 = 0.1	0.18772÷12.464	6.3438	3.9388
x3 = 0.2	0.21395÷11.938	5.745	3.8243
x4 = 0.3	0.31295÷12.736	6.1292	4.0321
x5 = 0.4	0.44137÷21.051	10.887	6.8754
x6 = 0.5	0.42947÷23.795	12.960	7.7634
x7 = 0.6	0.49763÷27.514	15.799	9.4631

- x – content compositions of W^{6+} - defining sample specimen
- predictions and analyse was related only to x5 = 0.4 sample compositions

Table 2. The main conditions and partial results of simulations

	Minimax decision procedure	Regression procedure
Size of training subset	N = 50	N = 50
Size of validation subset	N = 22	N = 22
Range of relative errors on validation subset	-10.7 ÷ 16.7 [%] based on Fig. 2	-53.2 ÷ 26.7 [%] based on Fig. 2
Range of relative errors between new predictions and new experimental measurements	-0.21 ÷ 8.23 [%] based on Fig. 3	-33.14 ÷ 19.43 [%] based on Fig. 3
Wavelengths for new predictions and experimental measurements [nm]: 895, 887, 879, 871, 863, 855, 847, 839, 831, 823, 815, 807, 799, 791, 783, 775, 767, 759, 751, 743.		
Kernel functions for *minimax decision procedure* $$K(\mathbf{x}_i, \mathbf{x}_j) = exp\left(-\left\|\mathbf{x}_i - \mathbf{x}_j\right\|^2 \big/ 2\sigma^2\right)$$		

Note: predictions and analyse was related only to x5 = 0.4 sample compositions.

Based on the accuracy in predictions of *minimax decision procedure* obtained *model* can be used to examine the effect of any individual input on the output parameter, whereas this may be difficult to do experimentally. Also the *model* can correlate thus reduce the number of experimental measurements and substitute them with predicted values related to entire database of experimental conditions (Table 2 and Fig. 3). To include the predictive uncertainty, a plot with error bars at typical 1σ error between the prediction

Physical properties predictions of indium based solid solution by artificial intelligence techniques

and "true" value of the output showing 95% confidence intervals on the diffuse reflectance was preferred (Fig. 2 and Fig. 3).
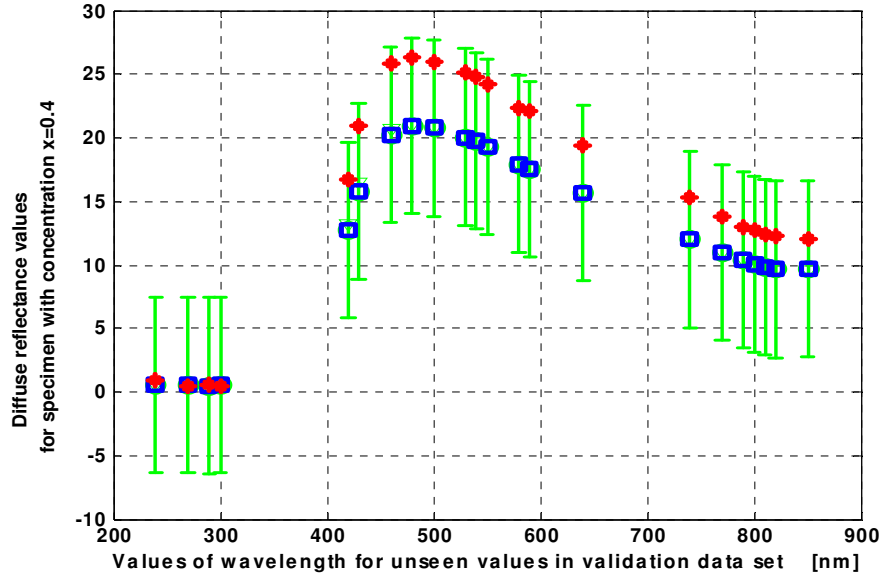
Figure 2. Comparative diffuse reflectance values in validation data set
o Original experimental values; ∗ Values from regression; □ Values from our procedure;
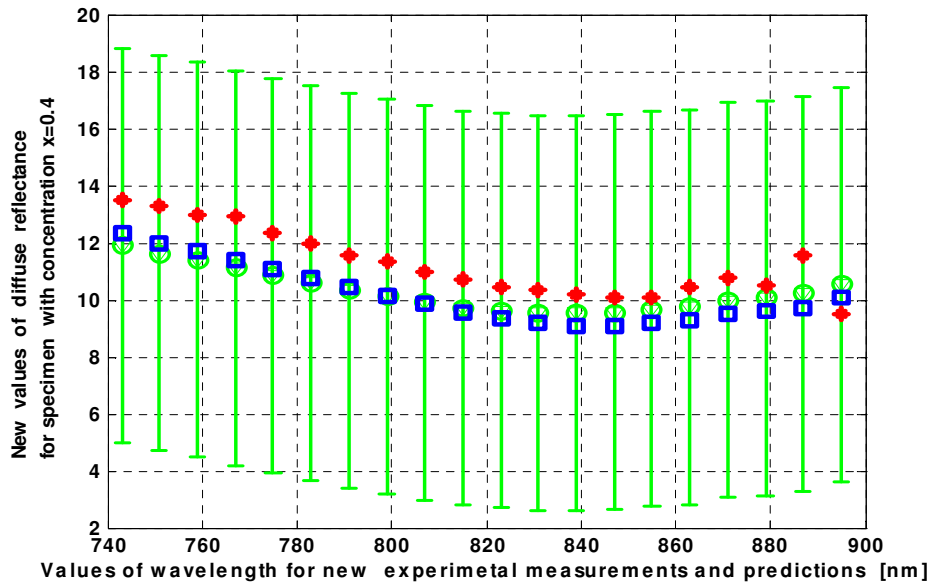
Figure 3. Comparative diffuse reflectance values for new experimental measurements
o New experimental values; ∗ Values from regression; □ Values from our procedure;

4. Conclusions

The applications reveal the accuracy and predictive power of this procedure and the opportunity of such approaches based on artificial intelligence methods in material science. Furthermore, used with an appropriate experimental policy our framework is not limited to reduce the number of laboratory experiments. Future research will focus on establishing other application in material science, for example to reduce the costs of future researches in order to improve the properties or to obtain new transparent conducting oxides.

5. Acknowledgements

References

[1]. W. S. Dabney, N. E. Antolino, B. S. Luisi, A. P. Richard, D. D. Edwards, Thin Solid Films, No. 411 (2002) 192.
[2]. T. Gaewdang, J. P. Chaminade, A. Garcia, C. Fouassier, M. Pouchard, P. Hagenmuller B. Jacquier , Mat. Letters No.18 (1993) 64.
[3]. A. P. Richard, D. D. Edwards, J. Solid State Chem., No. 177 (2004) 2740.
[4]. J. J. Lee, D. K. Kim, S. K. Chang, C. F .M. Nocete, Computational Materials Science No. 44 (2009) 988.
[5]. B. T. Chen, T. P. Chang, J. Y. Shih, J. J. Wang, Computational Materials Science 44 (2009) 913.
[6]. N. S. Reddya, A. K. Prasada Raoa, M. Chakrabortya, B. S. Murtyb, Materials Science and Engineering A No. 391 (2005) 131.
[7]. G. R. G. Lanckriet, EL. Ghaoui, C. Bhattacharyya, M. I. Jordan, in: T. G. Dietterich, S. Becker, Z. Ghahramani (Eds.), Advances in Neural Information Processing Systems No.14 Cambridge MA. MIT Press (2002a) 801. (http://robotics.eecs.berkeley.edu/~gert/)
[8]. G. R. G. Lanckriet, EL. Ghaoui, C. Bhattacharyya, M. I. Jordan, Journal of Machine Learning Research No.3 (2002b) 555. (http://robotics.eecs.berkeley.edu/~gert/).
[9]. T. R. Strohmann, G. Z. Grudic, in: S. Thrun S. Becker, K. Obermayer (Eds.), Advances in Neural Information Processing Systems No.15 (2003) Cambridge MA. MIT Press, 769.
[10]. C. I. Anghel, Al. Ozunu, Chem. Pap. No.60 (6) (2006) 410.
[11]. C. I. Anghel, Al. Ozunu, Computer-Aided Chemical Engineering, No.24, 1211, Elsevier 2007.
[12]. A. Bordes, S. Ertekin, J. Weston, L. Bottou, Journal of Machine Learning Research No.6, (2005) 1579.
[13]. J. Choisnet, L. Bizo, R. Retoux, S. Hebert, B. Raveau, J. Solid State Chem., No. 177 (2004) 3748.

Development of Decision Support System through Multi-Objective Optimization for Ill-posed Problems

Yoshiaki Shimizu, Yasumasa Kato and Takeshi Kariyahara

Department of Production Systems Engineering, Toyohashi University of Technology, 1-1 Hibarigaoka, Tenpaku-cho,Toyohashi 441-8580, Japan, shimizu@pse.tut.ac.jp

Abstract

To ensure customer-defined satisfaction, "Kansei" (emotion) has attracted interest recently in the design of artifacts in today's climate of global competition and short product life cycles. Accordingly, intelligent decision-making through multi-objective optimization (MOP) has been proposed as an efficient methodology for such human-centered manufacturing. From this viewpoint, in this study, we extend the application of a MOP method referred to as MOON^{2R} to an ill-posed problem that involves both qualitative and quantitative performance measures. Additionally, to facilitate portability of the proposed method, especially in multi-disciplinary decision-making environments, we implement the algorithm in an Excel spreadsheet and validate its effectiveness through a case study.

Keywords: Integrated design, Multi-objective optimization, Meta-modeling, Kansei, Excel spreadsheet.

1. Introduction

Developing intelligent technologies characterized by "Kansei" (emotion) has been proposed recently for the optimal design of artifacts. The aim of this approach is to develop an adaptive and human-centered decision support system toward customer-defined satisfaction. From this viewpoint, we aim in this study to develop an intelligent decision system through application of a multi-objective optimization (MOP) method referred to as MOON^{2R} (Shimizu et al., 2005). The method is known to be robust against unsteady and unstable subjective human judgments that characterize MOP problems. MOON^{2R} is also amenable to MOP relying on a particular meta-model. Such meta-models are often used when a reliable physical model cannot be readily obtained for empirical field operations and in the case of emerging technologies (Shimizu and Nomachi, 2008).

We cope with ill-posed MOP by meta-modeling techniques and subjective-oriented approaches such as the Kano method (Kano et al., 1984) and grey theory (Deng, 1989). Additionally, to facilitate its wide and simple application in multi-disciplinary decision-making environments, we propose implementation of the algorithm in an Excel spreadsheet. Finally, in a case study on a rolling machine for plastic sheets, we validate the effectiveness of this approach.

2. Outline of MOP with Meta-Modeling

Generally, MOP is described as follows.

(p. 1)   Min $\boldsymbol{f}(\boldsymbol{x}) = \{f_1(\boldsymbol{x}), f_2(\boldsymbol{x}), \ldots, f_N(\boldsymbol{x})\}$ subject to $\boldsymbol{x} \in X$,

where $\boldsymbol{x}$ is a decision variable vector, X is a feasible region and $\boldsymbol{f}$ is an objective function vector, certain elements of which are in conflict and incommensurable. Recently, in addition to the use of the conventional physical model-based approach, simulation-based approaches have been increasingly applied to cope with complicated artifact design. The present approach is intended to replace time-consuming and/or labor-intensive tasks performed iteratively in the cycle of prototyping, inspection, evaluation and improvement with integrated and structured computer-assisted procedures.
By applying such an idea in MOON^{2R}, we can restate (p. 1) as follows.

(p. 2)   *Max* $V_{\mathrm{RBF}}(\boldsymbol{f}(\boldsymbol{x}), Meta - \boldsymbol{f}(\boldsymbol{x}); \boldsymbol{f}^{\mathrm{R}})$ *subject to* $\boldsymbol{x} \in Meta - X$,

where *Meta-X* is the feasible region that can be partially described by a meta-model. Moreover, V_{RBF} is an overall value function that integrates each objective function involving meta-model Meta-$\boldsymbol{f}(\boldsymbol{x})$, and $\boldsymbol{f}^{\mathrm{R}}$ is an appropriate reference vector of the objective function that provides a basis for evaluation. In MOON^{2R}, V_{RBF} is modeled in terms of a radial basis function (RBF) network. Training data used to identify the RBF network is gathered through a pair-wise comparison that requires the decision maker (DM) to judge which of each pair is preferred and the extent to which it is preferred by using linguistic statements, as in the analytic hierarchy process (AHP; Saaty, 1980).
Eventually, (p. 2) refers to a typical or single-objective problem, to which a variety of conventional optimization methods can be applied. However, since the solution often may not be satisfactory to the DM, mainly due to insufficient accuracy of the meta-model (the inaccuracy of which arises from both the objective functions and the system model), an iterative procedure should be applied to obtain the final solution. For this purpose, the meta-model can be updated by adding new data around the tentative solution and deleting old data far from it. After rebuilding the system meta-model (*Meta-X*), the value function V_{RBF} can be updated through rebuilding of objective meta-model (*Meta-$\boldsymbol{f}(\boldsymbol{x})$*). After these consecutive revisions, the problem can be solved iteratively until a satisfactory solution is obtained. As long as the re-modeling is carried out in a cooperative manner, the MOP can be solved not only effectively but also satisfactorily (Shimizu et al., 2005).

3. MOP Involving Qualitative Evaluation

3.1. Method for Evaluating Quality

Although quality is an important factor for evaluating the design and operation of artifacts, it has scarcely been evaluated in a definite manner. This is partly because the evaluation of quality involves qualitative and fuzzy attributes related to appearance, emotion and aesthetics, in other words, attributes related to Kansei. To cope with this problem, we propose applying a decision-making method based on grey theory (the grey situation decision model) together with classifiers of the Kano method. Grey theory is designed to deal with uncertainty in a system and is considered to be more

Development of Decision Support System through Multi-Objective Optimization in Ill-posed Environment

general than the fuzzy set theory. On the other hand, the Kano method is amenable to defining customer-defined quality and deriving how to improve product development and commerce.

The foundation of the grey situation decision model is classification of an attribute into an upper effect measure, a lower effect measure and a medium effect measure. Then, the effectiveness of the attribute is evaluated as x_i/x_{max}, x_{min}/x_i and $x_i^R/(x_i^R+|x_i-x_i^R|)$, for the upper, lower and medium effect measures, respectively. Here, x_{max}, x_{min} and x_i^R denote the maximum, minimum and target effects of the i-th attribute, respectively.

Moreover, to evaluate quality, we must note the existence of the threshold (x_i^b, x_i^B), beyond which the product becomes irregular or falls outside the design specifications. In addition, it is more suitable to take into account the classifiers of the Kano method for evaluating quality. The Kano method defines three categories of customer needs, namely, *must-have*, *linear satisfier* and *delighter*. The *must-have* requirement represents product features that the customer expects to receive as a matter of course. When this *must-have* is not adequately addressed, the customer experiences dissatisfaction. The second type of need, *linear satisfier*, is characterized by a relationship where the greater the extent to which the need is fulfilled, the greater the customer satisfaction becomes. The third type of need, *delighter*, is not expected at first by the customer, but its fulfillment brings about great delight.

After modifying the grey situation decision model to take into account the classifiers of the Kano method, we define the level of quality of each attribute as follows.

(1) For the upper and lower effect measures,

$$Q_i = \begin{cases} \left(\dfrac{x_i - x_i^B}{x_i^R - x_i^B} \right)^q, & \text{if } (A): x_i^B \le x_i \le x_i^R \quad \text{or } (B): x_i^R \le x_i \le x_i^B \quad (1) \\ 0, & \text{if } (A) \text{ and } x_i \le x_i^B \quad \text{or } (B) \text{ and } x_i \ge x_i^B \end{cases}$$

(2) For the medium effect measure,

$$Q_i = \begin{cases} \left(\dfrac{x_i - x_i^b}{x_i^R - x_i^b} \right)^q, & \text{if } x_i^b \le x_i \le x_i^R \\ \left(\dfrac{x_i - x_i^B}{x_i^R - x_i^B} \right)^q, & \text{if } x_i^R \le x_i \le x_i^B \\ 0, & \text{if } x_i \le x_i^b, \ x_i \ge x_i^B \end{cases} \qquad (2)$$

Here, the parameter q takes a value greater than one for the *delighter*, one for the *linear satisfier* and less than one for the *must-have*.

Then, we evaluate the quality as a weighted sum of each Q_i for all attributes, that is, $Q = \Sigma_i w_i Q_i$, where w_i denotes the weight representing the relative importance of each

Fig. 1 Photograph of rolling machine

attribute, which can be decided, for example, by applying AHP.

3.2. Software Implementation

A general procedure of MOP based on the proposed method is implemented in a Microsoft Excel spreadsheet. The procedure comprises a problem definition phase, a meta-modeling phase, a pair-wise comparison phase, an optimization phase and a result phase, in addition to supplemental phases (Shimizu and Nomachi, 2008). To evaluate a qualitative objective like quality, we provide an additional phase. The phases are programmed in Visual Basic (VB) and connected to each other by spreadsheet macros. In addition to the portability of the solution at hand, another advantage of implementing the software in Excel is that the spreadsheet can be easily applied to other problems after slight modifications.

4. Case Study on a Rolling Machine for Plastic Sheets

Plastic sheets are a widely used chemical product and shipped in rolls from the factory. These rolls are shaped using a rolling machine such as that shown in Fig. 1. Although increasing take-up velocity can raise productivity, the quality of the product will consequently decrease, and vice versa. Hence, it is desirable to determine the operating conditions that can satisfy these two conflicting objectives in proportion to their relative importance. After a preliminary evaluation, we formulated the following bi-objective optimization problem.

(p. 3) Maximize { $f_1(x)$: Quality, $f_2(x)$: Productivity} subject to $\underline{x_i} \leq x_i \leq \overline{x_i}$, { $i = 1,2,3$}

with respect to $\begin{cases} x_1\text{: extension force of rolling [N/m]}, \\ x_2\text{: acceleration of rolling speed [m/min}^2\text{]}, \\ x_3\text{: press pressure of rolling machine [N/m}^2\text{]}, \end{cases}$

where $\underline{x_i}$ and $\overline{x_i}$ denote the lower and upper bounds for each operating condition, respectively. Here, productivity is the inverse of the unit production rate defined by

$$f_2 = \left(-\frac{P}{v} + \frac{v}{x_2} \right)^{-1}, \qquad (3)$$

where v denotes the velocity and P is a constant.

Table 1 Employed parameters for evaluating Q_i

Attribute	x_i^b	x_i^R	x_i^B	q
Wrinkle (*low, must*)	-	0	80	2
Temper (*med, linear*)	580	680	780	1
Smoothness (*low, must*)	-	0	40	2
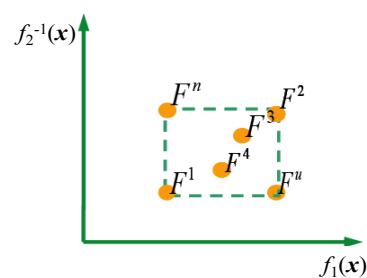
Fig. 2 Location of trial solutions

The quality is evaluated on the basis of three attributes, namely, wrinkle, temper and smoothness of the roll. However, since it is difficult to evaluate quality directly from these attributes, we apply the procedure described above. We assume that each property

Development of Decision Support System through Multi-Objective Optimization in Ill-posed Environment

associated with the grey theory and the Kano classifiers is described by the following pairs: (lower effect, *must-have*) for the wrinkle, (medium effect, *linear satisfier*) for the temper and (lower effect, *must-have*) for the smoothness; accordingly, we adopt the parameters listed in Table 1.

To obtain the response surface model of quality, we set three levels and define a set of experimental conditions, applying the $L_9(3^4)$ orthogonal array of the design of experiment (DOE). Table 2 presents the results of this experiment. Using these results, we derive the response surface model or meta-model of quality (f_1) as a RBF network model. In the last column of the table, we also show the subjective value supplied separately by the operator (DM). Here, we should note that both evaluations are quite similar except for the first run. Even for this value, in the post-analysis, the DM recognized that his score was slightly too high after reviewing all the evaluations again.

In the value function modeling phase, after setting the utopia F^u and nadir F^n, the system randomly generates four trial solutions within the objective search space enclosed in a rectangle defined by these two points, as shown in Fig. 2. Then, the system asks the DM to indicate a preference for every pair of trial solutions through pair-wise comparison. To address problems of inconsistency regarding the pair-wise comparison, if necessary, we can obtain a pair-wise comparison matrix into which the DM's preference information is integrated. Using the results thus obtained, we can train another RBF network to identify the value function as $V_{RBF}(\text{Meta-}f_1, f_2; \boldsymbol{f}^R)$, where Meta-$f_1$ represents the meta-model of quality derived from the above procedure.

Eventually, we can restate the above problem (p. 3) as follows.

$$(\text{p. 4}) \quad Max \ V_{RBF}(Meta - f_1(\boldsymbol{x}), f_2(\boldsymbol{x}); \boldsymbol{f}^R) \quad subject \ to \ \underline{x}_i \leq x_i \leq \overline{x}_i, \ (i = 1, 2, 3).$$

Now, since every value function can be evaluated for the arbitrary decision variable, we adopt differential evolution (DE) in the optimization phase, as implemented in the spreadsheet. DE, developed by Price and Storn (1997), is a powerful real number coding version of the genetic algorithm (GA). By applying the procedures embedded in the Excel spreadsheet, MOP can be carried out readily by setting only a few tuning parameters and the initial conditions. Table 3 presents the results obtained by using the software. Since the DM felt that the first result could be further improved, a second round was carried out to improve the solution by revising the meta-model in accordance with the first solution. Specifically, this is accomplished by adding five experiment points and deleting six experiment points, following a previously reported method (Shimizu et al., 2005). By selecting the proper sheet of the Excel for this task, users can revise the tentative solution appropriately. Notably, the second solution in Table 3 represents an improvement on not only the first solution, but also the conventional decision based on the experience and intuition of the operator.

5. Conclusions

In this study, we focused on MOP amenable even to an ill-posed problem, which is closely related to Kansei (emotion). In particular, we focused on devising a method for

numerically expressing measures such as the quality of a product. Then, we proposed applying the tenets of grey theory, in combination with classifiers of the Kano method, to evaluate quality. With the aim of developing a practical application for real-world manufacturing, we implemented the MOON^{2R} algorithm for MOP in a Microsoft Excel spreadsheet.

To demonstrate the application of the spreadsheet, we carried out MOP in a case study on a rolling machine for plastic sheets. Numerical experiments revealed that the proposed procedure is adequate and promising for problem solving by mutual collaboration between parties with different qualifications, for example, engineers and field operators. This was also confirmed preliminarily in comparison with the empirical results from a field trial. Finally, considering the remarkable advances being made in today's computer simulation technologies, we note that the proposed approach employing meta-modeling is particularly useful for flexible MOP in multi-disciplinary decision-making environments.

Table 2 Experimental results and evaluation of quality

	Factor of DOE (*Real value)			Observed attribute value			Evaluated quality (f_1)	
Run #	Extensi-on force ([N])	Accele-ration ([m/min^2])	Press pressure ([N])	Wrinkle [m]	Tempe r [-]	Smooth-ness [mm]	Metric (Q)	Subjective (DM)
1	-1 (*40)	-1 (*100)	-1 (*100)	95	592	1	28	40
2	-1	0 (*350)	0 (*200)	0	642	0	88	80
3	-1	1 (*600)	1 (*300)	50	685	0	72	75
4	0 (*60)	-1	0	70	742	0	57	65
5	0	0	1	0	729	0	84	75
6	0	1	-1	0	671	1	83	75
7	1 (*80)	-1	1	100	761	0	43	55
8	1	0	-1	120	705	0	59	55
9	1	1	0	50	732	0	58	60

Table 3 Comparison between the proposed solution and empirical solution

Decision		f_1 [-]	f_2^{-1} [min]	V_{RBF}	x_1: Extension force [N/m]	x_2: Acceleration [m/min^2]	x_3: Pressure [N/m]
Empirical		85	7.42	-	70	400	250
Proposed	1st	99	6.64	0.16	44	244	140
	2nd	96	5.89	0.33	42	447	172

References

T. L. Saaty, The Analytic Hierarchy Process, McGraw-Hill, 1980

Y. Shimizu, K. Miura, J-K. Yoo and Y. Tanaka, JSME, C-71, 296, 2005 (in Japanese)

R. Storn and K. Price, Journal of Global Optimization, 11, 341, 1997

J. Deng, Journal of Grey Syst., 1, 1, 1, 1989

N. Kano, N. Seraku, F. Takahashi and S. Tsuji, J. Jpn. Soc. Qual. Control, 31, 4, 147, 1984 (in Japanese)

Y. Shimizu and T. Nomachi, J. Chem. Eng. Japan, 41, 11, 1068, 2008

Global Optimization and Parametric Analysis of Large-Scale Extended Pooling Problems

Ruth Misener, Chrysanthos E. Gounaris, Christodoulos A. Floudas

Department of Chemical Engineering, Princeton University, Princeton, New Jersey, 08544-5263, USA, floudas@titan.princeton.edu

Abstract

The pooling problem addresses the storage of intermediate refinery streams under limited tankage conditions by optimizing profit subject product availability, storage capacity, demand, and product specifications. We extend the pooling problem by explicitly incorporating the Environmental Protection Agency (EPA) Title 40 Code of Federal Regulations Part 80.45: Complex Emissions Model into the constraint set. We globally optimize pooling problem networks with as many as fourteen feeds, five pools, and ten products (Misener et al., 2010; Misener and Floudas, 2009). We further extend this work by parametrically analyzing the economic impact varying of the different inputs to the EPA model.

Keywords: Pooling Problem, Bilinear MINLP, EPA Complex Emissions Model

1. Introduction

Developed in response to the Clean Air Act of 1990, the EPA model codifies and legally certifies a mathematical model of reformulated gasoline (RFG) emissions based on eleven measurable fuel components. The EPA RFG standard that regulates volatile organic, NO_X, and carcinogenic emissions impacts approximately 75 million people in the United States (EPA Website, 2009).

Table 1. EPA Complex Emissions Model Fuel Components during the Summer

	Variable	Fuel Quality	Baseline	Bounds	Units
1	OXY	oxygen	0	0 – 4	wt%
2	SUL	sulfur	339	0 -- 500	ppm
3	RVP	Reid Vap Press	8.7	6.4 – 10	psi
4	E200	200°F dist. frac.	41	30 – 70	vol%
5	E300	300°F dist. frac.	83	70 – 100	vol%
6	ARO	aromatics	32	0 -- 50	vol%
7	BEN	benzene	1.53	0 -- 2	vol%
8	OLE	olefins	9.2	0 -- 25	vol%
9	MTB	MTBE	0	0 – 4	wt% oxygen
10	ETB	ETBE	0	0 – 4	wt% oxygen
11	ETH	ethanol	0	0 – 4	wt% oxygen
12	TAM	TAME	0	0 – 4	wt% oxygen

Table 1 lists the eleven monitored fuel components that serve as inputs into the EPA Complex Emissions model (oxygen is defined by the EPA as the sum of the MBTE, ETBE, ethanol, and TAME). The fuel emissions are defined in relation to the *baseline* fuel, a representative fuel defined by the EPA. The baseline fuel properties vary

between the summer and winter (*e.g*., the distillation fraction of the baseline fuel is 50 volume percent in the winter).

The pooling problem, an optimization challenge of maximizing profit subject product availability, storage capacity, demand, and product specification constraints, has applications to petroleum refining, wastewater treatment, supply-chain operations, and communications (Floudas and Aggarwal, 1990; Tawarmalani and Sahinidis, 2002). The pooling problem is defined with a feed-forward network topology and a set of attributes, or qualities, that constitute fundamental properties of the flowing medium. Our mid-size test case for the extended pooling problem, which is pictured in Figure 1, has input feedstock nodes, intermediate storage nodes called pools, and output product nodes. We assume linear blending at each intermediate and output node except for the case of Reid Vapor Pressure, which blends according to a power law (Visweswaran, 2009).

Figure 1. Extended Pooling Problem Network
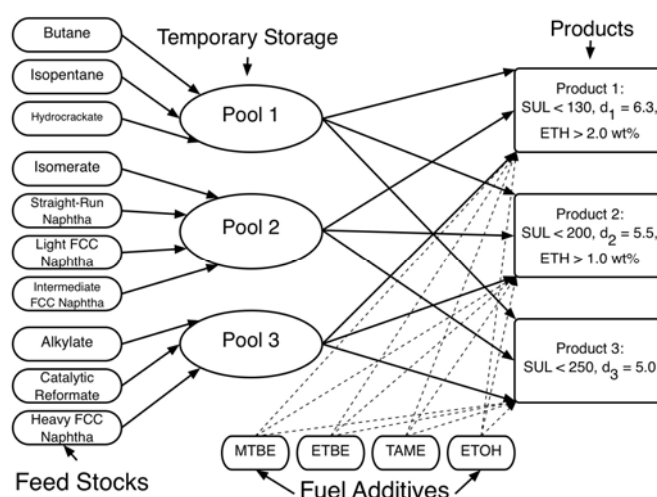
The extended pooling problem incorporates *Title 40 Code of Federal Regulations Part 80.45: Complex Emissions Model* and associated legislative bounds into the constraint set. The extended pooling problem restricts the volatile organic, NO_X, and toxics emissions of each RFG product by appending three sets of emissions model equations and constraints on the three regulated emissions to the standard pooling problem.

2. Mathematical Programming Model of the Extended Pooling Problem

As suggested in Section 1, the extended pooling problem is a standard pooling problem with EPA constraints appended (Misener et al., 2010; Misener and Floudas, 2009). It minimizes cost (equivalent to maximizing profit) on a network of (1) input feedstocks from a refinery, (2) intermediate storage nodes or *pools*, (3) fuel additives, and (4) final products. We incorporate the EPA Complex Emissions Model by computing and constraining the volatile organic, nitrous oxide, and toxics emissions for each product. The extended pooling problem optimization model is a mixed integer nonlinear programming model (MINLP) where the mixed integer portion arises from the logical

disjunctions in the EPA Complex Emissions Model that extend the accurate range of the model. The nonlinearities, most of which are also nonconvex, come from the linear and nonlinear blending at the intermediate nodes and the definitions of the EPA Complex Emissions Model. The model, which cannot be included due to space restrictions, can be found in its entirety in Misener et al. (2010). While Furman and Androulakis (2008) have previously introduced an MINLP model of the EPA Complex Emissions Model, our model of the extended pooling problem uses variables natural to the pooling problem and adheres to 40CFR80.45. Additionally, there are a few piecewise-linear functions that are modeled by Furman and Androulakis (2008) as piecewise-nonlinear because of obtuse wording in the EPA Complex Emissions Model. We model these functions as piecewise-linear.

To further study the EPA Complex Emissions Model, we find the emissions of the baseline fuel using the EPA spreadsheet model corresponding to 40CFR80.45 (EPA Website, 2009), increase each of the fuel qualities by 1%, find new emissions levels, and compare the two results. Table 2 records the percentage difference between the baseline fuel emissions and the new emissions level. The MBTE, ETBE, ethanol, and TAME fuel additives, which have a baseline level of 0 and are summed together to find the total oxygen content of the fuel, were varied from 0 to 1 weight percent oxygen.

Table 2. Response of the Region 1 Summer EPA Complex Emissions Model to Perturbation. Values record percentage difference between Baseline Fuel Emissions and Emissions levels after the fuel quality is increased by 1%.

	Volatile Organic Emissions (VOC)	NO_X	Total Toxics
Sulfur	0.04 %	0.08 %	0.11 %
Reid Vap Pres	1.40 %	0.03 %	0.18 %
E200	-0.13 %	0.04 %	-0.08 %
E300	-0.36 %	-0.04 %	0.06 %
Aromatics	0.08 %	0.03 %	0.32 %
Benzene	0.00 %	0.00 %	0.28 %
Olefins	-0.02 %	0.04 %	0.02 %
MBTE	-0.22 %	-0.10 %	-3.50 %
ETBE	-0.22 %	-0.10 %	-1.61 %
Ethanol	-0.22 %	-0.10 %	-2.07 %
TAME	-0.22 %	-0.10 %	-3.53 %

We can make several observations about the EPA Complex Emissions model based on Table 2. First, all four of the fuel additives serve to reduce the three emissions levels. Second, the volatile organic emissions are strongly related to the fuel RVP and secondarily to the distillation fraction of the fuel at 300^oF and 200^oF. This corresponds to Rhodes (1998) first-order recommendation of eliminating most hydrocarbon chains of length four and five from RFG. Next, nitrous oxide emissions vary most with changes in the sulfur level of the fuel. The total toxics emissions are sensitive to changes in the aromatics and particularly the benzene concentration of the fuel. Because the baseline fuel has 32 vol% aromatics but only 1.53 vol% benzene, small absolute changes in benzene, which correspond to large relative changes, strongly impact the toxics emissions. Finally, we have recorded the sensitivities of each fuel

quality to the three emissions types for the winter and Region 2 cases. Although the values do vary for the three other condition types, the relative importance of the fuel qualities remains similar.

3. Relaxation and Global Optimization of the Extended Pooling Problem

3.1. Relaxation Construction

In Misener et al. (2010), we relax the extended pooling problem using a number of recently introduced relaxation techniques. We employ the relaxation-linearization technique-based 'PQ'-formulation to tighten the termwise bilinear relaxation of the standard pooling problem portion (Tawarmalani and Sahinidis, 2002), closely underestimate the bilinear terms via piecewise relaxation (Meyer and Floudas, 2006; Karuppiah and Grossmann, 2006; Wicaksono and Karimi, 2008; Gounaris et al., 2009). To underestimate the equations in the EPA Complex Emissions model, we use piecewise techniques (Floudas, 1995), outer approximation, and the edge-concave method (Meyer and Floudas, 2005).

Figure 2 pictures the piecewise-linear underestimators that are used to relax the bilinear terms in the pooling problem. The bilinear function in the upper left of Figure 2 can be more closely represented using two or four piecewise segments than the original convex envelope.

Figure 2. Piecewise Linear Relaxations of a Bilinear Term
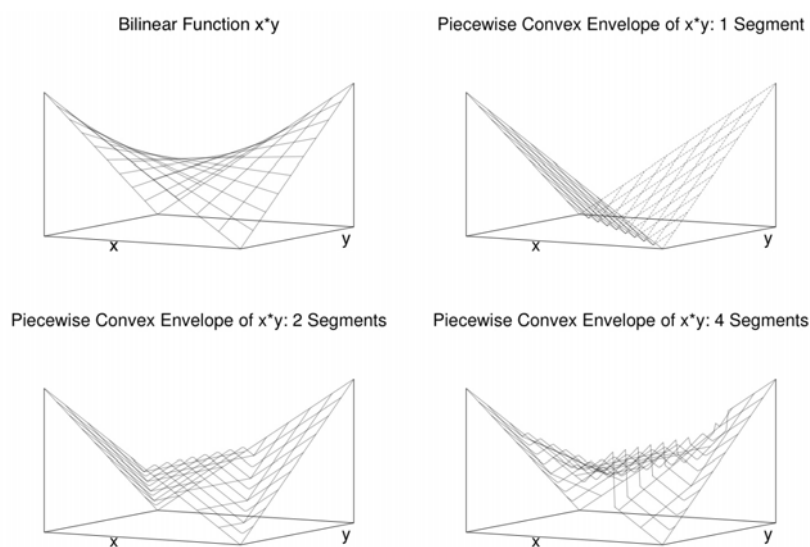
3.2. Global Optimization

Table 3. Topology of the Three Test Cases

	Feeds	Pools	Products	Contin Vars	Binary Vars	Nonlinear Terms
Small Case	7	1	2	214	30	108
Mid-Size Case	14	3	3	331	45	180
Large Case	14	5	10	1104	150	640

Table 3 records the topology and Table 4 has the optimization data of the three test cases that we optimized in Misener et al. (2010). The test cases were solved on a Linux workstation with an Intel Core 2 Duo processor containing two 2.40 GHz cores. The optimization process was run serially using CPLEX Version 11.0 (ILOG, 2007) for the node relaxations and MINOS Version 5.51 (Murtagh et al., 2004) for the local upper bounding solves. The global optimization algorithm terminates at a 0.5% gap between the lower and upper bounds. Note that even the large-scale test case recorded in Table 3 converges relatively quickly in Table 4.

Table 4. Computational Results for the Three Test Cases listed in Table 3

	Root Relax'n	Final LB	Final UB	CPU (s)
Small Case	$-4.731 * 10^2$	$-2.808 * 10^2$	$-2.794 * 10^2$	8.8
Mid-Size Case	$-4.598 * 10^3$	$-4.587 * 10^3$	$-4.567 * 10^3$	487
Large Case	$-1.500 * 10^4$	$-1.498 * 10^4$	$-1.490 * 10^4$	7987

Figure 3 records and branch-and-bound tree for the mid-size test case. Each node of the tree corresponds to solving a MILP relaxation of the MINLP model.

Figure 3. Branch-and-Bound Tree for the Mid-Size Test Case
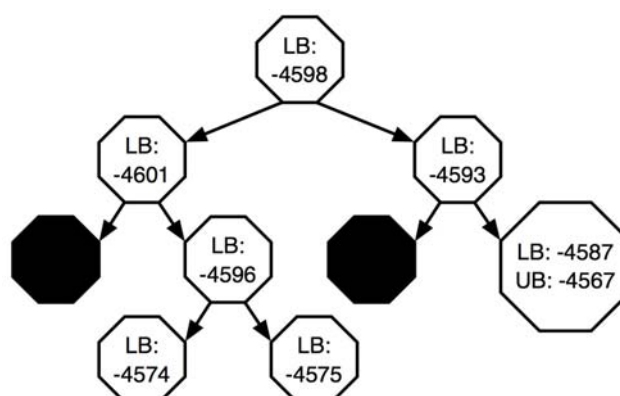
4. Parametric Analysis of the Optimization Results

Table 5. Global Response of the Mid-Sized Test Case to Parametric Pertubation

	Lower Bound	Upper Bound	CPU (s)
Sulfur	-4397	-4374	183.0
RVP	-4508	-4490	233.7
E200	-4574	-4540	458.6
E300	-4585	-4567	243.8
Aromatics	-4587	-4567	333.6
Benzene	-4413	-4377	196.3
Olefins	-4588	-4567	328.8

Table 5 records the global optimization results of the mid-size test case topology under uncertain fuel qualities where each of the fuel qualities was increased in turn by 10%. For each of the seven fuel qualities in Table 5, we assume that the sensors monitoring the qualities of the feedstocks may be 10% inaccurate. This effectively raises the

concentration of that fuel quality in the feedstock because we pessimistically estimate that we have to assume that the fuel quality is 10% poorer. Table 5 shows that, in the mid-size network, the sulfur and benzene concentrations most strongly impact the profitability of the network.

5. Conclusions

We have further explored the extended pooling problem, which appends the EPA Complex Emissions Model to a pooling problem network. The tight, effectively formulated relaxation equations reduce the number of nodes explored and the three test cases quickly solve to a 0.5% gap even for large-scale problems. We also investigated the influence of variable fuel quality parameters on the profitability of the network.

6. Acknowledgements

References

40CFR80.41, 2008, *Code of Federal Regulations: Standards and requirements for compliance*.

40CFR80.45, 2007, *Code of Federal Regulations: Complex Emissions Model*.

EPA Website, 2009, Reformulated Gas, http://www.epa.gov/OMS/rfg.htm, Accessed 13 Nov.

C. A. Floudas, 1995, *Nonlinear and Mixed-Integer Optimization: Fundamentals and Applications*, Oxford University Press, New York, NY.

C.A. Floudas, A. Aggarwal, 1990, A Decomposition Strategy for Global Optimum Search in the Pooling Problem. *ORSA J. on Comput.*, 2, 3, 225-234.

K.C. Furman, I.P. Androulakis, 2008 A Novel MINLP-Based Representation of the Original Complex Model for Predicting Gasoline Emissions. *Comput. Chem. Eng.*, 32, 2857-2876.

C.E. Gounaris, R. Misener, C.A. Floudas, 2009, Computational Comparison of Piecewise-Linear Relaxations for Pooling Problems, *Ind. Eng. Chem Res.*, 48, 12, 5742-5766.

C.A. Haverly, 1978, Studies of the Behavior of Recursion for the Pooling Problem. *ACM SIGMAP Bulletin*, 25, 19-28.

R. Karuppiah, I.E. Grossmann, Global Optimization for the Synthesis of Integrated Water Systems in Chemical Processes. *Comput. Chem. Eng.*, 30, 650-673.

C.A. Meyer, C.A. Floudas, 2005, Convex Envelopes for Edge-Concave Functions. *Math. Program.*, 103, 2, 207-224.

R. Misener, C.A. Floudas, 2009, Advances for the Pooling Problem: Modeling, Global Optimization, and Computational Studies, *Appl. Comput. Math.*, 8, 1, 3-22.

R. Misener, C.E. Gounaris, C.A. Floudas, 2010, Mathematical Modeling and Global Optimization of Large-Scale Extended Pooling Problems with the (EPA) Complex Emissions Constraints, Submitted for Publication.

A. K. Rhodes, 1998, U.S. Re ners Make Complex-Model RFG as They Prepare for the Next Hurdle. *Oil & Gas Journal*, 96, 22-27.

F. Tardella, 2003, On the Existence of Polyhedral Convex Envelopes. In C. A. Floudas and P. M. Pardalos, editors, *Frontiers in Global Optimization*, Kluwer Academic Publishers, 563-573.

M. Tawarmalani, N.V. Sahinidis, 2002, *Convexi cation and Global Optimization in Continuous and Mixed-Integer Nonlinear Programming: Theory, Applications, Software, and Applications. Nonconvex Optimization and Its Applications*. Kluwer Academic Publishers, Norwell, MA, USA.

V. Visweswaran, 2009, MINLP: Applications in Blending and Pooling. In C. A. Floudas and P. M. Pardalos, editors, *Encyclopedia of Optimization*, Springer Science, 2 edition, 2114-2121.

D.S. Wicaksono, I.A. Karimi, 2008, Piecewise MILP Under- and Overestimators for Global Optimization of Bilinear Programs. *AIChE J.*, 54, 4, 991-1008.

Biobjective optimization using Environmental and Economic Functions in Utility Plants

Pablo Enrique Martinez and Ana María Eliceche

Departamento de Ingeniería Química, Universidad Nacional del Sur, PLAPIQUI-CONICET, Camino La Carrindanga km 7, 8000 Bahía Blanca, Argentina.

Abstract

Environmental and economic objective functions are used simultaneously to select the operating conditions of a steam and power plant. Different methodologies to solve bi objective optimization problems were implemented successfully. The life cycle potential environmental impact and the operating cost of the utility plant are minimized simultaneously. A methodology is presented to estimate the potential environmental impacts considering the most important life cycle stages associated with imported electricity and natural gas in the utility plant. Mixed Integer Non Linear bi objective problems are formulated and different strategies are implemented and successfully solved in GAMS.

Keywords: bi objective optimization, life cycle environmental impact, utility plant.

1. Introduction

The formulation of multi objective optimization problems including environmental and economic metrics to support a decision making process can contribute to a sustainable development. Multi-objective optimization applied to environmental and economic objectives has been treated by authors like Ciric and Huchette [1] minimizing the amount of waste and the net profit of an ethylene glycol production plant. Dantus and High [2] proposed a method to convert a bi objective optimization problem into a single objective optimization problem; the method proposed is a variation of the utopia point distance minimization, including discrete variables to select the type of reactor to be used in the methyl chloride superstructure plant design.

In the present work the operating conditions of a steam and power plant are selected to minimize life cycle environmental impact and operating cost simultaneously solving a bi objective optimization problem. The environmental objective is the life cycle environmental impact associated with gaseous and liquid emissions, and solid wastes of an ethylene steam and power plant. In the life cycle context, the battery limits of the steam and power plant need to be extended in order to include emissions of imported natural gas and electricity generated by nuclear, hydroelectric and thermoelectric plants. The operating cost includes costs of imported electricity, natural gas, makeup water and water treatment. A Mixed integer non linear bi objective optimization problem is formulated and solved in GAMS [3].

2. Evaluation of environmental and economic objective functions

2.1. Potential Environmental Impact Evaluation

The Potential Environmental Impact (PEI) function considered is a multi objective function itself, since nine environmental impact categories are considered: global warming, acidification, ozone depletion, photo oxidant formation, eutrophication, fresh water ecotoxicity, human toxicity, source depletion and the impact due to ionizing radiation. The Potential Environmental Impact is calculated using the Guinée et al. [4] methodology. The contribution of the emission of a pollutant **k** to a given environmental impact category **j** is evaluated multiplying the pollutant **k** flow rate $\mathbf{F_k}$ emitted into the environment by a characterization factor γ_{kj} published by Guinée et al. [4]. This characterization factor represents the effect that chemical **k** has on the environmental impact category **j**. Hence, the Potential Environmental Impact, PEI, is calculated as follow:

$$\mathbf{PEI} = \sum_{\mathbf{j}} \sum_{\mathbf{k}} \alpha_{\mathbf{j}} \times \mathbf{F_k} \times \gamma_{\mathbf{k,j}} \tag{1}$$

Where $\alpha_{\mathbf{j}}$ represent the weighting factors for each environmental impact category **j**. More information can be found in Eliceche et al. [5]. Eq.1 transforms the pollutants emissions flow rates into potential environmental impacts.

2.1.1. Evaluation of the Utility Plant Environmental Impact

The emissions of the steam and power plant are evaluated from the modelling of the main processes formulated in GAMS. The emissions come mainly from the combustion in the boilers of a mixture of natural gas, $\mathbf{F_{ng}}$ and residual gas, $\mathbf{F_{rg}}$. Liquid emissions of purge streams, $\mathbf{F_p}$, in the boilers and cooling system are also considered. The pollutants emissions from the utility plant (UP) are calculated as follow:

$$\mathbf{F_k^{UP}} = \mathbf{F_{gn}} \times \mathbf{e_{k,ng}} + \mathbf{F_{rg}} \times \mathbf{e_{k,rg}} + \mathbf{F_p} \times \mathbf{e_{k,p}} \tag{2}$$

Where $\mathbf{e_{k,ng}}$ is the emission factor for pollutant **k** due to the combustion of natural gas, $\mathbf{e_{k,rg}}$ is the corresponding emission factor for residual gas combustion and $\mathbf{e_{k,p}}$ is the pollutant emission factor for liquid emissions. The emissions factors express the amount of pollutant **k** emitted by unit mass of natural gas, residual gas and liquid stream, respectively. A detailed analysis of each life cycle stage considered as well as the literature sources was presented in Martínez and Eliceche [6].

2.1.2. Life Cycle Environmental Impact Assessment of Imported Electricity and Natural Gas

Life cycle approach considers emissions during the entire life cycle of a product or service accounting by emissions from raw material extraction to waste disposal. In the utility plant, the life cycle emissions are considered for the natural gas feedstock and the imported electricity needed to move some electrical motors in the superstructure of the steam and power plant. Pollutant flow rate for natural gas (NG) life cycle $\mathbf{F_k^{NG}}$ is calculated in the following equation:

Biobjective Optimization using Environmental and Economic Functions in Utility Plants

$$F_k^{NG} = F_{ng} \times \sum_l e_k^l \qquad l = 1,...,l_{ng} \qquad (3)$$

Where e_k^l is the emission factor for pollutant k in the life cycle stage l, l_{ng} is the total number of life cycle stages considered for the natural gas fuel cycle: exploration, extraction and transportation stages. As the residual gas is produced in the ethylene plant, no life cycle stage has been considered for it.

The imported electricity (IE) life cycle emissions have been assessed through the life cycle of different electricity generation plants. The electricity generation sector in Argentina has contributions from thermoelectric, hydroelectric and nuclear plants. Thermoelectric power generation consumes coal, oil and natural gas as fuels; nuclear power generation consumes natural uranium fuel. The estimation of pollutant emissions in the electric power generation includes the following life cycle stages: extraction and processing of raw materials, transport, refining (where it is applicable) and electricity generation itself:

$$F_k^{IE} = \sum_q \sum_{l_q} W_q \times e_{k,q}^{l_q} \qquad l = 1,...,l_{ie} \qquad (4)$$

Where W_q is the electricity imported and generated with technology q, l_q superscript accounts life cycle stage l in electricity generated by option q, finally $e_{k,q}^{l_q}$ is the corresponding emission factor of pollutant k in electricity generated with option q, for the life cycle stage l_q. The life cycle stages considered are: (i) exploration, extraction, refining and transport of natural gas, oil, coal and uranium consumed in thermoelectric and nuclear plants; (ii) submerged biomass decay in hydroelectric plants (iii) waste treatment and disposal for nuclear plants and (iv) transport in the construction stage of hydroelectric and nuclear plants.

The utility plant potential environmental impact, PEI^{UP} is calculated as follows:

$$PEI^{UP} = \sum_j \sum_k \alpha_j \times F_k^{UP} \times \gamma_{k,j} \qquad (5)$$

The component k life cycle emissions F_k^{LC} are estimated adding the component k emissions in the utility plant, life cycle of imported natural gas and electricity:

$$F_k^{LC} = F_k^{UP} + F_k^{NG} + F_k^{IE} \qquad (6)$$

The life cycle potential environmental impact is evaluated as follows:

$$PEI^{LC} = \sum_j \sum_k \alpha_j \times F_k^{LC} \times \gamma_{k,j} \qquad (7)$$

Global warming due to combustion emissions is the most relevant environmental category for steam and power plants and for fossil fuels electricity generation.

2.1.3. Evaluation of Economical Objective Function

The operating cost of the utility plant includes costs of imported electricity (IW), natural gas feed (NG), makeup water (MW) and water treatment (WT); where c_{ng}, c_q, c_{MW} and c_{WT} are the cost coefficients:

$$C = F_{ng} \times c_{ng} + \left(\sum_q W_q \right) \times c_W + F_{MW} \times c_{MW} + F_{WT} \times c_{WT} \tag{8}$$

3. Formulation of Bi objective optimization problem

The multi objective (MO) optimization is a system analysis approach to problems with conflictive objectives. A key factor of MO optimization is that rarely exist a single solution that simultaneously optimizes all the objectives. In its place, there is a set of solutions where one objective cannot be improved except at expense of another objective. This set of compromise solutions are generally referred as non-inferior or Pareto optimal solutions. A variety of strategies to solve multi objective optimization problems exist, that can be found in Alves et al [7]. The general approach consists in converting the multiple objectives into a single objective. Some of these methods are: weighted sum, utopia point distance minimization, e-constraint method and global criteria method. The general formulation of a bi objective optimization problem considering continuous and discrete variables follows:

$$\text{Min}_{x,y} Z = Z[PEI(x,y),C(x,y)]$$

$$\text{s.t :} \quad h(x) = 0$$

$$g(x) + A(y) \le 0$$

$$x^{LB} \le x \le x^{UB} \tag{P1}$$

$$x \in R^n$$

$$y \in \{0,1\}^m$$

Where x and y are the continuous and binary optimization variables, respectively. Superscripts U and L, indicates upper and lower bounds on vector x, respectively. The equality constraints $h(x) = 0$ are the system of non-linear algebraic equations that represent the steady state modelling of the process plant, including mass and energy balances; enthalpy and entropy prediction. The inequality constraints $g(x) + A(y) \le 0$ represent logical constraints, minimum and maximum equipment capacities, operating and design constraints, etc. The A matrix includes linear relations between binary variables such as logical constraints. A detailed mathematical model of the utility plant is presented in Eliceche et al. [5].

Different strategies to solve bi objective optimization problems have been implemented successfully. The bi objective function Z in problem *P1* for the global method presented by Dantus and High [2] follows, with the nomenclature presented in section 2:

$$Z = \omega_1 \times \left[\frac{PEI - PEI^*}{PEI^{**} - PEI^*} \right]^p + \omega_2 \times \left[\frac{C - C^*}{C^{**} - C^*} \right]^p \tag{9}$$

Where ω_1 and ω_2 are weighting factors, these preference weights ω_i are used to represent the relative importance of each objective. The decision-maker's preferences are also expressed in the compromise index p $(1 \le p \le \infty)$, which represents the decision-maker's concern with respect to the maximal deviation from the utopia point. As a result, the non-inferior solutions defined within the range $1 \le p \le \infty$ correspond to the compromise set from which the decision maker still has to make the final choice to identify the best compromise solution [2]. The single asterisk indicates the minimum

Biobjective Optimization using Environmental and Economic Functions in Utility Plants

values of a given objective function solving a single objective optimization problem, while double asterisk indicates the alternative objective function value obtained. The objective functions used are life cycle potential environmental impact (PEI) given in Equation 6 and operating cost (C) given in Equation 8.

The convex weighted sum method presented by Westerberg [8] was also successfully implemented, using the nomenclature presented in section 2, as follows:

$$Z = \omega_1 \times PEI + (1 - \omega_1) \times C \tag{10}$$

Where ω_1 is the weighting factor, this factor is used to represent the relative importance of each objective. A similar method as those stated by Eq. (10) is the normalized weighted sum method [9], where each single objective function is normalized respect to the minimum value reached in single optimization of each function (e.g PEI* and C* in equation 9). The division by the minimum value avoids biases in the results generated by the different magnitude of each function.

4. Discussion of numerical results

A rigorous modelling of the utility plant is formulated in GAMS, including the power and steam demands of the ethylene plant [5]. The continuous operating variables selected are temperature and pressure of the high, medium and low pressure steam headers and the deareator pressure. Binary operating variables are introduced to represent discrete decisions such as the selection of: (i) alternative pump drivers such as electrical motors and steam turbines and (ii) boilers which are on or off, and their auxiliary equipment such as feed pumps and air fans. Thus a bi objective Mixed Integer Nonlinear Programming problem is formulated and solved in GAMS.

Different strategies were implemented to solve the bi objective problem. The solution point reported in Table 1 was obtained with the Dantus and High [2] method and the following parameters for Eq. (9): ω_1=0.1, ω_2=0.9, p=1. The following GAMS options were used: DICOPT as the outer approximation algorithm; CONOPT3 to solve the Non Linear Programming sub problem and CPLEX to solve the Mixed Integer Linear Programming sub problem. The same results were achieved using the convex weighted sum and the normalized weighted sum methods (with ω_1=0.25).

Significant reductions in the order of 12 % in life cycle environmental impact and 16 % in operating cost can be achieved selecting the operating conditions with the methodology proposed, as shown in Table 1.

Regarding the selection of pump's drivers, steam turbines are chosen rather than electrical motors, due to the fact that the environmental impact to power generated ratio is smaller in the steam and power plant than in the generation of the imported electricity.

Table 1. Multiobjective problem solution.

Objective Functions		Initial Point	Solution point	Reductions
PEI^{LC}	(PEI / h)	33627.33	29581.88	**12 %**
Cost	(U$ / h)	561.84	470.97	**16 %**

This is due to the fact that natural gas is burned with residual gas from the demethanizer column. The residual gas is a Hydrogen rich stream, having higher combustion heat and

lower combustion emissions than natural gas or any other fossil fuel. The operating cost is also cheaper with steam turbines than with electrical motors. The number of the boilers in operation is reduced from four to three, due to a proper selection of temperature and pressure of steam headers, mainly the high pressure steam header.

This is a process where improving process efficiency, environmental impact and cost are reduced simultaneously. They are not conflictive objectives in a life cycle approach. This is not the case if the environmental impact evaluation is reduced to the battery limits, where minimizing environmental impact and operating cost leads to different solutions mainly regarding the selection of alternative drivers.

5. Conclusions

A methodology has been presented to select the operating conditions minimizing simultaneously life cycle environmental impact and operating cost, solving a mixed integer nonlinear bi objective optimization problem. Different strategies to solve bi objective optimization problems were implemented successfully. Significant reductions in the in the order of 12 % in life cycle environmental impact and 16 % in operating cost can be achieved simultaneously, as shown in Table 1. The utility sector studied, has relevant contributions to combustion emissions, global warming, consumption of non renewable fossil fuels and water and also to operating cost. This is a plant where improving process efficiency, environmental impact and cost can be reduced simultaneously if a life cycle approach is followed. For these reasons it is very important to extend the battery limits to include life cycle analysis, when environmental objectives are used to support a decision making process.

References

[1] A. Ciric and S. Huchette, Ind. Eng. Chem. Res., 32 (1993), 2636-2646.
[2] M. Dantus, K. High, Comp. Chem. Eng. 23 (1999), 1493-1508.
[3] A. Brooke, D. Kendrick, A. Meeraus, R. Raman (eds.), GAMS, A user guide, GAMS
    Development Corporation, Washington DC, 2003.
[4] J. Guinée, R. Heijungs, G. Huppes, R. Kleijn, A. Koning, L. van Oers, A. Sleeswijk, S. Suh,
    H. Udo de Haes (eds.), Handbook on Life Cycle Assessment. Operational Guide to the ISO
    Standards. Kluwer Academic Publishers, Dordrecht, 2002.
[5] A. Eliceche, S. Corvalan, P. Martínez, Comp.& Chem. Eng., 31 (2007), 648-656.
[6] P. Martínez and A. Eliceche, Clean Technologies and Environmental Policy, 11 (2009), 49-57.
[7] M. Alves and J. Climaco, European J. of Operatinal Research, 180 (2007), 99-115.
[8] A. Westerberg and P. Clark, Computers and Chemical Engineering, 7 (1983), 259-278.
[9] K. Debb, Multiobjective optimization using evolutionary algorithms. John Wiley & Sons,
    LTD, New York, 2001.

Multiobjective optimization of industrial water networks with contaminants

Marianne Boix,[a] Ludovic Montastruc,[a] Luc Pibouleau,[a] Catherine Azzaro-Pantel,[a] Serge Domenech [a]

LGC-CNRS-INPT ; Université de Toulouse ; 4, Allée Emile Monso, BP 84234, F-31432 Toulouse, France
Marianne.Boix@ensiacet.fr

Abstract

This paper presents a multiobjective MILP formulation for optimizing industrial water networks. By expressing balance equations in terms of partial mass flows instead of total mass flows and concentrations, and because the contaminant mass flow (ppm) is very small compared to the water mass flow $(T.h^{-1})$, the problem becomes linear. The integer variables are related to the interconnections into the network. The biobjective optimization of the fresh water flow rate at the network entrance and the water flow rate at regeneration unit inlets, parameterized by the number of interconnections, is carried out according to a lexicographic procedure. A monocontaminant network involving ten processes and one regeneration unit illustrates the approach. Even if the results are specific, the methodology guide can be applied to a large panel of networks. On the one hand, this example shows that the Pareto front is a straight line where each point is a feasible solution, when the number of connections is maximal (120). On the other hand, the Pareto front is reduced with the number of connections (11) and constituted by isolated points located mainly on a straight line with the same slope as for 120 connections, but no feasible solution exists between these points.

Keywords: Water network, Multiobjective optimization, MILP, Monocontaminant.

1. Introduction

During the last decades, industrialization has contributed to the rapid depletion of natural resources such as water or natural gas. With the increasing interest for global environment preservation, the unlimited resources paradigm became little by little obsolete. In 2000, the global needs in fresh water were estimated to be 5000 Km^3 [1], among which 70% were used for agriculture, 20% were used by industry and 10% were consumed for domestic uses and have been increased by a factor of 4 for 50 years. Among the industrial consumers, the process industry is by far the most important user of fresh water. The environmental impact induced by the process industry is linked both to the high volumes involved and to the diversity of toxic products generated along the process chain. So, a real need to define optimized water networks to reduce the impact of contaminants on the environment, has recently emerged. This paper aims at defining a general methodology for taking into account the monocontaminant case.

2. Background

Water networks problems have been tackled by three main approaches. These techniques include graphical methodology [2-5], mathematical programming [6-9] and synthesis of mass exchange networks [10-12]. Due to the recent development of

efficient numerical toolboxes, the graphical methods pinch-based techniques have been replaced by mixed-integer programming approaches, either linear (MILP) or nonlinear (MINLP). The linear case is generally restricted to simple water networks involving only one contaminant, while the nonlinear one can theoretically be applied to more complex networks.

Huang et al. [8] defined a superstructure of a complex network involving processes using both water and regenerating units for water with a given output concentration of contaminants. Linear formulations implemented for maximizing the water regeneration and reuse into industrial processes have been first developed by Bagajewicz and Savelski [9] and El-Halwagi et al. [13]. Indeed, the maximization of the water recovery implies the simultaneous minimization of fresh water consumption and effluent emissions. A linear formulation is also given by Wang et al. [14] for monocontaminant networks. Quesada and Grossmann [15] and, later, Galan and Grossmann [16] develop a MINLP strategy based on the relaxation of the bilinear terms involved in the balance equations. Even if significant advances have been performed in the field on nonlinear mixed-integer programming, the search for a solution of a linear problem is always easier than the one of MINLP. This concerns the global optimality of the solution found, as well as the ease to initialize the search. Furthermore, MILP methods may support important numbers of variables and high combinatorial aspects. This can be particularly interesting when ecoparks are taking into account because of their greater number of variables and constraints. In this paper, only monocontaminant networks are considered. However, it appears that the proposed strategy could be easily extended to multicontaminant problems.

3. Solution procedure

3.1. Superstructure definition

Given a set of regeneration units and processes, the objective is to determine a network of interconnections of water streams among them so that both the overall fresh water consumption and the regenerated water flowrate are minimized. Water networks are defined as follows. All the possible connections between processes and regeneration units may exist, except regeneration recycling to the same regeneration unit or process. Each process admits maximal input and output concentrations, and in the same way, regeneration units have a given processing capacity. For each process using water, input water may be fresh water, used water coming from other processes and/or recycled water; the output water for such a process may join either the discharge, either other processes and/or to regeneration units. Similarly, for a regeneration unit, input water may come from processes or other regeneration units. Regenerated water may be reused in the processes or join other regeneration units. The generic problem to solve is built as a set of black-boxes, in order to adapt the formulation to a large variety of practical cases. In this black-box approach, the role of each process within the network is not taken into account. For each process input or output, contaminant mass fractions (in ppm) are imposed by the user, and constitute bounds for the problem.

Each task performed by a process contaminates its input water up to a given mass fraction. The amount of pollutant i generated by a process j is noted M_i^j and is expressed in mass flow (g.h^{-1}). For each practical example, the values of M have to be provided. A regeneration unit can be defined by two ways: 1) it has a given efficiency depending on the pollutant under treatment (in that case, E_i^l represents the efficiency of

the regeneration unit l for component i, $0 < E < 1$), 2) the mass fraction (in g.h^{-1}) of pollutant at the regeneration unit output is fixed. In most cases, for monocontaminant networks, the last definition, being more in agreement with the practical usage, is preferred.

3.2. Modeling equations

In the majority of previous works, the problem is generally stated in terms of concentrations and total mass flows, giving birth to bilinear formulations [6] due to products between concentrations and total mass flows. If partial mass flows are used instead of total mass flows (noted as follows), the balance equations are all linear. This approach is similar to that of Bagajewicz and Savelski [9].

- $wp_i^{j \to k}$ mass flow of component i going from process j to process k (T.h^{-1})

- $wr_i^{l \to m}$ mass flow of component i going from regeneration unit l to regeneration unit m (T.h^{-1})

- $wrp_i^{l \to j}$ mass flow of component i going from regeneration unit l to process j (T.h^{-1})

- $wpr_i^{j \to l}$ mass flow of component i going from process j to regeneration unit l (T.h^{-1})

- wd_i^{j} mass flow of component i going from process j to the discharge (T.h^{-1})

- wrd_i^{l} mass flow of component i going from regeneration unit l to the discharge (T.h^{-1})

- w_1^{j} mass flow of fresh water at the entrance of process j (T.h^{-1}).

The balance equations give the following set of six linear equations (for a process j, Eqns 1 and 2, for a regeneration unit l, Eqns 3 and 4 and for the network entrance, Eqns 5 and 6).

$$w_1^{j} + \sum_k wp_1^{k \to j} + \sum_l wrp_1^{l \to j} = wd_1^{j} + \sum_k wp_1^{j \to k} + \sum_l wpr_1^{j \to l} \tag{1}$$

$$\sum_k wp_{i>1}^{k \to j} + \sum_l wrp_{i>1}^{l \to j} + M_{i>1}^{j} = wd_{i>1}^{j} + \sum_k wp_{i>1}^{j \to k} + \sum_l wpr_{i>1}^{j \to l} \tag{2}$$

$$\sum_m wr_1^{m \to l} + \sum_j wpr_1^{j \to l} = wrd_1^{l} + \sum_j wrp_1^{l \to j} + \sum_m wr_1^{l \to m} \tag{3}$$

$$\sum_m wr_{i>1}^{m \to l} + \sum_j wpr_{i>1}^{j \to l} = wrd_{i>1}^{l} + \sum_j wrp_{i>1}^{l \to j} + \sum_m wr_{i>1}^{l \to m} \tag{4}$$

$$\sum_l wrd_1^{l} + \sum_j wd_1^{j} = \sum_j w_1^{j} \tag{5}$$

$$\sum_l wrd_{i>1}^{l} + \sum_j wd_{i>1}^{j} = \sum_j w_{i>1}^{j} \tag{6}$$

In these equations, index i can either represent water (if is equal to 1), or contaminants (if is greater than 1).

The equation governing the concentration conservation between the output streams of processes and regeneration units is:

$$\frac{wp_i^{j \to k}}{wp_1^{j \to k} + \sum_{i=2}^{N+1} wp_i^{j \to k}} \leq CM_{i,j}^{o} \tag{7}$$

N being the number of contaminants, $j \neq k$ and $CM_{i,j}^o$ is the maximal concentration of contaminant i at the output of process or regeneration unit j.

Insofar as $wp_i^{j \to k}$ (order of magnitude ppm) is far lower than $wp_1^{j \to k}$ (T.h⁻¹), the nonlinear Eqn (7) can be rewritten linearly as:

$$wp_i^{j \to k} - CM_{i,j}^o \times wp_1^{j \to k} \leq 0 \tag{8}$$

The same equation holds for the other output streams of process j:

$$wpr_i^{l \to j} - CM_{i,j}^o \times wpr_1^{l \to j} \leq 0 \tag{9}$$

In the same way, it comes:

$$wd_i^j - CM_{i,j}^o \times wd_1^j \leq 0 \tag{10}$$

The output streams of a given process must have the same pollutant concentration, that is to say:

$$wp_i^{j \to k} - CM_{i,j}^o \times wp_1^{j \to k} = wpr_i^{l \to j} - CM_{i,j}^o \times wpr_1^{l \to j} = wd_i^j - CM_{i,j}^o \times wd_1^j \tag{11}$$

However, these equalities only hold for existing streams. If the mass flow of water is null for a stream, this stream does not exist, that is to say:

if $wp_1^{j \to k} = 0$ then $wp_i^{j \to k} = 0$.

Finally, the constraints on the output streams of regeneration units are given by:
i) if the output concentration is fixed

$$wrp_i^{l \to j} - CM_{l,i}^o \times wrp_1^{l \to j} \leq 0 \tag{12}$$

$$wr_i^{l \to m} - CM_{l,i}^o \times wr_1^{l \to m} \leq 0 \tag{13}$$

$$wr_i^{l \to m} - CM_{l,i}^o \times wr_1^{l \to m} = wrp_i^{l \to j} - CM_{l,i}^o \times wrp_1^{l \to j} \tag{14}$$

ii) if the efficiency is fixed

$$\frac{wrin_i^l - wrout_i^l}{wrout_i^l} = E_i^l \tag{15}$$

3.3. Multiobjective optimization

Two objective functions have to be simultaneously minimized while the third is considered as a constraint:
- Fresh water flow rate at the network entrance (F1)
- Water flow rate at inlet of the regeneration unit (F2)
- Number of interconnections into the network (F3).

For the example presented below, the number of interconnections in the network is defined in the reduced integer range [11-120]; the methodology consists in solving the biobjective problem (F1, F2) parameterized by the number of interconnections. A lexicographic optimization [17] based on the ε-constraint strategy is implemented. During the first phase, the first objective is minimized alone, while the second one is introduced in the form of a bounded constraint. The second objective is minimized in the second step, where the first one can vary in a closed interval whose the optimal value obtained in the first phase is the median. When the solutions obtained in the two phases are identical, an optimal solution for the biobjective problem is reached. So, for

each particular value of the number of interconnections, a Pareto front can be generated.

4. Numerical example

This example involving ten processes, one regeneration unit and one contaminant, was already proposed by Bagajewicz and Savelski [9]. The corresponding MILP involves 143 binary variables related to interconnections, 332 continuous variables and 351 constraints, and it solved with the solver CPLEX of the GAMS package.

In order to set the problem limits, each objective was first minimized alone (see table 1), the minimum number of interconnections is 11, the minimum fresh water is 10 T.h^{-1}, and the amount of regenerated water is null (all the used water is discharged, case without any practical interest). These values are reported in bold in table 1.

Table 1. Results of the mono-objective optimization

	Interconnections number	Fresh water flow rate (T/h)	Regenerated water flow rate (T/h)
Interconnections number	**11**	259.9	10
Fresh water flow rate (T/h)	120	**10**	285.7
Regenerated water flow rate (T/h)	120	289.4	**0**

Then, the biobjective optimization is performed for different values of the interconnection number. So, for 120 (respectively 11) interconnections, the Pareto fronts are reported in Fig. 1a (respectively 1b), Fig. 1c giving a zoom of Fig. 1b. Fig. 1d shows the minimum value of fresh water flow versus the number of interconnections (the regenerated water flow is not taken into account). As it was already shown in [18], the Pareto fronts of Fig. 1a, b and c are linear. For 120 interconnections, all the points located on the straight line are feasible solutions; for the minimum number of interconnections (11), the Pareto front is reduced and composed of a finite number of points located on a straight line with the same slope than the one corresponding to 120 interconnections. From Fig. 1b, no solution exists for a regeneration flow greater than 80 T.h^{-1}. The results presented in Fig. 1c, obtained by taking a step length of 1 instead of 20 in the lexicographic procedure, shows no additional solution located between the points. From Fig. 1d, it can be observed that for 11, 12, 13 and 14 connections, the fresh water flow rate is respectively 90, 50, 25 and 10 T.h^{-1} and remains fixed at 10 T.h^{-1} when the number of interconnections varies from 15 up to 120. The values obtained for the example problem are identical with the ones reported in the literature [6, 9], so the solution procedure is numerically validated.

5. Conclusion and future works

From this example, the following items can be pointed out: (i) for the maximum number of connections (120), the Pareto front is a straight line and all the points located on it are feasible solutions, (ii) the Pareto front is reduced with the number of connections, (iii) for low numbers of connections corresponding to highly constrained problems, the Pareto front is constituted by isolated points located on a straight line, and no feasible solution exists between the points. So the number of feasible solutions decreases with

the number of connections. Finally, since it only requires a standard initialization phase and can tackle large scale problems, this approach will be implemented in the next future on the one hand to optimize ecopark networks, and on the other hand to solve multicontaminant problems.
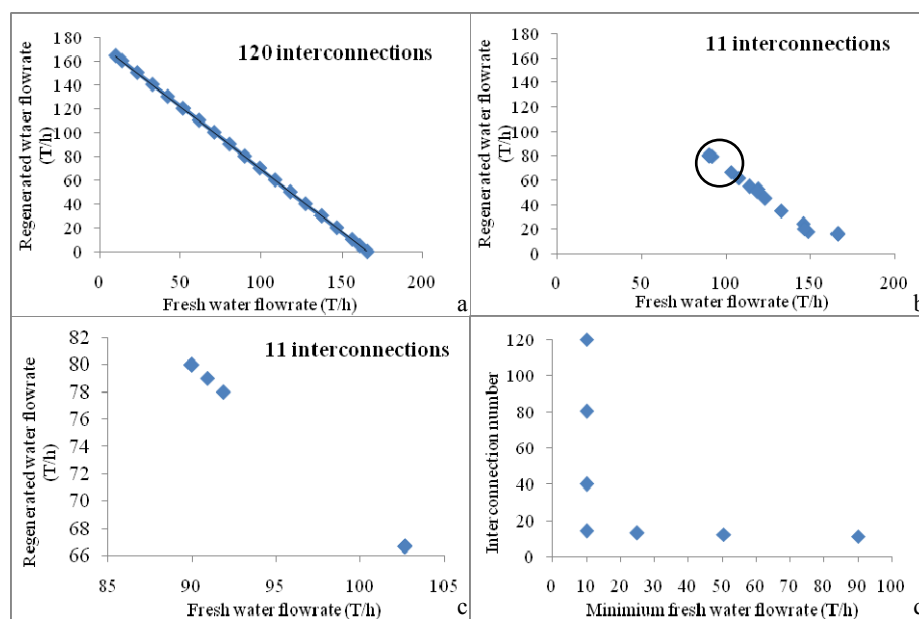
Fig.1 Pareto fronts obtained with the biobjective optimizations

References

[1] B. Dave and O. Nalco., National Academy Press, (2002).
[2] Z.A Manan, S.R. Wan Alwi and Z. Ujang, Desalination, 194 (2006) 52.
[3] R.F. Dunn and M.M. El-Halwagi, J. Chem. Technol. Biotechnol., 78 (2003) 1011.
[4] J. Jacob, C. Viviant, H.F. Houle and J. Paris, Pulp Pap. Can., 103 (2002) 24.
[5] B. Linhoff and R. Smith, Pro. Sys. Eng., 1984.
[6] X. Feng, J. Bai, H. Wang and X. Zheng, Comput. Chem. Eng., 32 (2008) 1892.
[7] M. Savelski and M. Bagajewicz, Chem. Eng. Sci., 58 (2003) 5349.
[8] C.H. Huang, C.T. Chang and C.C. Chang, Ind. Eng. Chem. Res., 38 (1999) 2666.
[9] M. Bagajewicz and M. Savelski, Chem. Eng. Res. Des. 79 (2001) 600.
[10] S. Shafiei, S. Domenech, R. Koteles and J. Paris, J. Clean. Prod., 12 (2003) 131.
[11] N. Hallale and D.M. Fraser, Trans. Inst. Chem. Eng., 78 (2000) 202.
[12] M.M. El-Halwagi, Academic Press, CA, USA (1997).
[13] M.M. El-Halwagi, F. Gabriel and D. Harell, Proc. Des. Cont., 42 (2003) 4319.
[14] Y.P. Wang and R. Smith, Chem. Eng. Sci., 49 (1994) 981.
[15] I. Quesada and I.E. Grossmann, Comp. Chem. Eng., 19 (1995) 1219.
[16] B. Galan and I.E. Grossmann, Ind. Eng. Chem. Res., 37 (1998) 4036.
[17] G. Mavrotas and D. Diakoulaki, Ap. Math. Comp., 171 (2005) 53.
[18] X. Feng, J. Bai and X. Zheng, Comput. Eng. Sci., 62 (2007) 2127.

Broke management optimization in design of paper production systems

Aino Ropponen[a], Risto Ritala[a], Efstratios N. Pistikopoulos[b]

[a]*Tampere University of Technology, Department of Automation Science and Engineering, P.O. Box 692, FI-33101 Tampere, Finland,
aino.ropponen@tut.fi/risto.ritala@tut.fi*
[b] *Imperial College, Department of Chemical Engineering, London SW7 2BY, U.K., e.pistikopoulos@ic.ac.uk.*

Abstract

This paper presents an optimization strategy for a broke system design and operation in papermaking process. A multiobjective stochastic optimization model is presented featuring (i) a stochastic two-state Markov process based submodel for the broke tower, (ii) an operational submodel for the optimization of the broke dosage, and (iii) a multiobjective design problem. An efficient optimization strategy is also proposed involving a quadratic optimization step for the operational subproblem and an effective multiobjective design optimization step.

Keywords: Papermaking process, broke management, multiobjective optimization, stochastic process, process design

1. Introduction

In papermaking, broke is discarded production, which is generated mainly during the web breaks when all the production becomes unused [1]. Broke is stored in a tower and reused as raw material for economic reasons [1]. The reuse may cause disturbances to the process as the material components and their composition can be different from the components and composition of the fresh pulp. Fast changes of the recycled broke, the dosage, are likely to cause the quality of the paper produced to deteriorate and the higher the dosage the greater the probability for a new break. Web breaks and thereby the generation of the broke are random events, hence in broke management the task of controlling the dosage of the reused broke to the manufacturing papermaking system becomes stochastic. As the volume of the broke tower increases, broke management becomes easier, albeit at the expense of higher capital costs due to an increase of the size of the broke tower.

In this paper we address the broke management through a multiobjective [2-3] stochastic optimization model which involves two levels – an operational inner optimization subproblem featuring a stochastic broke tower model, and a design outer optimization subproblem. Multiple objectives are considered at both levels.

This paper is organized as follows. In section 2, the stochastic broke tower model is presented whereas section 3 introduces an effective optimization solution strategy suitable for both the operational and design broke management optimization levels. An example is used throughout to demonstrate some key features of the proposed approach. Finally, section 4 provides a summary of the key points and highlights future/ongoing research directions.

2. Multiobjective stochastic design/operational optimization model

2.1. Broke tower model

At the operational level, the objective of the broke management task is to optimize the schedule of the break dosage, the amount of the broke recycled to the system. At any given time step, the optimal dosage depends on three factors – the current amount of the broke in the broke tower, the previous dosages and the (critical) information of whether a break occurs or not. The occurrence of the breaks is a random event and is typically provided by a break probability function, which is assumed to depend on the broke dosage.

By considering discrete time steps, at time $n+1$ the amount of broke in the tower can be described as follows

$$V(n+1) = V(n) - u(n) + (1-b(n))v_0 + b(n)v_1 \tag{1}$$

where $V(n)$ is the current amount of the broke in the tower, $u(n)$ is the dosage from the tower to the system, v_0 is the amount of broke generated per time step when there is no break, v_1 is the amount of broke generated during a break ($v_1 \gg v_0$), and $b(n)$ is a binary break variable as

$$b(n) = \begin{cases} 1 & \text{íf break on} \\ 0 & \text{else} \end{cases} \tag{2}$$

with transition probability from 0 to 1 as $q_1 = q_1(u_{eff}(n))$, and constant transition probability from 1 to 0 as q_2 (i.e. a two-state Markov chain).

Breaks occur with probabilities

$$\begin{vmatrix} p(b(n+1)=0) \\ p(b(n+1)=1) \end{vmatrix} = \begin{vmatrix} 1-q_1(u_{eff}(n)) & q_2 \\ q_1(u_{eff}(n)) & 1-q_2 \end{vmatrix} \begin{vmatrix} p(b(n)=0) \\ p(b(n)=1) \end{vmatrix}$$

$$q_1(u_{eff}(n)) = S(u_{eff}(n)) = q_{min} + \frac{q_{max} - q_{min}}{1 + \exp\left(-\dfrac{u_{eff}(n) - u_{th}}{\sigma_w}\right)} \tag{3}$$

$$u_{eff}(n) = \sum_{i=1}^{\infty} s(i)u(n-i) \quad \sum_{i=1}^{\infty} s(i) = 1$$

where $S(u_{eff}(n))$ is the break model known to the decision maker, with parameters q_{max}, q_{min}, u_{th} and σ_w, and u_{eff} is the effective dosage with s as a vector of coefficients defining the dynamics between the dosage and the break probability. The volume of broke at time n can be then calculated as

$$V(n) = V(0) + nv_0 - \sum_{n'=0}^{n-1} u(n') + Z(n,b(0))(v_1 - v_0) \tag{4}$$

where $Z(n,b(0))$ is the number of time steps with break on after n time steps when the initial break state has been $b(0)$. Fig. 1 shows the effect of the initial state $b(0)$ to the distribution of $Z(n,b(0))$ for four n-values.

Broke management optimization in design of paper production systems
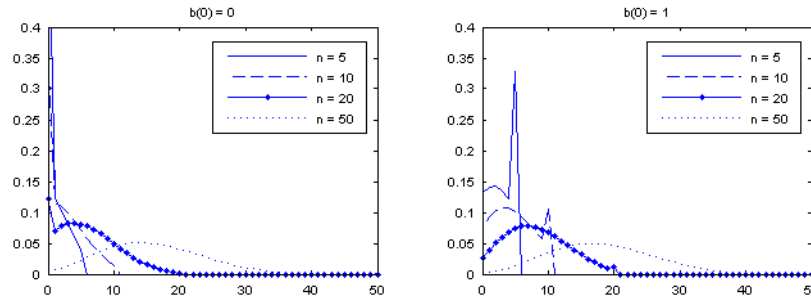
Fig. 1. The distribution of the break instants for the initial states $b(0)=0$ and $b(0)=1$, using constant break probability $q_1=0.1$. The probability that break ends is $q_2=0.2$.

2.2. Operational model – inner optimization subproblem

In broke management, the target is to optimize the schedule of the broke dosage to the process K_H time steps ahead. The objectives are (i) to minimize the probability of broke tower overflow, and to maximize (ii) the effective production time, (iii) the uniformity of quality of the end product, and (iv) the smoothness of the dosage. The operational problem can be then formulated as follows.

$$\min_{\{u(n+k)\}_{k=0}^{K_H-1}} \begin{cases} \max_{k=1,..K_H} \left\lfloor \dfrac{p(V(n+k) > V_{max})}{p_0(k)} \right\rfloor & \text{(i)} \\[2mm] \displaystyle\sum_{k=1}^{K_H} \gamma(k) q_1(u_{eff}(n+k)) & \text{(ii)} \\[2mm] \displaystyle\sum_{k=1}^{K_H} \gamma(k) c_f(n+k)^2 & \text{(iii)} \\[2mm] \displaystyle\sum_{k=1}^{K_H} \gamma(k)\big(u(n+k-1) - u(n+k)\big)^2 & \text{(iv)} \end{cases} \tag{5}$$

$$s.t. \quad P(V(n+k) < 0) = 0 \quad k = 1,..,K_H$$
$$0 \le u(n+k) \le u_{max} \quad k = 1,..,K_H$$

given $\ V(n), b(n), \{u(k)\}_{k=-\infty}^{n-1}$

where K_H is the optimization horizon, $p_0(k)$ is the accepted risk of overflow k time steps from the present time, $\gamma(k)$ is a time-wise weighting factor for the objectives, and c_f is a filler content variation that is consider as an indicator for uniformity of quality of the end product. The filler content deviation is defined as

$$c_f(n) = \sum_{i=1}^{\infty} h(i) u(n-i) \quad \sum_{i=1}^{\infty} h(i) = 0 \tag{6}$$

As the filler content is controlled further down the production line, the coefficients of the filler response, $h(n)$, sum up to zero, and the effect of the broke dosage on the filler is transient.

2.3. Design model – outer optimization problem

The objectives in the broke tower design optimization are to minimize the investment cost of the tower volume and to maximize the performance of the design (i-iv). As the behavior of the system is stochastic, the design objectives address the expected values

of behavior. Let us assume we have a dosage policy. When this policy is applied, together with the system dynamics Eqs. (1-4, 6), this defines a stochastic process on system states. We shall denote this stochastic process as Ψ and the expectation value of system performance measures with respect to it as $E_\Psi\{\ \}$. The design problem can then be formulated as follows

$$\min_{V_{\max}} \left[H(V_{\max}), -E_\Psi\{T_{of}\}, E_\Psi\{q_1\}, E_\Psi\{c_f^2\}, E_\Psi\{(u(n+1)-u(n))^2\} \right] \tag{7}$$

$$s.t.\ V_{\max} \geq 0$$

where $H(V_{\max})$ is the investment cost of a tower volume $V_{\max}$ and T_{of} is the broke tower overflow time as $T_{of} = \min_n \{n \mid V(n) > V_{\max}\}$.

3. Optimization solution strategy

At the inner operational optimization level, we reformulate Eq. (5) by considering a weight average of the objectives (ii), (iii) and (iv) [see section 2.2], while objective (i), the probability of broke tower overflow, is considered as a constraint. This leads to the following optimization problem

$$\{u^*(n+k)\}_{k=0}^{K_H-1} = \arg\min_{\{u(n+k)\}_{k=0}^{K_H-1}} \sum_{k=1}^{K_H} \gamma(k)\left(c_f(n+k)^2 + \alpha(u(n+k)-u(n+k-1))^2 + \beta u(n+k-1)^2\right)$$

$$s.t.\ P(V(n+k) > V_{\max}) < 1-(1-p_0)^k \equiv p_0(k) \quad k=1,..,K_H$$

$$P(V(n+k) < 0) = 0 \quad k=1,..,K_H \tag{8}$$

$$0 \leq u(n+k) \leq u_{\max} \quad k=1,..,K_H$$

where α and β are scalar parameters, considered as design parameters at the design level. As Eq. (8) cannot be solved directly, we further consider the following: First, the probability of broke tower constraint can be rewritten as

$$-\sum_{k=0}^{K_H-1} u(n+k) \leq \frac{V_{\max}-V(n)}{\Delta t} - kv_0 - (v_1-v_0)F_{Z(k,b(n))}^{-1}(1-p_0(k)) \tag{9}$$

where F_Z is the cumulative distribution of Z. Eq. (9) can be further simplified by assuming that over the optimization horizon, the probability of break q_1 is known at any given step based on which Eq. (8) becomes a quadratic problem/programming. An algorithm then can be constructed involving the following steps

Step 1. Initialize the dosage, calculate the effective dosage and break probability $\{q_{1,0}(u_{eff,0}(n+k))\}_{k=0}^{K_H-1}$ over the optimization horizon. Set $j=0$.

Step 2. Solve the quadratic-linear optimization problem defined by $\{q_{1,j}(u_{eff,j}(n+k))\}_{k=0}^{K_H-1}$ and giving $\{u_{j+1}^*(n+k)\}_{k=0}^{K_H-1}$.

Step 3. If $\max|u_{j+1}^* - u_j^*| < \varepsilon$ stop, else continue.

Step 4. Calculate the new break probability as $q_{1,j+1}(u_{eff,j+1}(n)) = S(u_{eff,j+1}(n))$

with $u_{eff,j+1}(n) = \sum_{i=1}^{\infty} s(i)u_{j+1}^*(n-i)$. Set $j \leftarrow j+1$ and go back to step 2.

Fig. 2 shows an example of the broke dosage schedule optimization at the time n for the optimization horizon 30 for varying initial conditions. It can be seen that the current state of the system strongly affects the optimal solution.
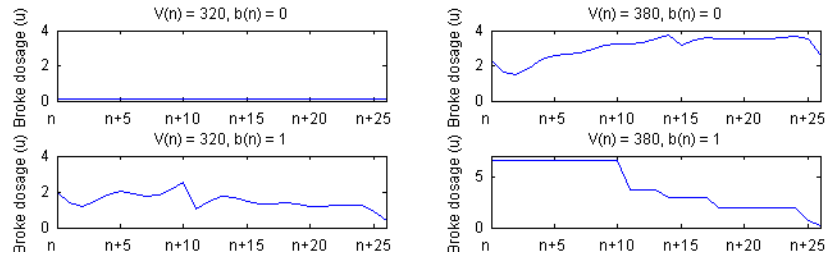
Fig. 2. The dosage schedule optimized at the time n for the optimization horizon $K_H=30$ for initial states $b(n) = 0$ (no break) and $b(n) = 1$ (break), and broke volumes $V(n)=320$ and $V(n)=380$, and initial. The tower volume in all the cases is $V_{max} = 400$, and the maximum dosage is set to $u_{max} = 4$. There is a delay of 3 time steps in the response, thus the optimization is done until 27 only. In the case $V(n)=380$ and $b(n) = 1$ (bottom right) no feasible solution is found.

Fig. 3 shows two examples in which the broke tower model is simulated, and the dosage is optimized and $u^*(n)$ executed at each time step.
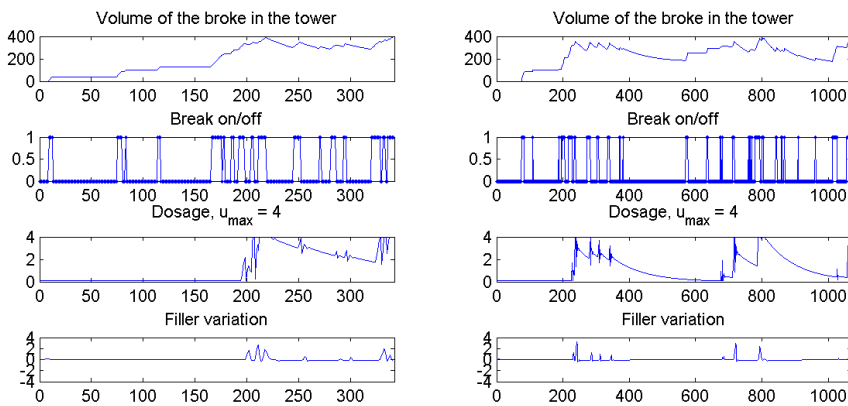
Fig. 3. Example of two broke tower simulations starting from the same initial conditions and ending up to a relatively different result. Both simulations are run until overflow occurs.

At the design level the minimization of the objectives (Eq. (7)) is done with respect to the broke tower volume V_{max} and the parameters β and p_0. The goal on smooth dosage has been removed, and the corresponding scalar parameter α is determined intuitively. Fig. 4 shows a solution set in which for a set of 126 designs (V_{max}, β, p_0), 20 simulations were carried out and the expectation value was estimated as the mean of the 20 observations. Then, the most preferable solution can be chosen from this set based on the decision maker's subjective assessment.
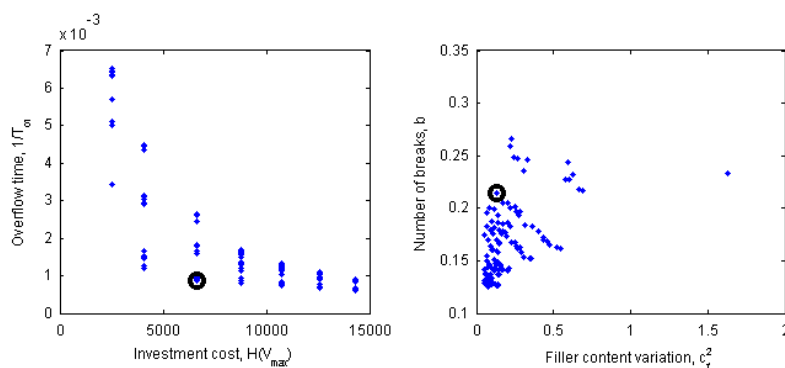
Fig. 4. The set of designs in respect to the investment cost and the tower overflow (left), and the filler content variation and the number of breaks (right). The decision maker's choice is marked with a circle.

4. Conclusions and future work

In this paper we have formulated the stochastic broke management as a multiobjective bilevel optimization problem with design and operational levels. For the operational level we have presented an efficient a strategy to solve the problem online without a decision maker by formulating the problem in a quadratic programming form and using scalarization of the multiple objectives. For the design problem we have presented a straight forward method to generate Pareto optimal solutions through simulations. This can be generalized for the subsystem methodology presented in [4,5].

In this study, we have assumed that the break model, i.e. how the dosage affects the break probability, is known to the decision maker. Usually there is relevant uncertainty about such a model. Hence, our future work will analyze the robustness of the results. We are going to study more effective methods to generate Pareto optimal solutions, and to expand the strategies presented in this paper to cover the entire material flow in papermaking, especially the optimization of the four mass/water tower operation and designs. Our future studies will be based on the operational and design optimization strategies presented in this paper.

5. Acknowledgements

References

[1] H. Paulapuro, 2008, Papermaking Science and Technology, Papermaking Part 1, Stock Preparation and Wet End, Second Edition, Finnish Paper Engineers' Association.
[2] K. Miettinen, 1999, Nonlinear multiobjective optimization, Kluwer.
[3] P. Clark, A. Westerberg, 1983, Optimization for design problems having more than one objective, Computers & Chemical Engineering, Vol. 7,No. 4, pp. 259-278.
[4] A. Engau, M. Wiecek, 2007, 2D decision-making for multicriteria design optimization, Struct Multidisc Optim, Vol. 34, pp. 301-315.
[5] A. Engau, M. Wiecek, 2008, Interactive Coordination of Objective Decompositions in Multiobjective Programming, Management Science, Vol. 54, No. 7, pp. 1350-1363

Recycle and reuse mass exchange networks based on properties using a global optimization technique

Fabricio Nápoles-Rivera[a], Arturo Jiménez-Gutiérrez[a], José María Ponce-Ortega[b], Mahmoud M. El-Halwagi[c]

[a] Instituto Tecnológico de Celaya, Departamento de Ingeniería Química, Celaya, Gto. México
[b] Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mich. México
[c] Texas A&M University, Chemical Engineering Department, College Station, TX, USA

Abstract

This paper presents a rigorous model for the optimal design of mass exchange networks based on properties under a recycle and reuse scheme. The model features an in-plant treatment system and considers simultaneously process and environmental constraints for properties such as composition, toxicity, pH, chemical oxygen demand, density and viscosity. The model is a mixed integer non linear programming problem with bilinear terms on the property balances, and it is solved using a global optimization algorithm. The advantages of the proposed approach over sequential methodologies are shown with the solution of two cases of study.

Keywords: *Property Integration, Mass Integration, Environmental Constraints, Global Optimization*.

1. Introduction

Mass integration based on properties is a new approach to deal with environmental an economic issues in the chemical industry to reduce fresh sources consumption and waste generation. It was first proposed by Shelley and El-Halwagi (2000) to solve one of the limitations of mass integration strategies, in that they considered constraints on flowrates and compositions, but not on properties. This new paradigm was called property integration (El-Halwagi *et al.,* 2004). Ponce-Ortega *et al.* (2009a) addressed the problem of property integration under a direct recycle scheme. They considered simultaneously process and environmental constraints and found that this optimization policy resulted in network configurations with lower overall costs when compared with sequential approaches, in which the process network optimization and the treatment system optimization were treated as separate problems. Later, Ponce-Ortega *et al.* (2009b) proposed a formulation for a recycle and reuse configuration with distributed treatment facilities, so that the flowrates treated in each unit were reduced; in this formulation, most of the nonlinearities were avoided.

In this work, a rigorous model for a property integration network under a recycle and reuse configuration is presented (see Figure 1). The model considers simultaneously process and environmental constraints. The resulting model is a mixed integer non lineal programming problem (MINLP), which is solved using a global optimization technique.

2. Model Formulation

Mass balances in the splitting and mixing points of the network:
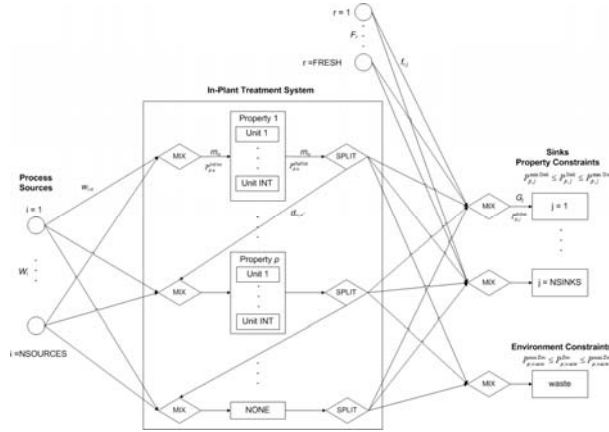
Figure 1. Rigorous Mass and Property Integration Network: Recycle and Reuse Scheme

$$F_r = \sum_{\substack{j \in NSINKS \\ j \neq waste}} f_{r,j}, \quad r \in FRESH \tag{1}$$

$$W_i = \sum_{u \in NINT} w_{i,u}, \quad i \in NSOURCES \tag{2}$$

$$\sum_{i \in NSOURCE} w_{i,u} + \sum_{\substack{u^1 \in NUNITS \\ u^1 \neq u}} d_{u^1,u}, = m_u \quad u \in NUNITS \tag{3}$$

$$m_u = \sum_{j \in SINKS} g_{u,j} + \sum_{\substack{u^1 \in NUNITS \\ u^1 \neq u}} d_{u,u^1}, \quad u \in NUNITS \tag{4}$$

$$G_j = \sum_{r \in FRESH} f_{r,j} + \sum_{u \in NUNITS} g_{u,j}, \quad j \in NSINKS \tag{5}$$

Property balances in the mixing points of the network:

$$\sum_{i \in NSOURCE} \left[w_{i,u} \psi_p \left(P_{p,i}^{InSource} \right) \right] + \sum_{\substack{u^1 \in NUNITS \\ u^1 \neq u}} \left[d_{u^1,u} \psi_p \left(P_{p,u^1}^{OutUnit} \right) \right] = m_u \psi_p \left(P_{p,u}^{InUnit} \right), \begin{array}{l} u \in NUNITS \\ p \in NPROP \end{array} \tag{6}$$

$$G_j \psi_p \left(P_{p,j}^{InSink} \right) = \sum_{r \in NFRESH} \left[\psi_p \left(P_{p,r}^{InFresh} \right) f_{r,j} \right] + \sum_{u \in NUNITS} \left[g_{u,j} \psi_p \left(P_{p,u}^{OutUnit} \right) \right], \begin{array}{l} j \in NSINKS \\ p \in NPROP \end{array} \tag{7}$$

Notice the existence of bilinear terms in the property balances.

A set of units with known unit costs and correction factors are available to treat the properties here considered. The following disjunction is used to choose which unit should be used for each property:

$$\bigvee_{I(U) \in NUNITS} \begin{bmatrix} Y_{I(U)}^{U(p)} \\ \psi_p \left(P_{p,U(p)}^{OutUnit} \right) = \psi_p \left(P_{p,U(p)}^{InUnit} \right) \left(1 - \alpha_{I(U)}^{U(p)} \right) \\ Cost_{U(p)}^{unit} = Cost_{I(U)}^{U(p)} m_{U(p)} \end{bmatrix}$$

The convex hull reformulation is used to model the disjunction as follows:

$$\sum_{I(U)} y_{I(U)}^{U(p')} = 1, \quad U(p') \in NUNITS \tag{8}$$

Recycle and reuse mass exchange networks based on properties using a global opt. technique

$$\psi_{p'}\left(P_{p',U(p')}^{OutUnit}\right) = \sum_{I(U)} \psi_{p'}\left(p_{p',U(p')}^{OutUnit,I(U)}\right), \quad U(p') \in NUNITS \tag{8}$$

$$\psi_{p'}\left(P_{p',U(p')}^{InUnit}\right) = \sum_{I(U)} \psi_{p'}\left(p_{p',U(p')}^{InUnit,I(U)}\right), \quad U(p') \in NUNITS \tag{9}$$

$$m_{U(p')} = \sum_{I(U)} m_{U(p')}^{I(U)}, \quad U(p') \in NUNITS \tag{10}$$

$$Cost_{U(p')}^{unit} = \sum_{I(U)} Cost_{U(p')}^{unit,I(U)}, \quad U(p') \in NUNITS \tag{11}$$

$$p_{p',U(p')}^{OutUnit,I(U)} = p_{p',U(p')}^{InUnit,I(U)}\left(1 - \alpha_{p',I(U)}^{U(p')}\right), \quad U(p') \in NUNITS, I(U) \in NINT \tag{13}$$

$$Cost_{U(p')}^{unit,I(U)} = Cost_{I(U)}^{U(p')} m_{U(p')}^{I(U)}, \quad U(p') \in NUNITS, I(U) \in NINT \tag{14}$$

$$p_{p',U(p')}^{OutUnit,I(U)} \le M^{p'} y_{I(U)}^{U(p')}, \quad U(p') \in NUNITS, I(U) \in NINT \tag{15}$$

$$p_{p',U(p')}^{InUnit,I(U)} \le M^{p'} y_{I(U)}^{U(p')}, \quad U(p') \in NUNITS, I(U) \in NINT \tag{16}$$

$$m_{U(p')}^{I(U)} \le M^{m} y_{I(U)}^{U(p')}, \quad U(p') \in NUNITS, I(U) \in NINT \tag{17}$$

$$Cost_{U(p')}^{Unit,I(U)} \le M^{Cost^{p'}} y_{I(U)}^{U(p')}, \quad U(p') \in NUNITS, I(U) \in NINT \tag{18}$$

Only one property can be treated in each unit while the rest of the properties remain unchanged; thus, the property balance for the properties not treated is:

$$\psi_{p'}\left(P_{p',U(p')}^{OutUnit}\right) = \psi_{p'}\left(P_{p',U(p')}^{InUnit}\right), \quad U(p') \in NUNITS, \forall p \ne p' \tag{19}$$

Process and environmental constraints are considered simultaneously:

$$P_{p,j}^{\min Sink} \le P_{p,j}^{inSink} \le P_{p,j}^{\max Sink}, \quad j \in NSINK, j \ne waste, P \in NPROP \tag{20}$$

$$P_{p,j}^{\min Env} \le P_{p,waste} \le P_{p,j}^{\max Env}, \quad j = waste, P \in NPROP \tag{21}$$

The objective function consists in the minimization of the total annual cost for the network, which accounts for fresh sources cost, treatment units cost and piping costs:

$$
\begin{aligned}
Min\ TAC = H_Y \sum_{r \in FRESH} Cost_r^{Fresh} F_r + H_Y \sum_{u \in NUNITS} Cost_u^{Unit} \\
+ H_Y \sum_{u \in NUNITS} \sum_{i \in NSOURCES} \left(pip_{i,u}^{In} w_{i,u}\right) + H_Y \sum_{u \in NUNITS} \sum_{u' \in NUNITS} \left(pip_{u,u'}^{Int1} d_{u,u'}\right) \\
+ H_Y \sum_{u \in NUNITS} \sum_{j \in NSINKS} \left(pip_{u,j}^{Exit} g_{u,j}\right) + H_Y \sum_{r \in NFRESH} \sum_{j \in NSINKS} \left(pip_{r,j}^{InFresh} f_{r,j}\right)
\end{aligned}
\tag{22}
$$

To solve the nonconvex MINLP, a deterministic global optimization technique similar to the one proposed by Karuppiah and Grossmann (2006) is used.

3. Results

Two cases of study are solved to show the applicability of the proposed methodology.
Case 1. The first example consists of two process sources, one fresh source and two process sinks. The properties for the process and fresh sources are shown in Table 1. The optimal configuration for this problem is shown in Figure 2. In this example, all the available treatment units were used to satisfy process and environmental constraints. The units selected to treat toxicity and pH were those with the highest correction factors, while the units used to treat composition and THOD were those with the lowest factors. Notice the use of the fictitious unit NONE, where the change in properties is only due to the mixing of streams.

Table 1. Process (W) and fresh (F) sources characteristics for Example 1

Stream	Flowrate	Composition (ppm)	Toxicity (%)	THOD (mg O_2/l)	pH	Density	Viscosity
W1	2900	0.033	0.8	75	5.3	2	1.256
W2	2450	0.022	0.5	88	5.1	2.208	1.22
F1	-	0	0	0	7	2.204	1.002
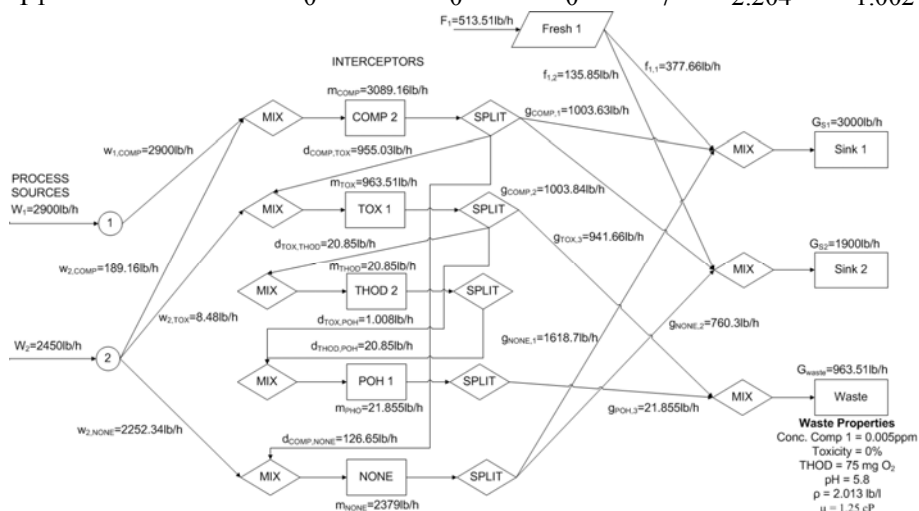
Figure 2. Optimal configuration for Example 1

Table 2 shows the most relevant results for this problem, as well as a comparison with the results obtained with a sequential approach. Notice that the network cost for the sequential approach is lower than the one obtained with the simultaneous approach; however, the waste generated does not satisfy environmental regulations, and the implementation of the additional treatment system raises the total cost of the network to $271807/year, which is 32% higher than the one obtained with the simultaneous strategy.

Table 2. Comparison between simultaneous and sequential approach for Example 1

Concept	Optimal Solution (Simultaneous Approach)	Sequential Approach
F_1 (lb/hr)	513.519	515.398
Waste (lb/hr)	963.519	965.398
Fresh sources cost ($/year)	36973.33	37108.681
Treatment Units Cost ($/year)	158764.128	72611.597
Piping Cost ($/year)	10170.596	9939.972
Network Cost ($/year)	205908.058	119660.249
Waste Treatment System ($/year)	-	152146.725
Total Annual Cost ($/year)	205908.058	271806.978

Case 2. The second problem consists of three process sources, two fresh sources and three process sinks. The characteristics of all the sources are given in Table 3. The optimal configuration for the network is shown in Figure 3.

Table 3. Process (W) and fresh (F) sources characteristics for example 2

Stream	Flowrate	Composition (ppm)	Toxicity (%)	THOD (mg O₂/l)	pH	Density	Viscosity
W1	8083	0.016	0.3	0.187	6.4	2.000	1.256
W2	3900	0.024	0.5	48.85	5.1	2.208	1.22
W3	3279	0.22	1.5	92.10	4.8	2.305	1.261
F1	-	0	0	0	7	2.204	1.002
F2	-	0.01	0.1	0.01	6.8	2.209	0.992
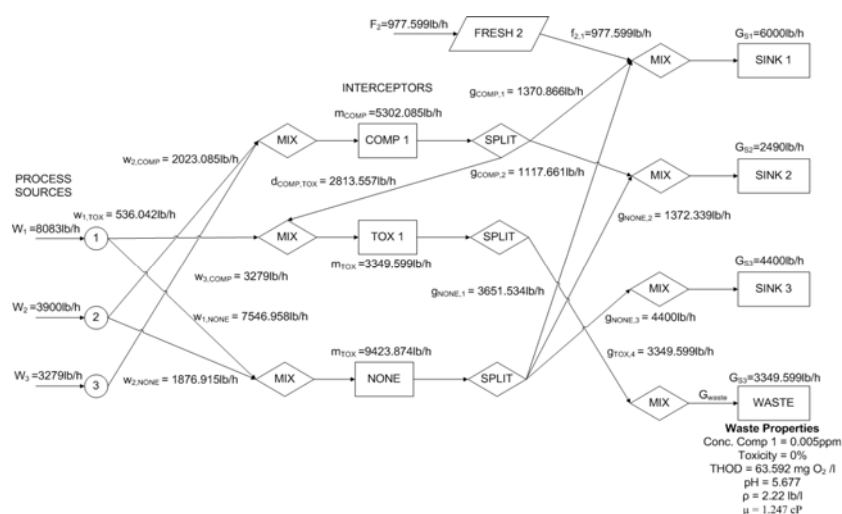
Figure 3. Optimal configuration for Example 2

Table 4 shows a comparison between the simultaneous and sequential approaches. Notice that the wastewater minimization policy is not always the best strategy when environmental constraints are considered. In this case, the simultaneous approach requires approximately 100 lb/h more of fresh source 2 (the cheapest utility) than the sequential solution, but provides an integrated network with a lower total annual cost.

Table 4. Comparison between simultaneous and sequential approach for Example 2

Concept	Optimal Solution (Simultaneous Approach)	Sequential Approach
F_2 (lb/hr)	977.599	879.834
waste (lb/hr)	3349.599	3251.834
Fresh sources cost ($/year)	46924.754	42232.035
Treatment Units Cost ($/year)	538316.968	155504.671
Piping Cost ($/year)	29297.187	28482.818
Total annual cost ($/year)	614538.909	65258.677

4. Computational results

Figure 4a and 4b show the evolution of the global optimization algorithm. OUB represents the upper bound (solution of the original MINLP) and LB represents the lower bound (solution for the relaxed MILP), and the algorithm finishes when the gap between the two solutions reaches a convergence criterion. Notice in Figure 4a that the

best OUB is reached in seven iterations; the algorithm converged in eight iterations with a CPU time of 131.29s for a final gap of 0.0118 % between the lower and upper bounds. For Example 2, Figure 4b shows that the best OUB of the root node remains constant throughout the nine iterations that were required to reach a final gap of 4.4% with the LB. The solution took 906.62s of CPU time; no further improvement was achieved in this case after exhausting all of the branching variables.

a)

b)
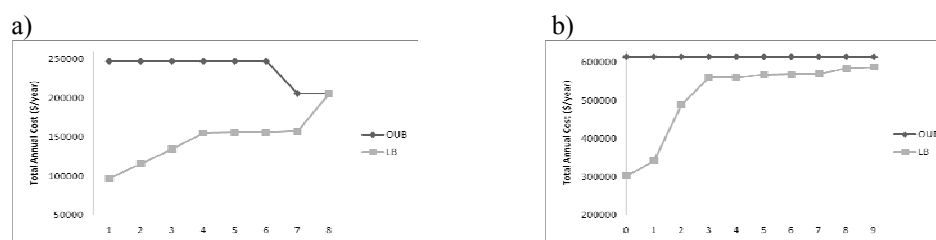
Figure 4. Evolution of the global optimization algorithm for a) Example 1 and b) Example 2

5. Conclusions

A rigorous model for a mass integration network under a recycle and reuse scheme was presented. The model includes process and environmental constraints simultaneously. This model is more rigurous with respect to previously reported methodologies because it considers the simultanoeus optimization of the treatment system based on properties. The resulting MINLP problem, which includes bilinear terms in the property balances, was solved using a global optimization technique. The results demonstrate the advantages of the simultaneous approach, yielding lower overall network costs when compared with conventional sequential apporaches. It is important to remark that as a consequence of the proposed methodology, a reduction in fresh water consumption is obtained by properly recycling waste process streams.

6. References

El-Halwagi, M. M., I. M. Glaswog, X. Qin and M. R. Eden (2004). "Property integration: Componentless design techniques and visualization tools" AIChE Journal. **50**(8): 1854-1869.

Karuppiah, R. and I. Grossmann (2006). "Global optimization for the synthesis of integrated water systems in chemical processes" Computer and Chemical Engineering. **20**(4): 650-673.

Ponce-Ortega, J. M., A. C. Hortua, M. M. El-Halwagi and A. Jiménez Gutierrez (2009a). "A property-based optimization of direct recycle networks and wastewater treatment processes" AIChE Journal. **55**(9): 2329-2344.

Ponce-Ortega, J. M., A. Jiménez Gutierrez and M. M. El-Halwagi (2009b). "Simplified formulation for the mass and property interception networks considering environmental constraints" Accepted in Computers and chemical Engineering.

Shelley, M. D. and M. M. El-Halwagi (2000). "Component-less design of recovery and allocation systemas: A functionality-based clustering approach" Computers and Chemical Engineering. **24**: 2081-2091.

Energy Management for Heat Intensive Production Plants using Mixed Integer Optimization

Steffen Sommer,[a] Jens Böhm,[b] Achim Kienle[a,c]

*aOtto-von-Guericke-University, Uni-Platz 2, 39106 Magdeburg, Germany,
steffen.sommer@ovgu.de*
*bRothenseer Generatorenfertigung GmbH, August-Bebel-Damm 24-30, 39126
Magdeburg, Germany, jens.boehm@enercon.de*
*cMax-Planck-Institute for Dynamics of Complex Technical Systems, Sandtorstr. 1,
39106 Magdeburg, Germany, achim.kienle@mpi-magdeburg.mpg.de*

Abstract

An energy management system for heat intensive production plants is presented. The aim is to minimize energy costs with respect to additional constraints concerning workshop temperature and air ventilation. The system realizes a dynamic load balance, which prevents unneeded load peaks. The problem is solved by means of mixed integer optimization. Therefore, the production plant is modeled mathematically. The energy management system provides considerable savings in energy costs.

Keywords: Energy management, production plant, energy costs, scheduling, mathematical modeling, mathematical optimization, MILP

1. Introduction

In this work, an energy management system for heat intensive production plants is presented. Such a system reduces electrical energy costs. High energy costs result from unneeded load peaks and high energy consumption during the production process.

A process is considered, where rotor blades for wind turbines are produced in a factory work room. In general terms, plastic parts are thermally treated. Heat is released during the production of the rotor blades.

The aim is to minimize the energy costs of the above mentioned process with respect to constraints concerning the sequence of all operations, workshop temperature and air ventilation. The minimization of energy costs can be achieved by means of a dynamic load balance, which prevents unneeded load peaks. The appropriate turn-on and turn-off times of every load have to be found. This problem is solved by means of mathematical optimization.

To solve the optimization problem, the production plant is modeled mathematically. A time-discrete model is built. It consists of sequence constraints and elimination criteria regarding the production process, temperature and ventilation constraints and a dynamic energy balance of the complete factory work hall. All in all, a mixed integer linear program (MILP) is obtained. The optimization result represents the above mentioned dynamic load balance with optimal turn-on and turn-off times. Thus, unnecessary load peaks are prevented, and the electrical energy consumption is reduced simultaneously.

Energy costs are minimized often by means of heuristic strategies. For example in [1] an energy management system for cooling, heating and power technologies using demand forecasting with historical records as ambient temperature or ambient heat radiation is presented.

In contrast to this, here, a rigorous approach is proposed based on mathematical optimization leading to a scheduling problem [2] with path and dynamic constraints arising from temperature and ventilation requirements.

Further publications regarding energy cost minimization in industrial processes using energy management systems are [3-5]. The optimization is also based on mathematical optimization. The main focus is on mathematical modeling of specific processes without considering additional constraints. In [3], a developed industrial load management is applied to a flour mill. The optimal operating schedule that minimizes the operating cost in an energy intensive air separation process is found in [4]. An optimal energy management of the power plant in pulp and paper mills is introduced in [5]. The goal of this study is the fulfillment of the requirements in energy and steam with the minimum possible cost.

The paper is organized as follows. In the next chapter, the considered heat intensive production plant is described and modeled mathematically. After this, the optimization results are presented and compared. A summary and an outlook are given in Chapter 4.

2. The Production Process

2.1. Process description

The production of one rotor blade consists of four ($k=1,...,4$) operations (production steps) with different durations. They are equal to *1h*, *2h*, *1h* and *3h*. Five rotor blades ($b=1,...,5$) have to be manufactured within 48 hours. The completion of the five parts or jobs respectively and the execution of the operations are carried out sequentially. All operations are allocated to a single machine. During the production process, an internal temperature of *18-35°C* and a complete air exchange per hour on average are required. For these purposes, two ($l=1,2$) air ventilations and one heating are available.

The production steps represent the main loads. Additional loads are the heating and the air ventilations for cooling and air change. All loads consume electrical energy.

2.2. Mathematical process model

In this section, the production process including factory hall is modeled mathematically. Due to space restrictions, the detailed mathematical model can not be presented here. A brief summary is given below.

A discrete-time formulation is used because of the path constraints which have to be satisfied over the whole time horizon. The production period of *48h* is divided into *48* time intervals $i=1,...,n_i=1,...,48$ with the duration of $\Delta t = 1h$.

The loads are described by the following variables:

$$IPP(b,k,i) = \begin{cases} 1 & \text{Production active} \\ 0 & \text{else} \end{cases} \tag{1}$$

$$IH(i) = \begin{cases} 1 & \text{Heating active} \\ 0 & \text{else} \end{cases} \tag{2}$$

$$IL(l,i) = \begin{cases} 1 & \text{Air ventilation active} \\ 0 & \text{else.} \end{cases} \tag{3}$$

The production process can be classified as a flow shop scheduling problem. Appropriate sequence constraints and elimination criteria on *IPP(b,k,i)* (1) concerning

the considered time-discrete case have to be implemented. Four sequence constraints are needed to mathematically describe the production flow. The first constraint determines the time interval where the first operation of the first job (*k=1, b=1*) starts. The second and the third constraint are necessary to guarantee that an operation *k* and a job *b* start only if the previous operation *k-1* and the last operation of the previous job *b-1* have been finished. The fourth constraint determines the time interval where the last operation of the last job starts. This time interval depends on the duration of the last operation. Elimination criteria are additional constraints which can reduce the solution set to improve the computational performance. For other scheduling problems like job shop problems for example, these conditions have to be adapted appropriately.

Furthermore, the model consists of the temperature constraint

$$291.15K \leq T(i) \leq 308.15K \tag{4}$$

and the air exchange constraint. The number of internal air exchanges is

$$n_{Ex} \geq 1 \tag{5}$$

per hour on average.

A dynamic energy balance, from which the internal temperature $T(i)$ can be computed in every time interval, is implemented in discrete form:

$$\rho_{Air}\, V_{Hall}\, C_{P,\,Air}\, \frac{T(i+1)-T(i)}{\Delta t} = \dot{Q}_P(i+1) + \dot{Q}_H(i+1) - \dot{Q}_{Air}(i+1) - \dot{Q}_T(i+1) + \dot{Q}_R(i+1). \tag{6}$$

It contains the elements heat power regarding the heating $\dot{Q}_H(i)$ and the thermal operations of the production process $\dot{Q}_P(i)$, cooling power concerning the ventilators $\dot{Q}_{Air}(i)$, heat transfer of the building $\dot{Q}_T(i)$ and ambient heat radiation $\dot{Q}_R(i)$. ρ_{Air}, V_{Hall} and $C_{P,Air}$ are the density of air, the volume of the factory building and the specific heat capacity of air.

The objective function is equal to the energy costs

$$J = E_{Tot}ER + P_{max}DC, \tag{7}$$

which depends on the total consumption of energy E_{Tot} and the maximum power value P_{max} during a specific time period. *ER* is the Energy rate or the price per *kWh* in *€/kWh*, and *DC* is the demand charge in *€/kW*. The following equations are needed:

$$E_{Tot} = \sum_i P_{Tot}(i)\Delta t, \tag{8}$$

$$P_{Tot}(i) = \sum_b \sum_k IPP(b,k,i)P_{el,\,P}(k) + \sum_l IL(l,i)P_{el,\,Air}(l) + IH(i)P_{el,\,H}, \tag{9}$$

$$P_{max} = \max_i \left(P_{Tot}(i) \right). \tag{10}$$

$P_{Tot}(i)$ is the total power of one time interval. It is equal to the sum of all load powers (9). $P_{el,\,P}(k)$ is the electric power of operation *k*, $P_{el,\,Air}(l)$ is the electric power of air ventilation *l*, and $P_{el,\,H}$ is the electric power of the heating.

The mathematical model of the heat intensive production process including optimality criterion (7) represents a mixed integer linear program (MILP).

3. Optimization Results

The optimization problem from Chapter 2 is solved using GAMS/CPLEX [6,7]. To demonstrate the energy management system, a winter scenario is considered. In this context, a typical ambient temperature curve with

$$-3°C \leq Text(i) \leq 2°C \tag{11}$$

is assumed. Some important process parameters are summarized in Table 1. The optimal production process is presented by the Gannt-Chart in Fig. 1. The time intervals are marked, where the loads are active. This solution ensures minimal energy costs subject to the mentioned constraints. The temperature curves are shown in Fig. 2. The computational characteristics are: Intel Core 2 Quad CPU, Q8400, *2.66GHz*, *#* of binary/continuous variables: *2784/292*, *#* of equations/constraints: *7167*, CPU time: *4000s*.

The active production intervals are evenly distributed over the available production period of *48h*. Thus, the heat release is used efficiently in the case of low ambient temperatures to hold the desired minimum internal temperature. The use of the heating is necessary. It is active in *33* of *48* time intervals. It can be inactive during the outstanding heat intensive production steps two and four. The air ventilations are only activated to provide the minimum air exchange of $n_{Ex}=1$ (5). Therefore, *24* active intervals concerning air ventilation *l=1* or *12* active intervals concerning air ventilation *l=2* are necessary. The air ventilation is primarily in operation, if no production is activated. The reason is that the heat effect of the production process must not be compensated by cooling air ventilators. A concurrent activity of production and air ventilation takes only place in intervals where the outstanding heat intensive production steps two and four are active. Thus, the cooling influence on the internal temperature is minimized.

The average internal temperature of $\overline{T}=294K$ is approximately equal to the desired minimum temperature of $T_{min}=291.15K$.

During the *48h* winter scenario the electrical energy of $E_{Tot}=1027.3kWh$ (8) with a maximum power of $P_{max}=30.1kW$ (10) is consumed. The overall energy costs (7) are equal to *352.36€*.

The optimal solution is compared with a conventional solution (Fig. 3, CPU time *62s*), where only the total energy consumption is minimized. That means the optimality criterion is

$$J=E_{Tot}. \tag{12}$$

During *48h*, the total energy of $E_{Tot}=1019.1kWh$ (8) with a maximum power of $P_{max}=74.6kW$ (10) is consumed. From (7), energy costs of *796.95€* are computed. The energy management system provides savings in energy costs of *444.59€*.

Parameter	Value	Parameter	Value
ρ_{Air}	$1.184kg/m^3$	V_{Hall}	$25000m^3$
$C_{P,Air}$	$1.005KJ/kg/K$	$P_{el,P}(k)$	$(18\ 16.5\ 26\ 5)kW$
$P_{el,H}$	$4.1kW$	$P_{el,Air}(l)$	$(18\ 36)kW$
ER	$0.05€/kWh$	DC	$10€/kW$

Table 1. Process parameters

Energy Management for Heat Intensive Production Plants using Mixed Integer Optimization
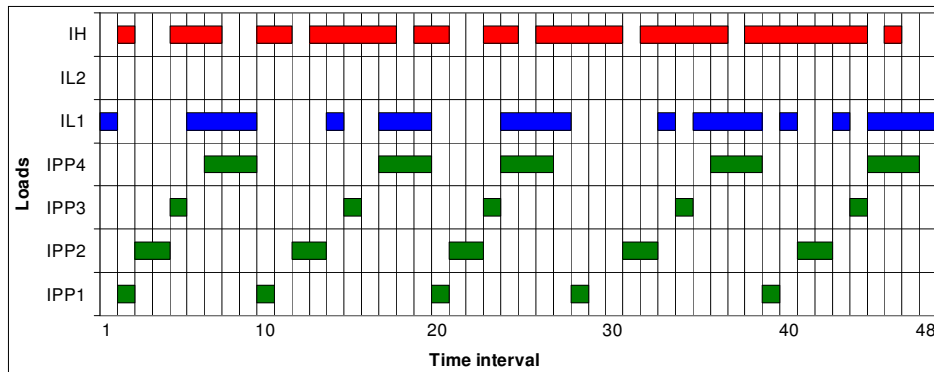
Figure 1. Gannt-chart, winter scenario, optimal solution

Finally, the power curves of the optimal production process and the conventional production process are compared (Fig. 4). In contrast to the conventional production process, no unneeded load peaks occur during the optimal process.

4. Conclusions

The resume of this contribution is as follows. An energy management system for heat intensive production plants, which can be regarded as a new scheduling problem with energy costs as optimality criterion and additional constraints concerning the climatic conditions in the factory work hall, has been introduced. The energy management system provides considerable savings in electrical energy costs.

In future work, we plan to derive simplified procedures from the rigorous optimization results because the optimization process may be quite time-consuming.
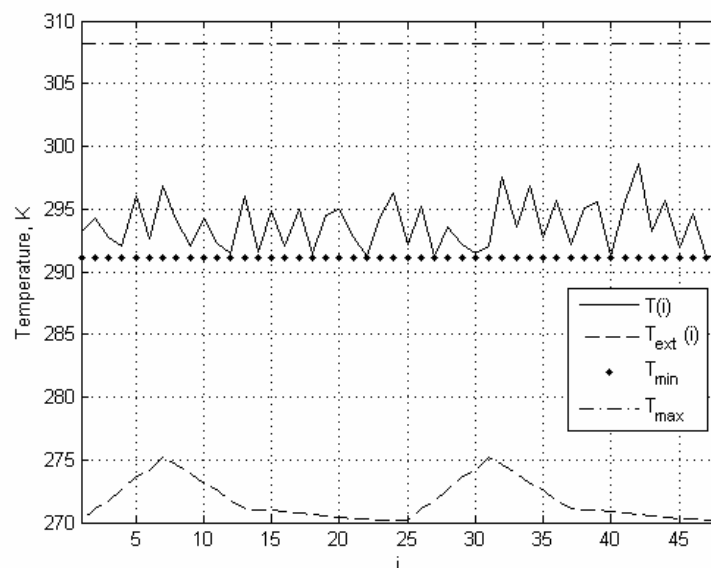
Figure 2. Temperature curves, winter scenario, optimal solution
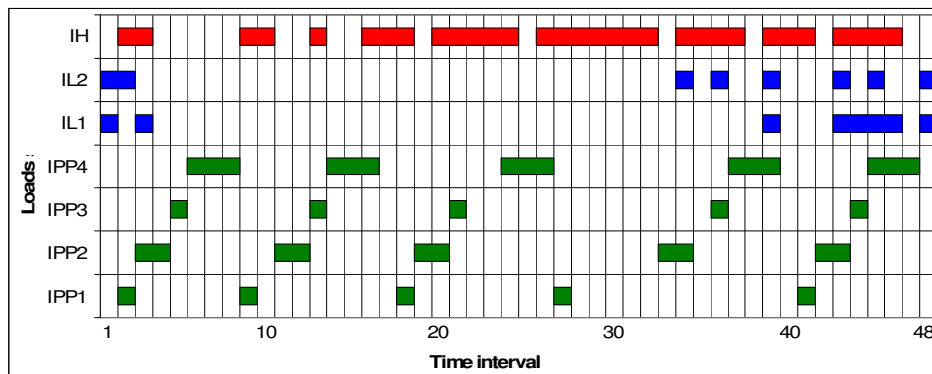
Figure 3. Gannt-chart, winter scenario, conventional solution
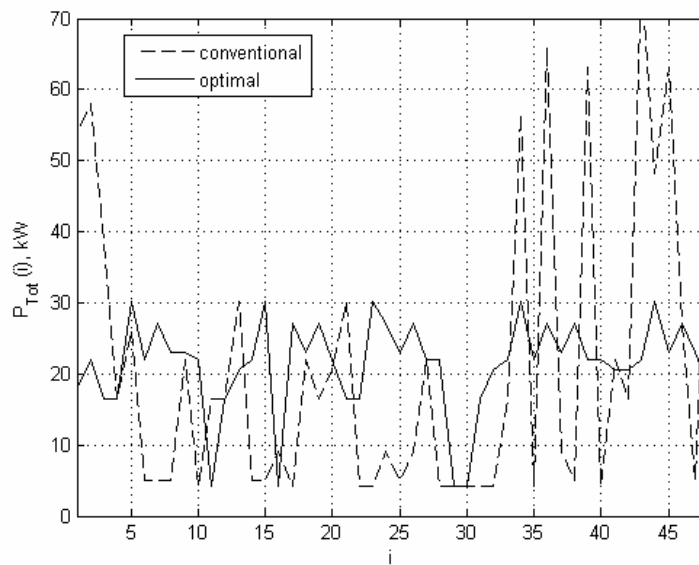
Figure 4. Power curves, winter scenario

References

[1] K. Mařik, Z. Schindler and P. Stluka, 2008, Decision support tools for advanced energy management, Energy, Vol. 33, 858–873

[2] R. L. Graham et al., 1979, Optimization and approximation in deterministic sequencing and scheduling: A survey, Annals of Discrete Mathematics 5, 287-326

[3] S. Ashok and R. Banerjee, 2001, An Optimization Mode for Industrial Load Management, IEEE Transactions on Power Systems, Vol. 16, No. 4, 879-884

[4] M. G. Ierapetritou et al., 2002, Cost Minimization in an Energy Intensive Plant Using Mathematical Programming Approaches, Ind. Chem. Res., Vol. 41, No. 21, 5262-5277

[5] H.K. Sarimveis et al., 2003, Optimal energy management in pulp and paper mills, Energy Conversion and Management, Vol. 44, 1707–1718

[6] R. Rosenthal, 2008, GAMS – A User's Guide, GAMS Development Corporation Washington

[7] GAMS Development Corporation Washington, 2008, GAMS – The Solver Manuals

A Novel Optimization Method to Automated Wet-Etch Station Scheduling in Semiconductor Manufacturing Systems

Adrián M. Aguirre, Carlos A. Méndez[*]

INTEC (UNL-CONICET), Güemes 3450, 3000 Santa Fe, Argentina
*E-mail: cmendez@intec.unl.edu.ar

Abstract

This work addresses the short-term scheduling of one of the most critical stages in the semiconductor industry, the automated wet-etch station (AWS). An efficient MILP-based computer-aided tool is developed in order to achieve a proper synchronization between the activities of sequential chemical and water baths and limited automated wafer's lot transfer devices. The major goal is to find the optimal integrated schedule that maximizes the whole process productivity without generating wafer contamination.

Keywords: MILP, Scheduling, Semiconductor, Automated Wet-etch Station (AWS).

1. Introduction

The efficient operation of complex semiconductor manufacturing facilities has attracted an increasing research interest in the recent years. This industry is today in a growth expansion, immersed in a series of highly competitive global markets that are characterized for being intensely technological and dynamic. This current situation forces the wafer fabrication facilities to focus their efforts on providing to their costumers high-quality of affordable products, with minimum delivery times and lower processing times. As a consequence, the development of efficient short-term scheduling strategies becomes a potential alternative to reach competitiveness, answering agilely to the requirements of highly demanding markets and customers.

The Automated Wet-etch Station (AWS) is a key part of a modern semiconductor production system, which has to simultaneously deal with many complex constraints and limited resources. This station is composed of a series of successive chemical and water baths and a shared automated lot transfer system, in which mixed intermediate storage policies must be strictly followed in order to avoid very expensive wafer contaminations. Thus, the efficient operation of this stage will offer a substantial reduction of the processing time required to complete all the jobs, providing, at the same time, a better utilization of critical limited resources.

Motivated by the inherent features of this challenging problem, a heuristic procedure based on Tabu Search was firstly developed by Geiger et al. (1997) to find a near optimal production schedule to the AWS. Subsequently, Bhushan & Karimi (2003a), introduced a slot-based MILP mathematical model to minimize the makespan in a AWS station. Later, Bhushan & Karimi (2003b) studied the proper combination of Tabu Search strategies and new algorithms of robot scheduling. A novel effective continuous-time MILP formulation to solve short-term scheduling problems in the AWS stage is introduced below.

2. Problem Statement

Typical wafer fabricatio n plants in volve fou r m ain stag es: Fabricatio n, Pro bing, Assembly o r Pack ing and Final Testin g. W et-Etching rep resents one of th e m ost complex op erations carried out i n th e wafe r fabricatio n stage. It u tilizes automated transfer of wafer's lots ac ross a pre defined sequence of successive chemical and water baths, with strict and deterministic exposure times in the chemical ones (see Figu re 1). An au tomated material-h andling device, lik e a robot, is u sed as a s hared resource for transferring lots bet ween consecutive baths. Transfer tim es between bat hs are deterministic. A robo t cann ot h old a wafer's lo t more than a pre defined trans fer tim e and cannot carry more than a single lot at a ti me. In addition, a zero wait storage policy is followed from the chemical baths while local storage is allowed for water baths.

Baths must process only one lot at a time and the overexposure to the chemical ones can seriously damage or contaminate the wafer. This kind of operational constraints, known as a Zero-Wai t (Z W) and L ocal Stora ge (L S) po licies, beco mes the Au tomated W et-Etch Station (AWS) in a serial flowshop multiproduct process with Mixed Intermediate Storage p olicies (M IS) (Bhushan & Karimi, 200 3a). Th us, t he A WS sc heduling problem p rovides a co mplex in terplay b etween m aterial-handling and pro cessing constraints with the application of mixed intermediate storage (MIS) policies.
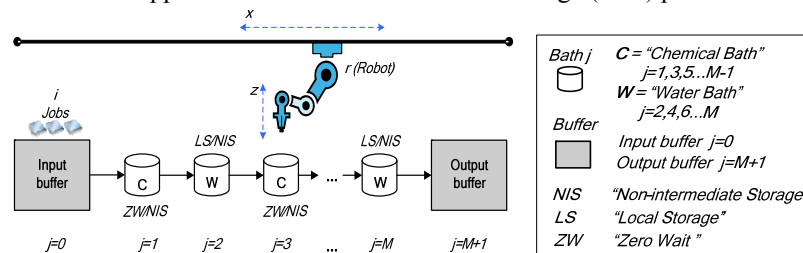
Figure 1. Automated Wet-etch Station (AWS) process scheme

In t his wo rk, it is assu med th at N lo ts $(i=1,2,...,N)$ have to be pr ocessed fol lowing predefined recip es, that in dicate th e seq uence of baths t o be visited. In ad dition, t he problem considers that a sing le robot is ava ilable, which supposes that all the lots have to follow the same sequence $(j=0,1,2,...,M+1)$ in all baths. In consequence, the material movement control scheduling of the robot adopts a central relevance.

The problem to face co rresponds t o t he scheduling of N j obs i n M baths, in a s erial flowshop m ultiproduct with ZW/LS/NIS policies with sh ared resou rces with limited capacity for the wafer movement.

This work pr esents the devel opment and application of a r igorous MILP mathematical formulation t o t he A WS sc heduling problem. It pr ovides t he optimal sequence a nd timing for t he pr ocessing operations t o be pe rformed i n a given t ime h orizon an d, simultaneously, d etermines t he d etailed p ick-up an d delivery activ ity p rogram fo r th e robot, with the m ain o bjective of m inimizing the time requ ired to fi nish all th e wafer lots.

3. The MILP mathematical formulation

3.1 Timing constraints. The processing tim e $t(i,j)$ d epends on th e j ob *"i"* and the bath *"j"*, where J_{odd} *(j=1,3,5,...,M-1)* represents the chem ical baths and J_{even} *(j=2,4,6,...,M)* denotes th e water b aths. Furthermore, va riables $Ts_{(i,j)}$ and $Tf_{(i,j)}$ d etermine th e start an d finish time of every job *"i"* in each bath *"j",* respectively. Equations (1) and (2) define that ex act ti ming decisions of jo bs in ch emical an d water b aths, allowi ng only water baths t o del ay every job *i* more than the pre defined m inimum reside nce tim e. Th e

A Novel Optimization Method to Automated Wet-Etch Station Scheduling in Semiconductor Manufacturing Systems

intermediate storage policies (SP_j) "Zero Wait" (ZW) and "Local Storage" (LS) must be followed in the corresponding bath j.

$$Tf_{(i,j)} = Ts_{(i,j)} + t_{(i,j)} \qquad \forall i \in I, j \in J_{odd} \ : SP_j = ZW \qquad (1)$$

$$Tf_{(i,j)} \geq Ts_{(i,j)} + t_{(i,j)} \qquad \forall i \in I, j \in J_{even} : SP_j = LS \qquad (2)$$

3.2 Transfer time between consecutive baths. The deterministic time for transferring lots between s uccessive baths $j-1$ and j, is p redefined by th e parameter π_j, where $j=1,2,...,M+1$, bei ng $M+1$ t he "Output B uffer". Tra nsfer tasks m ust be carried out enforcing a Non-Intermediate Storage (NIS) p olicy b etween $j-1$ an d j due t o a r obot cannot hold the w afer lots m ore th an a know n tran sfer time (see Eq . 3). Add itionally, equation (4) determines the starting of any job i in the first chemical bath.

$$Ts_{(i,j)} = Tf_{(i,j-1)} + \pi_j \qquad \forall i \in I, j \in J : j > 1 \ , SP_j = NIS \qquad (3)$$

$$Ts_{(i,j)} \geq \pi_j \qquad \forall i \in I, j = 1 \qquad (4)$$

3.3 Sequencing constraints. The binary variable $X_{(i,i')}$ defines the seque nce of every pair of j obs (i,i') in the system . If $X_{(i,i')} = 1$, t he j ob i' wi ll b e started before i in th e processing se quence of e very bath j. In the other ca se, j ob i' will b egin after i in th e sequence. Since a NIS policy with a si ngle robot m ust be en forced f or wa fer's movements, if $X_{(i,i')} = 1$, the beginning of job i in any bath j $(j=1...M+1)$ will be at least greater than the end time of a preceding job i' in the same bath plus the transfer time of job i' fro m j to $j+1$ a nd the m ovement of j ob i from $j-1$ to j, as sh own in Figure 2. Equations (5) and (6) determine timing decisions taking explicitly into account transfer times between consecutive baths. The parameter M_T is a large value.

$$Ts_{(i,j)} \geq Tf_{(i',j)} + \pi_j + \pi_{j+1} - M_T(1 - X_{(i,i')}) \ \forall i,i' \in I : (i > i'), j \in J \qquad (5)$$

$$Ts_{(i',j)} \geq Tf_{(i,j)} + \pi_j + \pi_{j+1} - M_T X_{(i,i')} \qquad \forall i,i' \in I : (i > i'), j \in J \qquad (6)$$
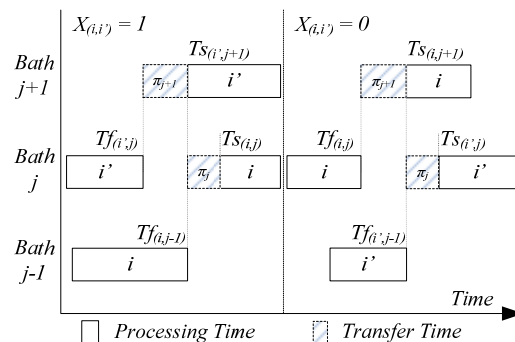
Figure 2. Example of alternative jobs sequence between i and i' in a bath j.

3.4 Transfers sequencing. Since a single robot is only ava ilable to perform all tran sfer operations, the use of this shared critical resource must be explicitly scheduled. For that reason, the transfer (i,j), i.e. job i from j to $j+1$, may be performed before or after the transfer (i',j'), i.e. job i' from j' to $j'+1$, whenever $i \neq i'$ and $j \neq j'$. It is easy to observe in equation (3) that if $i = i'$ and $j < j'$, then the transfer (i,j) will always o ccur before than transfer (i',j'). Similarly, if $i = i'$ and $j > j'$, then transfer (i,j) will be after trans fer (i',j') in th e tran sfer op eration's sequence. Furtherm ore, if $i \neq i'$ and $j = j'$ then t he corresponding tran sfers will b e ex ecuted in th e sam e o rder d etermined b y $X_{(i,i')}$ in

equations (5) and (6). Thus, a single binary variable $Y_{(i,j,i',j')}$ is only defined for every pair of transfers *(i,j)* and *(i',j')*, whenever $i > i'$ and $j \neq j'$. Then, $Y_{(i,j,i',j')} = 1$ if transfer *(i',j')* is performed before transfer *(i,j)*, as stated by Eq. (7). In the opposite case, $Y_{(i,j,i',j')}$ will be 0 and Eq. (8) will be defining the order between transfer operations *(i,j)* and *(i',j')*.

$$Ts_{(i',j')} \leq Ts_{(i,j)} - \pi_j + M_T(1 - Y_{(i,j,i',j')}) \quad \forall i,i' \in I : (i > i'), j,j' \in J : (j \neq j') \quad (7)$$

$$Ts_{(i,j)} \leq Ts_{(i',j')} - \pi_{j'} + M_T Y_{(i,j,i',j')} \quad \forall i,i' \in I : (i > i'), j,j' \in J : (j \neq j') \quad (8)$$

3.5 Assignment and sequencing of alternative transfer resources. So far, it was assumed that a single robot is always available to transfer wafers between consecutive baths. In some cases, however, several robots may be considered. Thus, it will be necessary to determine the assignment of each transfer operation to the corresponding robot *r*. To do that, the new {0-1} variable $W_{(i,j,r)}$ is introduced in order to define if transfer *(i,j)* utilizes the resource *r* during the transfer time π_j (see Eq. (9)).

$$\sum_{r \in R} W_{(i,j,r)} = 1 \qquad \forall i \in I, j \in J, r \in R \qquad (9)$$

Once the transfer resource *r* has been assigned to each operation *(i,j)*, it is necessary to determine the strict order in which they are performed. Consequently, original sequencing transfer constraints (7) and (8) defined for a single resource can be easily reformulated for every available resource *r* belonging to the set *R*, as denoted by equations (10) and (11).

$$Ts_{(i',j')} \leq Ts_{(i,j)} - \pi_j + M_T(1 - Y_{(i,j,i',j')}) + M_T(2 - W_{(i,j,r)} - W_{(i',j',r)})$$
$$\forall i,i' \in I : (i > i'), j,j' \in J : (j \neq j'), r \in R \qquad (10)$$

$$Ts_{(i,j)} \leq Ts_{(i',j')} - \pi_{j'} + M_T Y_{(i,j,i',j')} + M_T(2 - W_{(i,j,r)} - W_{(i',j',r)})$$
$$\forall i,i' \in I : (i > i'), j,j' \in J : (j \neq j'), r \in R \qquad (11)$$

3.6 Objective Function. The objective function used in the proposed model aims at minimizing the total time required to complete all the jobs in the wet-etch station. This goal can be reached by minimizing the variable makespan (*MK*), as shown in Eq. (12).

$$Min \quad MK \geq Ts_{(i,j)} \qquad \forall i \in I, j = M + 1 \qquad (12)$$

4. Results and discussions

Bhushan & Karimi (2003a) proposed a simplified two-step heuristic methodology named RCURM, which is based on an MILP model that can solve moderate size problems with reasonable computational effort in comparison with pure mathematical models. Two alternative models were solved by the authors, being the first one named URM ("unlimited robot model") and the second ORM ("one robot model"). The former aims at generating the optimal job sequence ignoring the robot restrictions while the latter also considers the impact of limited transfer resources in the objective function *(MK)*. The case study presented below is a slightly modified version of the one introduced by Bhushan et al. (2003a). It corresponds to a AWS scheduling problem considering four consecutive baths *NxM*=[4x8] and eight wafer lots. Table 1 shows the processing time of every wafer lot in each bath as well as the predefined transfer times. Transfer times are 10 times larger than the original case study in order to analyze their importance and impact on the scheduling decisions.

A Novel Optimization Method to Automated Wet-Etch Station Scheduling in Semiconductor Manufacturing Systems

Table 1. Wafer lots processing times in baths *j* and transfer times between consecutive baths

job	Bath1	Bath 2	Bath 3	Bath 4
i1	11.10	6.68	5.24	6.92
i2	8.47	6.35	10.10	7.02
i3	9.19	6.35	4.60	6.71
i4	10.80	7.12	10.20	6.83
i5	7.40	7.05	4.07	6.58
i6	10.80	6.76	1.01	6.37
i7	3.48	6.67	1.41	6.46
i8	2.51	6.23	8.00	6.23
$\pi_1=1.0$	$\pi_2=2.0$	$\pi_3=1.5$	$\pi_4=1.75$	$\pi_5=2.5$

The proposed case study wa s solved using the basic concepts previously introduced by Bhushan et al. (2 003a), i .e. the UR M and the OR M solution st rategies. An additional solution strategy where the entire problem is solved in a seq uential manner, here called RCURM, was also tested. The central idea is to first so lve the problem using the URM model, t o t hen fi x t he p roduction se quence obt ained by UR M model and s olve t he detailed robot schedule through the ORM formulation. The solutions generated by both strategies are depicted in Figures 3 and 4, respectively. Model statistics are summarized in Table 2. Here it is worth to remark that the URM model just only takes explicitly into account the predefined transfer times, assuming that a robot will be always available t o perform the transfer operations. In the m ore restricted case, the ORM m odel also takes into co nsideration t he sequ ential u se of the sing le tran sfer m ovement d evice, wh ich enforces a proper synchronization of bath s chedules and robot activ ities. By looki ng at the optimal Gantt Charts, we can easily observe that the same sequencing decisions are made in th e so lution of bo th prob lem in stances. Although t his t ype o f be havior i s relatively co mmon fo r sm all p roblems, it is u sually d ifficult to fi nd wh en larg er problems are faced. This is because se quencing decisions at the production level a re closely related to the sequencing decisions made at the transfer level. Consequently, the fact of using the un limited o r th e one robot m odel will h ave a direct i mpact o n sequencing decisions. A nother i mportant f eature t hat ca n be o bserved by com paring Figures 3 an d 4 i s t he co nsiderable di fference i n t erms of t iming d ecisions. The combination of strict in termediate sto rage po licies an d limited tran sfer devices m akes the pr oblem very rest ricted, whi ch en forces a very t ight coo rdination bet ween t he production system and the material handling device, i.e. the use of a single robot.

By an alyzing the m odel statistics, it is wo rth to rem ark the sig nificant di fference between the n umber of bi nary variables a nd constraints req uired by O RM and UR M models. The l arge differe nce in term s of model size mainly arises because of t he additional sequ encing con straints an d b inary decisions v ariables th at n eed to be included in the MILP formulation to manage the inherent ORM limitations.

Table 2. Model statistics and computational cost

MxN	Statistics	ORM (with robot limitations)	URM (without robot limitations)	RCURM (sequential solution)
4x8	Binary Variables	588	28	560
	Cont. Variables	73	73	101
	Constraints	1512	392	1512
	Makespan	120.47	114.85	120.47
	[1]CPU Time [sec.]*	44.725	1.090	1.090+0.163
	[2]CPU Time [sec.]*	12.271	0.418	0.418+0.135

*Using GAMS with Cplex[1] and Gurobi [2] in a Intel PC Core 2 Quad

Figure 3. Optimal schedule assuming unlimited robots (URM)

Figure 4. Optimal schedule assuming a single robot (ORM)
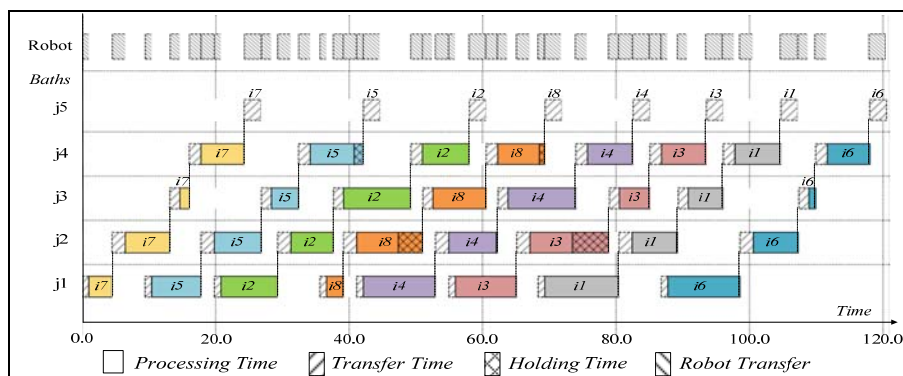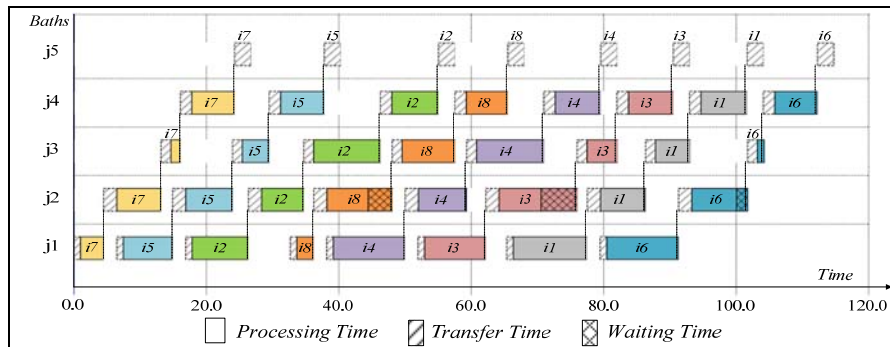
5. Conclusions

A n ovel M ILP co ntinuous-time form ulation has been proposed t o t he sh ort-term scheduling o f A WS pr ocesses i n t he sem iconductor i ndustry. I n co ntrast t o t ypical scheduling problems, this model is ab le to simultaneously generate a d etailed schedule of production activities an d tran sfer operations fo llowing strict in termediate sto rage policies. Also, it was demonstrated that the proposed model can be easily u sed to so lve the whole problem in a seq uential manner, i.e. making first the schedule of production activities to t hen so lving t he sch edule of tran sfer operations assu ming th e fixed production sequence defined in the first step. In all the problem instances, the case study was optimally solved with short computational effort. Future work will be focused on solving industrial-scale problems using efficient MILP-based decomposition strategies.

6. Acknowledgments

References

Bhushan, S. and Karimi, I.A. *IEC Research* 42 (7), 1391-1399 (2003a).
Bhushan, S. and Karimi, I.A. *Computers & Chemical Engineering*, 28(3), 363-379 (2003b)
Geiger, C.D., Kempf, K.G., Uzsoy, R. *Journal of Manufacturing Systems*, 16(2), 102-116 (1997).
Méndez, C., Cerdá, J, Grossmann, I.E., Harjunkoski, I. and Fahl, M. *Computers & Chemical Engineering,* 30 (6-7), 913-946 (2006).

Hybrid Compressor Model for Optimal Operation of CDA System

Kiwook Song, Changhyun Jeong, Chonghun Han

School of Chemical and Biological Engineering, Seoul Natioanl University, San 56-1, Shillim-dong, Kwanak-gu, Seoul 151-742, Korea
kiwook18@snu.ac.kr

Abstract

CDA(Compressed Dry Air) is an essential utility to be used in a variety of processes in LCD(Liquid Crystal Display) production industry. Since the demands for CDA fluctuate largely from moment to moment, it is common to supply compressed air with a number of small-capacity compressors rather than few large-capacity ones. In order to meet the varying demands, operation conditions of compressors change at every second and hence the efficiency of compressors is low. To find optimal operating strategy of such compressor network, first a hybrid modeling technique of ideal model and empirical model is developed to predict efficiency and power consumption of each compressor in the network. Then, optimization procedure is applied to search optimal operation strategy. The proposed method was applied to actual off-line data of LCD production industry and about 5% of annual power consumption was saved by optimization.

Keywords: compressor network, modeling, optimization, CDA, LCD

1. Introduction

Compressed air is widely-used in various chemical processes. Especially CDA (Compressed Dry Air) is an essential utility needed in a variety of processes in LCD (Liquid Crystal Display) production industry, such as in air knife or air curtain to be used in clean room. CDA system is composed of compressors, filters (coalescent, adsorbent, oil removing, etc), and driers. Compressors far and away take the majority of energy consumption.

Compressors can be categorized into centrifugal compressors, reciprocating compressors, and rotary compressors, which are chosen for use subject to given process conditions. When large amount of constant-pressure air is needed such as in the case of CDA production processes, it is common to use multistage centrifugal compressors of various capacities. However, since the demands for CDA fluctuate largely from moment to moment, it is preferred to supply compressed air with a number of small-capacity compressors rather than few large-capacity ones. These small-capacity compressors are interconnected to each other and form a network.

Among the major utilities consisting of boilers, cooling towers, turbines, etc, compressors typically take a large portion of the total energy usage. Hence, reducing the energy consumption of compression through modeling and optimization is needed. For industrial compressors, modeling for measurement and prediction of compressor performances is not straightforward unfortunately. This is because of aging components

that constitute the system, lack of sensors needed for measuring key process variables, and because the operating conditions are usually far from the original design.

This paper first presents modeling accurately the multistage compressors using hybrid technique of ideal thermodynamic models and artificial neural network as nonlinear empirical modeling method. The model is developed to predict efficiency and power consumption of each compressor in the network. Then, a constrained optimization procedure is applied to search optimal operating conditions of a compressor network in LCD production industry. Finally, the results of field application are presented and discussed.

2. Modeling of Compressors

Assuming the compressibility factor of air to be always equal to 1, minimum compression power required for all of the compression stages under an adiabatic and reversible compression process is given by the following equation. The subscript (i) refers to the stage number of the N-stage compressor. The superscript s refers to the suction and d refers to the discharge of each stage.

$$W_{ideal} = \sum_{i=1}^{N} \frac{k \, F_{(i)} \, \rho_a \, R \, T_{(i)}^s}{(k-1)M_{wa}} \left[\left(\frac{P_{(i)}^d}{P_{(i)}^s} \right)^{\frac{k-1}{k}} - 1 \right]$$

(1)

In most industrial cases however, temperature, pressure, and flow rate at each stage are not measured. In addition, the equation above is valid for dry air and hence modification of the molecular weight and density is needed for humid air. In this case, a simplified and modified ideal thermodynamic equation is needed. Let us assume that (1) the temperature of compressed air entering each compression stage is equal to the ambient temperature by perfect cooling. (2) The compression ratios of all of the compression stages of a multistage compressor are equal to each other. (3) The suction pressure and temperature of first stage is equal to the ambient conditions. Then, ideal power consumption is calculated by the following equation.

$$W_{ideal} = \frac{N \, k \, F^s \, \tilde{\rho}_a \, R \, T_e}{(k-1) \, \widetilde{M}_{wa}} \left[\left(\frac{P_{(N)}^d}{P_e} \right)^{\frac{k-1}{N \, k}} - 1 \right]$$

(2)

Here, only the first suction and last discharge of the compressor measurement is required. The mean molecular weight and mean density are used for humid air. The ideal thermodynamic model cannot accurately predict the actual consumption power of compressors. The actual power consumption is given in the following equation, which divides the ideal power by compressor efficiency.

$$W_{actual} = \frac{W_{ideal}}{\eta}$$

(3)

The efficiency of compressor varies due to operating conditions. Manipulated variables of compressors are mainly the flow rate and discharge pressure. Ambient condition variables consist of ambient temperature and relative humidity. The ambient pressure is typically assumed to be constant as 1 atm.

To predict the efficiency and power consumption of compressors, hybrid technique of ideal thermodynamic models and empirical modelling methods is applied. Ideal compression work is calculated by equation (2). Efficiency of compressor is function of manipulated variables and ambient condition variables, and the relation between them is highly nonlinear. An empirical modelling tool such as ANN (Artificial Neural Network) can be used here as a nonlinear modeller. In this study, a feed-forward back-propagation network with one hidden layer is employed to model the relation between the input variables (flow rate, discharge pressure, ambient temperature, and relative humidity) and target variables (efficiency). Since the ideal work and the efficiency is known, actual power consumption can be calculated by equation (3).

3. Case Study

The modeling method is applied to CDA system in LCD industry consisting of a network of 32 centrifugal compressors. The compressors are all 3-stage compressors. Models are constructed for each compressor for the prediction of efficiency on the basis of the operating data for the past eight months, and are combined with thermodynamic equation (2) to form a hybrid model for predicting the actual power consumption. The variables used are the discharge pressure at the last stage, air flow rate, ambient temperature, and relative humidity.

Performance of the model of a compressor is demonstrated by Figure 1 for about 200 observations. The modeling results show excellent agreement between the measured and predicted values of the electric power consumption with average error of less than 1.2%. Similar performances were obtained for other 31 compressors with average error ranging between 0.5~2 %.
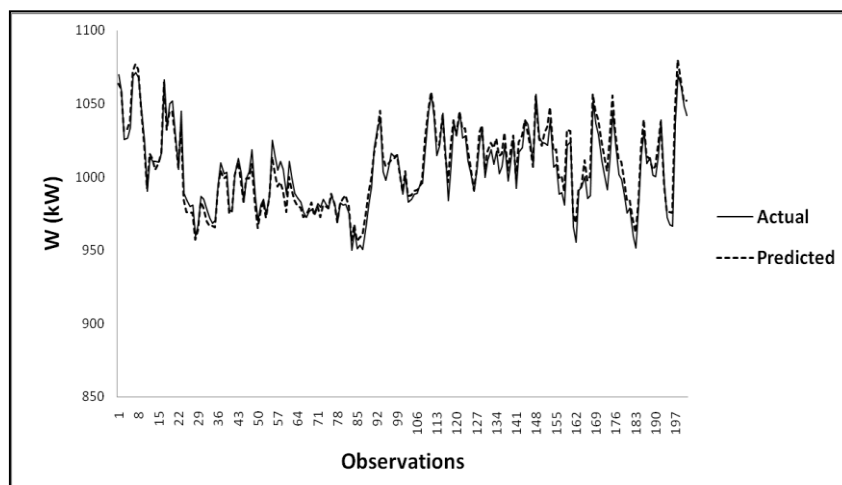
Figure 1. Comparison of the measured (solid lines) and predicted (dotted lines) values of the electric power consumption of Comp # 1.

4. Optimization of Compressor Network

The objective of the optimization is to minimize the total electric power consumption of the compressor network. As the load of the compressors decreases, the efficiency of the compressor also decreases. Therefore it is optimal to operate the compressors full-loaded. However, due to fluctuating demand of CDA, operation of the compressor network requires redundancy to supply compressed air at constant pressure.

For the CDA system in the case study, among the 32 compressors 25 compressors are used to supply compressed air ranging from 160 kCMH(1000 Cubic Meters per Hour) to 190 kCMH. The total demand of CDA changes from moment to moment in this range. Before optimization, the operating conditions of every compressor change as the total demand fluctuates and hence the efficiencies of the compressors are kept low.

Optimization procedure requires constant full load operation of compressors to keep the maximum efficiency level. On the other hand, enough safety margins are needed. Operating 25 compressors is a waste of energy when the demand for CDA is low such as 160~170 kCMH, since only 24 compressors are needed actually for until 175 kCMH. However, there should always be capacity redundancy to flexibly follow up situations when the demand increases rapidly. Therefore, let us fix the number of compressors to be used to 25 for safety.

In this study we categorize the compressors into two groups: (1) Full-load operating group and (2) Part-load operating group. The Full-load operating group is fixed to operate full-loaded so that they are not affected by the varying demand. The Part-load operating group will be operated in lower efficiency but can follow the demand fluctuation and keep the overall discharge pressure of the compressors constant.

4.1 Optimization Case I
Among the 32 compressors in the network, 25 compressors are operated. These compressors are further divided into part-load operating group and full-load operating group, 20 and 5 compressors respectively. In optimization case I, let us operate the most efficient 20 compressors (efficiency rank 1 to 20) as the full-load operating group and 5 next ones (efficiency rank 21 to 25) as the part-load operating group. The efficiency rank is calculated by the model at given operating conditions (discharge pressure, ambient temperature, and relative humidity).

4.2 Optimization Case II
In optimization case II, let us operate the most efficient 5 compressors (efficiency rank 1 to 5) as the part-load operating group and 20 next ones (efficiency rank 6 to 25) as the full-load operating group.

4.3 Optimization Results
The optimization results are shown in the next table.

Table 1. Optimization Results of Case I and II.

	Before Optimization	Optimization Case I	Optimization Case II
Total Work (kW)	25866	24812	24650
Average Efficiency	0.45543	0.4726	0.4781
Percentage Saved (%)	-	4.07	4.70

As can be seen from the result table, the optimization case II is the better solution. The result shows about 5 % energy saving. Hence the optimization strategy of a multistage compressor network is to operate the most efficient 5 compressors part-loaded to flexibly meet the fluctuating demand and the next efficient 20 compressors full-loaded.

5. Conclusion

This paper presented hybrid modeling technique of ideal thermodynamic models and empirical modeling method to predict the efficiency and actual power consumption of multistage centrifugal compressors. The modeling methodology was applied to CDA system of LCD industry and proved its excellence in prediction. Using the models developed, strategic optimization of compressor network was done. Minimizing the total power consumption of compressors while meeting the safety margin constraint can be achieved by decomposing the compressors in the network into part-load group and full-load group. The optimization technique was also applied to CDA compressor network in LCD industry and about 5 % of power consumption was saved.

Appendix

Calculations of mean molecular weights and mean density are as follows. Relative humidity is the ratio of vapor pressure of water to the saturation vapor.

$$P_v = \phi \times P^{sat}$$

(4)

Saturation vapor pressure can be calculated using the Antoine's equation[6].

$$ln\, P^{sat}/kPa = A - \frac{B}{t/\,^{\circ}\!C + C}$$

(5)

For water, A = 16.3872, B = 3885.70, C = 230.170 and these parameters are valid for temperature range 0 ~ 200 ℃ [7].

When we know the atmospheric pressure, usually assumed constant to be 1 atm, and the vapor pressure of water, partial pressure of dry air is just the difference of the two.

$$P_d = P_e - P_v$$

(6)

Now, the mean molecular weight and density is calculated by the following equations.

$$\tilde{\rho} = \frac{P_d}{R_d\,T} + \frac{P_v}{R_v\,T}$$

(7)

R_d is the specific gas constant for dry air which is equal to 287.05 J/kg K. R_v is the specific gas constant for water vapor which is equal to 461.495 J/kg K.

$$\tilde{M}_{wa} = \frac{P_d}{P_e} M_{wa} + \frac{P_v}{P_e}\, M_{ww}$$

(8)

M_{wa} and M_{ww} refers to the molecular weight of dry air and water vapor respectively.

Nomenclature

$F_{(i)}$ = flow rate of air at the compression stage i [m^3/s]

F^s = flow rate of air at the suction of a multistage compressor [m^3/s]

k = adiabatic exponent of air, constant as 1.398

M_{wa} = molecular weight of dry air [28.96 kg/kg mol]

$\widetilde{M}_{wa}$ = mean molecular weight of air [kg/kg mol]

M_{ww} = molecular weight of water [18.02 kg/kg mol]

N = total number of compression stages of a multistage compressor

$P_{(i)}^d$ = discharge pressure of the air at the compression stage i [kPa]

$P_{(i)}^s$ = suction pressure of the air at the compression stage i [kPa]

P_e = atmospheric pressure [kPa]

P^{sat} = saturation vapor pressure [kPa]

R = universal gas constant [8.314 kJ/kg mol K]

R_d = specific gas constant for dry air [287.05 J/kg K]

R_v = specific gas constant for water vapor [461.495 J/kg K]

T_e = ambient temperature [K]

Greek Symbols

η = efficiency of a compressor [0-1]

ρ_a = density of air

Φ = relative humidity of ambient air

References

[1] O'Neill, P.A., Industrial Compressors, Butterworth-Heinemann, Oxford, UK, (1993).

[2] Bloch, H.P., A Practical Guide to Compressor Technology, McGraw-Hill, New York, USA, (1996).

[3] Han, I. S., Lee, Y. H., Han, C., Modeling and Optimization of the Condensing Steam Turbine Network of a Chemical Plant, Ind. Eng. Chem. Res., 45, 670-680, (2006)

[4] Iyer, R. R., Grossmann I. G., Optimal Multiperiod Operational Planning for Utility Systems, Comput. Chem. Eng., 21, 787. (1997).

[5] Han, I.-S., Han, C., Modeling of multistage air-compression systems in chemical processes, Ind. Eng. Chem. Res., 42, 2209-2218, (2003).

[6] Smith, J. M., Van Ness, H. C., Abbott, M. M., Introduction to Chemical Engineering Thermodynamics, McGraw-Hill, (2005).

[7] Poling, B. E., Prausnitz, J. M., O'Connell, J. P., The Properties of Gases and Liquids, 5th ed., App. A, McGraw-Hill, New York, (2001).

Simulation and Optimization of Full Scale Reverse Osmosis Desalination Plant

Kamal M. Sassi and Iqbal M. Mujtaba

School of Engineering Design and Technology , University of Bradford,Bradfor, West Yorkshire BD7 1DP,UK. E-mail: I.M.Mujtaba@bradford.ac.uk

Abstract

This paper focuses on steady state performance predictions and optimization of the Reverse Osmosis (RO) process utilizing a set of implicit mathematical equations which are generated by combining solution-diffusion model with film theory approach. The simulation results were compared with operational data which are in good agreement having relative errors of 0.71% and 1.02%, in terms of water recovery and salt rejection, respectively. The sensitivity of different operating parameters (feed concentration, feed flow rate and feed pressure) and design parameters (number of elements, spacer thickness, length of filament) on the plant performance were also investigated. Finally a non linear optimization framework to minimize specific energy consumption at fixed product flow rate and quality while optimizing operating variables (feed flow rate, feed pressure) and design parameters (height of feed spacer, length of mesh filament). Reduction in operating costs and energy consumption up to 50 % can be reached by using pressure exchanger as energy recovery device.

Keywords: reverse osmosis, spiral wound membrane, simulation, optimization, energy recovery

1. Introduction

The shortage of fresh water resources and growth of industrialization have increased the reliance on water production using desalination technology. Thermal and membrane processes are, by far, the major desalination systems used now-a-days. Pressure driven membrane processes are less energy intensive than thermal.

Designing an efficient RO desalination system remained an elaborate work (Lu et al., 2006). It is connected to many variables, such as, feed flow rate, operating pressures, recovery rate, the type of membrane element and its geometry (i.e. spacer geometry) and RO system configuration. However, the prediction of RO process performance completely relies on the mathematical model accuracy. Kim et al. (2009) reviewed the analytical design methods of industrial RO plants, and the recent optimization techniques for predicting the optimal parameters values for RO plants using different mathematical programming.

In this work, the effect of different operating and design parameters such as feed pressure, salinity, spacer geometries, and number of membrane elements in the pressure vessel on the performance of RO performance is studied. An optimization problem incorporating a process model is formulated to optimize the design and operating parameters in order to minimize specific energy consumption constrained with fixed product demand and quality. Finally, energy recovery from brine is considered. Two different energy recovery devices are studied: hydrodynamic turbines and pressure exchanger.

Prediction of solute concentration polarization on the membrane surface in crossflow membrane processes has vital role in designing RO processes and estimating their performances. A film theory approach which was developed originally by Michaels (1968) is used in this work to describe the concentration polarization. It is simple, analytical, and (reasonably) accurate for most RO separations. Further, film theory can be extended to describe the effect of spacer-filled RO modules on concentration polarization which is inherently used in design and evaluation of the membrane processes. Solution-Diffusion model is used to illustrate solvent and solute transport through the membrane. This model is the most used and is able to provide an accurate prediction of the flow of water and salt through the membrane (Marcovecchio et al., 2005).

2. Reverse Osmosis Process Model

Fig. 1 summarizes the model equations for RO based on the following assumptions:

- Pressure drop along the permeate channel is neglected, this assumption is reasonable for 8 inches spiral wound module that has 37 membrane leafs with a length of 1 m (Geraldes et al., 2005).
- The feed channels of spiral wound element are flat. Feed stream flows along the channel parallel to the central line of the module and the curvature of membrane module was reported to have insignificant effect on system's performance (Meer et al., 1997). Therefore, an unwound flat sheet membrane with same channel height and spacers would adequately represent characteristics of the corresponding spiral-wound RO module.
- The feed concentration varies linearly along the feed side channel.

3. Optimization Problem Formulation

The performance of a membrane process is limited by the magnitude of the chemical potential driving force for mass transfer. This driving force can be maximised by manipulating temperature or pressure of the feed stream which require significant energy. This spending should be balanced against other costs in designing the membrane system. Consequently, proper optimization techniques are required to determine the best values for the various operating and design parameters. An optimization strategy which considers both operating and design parameters is shown in Fig. 2. This results in a non linear optimization problem solved using SQP method within gPROMS software. As shown in Fig. 2, there are four decision variables (P_f, Q_f, L_f, d_f). The bounds on each variable are specified in each case.

4. Case Study

In this work, a three-stage RO process described by Abbas (2005) is considered (Fig. 3). The plant nominal operating and design parameters are given in Table 1. Commercial Film Tec spiral wound RO membrane elements with three elements in each pressure vessel (connected in series) is considered. Each element is modeled by a set of nonlinear algebraic equations (Fig. 1). Operational data from Abbas (2005) are used to validate the model. The model yielded an overall 58.0 % water recovery and 98.6% salt rejection. The relative deviations of the simulated results compared to Abbas (2005) are 0.71% and 1.02%, respectively.

Then, the effect of different operating and design parameters on membrane performance was studied by varying one parameter and keeping the others constant (Table 1) as follow.

Simulation and optimization of full scale reverse osmosis desalination plant

---

Water flux: $Jw = A(\Delta P - \Delta \pi)$; Solute flux: $Js = B(C_m - C_p)$; Recovery %:

$R = (Q_p / Q_f) \times 100$; Concentration polarization: $CF = \dfrac{Cm - Cp}{CB - Cp} = \exp(Jw / K)$; Salt rejection

%: $SR = (1 - \dfrac{C_p}{C_f}) \times 100$; Pressure drop: $\Delta P_f = \dfrac{\rho u^2 LC_{td}}{2d_h}$; Mass transfer coefficient:

$Sh = \dfrac{Kd_h}{D} = 0.0664 K_{dc} \, \mathrm{Re}^{0.75} Sc^{0.33} \left(\dfrac{2d_h}{L_f} \right)^{0.5}$; Mass balance: $Q_f C_f = Q_p C_p + Q_r C_r$; Flow

balance: $Q_f = Q_p + Q_r$; Average bulk concentration: $Q_B = \dfrac{Q_f + Q_r}{2}$; Average velocity in feed

side: $U = \dfrac{Q_B}{Wh_{sp}\varepsilon}$; Water flow via membrane: $Q_p = j_w S$; Material balance around

membrane: $C_p = J_s / J_v$; Specific energy: $E = \dfrac{\Delta P_f Q_f}{\eta Q_p}$; With turbine energy recover (ER):

$E = \dfrac{\dfrac{\Delta P_f Q_f}{\eta} - P_r Q_r \eta_T}{Q_p}$; With pressure exchange (PX) ER: $E = \dfrac{\Delta P_F Q_P}{\eta Q_p}$

Fig. 1 RO process model

Given: Feed water conditions; membrane properties and specifications

Determine: The optimal feed pressure; feed flow; the optimum design decisions (feed spacer filament length and diameter, feed spacer thickness)

So as to minimize: Specific energy consumption E ((kwh/ m^3)

Subject to: Equality and inequality constrains

Mathematically optimization problem can be represented as:

$$\underset{P_f, \, Q_f, \, L_f, d_f}{\text{Minimize}} \quad E$$

Subject to: Equality constraints: Process model; Product demand; Product specification

Inequality constrains: $P_f^{lower} \le P_f \le P_f^{upper}$;

$Q_f^{lower} \le Q_f \le Q_f^{upper}$; $L_f^{lower} \le L_f \le L_f^{upper}$; $d_f^{lower} \le d_f \le d_f^{upper}$

Fig. 2 Optimization problem formulation
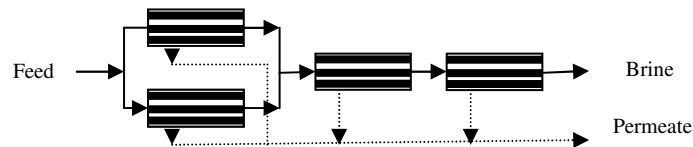
Fig. 3 Schematic diagram of reverse osmosis

Table 1 Membrane parameters and process operating conditions

Feed conditions: Q_f (m^3/h) 20.4; C_f (kg/m^3) 2540 ppm; P_f (bar) 12.2; T_f (○C) 28.8

Membrane and spacer characteristics: A (m/bar s) 9.39×10^{-7}; B (m/s) 5.65×10^{-8}; L$_f$ (m) 2.77×10^{-3}; L (m) 1; w (m) 37.2; S (m^2) 37.2; h_{sp} (m) 5.93×10^{-4}; d_h (m) 8.126×10^{-4}

4.1 Sensitivity analysis of operating parameters

4.1.1   Pressure

Fig. 4a shows the effect of operating pressure on RO plant performance. Salt rejection increases linearly at low to moderate pressure. At high pressure, salt rejection decreases dramatically due to the increase in osmotic pressure along the feed channel. Average permeates flux curve is divided into two regions. In the lower pressure region, water flux increases linearly which illustrates a linear relationship between the permeate flux and the driving pressure. In the higher pressure region water flux starts to level-off at 16 bar (corresponding to flux 1.2×10^{-3} m/s). This may be due to the accumulation of the salt along the membrane channel that exerts an increasing osmotic pressure. The limiting flux is (1.4×10^{-4} m/s) where the flux can not be increased even when the applied pressure increases. Variations of specific energy consumption (kwh/m^3) and concentration polarization factor (CF) are shown in Fig. 4b for operating pressure ranging from 6 to 25 bar. Higher pressure required less pumping energy. The minimum specific energy consumption is observed at 12 bar corresponding to water recovery rate 57 %, followed by increase in specific energy due to the stabilization in the permeate production despite increasing applied pressure. As expected CF increase with increase in operating pressure due to the increase in water flux.
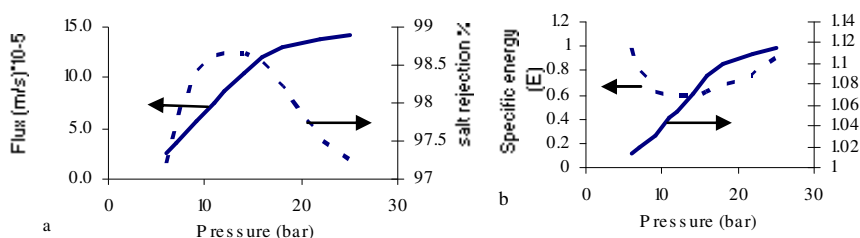
Fig. 4 Dependence of RO process performance on operating pressure

4.1.2   Number of elements in pressure vessel

It was observed that the water recovery ratio increases with the number of elements in the pressure vessel due to increased membrane area. There was a sharp increase at lower number of elements and a slow increase at higher number of elements. This was due to the salt build up on the brine channel as flux increases. Therefore adding more elements after certain limit not worthy.

4.1.3   Feed salinity

The effect of feed salinity on the total recovery ratio is shown in Fig. 5a. Two alternative feeds with 2500 ppm and 5000 ppm salt have been studied. Feed with low salt concentration produced 40 % higher recovery ratios compared to that produced by high feed (5000 ppm) salinity. This is a consequence of the much higher driving force for the same exerted pressure to the feed. This is due to the fact that the osmotic pressure is proportional to the feed salt concentration.

4.2   Feed spacer

Feed spacer channel can affect RO performance significantly, compared to that with slit feed channel. Even though the pressure drop is increased from 0.122 bar for empty channel to 1.23 bar, the mass transfer is enhanced by 80%, CF on membrane surface is reduced by about 27 %, and the specific energy consumption is reduced by 10%.

Simulation and optimization of full scale reverse osmosis desalination plant

4.2.1 Length of filament mesh in feed spacer

Fig. 5a,b show the recovery of fresh water and pressure drop when mesh length is varied for the two transverse filament thicknesses. It can be seen the recovery rate increases with the increase of mesh length until a turning point at mesh length 3 mm, after which the recovery rate remained relatively stable regardless of further increase of mesh length. Small mesh length has the advantage of more turbulent flow and consequently the polarization phenomenon is decreased. On the other hand smaller mesh length has the drawback of higher pressure drops along feed channel and therefore less water flux as in Fig. 5a.
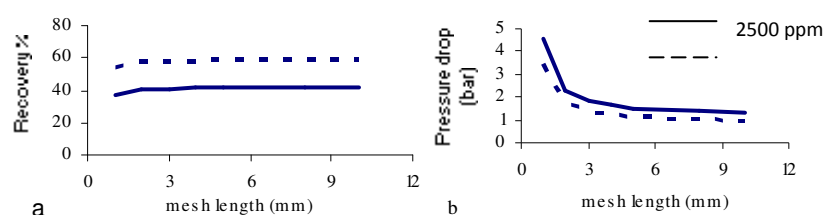
Fig. 5 Effect of mesh length on water recovery and axial pressure drop at different feed salinity

4.2.2 Filament diameter to feed spacer spacing ratio

Fig. 6 presents the water recovery, average concentration polarization factors and axial pressure drops for filaments of different diameter to feed spacer ratio. Pressure drop is significantly affected (increase by 342 %) at filament ratio 0.6 while the concentration polarization is reduced by 8 % at filament ratio 0.6. In general, larger filaments slightly enhance mass transfer by reducing concentration polarization, but significantly increase hydraulic pressure losses and consequently more expenditure. Therefore, spacers design should be optimized specifically for the particular operating conditions of the real application.
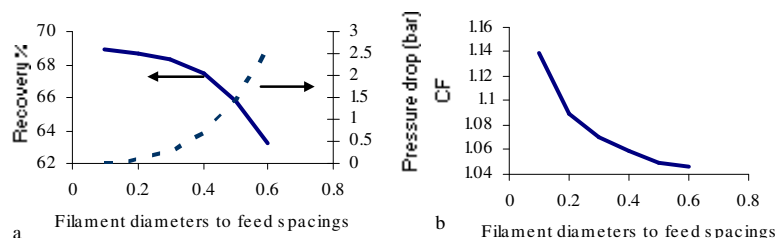
Fig. 6 RO performance for different spacer diameters and filament spacing

4.3 Optimization (Minimum specific energy consumption)

The optimized values for operating parameters (case 1) and both operating and design parameters (case 2) (at fixed product demand 10.8 m^3/h and permeate salt concentration less than 100 ppm) are shown in Table 2. A substantial saving of about 20 % which is equivalent to 1.7 kWh can be acquired by only optimizing operating parameters. Reduced feed flow and slightly increased operating pressure yields higher driving force in the brine channel. This result is concordant with sensitivity analysis presented earlier. Further reduction in specific energy can be achieved in case 2 by enlarged feed spacer

thickness and shorter filament length. This gives less pressure drop and consequently, more water flux.

Table 2 Optimization results

	Optimized parameter			Objective function E (kwh/ m^3)
	Parameter	Optimized Value	Design value	
Base case	Design conditions (Table 1)			0.7304
Case 1	Feed pressure (bar)	13.6	12.20	0.5865
	Feed flow (m3/h)	14.6	20.43	
Case 2	Feed pressure (bar)	13.10	12.20	0.5781
	Feed flow (m3/h)	14.91	20.42	
	Spacer thickness (mm)	2.20	0.59	
	Mesh length (mm)	2.37	2.77	

4.4   Energy recovery

Three different options were considered in this work: (a) No energy recovery, (b) Energy recovery by turbine and (c) Energy recovery using pressure exchanger (PX). The efficiencies for the feed pump, turbine and pressure exchanger were assumed to be 0.8, 0.8 and 0.97, respectively. Pressure exchanger was found to be the most profitable option as the pumping cost (reflected in the calculation of E, see Fig. 1) will reduced up to 50 % compared with 20 % when turbine was used as energy recovery choice.

5. Conclusion

In this work RO process model based on solution-diffusion model and thin film theory have been developed to investigate the effect of different operating and design parameters on the performance of the system. The model is verified against the operational data and a good agreement was found.

Optimization problem formulation is presented to minimize an objective function while optimizing design and operating parameters of the process. It is found that considerable reduction in pumping cost around 20 % is achievable. Furthermore, commercial module designs might be further refined in order to reach more economic improvements for RO processes subject to technical limitations.

Comparison of the two energy recovery alternatives including turbine and pressure exchanger showed that energy recovery by pressure exchanger yields the best results by 50 % reduction in the pumping cost.

References

Abbas, A. (2005). Chemical Engineering and Processing, 44, 999-1004.

Geraldes, V., Pereira, N. E. & De Pinho, M. N. (2005. Industrial & Engineering Chemistry Research, 44, 1897-1905.

Kim, Y. M., Kim, S. J., Kim, Y. S., Lee, S., Kim, I. S. & Kim, J. H. (2009). Desalination, 238, 312-332.

Lu, Y. Y., Hu, Y. D., Xu, D. M. & Wu, L. Y. (2006). Journal of Membrane Science, 282, 7-13.

Marcovecchio, M. G., Aguirre, P. A. & Scenna, N. J. (2005. Desalination, 184, 259-271.

Michaels, A. S. (1968). Chem. Eng. Prog., 64, 31−43.

Van der Meer, W. G. J. & Van Dijk, J. C. (1997). Desalination, 113, 129-146.

Strategies for the Global Optimization of Integrated Process Water Networks

Elvis Ahmetović,[a] Ignacio E. Grossmann[b]

[a]*University of Tuzla, Faculty of Technology, Department of Process Engineering Univerzitetska 8, Tuzla 75000, Bosnia and Herzegovina, elvis.ahmetovic@untz.ba*
[b]*Carnegie Mellon University, Department of Chemical Engineering, 5000 Forbes Avenue, Pittsburgh 15213, Pennsylvania, grossmann@cmu.edu*

Abstract

In this paper, we propose special strategies for obtaining the global or near global optimum solution from a general superstructure proposed recently by the authors for the design of integrated process water networks. The proposed model of the integrated water network is formulated as a Nonlinear Programming (NLP) and as a Mixed Integer Nonlinear Programming (MINLP) problem for the case when 0-1 variables are included to model the cost of piping and/or selection of technologies for treatment. The MINLP model can be used to find optimal network designs with different number of streams in the piping network. The proposed strategies rely on bounds on the variables that are derived as general equations obtained by physical inspection and using logic specifications needed for solving the model. The cut proposed in [1] and some variations of it are also used to significantly improve the strength of the lower bound for the global optimum. It is shown that the proposed strategies can effectively solve large-scale problems, and in most cases, to global optimality. Furthermore, the proposed strategies allow to readily obtain networks of varying degrees of complexity by limiting the number of piping connections

Keywords: Integrated water network, Superstructure, Model, Optimization strategies.

1. Introduction

Large freshwater consumption in industry, the shortage of freshwater, its increasing cost, as well as strict environmental regulations on the industrial effluents, provide a strong motivation for developing approaches and techniques to design more efficient process water networks. The two major approaches for the optimal design of water network systems are water pinch technology and mathematical programming. A comprehensive review of these approaches is given in the literature [2-8].

In this paper, we address the optimization of integrated process water networks. The problem is formulated as a Nonlinear Programming (NLP) and as a Mixed Integer Nonlinear Programming (MINLP). Two-stage solution strategies are proposed for solving the problem to global or near global optimality as well as to control the water networks complexity. We present an example of integrated water network consisting of water-using and wastewater treatment units to illustrate the proposed method.

2. Problem statement

The problem addressed in this paper can be stated as follows. Given is a set of single/multiple water sources with/without contaminants, a set of water-using units and wastewater treatment operations, fixed water demands of process units, maximum concentrations of contaminants in inlet streams at process units, mass loads of contaminants in process units, the costs of water sources and wastewater treatment units, % removal for each contaminant in treatment units and the maximum contaminant concentrations in the discharge effluent to the environment. The problem consists in determining the interconnections, flowrates and contaminants concentration of each stream in the water network, the freshwater consumption and wastewater generation, and the total annual cost of the water network.

3. Superstructure

The proposed superstructure of the integrated water network, which is an extension and generalization of the one by Karuppiah and Grossman [1], is given in Fig. 1. The superstructure consists of one or multiple sources of water of different quality, water-using processes and wastewater treatment operations. The unique feature is that all feasible connections are considered between them, including water re-use, water regeneration and re-use, water regeneration recycling, local recycling around process and treatment units and pre-treatment of feedwater streams.
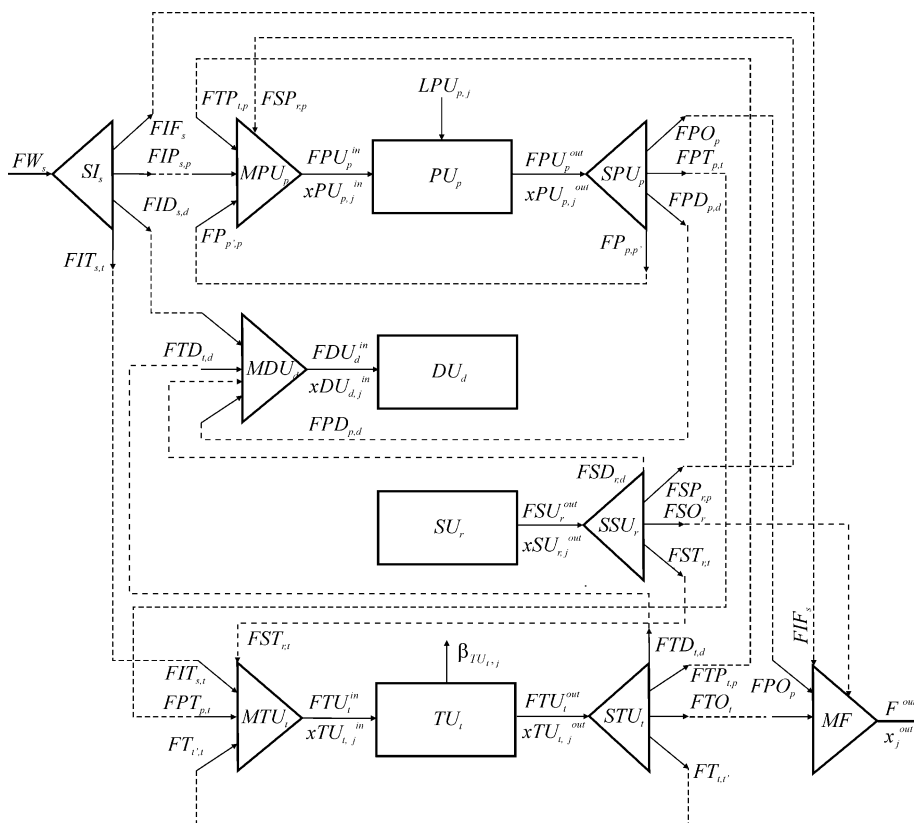
Fig. 1. Generalized superstructure for the design of integrated water networks.

Multiple sources of water include water of different quality that can be used in the various operations, and which may be sent first for pre-treatment. Local recycling [9] can be used to satisfy the flowrate constraints and in these cases it is possible to have an additional reduction in water consumption. As industrial water network systems usually consist of different types of water-using operations that can be classified as mass-transfer operations and non-mass transfer operations [4], both types of these operations are included in the superstructure. In addition to this, in many processes there is loss of water that is not available for re-use in a water-using operation. Hence, this unit involves water demand unit and is a water sink. Moreover, from some water-using operations water is available for re-use in other operations and they represent sources of water. According to this, the proposed superstructure can be used to represent separate subsystems as well as an integrated total system.

4. Model

The water network superstructure model is formulated as a nonconvex Nonlinear Programming (NLP) and as a nonconvex mixed-integer nonlinear programming (MINLP) for the case when 0-1 variables are included to model the cost of piping and/or selection of technologies for treatment. The model consists of mass balance equations for water and the contaminants for every unit in the network. The nonlinearities in the model appear in the mass balance equations in the form of bilinear terms (concentration times flowrate). In addition to this, nonlinearities appear in the objective function as concave terms of the cost functions for the water-treatment operations and for the investment cost of the pipes in the network. The objective function is to minimize the total network cost consisting of the cost of freshwater, the investment and operating costs of the treatment units. In most papers, the cost of the network piping and the cost of water pumping through pipes are not considered. Here, we introduce these costs in the objective function when the water network problem is formulated as an MINLP problem. Using the binary variables for the existence of pipe connections, the design specification for a maximum number of these connections in the network is formulated and used to establish a trade-off between cost and complexity of the piping network.

5. Solution strategy

To significantly improve the strength of the lower bound for the global optimum we incorporate the cut by Karuppiah and Grossmann [1] in the model. The bound strengthening in the nonlinear model corresponds to the contaminant flow balances for the overall water network system where bilinear terms are involved for the treatment units and final mixing points. It is also worth pointing out that when solving nonconvex water network problems by global optimization solvers, it is important to specify good variable bounds for all flowrates and concentrations in the water network. The reason is that these bounds are used in the convex envelopes for under and overestimating the nonconvexities (e.g. secant for concave function or McCormick envelopes for bilinear terms). In the proposed model the bounds on the variables are represented as general equations. They are obtained by physical inspection of the superstructure and by using logic specifications. Using the proposed model with the cuts in [1] and the proposed bounds we can effectively solve to global optimality the NLP water network problems for large-scale problems with multiple sources of water, multiple contaminants and more process and treatment units (large-scale problems). Also, the MINLP water network problems with a modest number of process units, treatment units, and

contaminants can be effectively solved to global optimality using the proposed model. However, for large-scale MINLP problems the global optimization solvers cannot find the global optimal solution in reasonable computational time.

To circumvent this problem, we propose a solution strategy that can be used for solving large-scale industrial water network problems. When the objective is to minimize the total network cost without specifying a maximum number of piping connections, we solve the NLP problem in which the 0-1 variables and the upper and lower bound constraints are excluded. Once we obtain the solution of the NLP, we fix all zero flowrates in the network and update the variable bounds before solution of reduced the MINLP. In the case when we specify a maximum number of pipe segments, we solve first the relaxed MINLP problem. The 0-1 variables of the streams in the network with zero value we fix at zero and then solve the reduced MINLP. With this solution method we can control the complexity of the water network. After solving the MINLP problem we can solve it again by restricting the number of piping connections in order to establish the trade-off between cost and network complexity. It that case, we assign to the model a new number of piping connections. Both, the NLP and reduced MINLP models are solved by a global optimization solver. While the rigorous global optimum cannot be guaranteed with the two-stage solution strategy, in our experience the optimality gaps are very small and the global optimum is still obtained in most cases.

6. Example

In this a MINLP example, we illustrate the advantage of using local recycles around the process units in the network and how the complexity of the water networks can be controlled by restricting the number of piping connections. The water network superstructure is shown in Fig. 2.
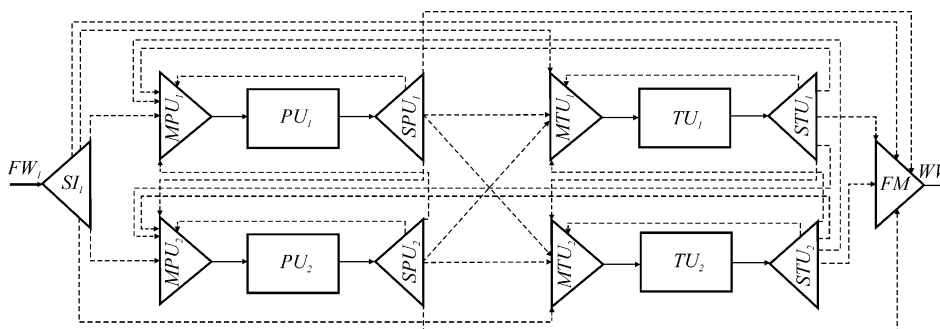
Fig. 2. Water network superstructure.

It consists of two process units (PU), two treatment units (TU), single source of water (SI), and two contaminants. Data for this example are taken from [1, 10]. Each treatment unit can remove only one contaminant. The environmental discharge limit for contaminant A and contaminant B is 10 ppm.

This example was implemented in GAMS 23.0 [11] and solved on a HP Pavilion Notebook PC with 4 GB RAM memory, and Intel Core Duo 2 GHz processor. The MINLP problem with local recycle involves 89 constraints, 79 continuous variables, and 22 discrete variables, and without local recycle 85 constraints, 75 continuous variables and 20 discrete variables. The optimality tolerance selected for optimization was 0.0 and BARON [12] was used for solving of this example to global optimality. The optimal network cost for the option without local recycle is $606,760.55/year (Fig.

3), and with local recycle $593,991.11/year (Fig. 4). In the first case the number of removable connections is 8 and in the second it is 9.

In order to simplify the water network, we assigned to the design constraint a new number of removable streams, and solved the MINLP corresponding problems. By application of the two stage solution strategy we solved all cases and computational times were less than 2 CPUs (solving directly the MINLPs required up to 94.2 secs and gave the same solutions). The results of the optimization by restricting the number of piping connections are shown in Table 1.
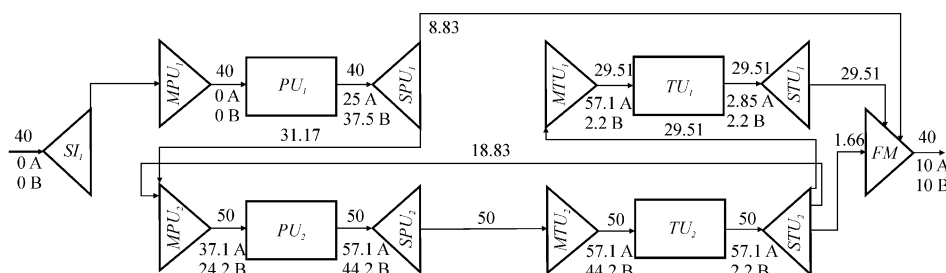
Fig. 3. Optimal solution for the MINLP problem without local recycle.
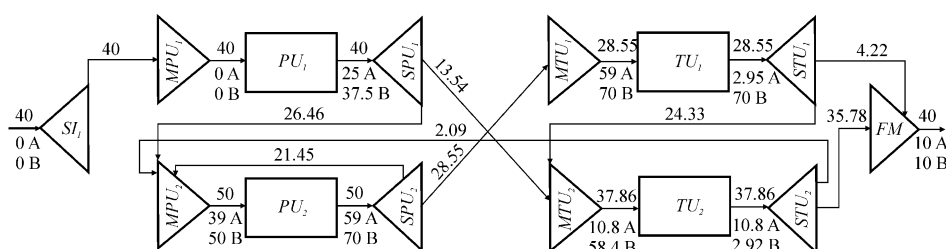
Fig. 4. Optimal solution for the MINLP problem with local recycle.

Table 1. Results of controlling the piping network complexity.

Total cost ($/year)		Number of removable streams in the network	
Without local recycle	With local recycle	Without local recycle	With local recycle
-	593,991.11	-	9
606,760.55	596,012.94	8	8
620,857.57	613,610.77	7	7
695,456.90	691,610.36	6	6

For instance, we can greatly simplify the water network with 6 removable piping connections as shown in Table 1. While we still keep the freshwater consumption at 40 t/h, the cost is significantly increased due to the increased cost of the treatment units..

7. Conclusions

We have presented a general superstructure and a global optimization strategy for the design of integrated process water network. To expedite the global optimization search we have represented the bounds on the variables as general equations obtained by physical inspection of the superstructure and used the cut in [1] to significantly improve the strength of the lower bound for the global optimum. Furthermore, we proposed a two-stage procedure for solving large-scale models. The proposed approach can be used for solving industrial water network problems as well as for controlling the water networks complexity.

8. Acknowledgments

References

[1] R. Karuppiah and I. E. Grossmann, Global optimization for the synthesis of integrated water systems in chemical processes, Computers & Chemical Engineering, 30 (2006), 650.

[2] A. P. Rossiter, Waste minimization through process design, McGraw-Hill, New York, USA 1995.

[3] M. M. El-Halwagi, Pollution prevention through process integration, Systematic design tools, Academic Press, San Diego, USA, 1997.

[4] J. G. Mann and Y. A. Liu, Industrial water reuse and wastewater minimization. McGraw-Hill, New York, USA, 1999.

[5] M. Bagajewicz, A review of recent design procedures for water networks in refineries and process plants, Computers & Chemical Engineering, 24 (2000), 2093.

[6] J. Jeżowski, Review and analysis of approaches for designing optimum industrial water networks, Chemical and Process Engineering, 29 (2008) 663.

[7] M. Bagajewicz and D. C. Faria, On the appropriate architecture of the water/wastewater allocation problem in process plants, (Editors J. Jeżovski and J. Thullie), 19th European Symposium on Computer Aided Process Engineering – ESCAPE19, Elsevier, 2009.

[8] D. C. Y. Foo, State-of-the-Art Review of Pinch Analysis Techniques for Water Network Synthesis, Industrial & Engineering Chemistry Research, 489(2009), 5125.

[9] Y. P. Wang and R. Smith, Wastewater Minimization with Flowrate Constraints, Chemical Engineering Research and Design. 73 (1995) 889.

[10] R. Karuppiah and I. E. Grossmann, Global optimization of multiscenario mixed integer nonlinear programming models arising in the synthesis of integrated water networks under uncertainty, Computers & Chemical Engineering, 32 (2008), 145.

[11] A. Brooke, D. Kendrick, A. Meeraus and R. Raman, GAMS: A user's guide, release 2.50. GAMS Development Corporation, 1988.

[12] N. V. Sahinidis, BARON: A general purpose global optimization software package. Journal of Global Optimization 8 (1996) 201.

On the Systematic Design and Optimization under Uncertainty of a Hybrid Power Generation System Using Renewable Energy Sources and Hydrogen Storage

G. Giannakoudis[b], A.I. Papadopoulos[a,b], P. Seferlis[a,b], S. Voutetakis[a]

[a]*Chemical Process Engineering Research Institute, Centre for Research and Technology-Hellas, 6th km Charilaou-Thermi Road, 57001, Thermi-Thessaloniki, Greece, spapadopoulos@cperi.certh.gr*
[b]*Department of Mechanical Engineering, Aristotle University of Thessaloniki, 54124, Thessaloniki, Greece, seferlis@auth.gr*

Abstract

This work addresses the optimal design and operation of a hybrid power generation system that uses renewable energy sources (RES) and hydrogen storage, while simultaneously accounting for associated design uncertainties in the form of stochastic variations in operating conditions. The considered hybrid system consists of photovoltaic panels, wind generators, accumulators, an electrolysis apparatus, hydrogen storage tanks, a compressor, a fuel cell and a diesel generator. The proposed design methodology involves the development of a power management strategy to determine all the feasible ways of dispatching power among the employed sub-systems. This is used in conjunction with an optimization method that considers design variables in the form of structural and operating parameters, as they directly affect the system performance. To emulate realistic operating conditions, the proposed design methodology is coupled with a systematic method to enable the efficient incorporation of uncertainties stemming from frequent external and internal system variations. Such uncertainties involve the fluctuating solar radiation and wind speed as well as the efficiency of the various energy conversion subsystems. The implementation of the design methodology results to robust and practically realizable system schemes, able to achieve high performance under a wide range of operating conditions.

Keywords: Renewable energy, hydrogen, systems design, uncertainty, optimization

1. Introduction

Systems that generate power using renewable energy sources (RES) are rapidly replacing power production based on conventional fuels. A major challenge that needs to be addressed in such systems is to facilitate the transformation into dependable and undisrupted power flows, of energy from sources that rely heavily on largely unpredictable natural phenomena. This goal is best served by the use of hybrid systems that combine multiple power generation units and storage media of diverse functionalities under an integrated power generation scheme. Such systems often involve photovoltaic panels and wind turbines to generate power that meets the demands of a targeted application. The intermittent nature of RES require the incorporation of batteries to store excess energy and release it under environmental condition variations. Whereas batteries are only able to serve short-term and limited power demands, hydrogen is emerging as an alternative, long-term and flexible energy storage medium that requires a rich infrastructure, to be integrated within hybrid power generation systems. Clearly, the involvement in the design of such systems of numerous

components with diverse operating characteristics causes significant complexities that require the implementation of efficient decision making methods. The intense presence of uncertainty due to fluctuating and unpredictable weather conditions or changes in the operational efficiency of the individual energy conversion units places additional requirements that need to be considered simultaneously with synergies developed among the employed sub-systems. Reported efforts to design systems with similar complexities are merely focused on the implementation of optimization algorithms on arbitrarily defined system characteristics [1-3], while the significant effects of uncertainty in the system performance have yet to be considered.

2. System description

The considered hybrid RES-based power system is shown in Figure1. It consists of PV panels and wind generators for power generation. Surplus energy is supplied to an electrolyzer after the specified load demand for a targeted application is satisfied.
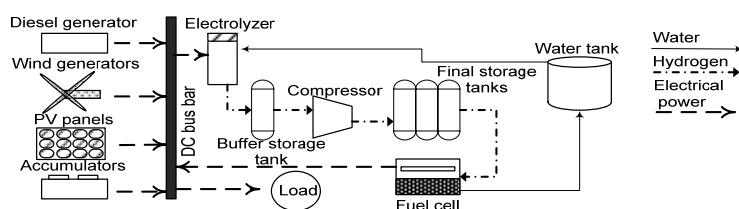
Figure 1: Block diagram of the proposed hybrid RES-based power system.

The produced hydrogen is stored in pressurized cylinders and in cases of energy deficit, is utilized in a fuel cell to provide the needed power to the system. Lead-acid accumulators (batteries) are used to regulate the power flows in the system through frequent charging and discharging cycles induced by the RES variability. In case of energy excess, units such as the hydrogen compressor utilize this energy to store hydrogen in long-term storage tanks. A diesel generator is also attached to the system and utilized only in cases of emergency (i.e. power demands of the application can not be covered by RES or stored hydrogen).

2.1. Power management

The system of Figure 1 involves several sub-systems with diverse requirements, that give rise to numerous operating options with regards to power utilization during system operation. The efficient integration of such subsystems requires the development of a power management strategy (PMS) to identify efficient operating decision alternatives, while maintaining a smooth system operation and protecting the individual components from irregular operating patterns that would eventually compromise their efficiency. Major operating parameters steering the generation of such alternatives are the state of charge (SOC) limits of the accumulator, the available system power (P) at any given time and the hysteresis band range (HBR), which determines the appropriate instance for initiation or termination of the operation of subsystems such as the electrolyzer and the fuel cells [4, 5]. The considered PMS involves the operating decision alternatives shown in Figure 2 for cases of power excess or deficit, depending on SOC values. In cases of power deficit ($P \leq 0$), the power required to meet the load demand is provided directly by the accumulators, if there is sufficient power to avoid utilization of the fuel cell (SOC > SOC_{fc}). If there is no available power in the accumulators (SOC $\leq SOC_{min}$) the fuel cell is utilized, provided that hydrogen is available in storage. If $SOC_{min} < SOC$

$<$ SOC_{fc} and the fuel cell was operating in the previous time step ($I_{fc}(t-1)\neq0$), the hysteresis band range (HBR=2%) enables prolonged operation of the fuel cell until the SOC reaches the limit SOC_{fc} (defined as SOC_{min}+ HBR). If the fuel cell did not previously operate ($I_{fc}(t-1)=0$), then it is not initiated to avoid frequent start-ups and shut-downs that could potentially lead to malfunction, and the load is covered by the accumulators. In case of power excess (P>0), the surplus power is used in the electrolyzer (SOC $\geq$ SOC_{max}) for the production of hydrogen, provided that the accumulators are charged to SOC_{max}. If the available power is higher than the power required from the electrolyzer (P > P_{max_elec}), the excess part is stored in the accumulators. This is possible due to the additional power storage available in the accumulator (over SOC_{max}) that is reserved to implement the hysteresis band policy. This extended SOC limit is in effect the sum of SOC_{max} and HBR and determines the point where only the auxiliary units may use the power excess. In case of insufficient power to enable operation of the electrolyzer (P < P_{min_elec}), the accumulators are used until P_{min_elec} is reached. In case of insufficient charge in the accumulators to enable operation of the electrolyzer (SOC < SOC_{elec}), the accumulators are charged. In case of $SOC_{elec} \leq SOC < SOC_{max}$, if the electrolyzer did not operate at a previous instance then it is not initiated and the available power is routed to the accumulators.
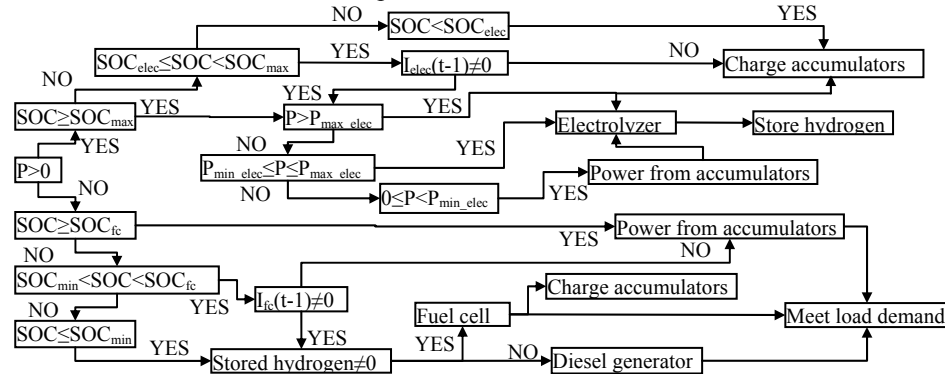
Figure 2: Operating decision alternatives represented by PMS

2.2. Uncertainty characterization

The design of a RES-based power generation system using solar and wind energy sources involves either the use of historical weather data or weather forecast methods to predict the future temporal evolution of the RES. Despite the use of such methods, the behavior of weather conditions always involves high uncertainty. Unless such uncertainty is accounted for during the system design, the performance of the RES-based system will only be optimum within the range of the considered weather conditions. Potentially unpredictable weather fluctuations will inevitably result to suboptimal system operation. In addition to the external variations propagated in the system operation, the efficiency of several subsystems, comprising the integrated power system, varies due to intense utilization or other exogenous factors throughout their anticipated life-term. For example, the efficiency of the electrolyzer and the fuel cell may faint during a prolonged system operation due to the utilization of delicate and prone to wear and tear materials such as membranes. The rate of efficiency degradation involves uncertainty as it is linked to the variable mode of equipment utilization, which is in turn affected by the considered weather conditions and the PMS employed to address them. The consideration of a constant system efficiency is expected to result in

suboptimal system performance, hence variations in efficiency should be considered during system design. Although these two types of uncertainty, namely external and system-inherent, are considered in this work for the parameters of solar radiation, wind speed and subsystem efficiencies, other types of uncertain parameters can be addressed using the design methodology presented in the following section.

3. Proposed design methodology

The design methodology illustrated in Figure 3 is based on the algorithm of Stochastic Annealing [6], which is used to address uncertainty in systems design and optimization. After the problem initiation, values must be assigned to the uncertain parameters. In case that uncertainty is considered to be of deterministic nature, it is described either by specific bounds imposed on the value range of the considered uncertain parameters or through a finite number of fixed parameter values. In case it is considered stochastic, uncertainty is represented by a probability distribution, showing the potential range of values for a parameter, in addition to accounting for the probability of occurrence of each value in the considered range. This attribute is suitable for the weather-associated uncertainty considered in this work. Intense unexpected variations in weather conditions are expected to be less frequent than milder ones, hence such type of uncertainty can successfully be represented through a normal distribution. As a probability distribution is a continuous function, it is represented by a finite set of discrete samples obtained using an appropriate sampling method [6].
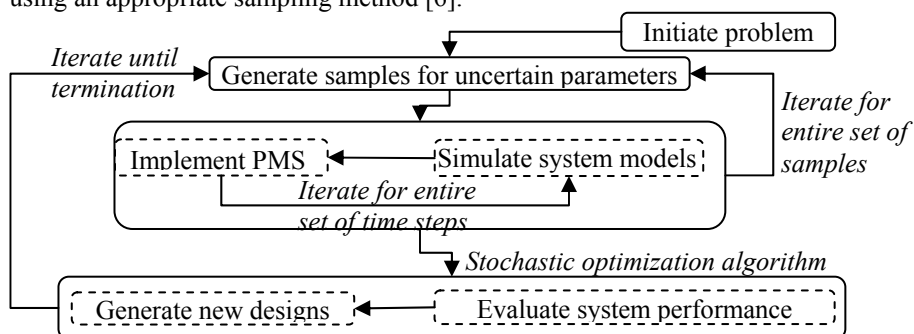
Figure 3: Methodology for optimization under uncertainty of hybrid power generation system

The system is subsequently simulated for each sample corresponding to the uncertain parameters. The temporal dependence of parameters such as solar radiation and wind speed require discretization into several time steps of the entire time period for which calculations are desired. The number of required time steps can be equal to the rate for which time-dependent data are available. In this respect each sample, represented by a randomly drawn value from the probability distribution, is added to the value of the time-dependent parameter for the entire set of considered time steps, hence emulating the unpredictable behavior of the uncertain parameters for the entire desired time period. As each point of the time-dependent data set is different in each time step (e.g. different solar radiation), the produced power is also of different intensity, hence the rules imposed by the PMS result to utilization of different subsystems in each time step. This procedure is iterated for the entire set of samples drawn from the probability distribution and correspond to a certain set of design variables. After termination of this set of iterations, the optimization algorithm invokes the generation of a new set of design variables and the procedure is repeated, until the overall algorithmic termination criteria are satisfied.

On the Systematic Design and Optimization under Uncertainty of a Hybrid Power
Generation System Using Renewable Energy Sources and Hydrogen Storage

4. Implementation

4.1. Background

The aim of the performed optimization is to minimize the net present value (NPV) of investment for a ten year operating period. The considered costs involve initial purchase and installation capital and additional expenditures for the operation, maintenance and replacement of the employed equipment. The highest possible system autonomy from fossil fuels is also required, hence the use of the diesel generator is penalized during system optimization. The considered decision variables involve the number of PV panels (n_{pv}), the number of the wind power generators (n_{wg}), the nominal capacity of the accumulators (n_{acc}), the maximum operating power of the electrolyzer ($P_{max,e}$), the capacity of the intermediate (V_b) hydrogen storage tanks, the nominal power of the fuel cell ($P_{op,fc}$) and the upper (SOC_{max}) and lower (SOC_{min}) limits of the stage of charge of the accumulators. The considered uncertain parameters involve the solar radiation ($u_{s,t}$), the wind speed ($u_{w,t}$) as well as the efficiencies of the electrolyzer (u_e) and of the fuel cell (u_{fc}). The weather data are taken from a database in the form of hourly averaged data for a year, corresponding to the conditions observed in a particular geographical area. The uncertainty in the weather data is implemented through samples drawn from a normal distribution that enable a random deviation of maximum ±12% from the hourly averaged points. The uncertainty in the efficiencies is implemented through an one-sided normal distribution that enables random deviation of maximum -8% from a pre-specified efficiency. In this case the uncertainty is not time-dependent and is only allowed to decrease, to emulate the probability of fainting equipment performance. The load demand of the targeted application is allowed to vary within the range of 0.6-1.3 kW throughout the year.

4.2. Discussion of results

The addressed cases involve system design without consideration of uncertainty (C1) and design considering uncertainty (C2) in order to compare the system performance. Table 1 shows the design results obtained for both cases. The major differences between the two involve the significantly larger number of PV-panels (n_{pv}) required in C1 and the significantly larger intermediate storage tanks (V_b) required in C2.

Table 1: Design results for cases C1 and C2

Case	n_{pv}	n_{wg}	n_{acc}	$P_{max,e}$(kW)	V_b(Nm³)	$P_{op,fc}$(kW)	SOC_{max}(%)	SOC_{min}(%)
C1	49	5	7	0.5	0.2	1	96	66
C2	45	6	6	0.5	0.9	2	93	68

Based on Table 1 the aim is to compare the economic performance of the designs corresponding to cases C1 and C2 in response to variations, hence a sample of variations is imposed only in the wind speed, the solar radiation and the fuel cell efficiency, and the results are shown in Figure 4a. The negative NPV of the investment is due to the widely acknowledged fact that such technologies are still under intense development and unable to compete the prices of conventional energy sources. In any case, it appears that under intensely favorable wind speed variations, unfavorable solar variations and high fuel cell efficiency, the design of C1 performs better than C2. However, as the wind speed deviates towards an increasingly negative direction and the fuel cell efficiency drops, a steep decrease is observed in the NPV for the design of C1, while the design of C2 remains unaffected and even presents a slight improvement. This shows that highly unfavorable operating conditions are addressed with increased efficiency when uncertainty has been accounted for during the design stages.
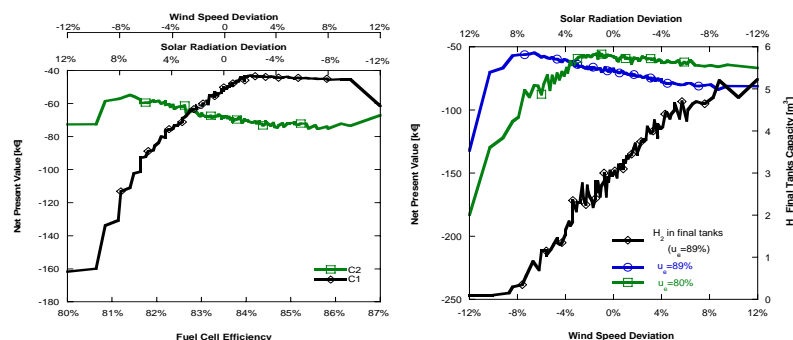
Figure 4: System performance under uncertainty in variations of wind speed, solar radiation and
a) fuel cell efficiency, b) electrolyzer efficiency.

Figure 4b shows the effects of varying wind speed and solar radiation considered simultaneously with variations in the efficiency of the electrolyzer for C2. For negative variations in the solar radiation and positive variations in the wind speed, high efficiency in the electrolyzer leads to increased production of hydrogen that is stored in the long-term (final) storage tanks, which result in a higher required investment. In case of unfavorable wind conditions and favorably high solar radiation, the system requires a high efficiency in the electrolyzer in order to maintain decreased investment costs. The storage and consumption profile of hydrogen is also shown in Figure 4b. In the case of 89% electrolyzer efficiency, the capacity of the long-term hydrogen storage tanks is over $5m^3$ (for compressed hydrogen), while as the wind conditions deteriorate the available hydrogen is gradually consumed.

5. Conclusions

The presented work addresses the optimal design of a hybrid RES-based power production system under uncertainty. The proposed design methodology determines the values for structural (i.e. equipment capacity) and operational (i.e. parameters of the PMS) design variables that optimize a performance criterion over the life span of the system under uncertainty due to unpredictable variation of weather phenomena and potential fainting equipment efficiency. In particular, the temporal variation of the weather conditions is addressed through an efficient power management strategy. The proposed methodology provides a reliable tool for the design of realistic and highly performing hybrid energy systems of complex structure in any geographical location.

6. Acknowledgements

References

[1] Bernal-Augustin, J.L., Dufo-Lopez, R., Renew. Sust. Energ. Rev.,13 (2009)2111-2118.
[2] Vosen, S.R., Keller, J.O., Int. J. Hydrogen Energ., 24 (1999), 1139-1156
[3] Dufo-Lopez, R., Bernal-Augustin, J.L., Renew. Energ., 33 (2008), 2559-2572.
[4] Ipsakis, D., Voutetakis, S., Seferlis, P., Stergiopoulos, F., Elmasides, C., Int. J. Hydrogen Energ., 34 (2009), 7081-7095.
[5] Ipsakis, D., Voutetakis, S., Seferlis, P., Stergiopoulos, F., Papadopoulou, S., Elmasides, C., Energy, 33 (2008), 1537-1550
[6] Chaudhuri, P., Diwekar, U., AICHE J., 45 (1999), 1671-1687

Optimization of Overall Network of Water-Using and Treatment Units

Raluca Tudor, Vasile Lavric

University Politehnica of Bucharest,Chemical Engineering Department, RO-011061, Polizu 1-7, Bucharest, e-mail: r_tudor@chim.upb.ro, v_lavric@chim.upb.ro

Abstract

The concept of process integration has gained more and more attention in the last years due to its beneficial results in the framework of freshwater minimization, costs reduction and environmental impact diminution. The optimization of an integrated water-using (WU) and treatment units (TU) network is addressed in this paper, by targeting for maximum treated water reuse as a considerably better alternative to fresh water consumption. The mathematical model of the integrated water network (IWN) is based upon total and contaminant species balances. They are written for each and every unit and are subject of restrictions of inlet and outlet contaminant concentrations. The optimization problem is carried via Genetic Algorithms. An industrial example was analyzed with respect to contaminants' mean availability and network reuse index to point out the benefits of integration.

Keywords: integrated network, fresh water minimization, contaminant mean availability, network reuse index, genetic algorithms.

1. Introduction

Process integration, as defined by El-Halwagi [1], is the "holistic approach to process design, retrofitting, and operation which emphasizes the unity of the process". In the framework of process synthesis, water network integration emerges as an important part of this field. There are two main approaches available for the optimization of integrated water networks: graphical techniques and mathematical programming.

The most recent comprehensive review of the various graphical techniques to design and retrofit continuous water networks has been studied by Foo [2]. *Water pinch analysis* is a useful graphical method, which shows simple solutions and valuable results when applied to water-using networks with single contaminant.

Mathematical approaches are more powerful in solving more complex systems with many water-process units, with respect to the optimal solution for water and utilities minimization [3], capital cost estimation [4] or evaluation of zero discharge possibility [5] for mono and multi-contaminant situations.

Generally water network design techniques were originally developed for water reuse/recycle networks and were then extended to accommodate the cases with regeneration and afterward with wastewater treatment network.

The search for more economic solutions has led to the use of evolutionary optimization methods: genetic algorithms [6], ant colony optimization [7], pinch multi-agent genetic algorithms [8] or particle swarm optimization [9].

Optimization of multiple contaminants water network by Genetic Algorithms (GA) was addressed by Lavric et al. [6]. A technique able to find the minimum water supply at the same time with the network topology which ensures the maximum water reuse was de-

veloped. This study was later applied on a network which allowed internal regeneration of the wastewater streams to find the optimal solution of the system [10].

This paper addresses the problem of optimality of an IWN, focusing on three main directions: firstly, optimization of the WN without the treatment network included using supply water as objective function, secondly, optimization of the IWN and comparison with the previous case, and thirdly, data analysis with respect to contaminants' mean availability and the networks' reuse index.

2. Mathematical model

An IWN can be abstracted as an oriented graph, with respect to water flow, starting with the WU operations which are supplied with fresh water only due to their inlet constraints until the last WUs which have the most permissive restrictions. Two ranking criteria are envisaged: by maximum fresh water consumption (FWC) and by maximum contaminant load (MCL). Each WU is assigned to a TU, depending on the stream's exit concentration. The TUs have inlet restrictions and fixed outlet concentrations. They are ranked in cascade, starting with the most permissive inlet TU which receives the heaviest contaminated wastewater streams and continuing with more restrictive inlet constraints until the last TU. The outlet concentrations of the latter comply with legal regulations for water discharge in nature.

Each TU is considered as a contaminated supply water source, available for the WUs; thus the optimization of water-using network reduces to the case of different level polluted sources [11]. The efficiency of such an integrated treatment system is twofold: on one hand, each polluted stream enters the correct TU, avoiding supplemental pumping costs and dilution of heavily polluted streams treated in the previous units, and on the other, permits a possible reuse of each TU exit into the appropriate WU.

The mathematical model is based upon overall and partial mass balances together with inlet and outlet restrictions for the WUs $C_{ki}^{U,in,\max}, C_{ki}^{U,out,\max}$ and inlet restrictions $C_{ki}^{T,in,\min}$ and fixed outlet concentration for the TUs. The optimization procedure implying the minimization of the fresh water flow seen as objective function uses GA as implemented in Matlab™ and is completely detailed in [11].

3. Case study

A synthetic example was used to study both the optimal topology of an IWN and which are the effects of the integration of TUs upon the sub-network of WUs. An original network of 6 WUs and 3 contaminants (the information regarding each unit's load and as-

Table 1: Primary data (mass loads, inlet and outlet restrictions) of the water-using units of the network

UNIT No.	Load (kg/hr)			Inlet concentration restrictions (ppm)			Outlet concentration restrictions (ppm)		
	Contaminant			Contaminant			Contaminant		
	1	2	3	1	2	3	1	2	3
U1	0.35	0.25	0.15	0	0	0	35	45	55
U2	0.15	0.35	0.25	15	20	25	70	90	120
U3	0.35	0.45	0.55	15	35	0	75	95	125
U4	0.45	0.15	0.45	25	45	45	95	85	135
U5	0.25	0.45	0.35	45	35	55	90	100	120
U6	0.15	0.45	0.85	35	20	25	85	80	95

sociated restrictions are presented in Table 1 and 3 TUs (the inlet restrictions and the fixed outlet values of the concentrations of each contaminant are presented in Table 2) is analyzed under the aforementioned scenarios having as objective to find the optimum topology which reduces fresh water consumption while increasing treated wastewater use. This topology is compared against the optimal topology found for the WN given in Table 1 when using supply water minimization as optimization criterion, but without the treatment network integrated.

The optimized WN topology for the latter case when considering FWC ranking and fresh water consumption minimization is depicted in Figure 1 (the MCL ranking case is disregarded due to the lack of space). Both networks consume approximately the same amount of fresh water and their topology differs depending on the ranking criteria used. The network with the WUs ordered by FWC is less complex, which implies lower operational costs.

Table 2: Primary data of the wastewater treatment units of the integrated network

Treatment Unit No.	Inlet concentration restrictions (ppm)			Outlet concentration (ppm)		
	Contaminant			Contaminant		
	1	2	3	1	2	3
TU 1	45	45	55	30	40	50
TU 2	25	35	45	20	25	30
TU 3	15	20	25	2	4	5

The IWN topology obtained using FWC as ranking criterion is presented in Figure 2 (the MCL case not shown). Both networks use the same fresh water supply (14.7347 t/hr) meaning they discharge the same flow (14.7411 t/hr) in environment; the difference inlet and outlet flows is given by the load removed in the IWN. In Table 3 are listed the inlet and outlet actual contaminants concentrations, emphasizing when the critical concentrations were attained. WUs 4 and 5 are fed with partially decontaminated water which, mixed with the upward streams, reaches the maximum concentration for at least one contaminant (Table 3).
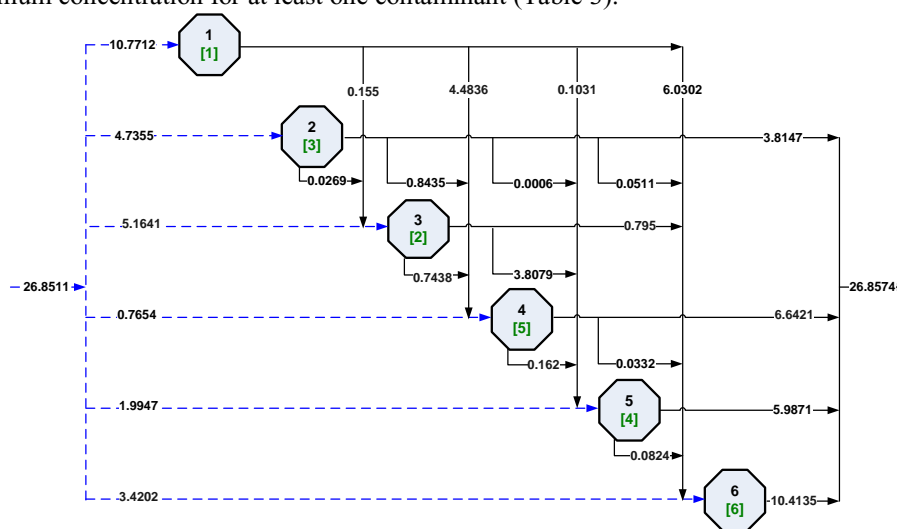
Figure 1. The optimal topology of the WN - ranking by FWC. The rank is given by the numbers in square brackets, the initial arrangement is given by the plain figures, the dashed arrows represent the fresh water entering each WU, the numbers on the arrows stand for the flow (in t/hr), while the rest of the arrows give the internal flow distribution.

The topology of the network has changed significantly in comparison with Figure 1; lesser internal water reuse and feed of WUs primarily with partially decontaminated

water, according to their restrictions. We considered two other optimization scenarios which disregard either the primary or the secondary TU but keep the two ranking criteria. The data obtained are analyzed and compared with all the other scenarios presented in this paper in Table 4.
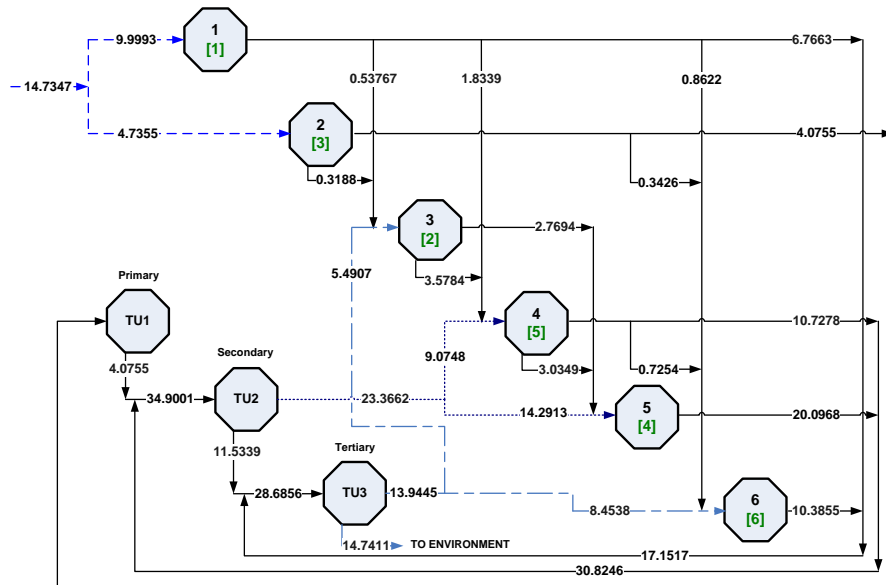
Figure 2. The optimal topology of the IWN ranked by FWC
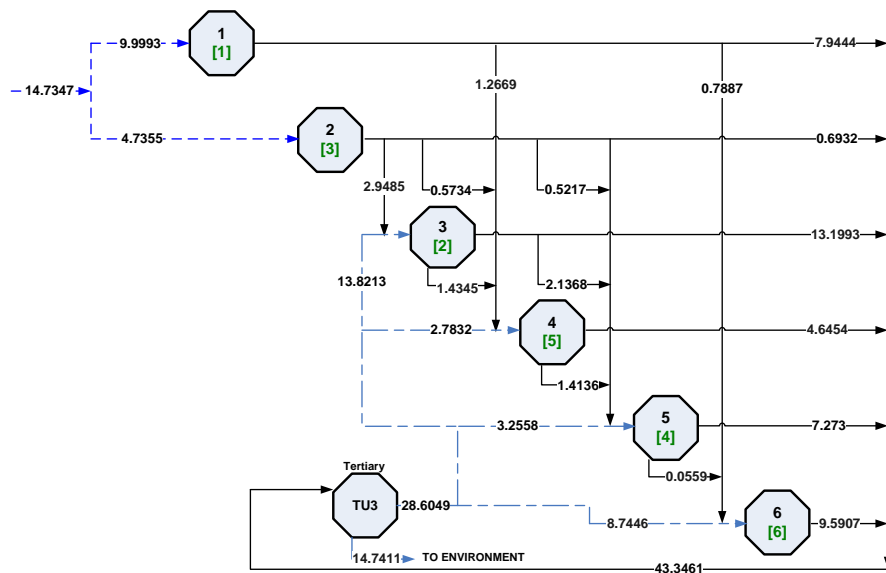
Figure 3. The topology of the optimal restricted IWN ranked by FWC with only the final treatment unit (TU3).

The last optimization scenario considers the IWN with only the final TU, for both ranking criteria. When applying FWC, the IWN uses more internal streams, but with lower values and the flow sent to treatment is higher, than in the other case when the internal flow (43.3461 t/hr vs. 34.0957 t/hr), although the internal streams are less in number but have higher values. More internal streams means more pipes, thus more energy consumed for pumping. At the same time, the water reused by the optimal IWN ranked by FWC is 28.6049 t/hr, while for the other situation, the water reused is 19.3546 t/hr.

Table 3. Contaminants' concentrations for the optimal networks. The values written in bold represent the critical concentrations.

UNIT No.	Actual inlet concentration (ppm)			Actual outlet concentration (ppm)			Network type
	Components			Components			
	1	2	3	1	2	3	
U 1	0	0	0	32.49	23.21	13.93	WN (Figure 1)
U 2	1.31	1.15	0.99	29.37	66.61	47.75	
U 3	0	0	0	73.89	**95**	116.11	
U 4	20.86	44.87	32.34	**95**	69.58	106.48	
U 5	33.62	34.19	28.65	70.18	**100**	79.84	
U 6	22.4	19.86	13.38	36.8	63.07	**95**	
U 1	0	0	0	35	25	15	IWN (Figure 2)
U 2	8.41	10.35	11.43	32.04	65.49	50.81	
U 3	0	0	0	73.89	**95**	116.11	
U 4	**25**	36.78	37.01	47.39	44.24	59.4	
U 5	24.87	**35**	33.24	42.13	66.06	57.4	
U 6	9.92	13.08	13.16	24.36	56.41	**95**	

An advanced analysis of the results could be done using the concept of *mean availability* as defined by Iancu et al. [10] and a new concept, *network reuse index* (NRI). The former means the overall mean pseudo-driving force of the mass transfer of the contaminant k, defined as the average of concentration differences computed at the entrance and at the exit of each WU unit. The latter is computed as the ratio between the actual water streams and the inlet and outlet water stream over the whole topology; it points out the level at which the outlet wastewater streams of each WU is reused by the next ones in sequence. A low value means scarce internal water reuse (treated or coming from the previous WUs) instead of fresh water consumption, while a high value

Table 4. Mean availability and network reuse index computation

No.	Network type			Contaminants			NRI
				1	2	3	
1	WN	FWC		**14.09**	**11.038**	22.128	2.4
2	WN	MCL		**9.822**	**9.503**	19.957	2.27
3	IWN	FWC		21.832	**16.882**	25.953	7.3
4		MCL		**16.939**	**17.045**	19.12	6.67
5		TU1 & TU3	FWC	**17.999**	**16.489**	23.175	7.33
6			MCL	**16.423**	**17.031**	23.012	8
7		TU2 & TU3	FWC	17.868	**14.043**	19.695	7
8			MCL	19.362	**14.669**	21.843	6.67
9		TU3	FWC	18.129	**15.367**	21.388	7
10			MCL	**14.322**	**16.077**	20.762	6.33

indicates an enhancement of the internal water reuse. As an example, NRI for the topology depicted in Figure 1, which is 2.4 (see Table 4), comes as the ratio of 24 active streams and 10 inlet and outlet streams. The results obtained are compared among each other in terms of contaminant mean availability and network reuse index (Table 4). For each case are highlighted the values of the critical contaminants, which control either the supply water consumption or part of the internal reuse due to the mass transfer bottlenecking. The third contaminant of the system is never critical, while the first and the second ones are critical either together, forming a *bottleneck island* [4], or the second one only (Table 4), depending on the network type (WN and IWN). In the first two cases (WUs network only, no treatment included), the first two contaminants form a bottleneck island with the lowest values for both the minimum availability and NRI.

When ranking the WUs by FWC, NRI gets rather high constant values irrespective of how many TUs are considered (Table 4). This could be an indication that the mass transfer driving force is well balanced from the beginning of the network (the contaminant input free WUs) till its last WU. When ranking the WU by MCL the highest NRI is obtained for the IWN with primary and tertiary TU while, the IWN with tertiary TU only accounts for the lowest NRI, but both having the first two contaminants forming a bottleneck island. The greater gap between these limits could be the result of a larger variation of the mass transfer driving force along the network.

4. Conclusions

The optimization of an integrated water network is studied in this paper. The mathematical model is solved using a GA, such that the algorithms guarantees minimization of fresh water consumption thus treated water maximization, observing all restrictions. The TUs are assimilated with potential partial contaminated water sources; the algorithm used allowing their assignation to the proper WU according to the closeness between their outlet and inlet concentration restrictions. The optimal topology depends on the order in which the WUs are arranged and, also, on the TUs available. It was observed that in an IWN 55% from the fresh water needed can be replaced by treated water which is important due to high cost of fresh water and sources availability. The performance of each type of network is evaluated in terms of critical contaminants' mean availability and network reuse index and the results are compared among each other, showing that optimization of the whole networks improves internal reuse.

5. Acknowledgements

References

[1] M. M. El-Halwagi, Academic Press; San Diego, 1997
[2] D. C. Y. Foo, Ing. Eng. Chem. Res., 48 (2009), 5125
[3] L. Savulescu, J. K. Kim, R. Smith, Chem. Eng. Sci., 60 (2005), 3291
[4] V. Lavric, P. Iancu, V. Plesu, Resour. Conserv. Recy., 50 (2007), 186
[5] B. Van der Bruggen, L. Braeken, Desalination, 188 (2006), 177
[6] V. Lavric, P. Iancu, V. Plesu, J. Clean. Prod., 13 (2005), 1405
[7] H. Maier, A. R. Simpson, Water Resour. Plng. and Mgmt., 129 (2003), 200
[8] K. Cao, X. Feng, H. Ma, Comput. Chem. Eng., 31 (2007), 1565
[9] J. Izquierdo, I. Montalvo, R. Perez, V. S. Fuertes, Comput. Math. Appl., 56 (2008), 777
[10] P. Iancu, V. Plesu, V. Lavric, Chem. Eng. Trans., 18 (2009), 851
[11] R. Tudor, V. Lavric, Ing. Eng. Chem. Res., under second revision, (2009)

Multiobjective optimization with economic and environmental objective functions using Modified Simulated Annealing

Florinda Martins[a], Carlos A.V. Costa[b]

[a] REQUIMTE, ISEP, Rua Dr. António Bernardino de Almeida, 431, 4200-072 Porto, ffm@isep.ipp.pt
[b] LEPAE – DEQ, Faculty of Engineering University of Porto, Rua Dr. Roberto Frias, s/n 4200-465 Porto, ccosta@fe.up.pt

Abstract

In this work an algorithm based on simulated annealing is presented to solve multiobjective optimization problems of chemical processes, considering Net Present Value as the economic objective function and Global Potential Environmental Impact as the environmental objective function. Toluene hydrodealkylation to produce benzene was used as case study. The results were analysed for: the algorithm performance considering the results average deviation, the Pareto optimal solutions and the comparison with monobjective optimization results.
Keywords: Multiobjective optimization; Simulated Annealing; Economic and Environmental objective functions;

1. Introduction

Real problems are usually multidimensional implying the search for more than one objective, most of them conflicting with each other. In multiobjective optimization there are mainly two approaches: one that establishes a hierarchy for the objectives, using that preference to form a composite objective function transforming the problem into a monobjective optimization one and a second one that tries to determine several trade-off solutions. The main disadvantages of the first method are the subjectivity associated with the weight vector and the incapacity of finding good optimal trade-off solutions in a non convex space. The work of Kim and Weck (2006) is an example of the weighting sum method proposing some strategies to overcome its disadvantages. There are other methods like Benson's (Benson et al., 1998; Ehrghott, 2000) , value function (Keeney and Raifa, 1993), goal programming (Deb, 1998) and Interactive Surrogate Worth Trade-off (Chankong et al. 1985). These algorithms must be applied several times, hoping that each time a different Pareto optimal solution is obtained, what is not guaranteed.

Besides these methods there are the stochastic methods, namely simulated annealing (SA), tabu search (TS) and the evolutionary algorithms. One can point out the works of Srinivas and Deb (1995) with the development of the NSGA algorithm (nondominated sorting genetic algorithm) and Deb et al. (2002) with the NSGA-II. These were applied to industrial hydrogen plants by Rajesh et al. (2001), to an incineration plant by Anderson et al. (2005), to an industrial unity by Bhutani et al. (2006) and to batch facilities by Dietz et al. (2006) .

The methods for multiobjective optimization based in Simulated Annealing present some advantages since they do not need derivatives, they easily solve continuous functions and combinatory problems and allow to obtain several optimal solutions in a

single optimization. Some of these algorithms are: MOSA developed by Ulungu et al. (1999), SMOSA developed by Suppapitnarm (2000), PSA, PDOSA and WMOSA (developed by Suman (2003) for problems with constrains). The strategy used in MOSA method uses a projection of the multidimensional space into a one-dimensional one through a weighted sum. Different scalarizing functions lead to different projection paradigms. The SMOSA algorithm uses the concept of archiving Pareto optimal solutions together with a strategy of returning to base . It does not consider a weighting vector in the acceptance criterion and uses multiple temperatures, one by objective (Suman, 2004). According to Suman (2004) in the PSA (Pareto Simulated Annnealing), Czyzak et al. (1994) and Czyzak and Jaszkiewicz (1998) introduced the concept of interactive solutions. In each iteration a set of solutions, named generating sample, is used to control the weights of the objectives in the multiple acceptance probability to insure a good dispersion of the generating sample. The PDMOSA algorithm introduces a strategy of Pareto-domination based fitness while WMOSA has as main characteristic the imbedding of restrictions in the main algorithm. The application of SA to full process optimization problems even of the LP or NLP type is relatively scarce. The number of works that include environmental concerns is also relatively scarce. In the present case SA will be applied to full process multiobjective MINLP problems with two objective functions. One of the objectives was to develop a multiobjective optimization algorithm based in SA and then analyze the performance of the algorithm in the optimization of flowsheets of different complexity mainly introduced by liquid recycling and heat integration, considering economic (NPV) and environmental (GPEI) objective functions and looking for the optimal Pareto set that represents the trade-offs between both criteria. The SA acceptance criterion was modified in order to obtain solutions from the optimal Pareto set. Another goal is to compare the results with those obtained by applying monobjective optimization (economical and environmental objectives) (Martins and Costa, 2009).

2. Algorithm and Objective Functions

To represent the full process model we used a commercial simulator, PROII (SimScience, 2005). An interface was developed to transfer data between the simulator and the optimization algorithm and vice-versa (Martins and Costa, 2009).

2.1. Modified SA Algorithm

The main characteristics of the proposed modified SA algorithm are the creation of an archive of solutions (at the end it is expected that the solutions in this archive are Pareto optimal solutions) that is refreshed during the optimization and a modified criterion of acceptance. The optimization is processed according to one of the objective functions but the acceptance criterion was modified in order to choose the solutions that correspond to movements that improve the other objective. A movement that improves both objectives is accepted with probability 1, being afterwards verified the relation of domination towards the solutions in the Pareto archive. If in this archive there is a solution that dominates the new solution, the solution is still used in the algorithm but is excluded from the optimal solutions archive. If the new solution dominates solutions in the archive it replaces those solutions. If it does not dominate and if it is not dominated it will be added to the archive. A movement is rejected if it improves the main objective but corresponds to a worst value to the other objective, although the domination relation is also verified. If it corresponds to a worst value to the first objective it is accepted with probability P, calculated as usual in the SA algorithm.

*Multiobjective optimization with economic and environmental objective functions
using Modified Simulated Annealing*

Some preliminary multi and monobjective optimizations considering NPV and GPEI as objective functions were performed, in order to tune SA parameters. The initial temperature (T0) is a function of the worst value of objective function (CM). So T0=0.1*CM. The Kirkpatrick (Kirkpatrick et al., 1983) cooling schedule was selected. The value used for the constant was 0.9 which corresponds to a slow cooling. The maximum number of iterations was determined by prior optimizations; we concluded that 26 cycles were sufficient (Martins and Costa, 2009).

2.2. Objective functions

Two objective functions were considered, one of the economic type (Net Present Value, NPV- maximization) and one of the global environment impact type (Global Potential Environmental Impact, GPEI - minimization). To determine the environmental performance a method based on the Guinée (2002) and on the IChemE (2007) methodologies was used. The impact categories considered were: climate change, acidification, eutrophication, photo-oxidant formation, human health and depletion of abiotic resources (Martins and Costa, 2009). A normalization procedure similar to the WAR algorithm (Cabezas et al., 1999) was adopted. The weight for each category was considered equal to 1.

3. Case study

As case study the non catalytic hydrodealkylation of toluene to benzene was chosen (HDA). Three topologies were chosen, E, Ee and Ge (Martins and Costa, 2009).
The optimization variables were chosen accordingly to the characteristics of the process, topologies and variable input set determined by the simulator. For each of the optimization variables a range was established (Martins and Costa, 2009).

4. Results

The modified SA algorithm was applied to the topologies E, Ee and Ge considering NPV and GPEI as objective functions using a number of iterations equal to the one used in monobjective optimization. For each topology 10 optimization runs were carried out , 5 with NPV as first objective function and 5 with GPEI as first objective function. By doing this it is possible to analyze the influence of considering a certain objective function as the first one. The results show that for topology E the number of non-dominated solutions per optimization does not vary significantly. For topology Ge a more significant variation was detected, particularly when using NPV as main objective function. In the case of topology Ee and for each optimization only one solution was usually found. Analyzing all the solutions for each topology it is possible to conclude that for topology E only 6 are really non-dominated solutions, for topology Ee only 2 are non-dominated and for topology Ge only 7 are non-dominated. It should be mentioned that for topology Ee the solutions with interest (economic and environmental) are very rare, which was already observed in monobjective optimizations (Martins and Costa, 2009). For each topology, the percentage deviation (%D) of NPV and GPEI of the several final solutions relatively to best value found to NPV and to GPEI, respectively was calculated (values in italic and bold style in tables 1, 2 and 3). Then the average percentage deviation (A%D) of each objective function was calculated and finally the global average percentage deviation (GA%D) was determined. These calculations were performed for two sets: all solutions and only the non-dominated solutions (bold values in tables 1, 2 and 3). For the topology Ee only the first set was considered since only two solutions are non-dominated.

Table 1 Performance of the multiobjective SA – Topology E

Number of the optimization		1st	1st	2nd	3rd	3rd	3rd	4th	4th	4th	5th	5th	5th	All solutions		Non –dominated solutions	
														A%D	GA%D	A%D	GA%D
Optimization NPV	NPV	12.3	3.2	3.9	4.5	8.4	0.6	7.7	2.6	6.5	2.6	7.1	1.9	5.1	9.2	3.5	9.0
	GPEI	13.2	18.8	11.2	15.7	3.0	18.3	1.0	13.7	11.7	14.2	*0.0*	26.4	13.4		14.4	
Optimization GPEI	NPV	5.8	*0.0*	3.9	3.2	3.9	0.0	5.8	6.5		6.5	3.9	1.9	4.1	7.7	3.0	8.5
	GPEI	0.0	26.4	15.2	10.7	7.1	26.4	8.6	4.1		3.6	9.6	13.2	11.4		14.0	

Table 2 Performance of the multiobjective SA – Topology Ee

Number of the optimization		1st	2nd	3rd	4th	4th	5th	All solutions	
								A%D	GA%D
Optimization NPV	NPV	31.4	147.4	*0.0*	31.4		40.6	62.7	99.0
	GPEI	92.6	425.1	**17.1**	29.8		111.7	135.3	
Optimization GPEI	NPV	126.9	**9.1**	126.9	133.1	128.6	26.3	91.8	188.6
	GPEI	298.0	*0.0*	344.8	277.9	411.7	95.0	285.5	

Table 3 Performance of the multiobjective SA – Topology Ge

N° of the optim		1st	1st	1st	2nd	2nd	2nd	2nd	3rd	3rd	3rd	4th	4th	4th	4th	5th	5th	All solutions		Non –dominated	
																		A%D	GA%D	A%D	GA%D
Optim NPV	NPV	1.3	4.6		3.3	9.8	*0.0*	2.6	0.7	0.7	17.0	1.3				9.8	0.7	4.7	16.6	8.8	18.7
	GPEI	43.2	15.4		15.8	15.1	61.0	23.6	47.9	12.7	12.0	18.5				13.5	64.1	28.6		28.6	
Optim GPEI	NPV	3.9	0.7	17.6	20.3	1.3	9.8	13.7	8.5	26.8	9.2	8.5	25.5	17.0	28.8	7.2	11.1	13.1	14.9	12.3	14.4
	GPEI	19.3	58.7	10.8	1.5	22.0	18.5	5.4	27.8	12.4	15.1	17.4	4.2	9.7	*0.0*	20.5	5.8	16.6		16.6	

For topology E the global average percentage deviation for the two series of optimizations is very similar when we consider the non-dominated solutions, meaning that optimizing according to NPV or to GPEI does not affect significantly the quality of the solutions obtained. Although when all the solutions are considered the results are different (optimizing according to NPV leads to higher global average percentage deviation). For topology Ge we verified that the opposite occurs: the global average percentage deviations are closer when all solutions are considered. The values of global average percentage deviation are usually higher than the ones obtained for topology E. For topology Ee the global average percentage deviations are very high and optimizing according to GPEI leads to higher deviations. If the values obtained for the two non-dominated solutions are considered it is possible to conclude that the deviations decrease significantly. So reasonable performance was observed for topology E (global average percentage deviation <10%). For more complex topologies the algorithm is not very efficient, in particular for topology Ee that presents very few interesting solutions in mono-objective optimization (Martins and Costa, 2009). The process we are using as case study aims at producing high purity benzene (purity above 0.990); by analyzing all the solutions we can conclude that all of them fulfill that requirement. Another issue is the quality of the non-dominated solutions, namely the trade-offs represented by those solutions. Figures 1 and 2 show the Pareto front obtained for topologies E and Ge. This was not shown for topology Ee since it has only two non-dominated solutions. Comparing the solutions obtained by multiobjective optimization with the results obtained by monoobjective optimization with NPV and GPEI as objective functions

Multiobjective optimization with economic and environmental objective functions
using Modified Simulated Annealing

(Martins and Costa, 2009), it is possible to conclude that for topology E only one solution obtained in the monobjective optimizations is non-dominated (using GPEI objective function). The best solution found in monobjective NPV optimizations is dominated while the best solution found in monobjective GPEI optimizations is non-dominated, being the best environmental solution found. For topology Ge it is possible to conclude that 2 solutions from monobjective GPEI optimizations are non-dominated (and dominated some solutions non-dominated found in multiobjective optimizations) and one solution from monobjective NPV optimizations is non-dominated. The best solution found in monobjective NPV optimizations is dominated while the best solution found in monobjective GPEI optimizations is non-dominated, being the best environmental solution found. For topology Ee the results obtained in multiobjective optimization are not very interesting since only two non-dominated solutions were found. When the solutions obtained by multiobjective optimization are compared with the solutions obtained in monobjective optimizations it is possible to see that they usually present higher values to GPEI and smaller values to NPV.
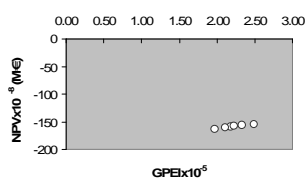
Figure 1 Pareto front – Topology E
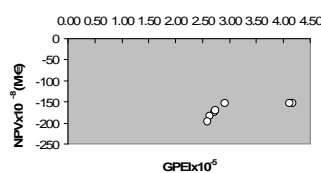
Figure 2 Pareto front – Topology Ge

5. Conclusions

This work analyzed the behavior of the modified SA multiobjective optimization algorithm developed, using two objective functions: net present value and global potential environmental impact. The hydrodealkylation of toluene to produce benzene was used as case study, considering three topologies with different complexity mainly obtained by including or not liquid recycling and heat integration.

The performance of the algorithm was observed from various angles:

a) the SA algorithm can obtain several optimal solutions belonging to the Pareto front to most of the topologies, although for some complex topologies it presents very few optimal solutions;

b) for the topologies E and Ge the optimizations with GPEI as main objective function contributed with more non-dominated solutions to the final optimal set;

c) the algorithm performs reasonably for the simplest topology; for complex topologies the performance of the algorithm is poor;

d) when comparing with monobjective results the multiobjective solutions of the simplest process topology are of good quality dominating almost every solution from monobjective optimizations. For more complex topologies the quality of the Pareto optimal solutions decreases and some solutions from monobjective optimization dominate the multiobjective solutions;

e) for all topologies there are non-dominated solutions from multiobjective optimization that are better than the best solution found in monobjective NPV optimizations, although the difference of NPV values between the non-dominated solutions from multiobjective optimization that are better and the best solutions from monobjective optimizations is less then 1.5% except for topology Ee. On the other hand for

topologies E and Ge the best solutions of monobjective GPEI optimizations are non-dominated solutions and have the best value for GPEI (the solution for topology Ee is dominated).

f) Since we are dealing with economic and environmental objective functions it is interesting to look for the trade offs, in the sense of having one objective function increasing while the other decreases. For the simplest topology the variations observed in the NPV objective function are relatively small but for GPEI they are meaningful. For complex topologies the variations for both objective functions are meaningful and usually higher.

References

Anderson, S. R., Kadirkamanathan, V., Chipperfield, A., Sharifi, V. and Swithenbank, J. (2005). "Multi-objective optimization of operational variables in a waste incineration plant." Computers & Chemical Engineering 29: 1121-1130.

Benson, H. P., Lee, D. and McClure, J. P. (1998). "Global Optimization in Practice: An Application to Interactive Multiple Objective Linear Programming." Journal of Global Optimization 12: 353 – 372.

Bhutani, N., Ray, A. K. and Rangaiah, G. P. (2006). "Modeling, Simulation, and Multi-objective Optimization of an Industrial Hydrocracking Unit." Ind. Eng. Chem. Res. 45: 1354-1372.

Cabezas, H., Bare, J. C. e Mallick, S. K. (1999). "Pollution prevention with chemical process simulators: the generalized waste reduction (WAR) algorithm-full version." Computers and Chemical Engineering 23: 623-634.

Chankong, V., Haimes, Y. Y., Thadathil, J. and Zionts, S. (1985). Multiple Criteria Optimization: A State-of-the_Art Review. Lecture Notes in Economics and Mathematical Systems Berlim, Springer – Verlag.

Deb, K. (1998). Non-linear Goal Programming Using Multi-Objective Genetic Algorithms. Dortmund, University of Dortmund.

Deb, K., Pratap, A., Agarwal, S. and Meyarivan, T. (2002). "A Fast and Elitist Multi-objective Genetic Algorithm: NSGA-II." IEEE Trans.Evol. Comput. 6(2): 182-197.

Dietz, A., Azzaro-Pantel, C., Pibouleau, L. and Domenech, S. (2006). "Multiobjective optimization for multiproduct batch plant design under economic and environmental considerations." Computers and Chemical Engineering 30: 599-613.

Douglas, J. M. (1988). Conceptual Design of Chemical Processes. Singapore, McGraw-Hill Chemical Engineering Series.

Ehrgott, M. (2000). Multicriteria Optimization. Berlin, Springer.

Guinée, J. B. (2002). Handbook on life cycle assessment. Operational guide to the ISO standards. Dordrecht, Kluwer Academic Publishers.

IChemE. (January/2007). "The Sustainability Metrics." www.icheme.org.

Keeney, R. L. and Raiffa, H. (1993). Decisions with Multiple Objectives : Preferences and Value Tradeoffs Cambridge, Cambridge University Press.

Kim, I. Y. and Weck, O. L. (2006). "Adaptative weighted sum method for multiobjective optimization: a new method for Pareto front generation." Struct. Multidisc. Optim. 31: 105-116.

Kirkpatrick, S., C. D. Gelatt, J. and Vecchi, M. P. (1983). "Optimization by simulated annealing." Science 220: 671-680.

Martins, F.and Costa, C. A. V. (2009). "Economic, Environmental and Mixed Objective Functions in Non-linear Process Optimization Using Simulated Annealing and Tabu Search "Computers & Chemical Engineering **in press (**peer-reviewed and accepted) http://dx.doi.org/10.1016/j.compchemeng.2009.10.015

Rajesh, J. K., Gupta, S. K., Rangaiah, G. P. and Ray, A. K. (2001). "Multiobjective optimization of industrial hydrogen plants " Chem. Eng. Sci. 56: 999-1010.

Srinivas, N. and Deb, K. (1995). "Multi-Objective Function Optimization using Non Dominated Sorting Genetic Algorithms." Evol. Comput. 2(3): 221 – 248.

Suman, B. (2003). "Simulated Annealing - Based Multiobjective Algorithms and their Application for System Reliability." Eng. Opt. 35: 391 – 416.

Suppapitnarm, A., Seffen, K. A., Parks, G. T. and Clarkson, P. J. (2000). "Simulated annealing: An alternative approach to true multiobjective optimization." Engineering optimization 33(1): 59 – 85.

Ulungu, L. E., Teghem, J., Fortemps, P. H. and Tuyttens, D. (1999). "MOSA method: A tool for Solving multiobjective combinatorial optimization problems." Journal of Multi-Criteria Decision Analysis 8: 221 – 236.

Run-to-run convergence analysis of model-based policy iteration algorithms for experimental optimization of batch processes

Mariano Cristaldi,[a] Smaranda Cristea,[b] Ernesto Martínez[c]

[a]*INTEC–CONICET/UNL, Güemes 3450, 3000 Santa Fe, Argentina*
[b]*Universidad de Valladolid, c/Real de Burgos, 47011 Valladolid, España*
[c]*INGAR–CONICET/UTN, Avellaneda 3657, S3002 GJC Santa Fe, Argentina*
 ecmarti@santafe-conicet.gob.ar

Abstract

Convergence analysis of iterative identification-optimization schemes is a key issue in modeling for optimization of batch processes. In this work, it is formally shown that for convergence is *sufficient* to guarantee that parametric uncertainty is increasingly reduced on a run-to-run basis. Convergence of a policy iteration algorithm to an optimal policy which satisfies the Hamilton-Jacobi-Bellman equation is thus assured as long as parametric uncertainty is iteratively reduced such that the performance prediction mismatch is driven to zero. The integration of global sensivity analysis with confidence interval boostrapping in the design of a convergent algorithm for model-based policy iteration is proposed. A simple bioprocess is used to exemplify run-to-run improvement.

Keywords: boostrap confidence intervals, experimental optimization, global sensitivity analysis, modeling for optimization, batch process.

1. Motivation and scope

Convergence analysis of iterative identification-optimization schemes based on imperfect models is a very difficult problem to solve as has been pointed out in some previous works (Brdyś & Roberts, 1997; Srinivasan & Bonvin, 2003). The standard procedure consists of iteratively using new measurements to increasingly bias an imperfect model by parameter re-estimation which feeds the cycle of continuous policy improvement. Key questions that should be addressed in the design of model-based policy iteration algorithms are: i) if this identification-optimization cycle converges to a policy in the face of modeling errors, and ii) under what circunstances the resulting policy is somewhat near to the optimal one for the real process? To compensate for parametric uncertainty, a model-based policy iteration algorithm requires data from carefully designed experiments using global sensitivity analysis (Martínez, et al., 2009). The key issue addressed here is how confidence intervals should be best represented upon scarce experimental data so as to guarantee run-to-run policy convergence. Boostrapping of model parameter distributions is proposed to describe uncertainty.

2. Model-based policy iteration

Let's assume the dynamic behavior of the batch process under study is modeled by the set of ODEs

$$\frac{dx}{dt} = f(x(t), \wp(w,t), \theta) \ 0 \le t \le t_f, \quad x(0) : \text{given} \tag{1}$$

and the optimization objective to be minimized is

$$J(t_f, x) = h(x(t_f)) + \int_0^{t_f} g(x, \wp(w, t)) dt \qquad (2)$$

where $x(t)$ is an n_s-dimensional vector of time dependent state variables, $w \in \Omega$ is an m-dimensional vector of parameters for the input policy $\wp$, $\theta \in \Theta$ is a p-dimensional vector of model parameters and t_f is the final time. For a given θ, the optimal policy parameterization $w*$ should satisfy the well-known *Hamilton-Jacobi-Bellman* (*HJB*) optimality condition

$$0 = \min_{w \in \Omega} \left\{ g(x(t), \wp(w^*, t)) + \frac{\partial J^*(t, x)}{\partial t} + \left[\frac{\partial J^*(t, x)}{\partial x} \right]^T f(x, \wp(w^*, t), \theta) \right\}, \qquad (3)$$

For any sub-optimal policy $\wp(\hat{w}, t)$, there exist an error term which can be characterized by rewriting the right-hand side of Eq. (3) as follows:

$$0 < \left[g(\hat{x}(t), \wp(\hat{w}, t)) + \frac{d(J(t, \hat{x}(t)))}{dt} \right] \qquad (4)$$

By integrating the right-hand side of (4), the Bellman residual BR is defined:

$$BR(\hat{\theta}, \hat{w}) = \left[\int_0^{t_f} g(\hat{x}(t), \wp(\hat{w}, t)) \, dt + h(\hat{x}(t_f) - J(0, x(0)) \right] > 0 \qquad (5)$$

By subtracting from the $BR(\hat{\theta}, \hat{w}^*)$ for the special case where $\hat{\theta} \neq \theta_{real}$ and $\hat{w}^*$ is the estimated "optimal" policy parameterization based on an imperfect model from the $BR(\theta_{real}, w^*)$ (which has the minimum possible value of zero for the optimal policy $\wp(w^*, t)$) the BR can be re-written as:

$$BR(\theta, \hat{w}^*) - BR(\theta_{real}, w^*) =$$

$$= \left[\int_0^{t_f} \left[g(\hat{x}(t), \wp(\hat{w}^*, t)) - g(x_{real}(t), \wp(w^*, t)) \right] dt + \left[h(\hat{x}(t_f) - h(x_{real}(t_f)) \right] \right]$$

$$BR(\theta, \hat{w}^*) = J(t_f, \hat{x}) - J^*(t_f, x_{real}), \quad \text{for any } t_f. \qquad (6)$$

Based on Eq. (6) above, the optimal policy $\wp(w^*, t)$ can be found by iteratively minimizing the performance prediction error resulting from applying an estimated optimal policy to the real process using a model-based policy iteration approach as it is shown in Fig. 1. Each iteration starts with a given policy $\wp(\hat{w}^*, t)$ from the previous iteration which is evaluated in a specifically designed dynamic experiment so as to bring the most sensitive information to reduce the performance prediction mismatch (Martínez et al., 2009) E using data from samples and a model with a given parameter vector θ as follows:

$$E(\theta) = \sum_{r=1}^{sp} \left(J^*(t_r, \hat{x}) - J(t_r, x_{obs}) \right)^2 \qquad (7)$$

The most important question for algorithm design in model-based experimental optimization is how this new information must be used so as to guarantee convergence of policy iteration while driving the performance prediction error $E(\theta)$ to zero.

Fig. 1. Model-based policy iteration cycle.
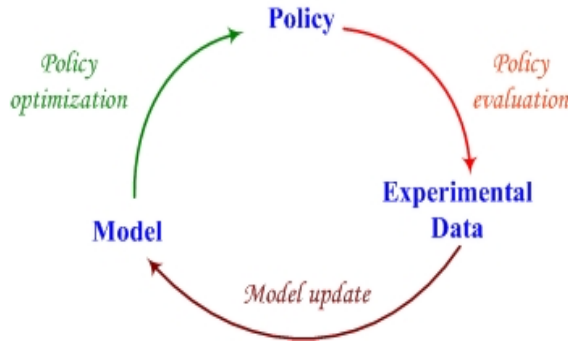
As soon as the performance prediction mismatch E can be driven to zero, policy iteration will converge to the optimal policy $\wp(w^*, t)$ which gives rise to optimal performance. Should model structure and parameters be perfectly known a priori model-based policy iteration is able to provide the optimal policy parametrization once the first iteration has been completed.

3. Run-to-run convergence

Consider the problem of minimizing the performance prediction error $E : \Theta \to \Re$ for a given model parameterization and its corresponding optimal policy, where Θ is a bounded Euclidean space over which any model parameterization is confined to as parametric uncertainty is increasingly reduced on a run-to-run basis. Let Γ denote the HJB optimality condition, and let $\Sigma \subset \Theta$ denote the set of model parameterizations for which Γ is satisfied. Also, let $\xi : \Theta \to \Re^-$ be a non-positive valued function such that $\xi(\theta) = 0$ if and only if $\theta \in \Sigma$. Such a function is called an *optimality function* associated with Γ (Polak, 1997). This function ξ must be defined so as to provide a quantitative measure of the extent to which a model parameterization along with its related optimal policy satisfy the condition Γ, namely the HJB equation in (3). The natural choice when defining an optimality function for convergence analysis of a policy iteration algorithm is resorting to the modulus of the *BR* in (5) for the special case where $\hat{w} = \hat{w}^*(\theta)$ corresponds to the optimal policy based on a model with parameter θ :

$$\xi(\theta) = -\left| BR(\theta, \hat{w}^*(\theta)) \right|, \ \theta \in \Theta \tag{8}$$

Definition 1. *Shrinking set* Θ. As the number n of runs increases the parameter space Θ which characterizes model parametric uncertainty shrinks (non-necessarily in a monotone way) towards a given accumulation point $\tilde{\theta}$ such that :

$$\lim_{n \to \infty} \Theta^n = \tilde{\theta}, \tag{9}$$

where the parameter space for any finite *n* always satisfies $\Theta^{n+1} \subset \Theta^n$.

Definition 2. *Sufficient descend*. A model-based policy iteration algorithm working over a shrinking parameter space Θ has the property of sufficient descend with respect to the chosen optimality function ξ if for every $\delta > 0$ there exists a $\eta > 0$ such that, for every $n = 1, 2, \ldots$, and for every iteration point (model parameterization) $\hat{\theta}^n$ computed by the algorithm, if $\xi(\hat{\theta}^n) < -\delta$, then

$$\mathrm{E}(\hat{\theta}^{n+1}) - \mathrm{E}(\hat{\theta}^n) < -\eta \tag{10}$$

where $\hat{\theta}^{n+1}$ is the next iteration point.

Proposition. *Suppose there exists a constant $D \in R$ such that $\mathrm{E}(\theta) \geq D$ for every $\theta \in \Theta$. If a model-based policy iteration algorithm has the property of sufficient descend, then for any infinite sequence $\{\hat{\theta}^n\}_{n=1}^{\infty}$, it computes, $\lim_{n\to\infty} \xi(\hat{\theta}_n) = 0$.*

Proof. The proof is immediate from Definition 2. □

To guarantee that a policy iteration algorithm effectively works over a shrinking set of model parametric uncertainty as required by Def. 2, it is mandatory to make the most sensible use of scarce data from each experimental run. To this aim, a key issue to be addressed is which sub-set of model parameters should be re-estimated at each iteration so as to increasingly reduce the performance prediction mistmatch. As proposed by Martínez, et al. (2009), the sub-set of parameters should be chosen based on a global sensitivity analysis (Saltelli, et al., 2008). But by doing this selective re-estimation, it is not sufficient to guarantee the shrinking set assumption above since it all depends on how new data is used to define confidence intervals and their associated probability distributions for model parameters as more experimental runs are made.

A method frequently used to describe confidence intervals of parameters estimated from scarce data is based on the Fisher-Information Matrix. The application of this traditional method has two important shortcomings to guarantee the convergence proposition: (i) it gives only lower bounds for the variance of a parameter if the solution of the underlying model equations is non-linear in the parameters; and (ii) each confidence interval is symmetric with respect to its estimated mean. To overcome these issues boostrapping confidence intervals along with distribution frequencies (histograms), mean and variance of each parameter is proposed to describe how uncertainty is increasingly reduced as data gathering is biased through optimal design of dynamic experiments.

The bootstrap method is a data-based simulation method for statistical inference(Joshi et al., 2006). To perform the analysis, an initial set **S** of experimental data is used as a database. The bootstrap approach now uses a large set of B-times artificially replicated experimental data $S_1^*, S_2^*, ..., S_B^*$ based on replacements to calculate statistical properties of the resulting distribution of the (re)-estimated set of parameters. These replicate data sets make possible to obtain histograms for re-estimated parameters. The confidence intervals for the parameters are then calculated by the percentile method: let $\hat{\theta}^{*(\alpha)}$ indicate the 100 (1-α) percentile of B bootstrap replications for a given parameter θ, then the percentile interval $(\bar{\theta}_{lo}, \bar{\theta}_{up})$ of intended coverage is obtained by

$$(\bar{\theta}_{lo}, \bar{\theta}_{up}) = [\hat{\theta}^{*(\alpha/2)}, \hat{\theta}^{*(1-\alpha/2)}] \tag{11}$$

4. Case study and results

To exemplify run-to-run convergence, a model of the semi-continuous (fed-batch) fermentation of baker's yeast is used. Assuming Monod-type kinetics for biomass growth and substrate consumption (including maintainance), the bioreactor model is:

$$\frac{dx_1}{dt} = (r - u_1 - \theta_4)x_1; \quad \frac{dx_2}{dt} = -\frac{rx_1}{\theta_3} + u_1(u_2 - x_2) - mx_1 \tag{12.a}$$

$$\frac{dV}{dt} = F; \quad r = \frac{\theta_1 x_2}{\theta_2 + x_2}; \quad u_1 = \frac{F}{V} \tag{12.b}$$

where x_1 is the biomass concentration (g l^{-1}); x_2 is the substrate concentration (g l^{-1}); u_1 is the dilution factor (h^{-1}); and u_2 is the substrate concentration in the feed (g l^{-1}). Both x_1 and x_2 can be sampled during each experiment; to simulate measurement noise in each sample a normal random noise with σ_1=0.25 and σ_2=0.18, for x_1 and x_2, respectively, was added to the "real" model output. Let's assume that for the "real" process unknown parameters are θ_1=0.31; θ_2=0.18; θ_3=0.55; θ_4=0.55; m=0.07, whereas initial parametric uncertainty is shown in Table 1. In each policy evaluation run, a number of 6 samples are taken at the most sensitive times determined following the optimizaiton procedure detailed in Martínez, et al. (2009). In each iteration, sampled data is used to boostrap mean values and confidence intervals. The boostrap parameter B was set to 3000.

The objective of interest is to maximize the final biomass concentration x_1 while minimizing the substrate concentration x_2 at the end of the run (t_f=12 h) by manipulating the dilution factor profile u_1 and the initial bioreactor volume:

$$Min \ J(u_1(t), V_i) = \left(- x_1 + x_2 \theta_3 \right) V, \ V \le 100, \ 0.05 \le u_1(t) \le 20 \tag{13}$$

The initial biomass and substrate concentration are assumed fixed to $x_1^\circ = 1$ g l^{-1} and $x_2^\circ = 20$ g l^{-1}, whereas the substrate concentration in the feed is chosen as u_2=35 g l^{-1}. The inflow rate parameterization is defined as follows:

$$u_1(t) = \sum_{i=1}^{4} a_i \left(\frac{t}{t_f} \right)^{i-1} \tag{14}$$

In Table 2, the performance of the optimal policy is compared with policy parameters and performance obtained for five iterations of the identification-optimization cycle in Fig. 1. As can be seen, despite the huge initial uncertainty in model parameters and measurement noise, model-based policy iteration quickly converges to a policy whose performance is very near to the performance provided by the optimal policy. Fig. 2 and Fig. 3 vividly demonstrates the importance of using boostrapping to achieve run-to-run shrinking of parametric uncertainty in a model-based policy iteration algorithm. It is worth noting that in the iteration 0 parameters are uniformly distributed in their intervals.

Table 1. Reduction of parametric uncertainty over five iterations

	Iter. # 0	Iter.# 1	Iter.# 2	Iter.# 3	Iter.# 4	Iter. # 5
θ_1	[0.05,0.98], 0.5	[0.27,0.51]	[0.31,0.37]	[0.27,0.61]	[0.33,0.38]	[0.27,0.72]
θ_2	[0.05,0.98], 0.5	[0.05,0.98]	[0.05,0.98]	[0.05,0.98]	[0.05,0.98]	[0.05,0.98]
θ_3	[0.05,0.98], 0.5	[0.23,0.98]	[0.23,0.98]	[0.29,0.97]	[0.29,0.97]	[0.37,0.97]
θ_4	[0.01,0.50], 0.25	[0.01,0.19]	[0.05,0.11]	[0.01,0.38]	[0.06,0.12]	[0.01,0.45]
m	[0.01,0.50], 0.25	[0.01,0.50]	[0.01,0.50]	[0.01,0.48]	[0.01,0.48]	[0.01,0.49]

5. Concluding remarks

Aimed at increasingly reducing the performance prediction mismatch under parametric uncertainty, run-to-run convergence conditions for policy iteration algorithms have been formally stated using the shrinking set assumption. Boostrapping has been proposed to guarantee *sufficient descend* (Def. 2) with scarce experimental data in policy iteration.

Table 2. Policy iteration and performance comparisons with the optimal policy

	Optimal policy	Iter.# 0	Iter.# 1	Iter.# 2	Iter.# 3	Iter. # 4	Iter. # 5
a_1	17.8438	4.457	7.191	5.188	18.3412	17.95	18.0828
a_2	-40.0267	33.379	50.418	58.295	-40.0886	-40.90	-41.0745
a_3	-6.02206	-44.423	-143.325	-143.40	-6.2522	-6.4934	-6.4933
a_4	0.15945	-16.064	-12.3811	-12.1765	-0.0130	-0.4205	-0.4016
V_i	54.2597	36.912	53.672	54.7972	51.8232	54.764	54.2823
J(real)	**-1139.94**	-228.548	-1132.55	-1135.91	-1059.47	-1136.24	**-1139.93**
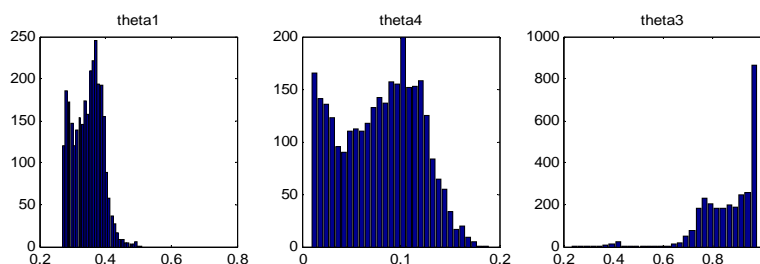
Fig. 2. Boostrapping frequencies for θ_1, θ_3 and θ_4 in the first iteration.
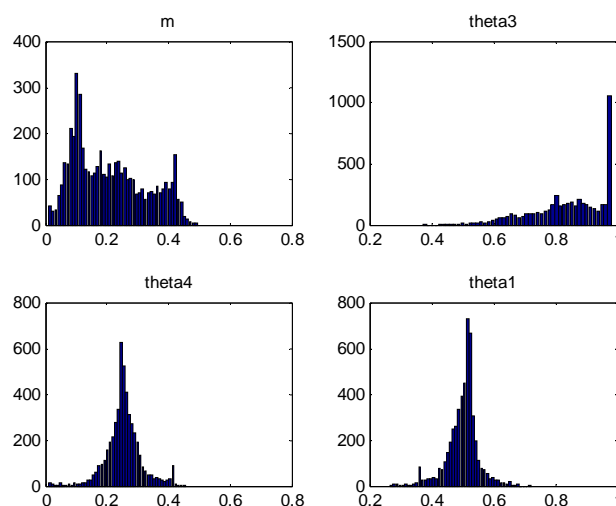
Fig. 3. Boostrapping frequencies for θ_1, θ_3, θ_4 and m (maintainance) in iteration #5.

References

M. Brdyś, P. Roberts, 1987, Convergence and optimality of modified two-step algorithm for integrated system optimisation and parameter estimation, Int. J. Syst. Sci., 1305–1322.

M. Joshi, A. Seidel-Morgenstern, A. Kremling, 2006, Exploiting the bootstrap method for quantifying parameter confidence intervals in dynamical systems, Metab. Eng, 8 447–455.

E. Martínez, M. Cristaldi M., R. Grau, 2009, Design of Dynamic Experiments in Modeling for Optimization of Batch Processes, Ind. Eng. Chem. Res, 48, 7, 3453-3465.

E. Polak, 1997, Optimization: Algorithms and Consistent Approximations,Springer: New York.

A. Saltelli, et al., 2008, Global sensitivity analysis. The primer, J. Wiley: Chichester, UK.

B. Srinivasan, D. Bonvin, 2003, Convergence analysis of iterative identification and optimization schemes. Proceedings of the ACC, DenVer, CO, USA.

Energy optimization in the process industries: Unit Commitment at systems level

Edwin Zondervan[a], Ignacio E. Grossmann[b], André B. de Haan[a]

[a]*Eindhoven University, P.O. Box 513 5600MB, the Netherlands*
[b]*Carnegie Mellon University, 5000 Forbes Avenue, P.A. 152130-3890, Pittsburgh U.S.*

Abstract

In previous work, the unit commitment problem has been formulated as a non-convex MINLP, which is computationally expensive to solve. To circumvent this problem, we reformulate in this paper the problem as a convex MIQCP and derive the relevant constraints using propositional logic. We consider a specific problem based on a network of gas and oil fired power generators. Numerical results are presented using several methods (e.g. CPLEX for MIQCP, and DICOPT, SBB for MINLP). It is shown that proposed convex MIQCP reformulation can be solved much faster than previous non-convex models reported in the literature.

Keywords: Energy optimization, MINLP, Unit Commitment problem.

1. Introduction

Energy production and management is becoming of increasing importance in the process industries. The energy market becomes deregulated and the size of the power grid increases (Dochain et al. 2006). On top of that the so called renewable power sources find their way into the world. Long since, the chemical engineering community has been concerned with process synthesis of utility systems in plants, generation, co-generation and poly-generation (Iyer and Grossmann (1998) and El Mahgary and Tamminen (1993)), and now the challenge of energy dispatch has been found by the process systems engineers; the link between power production and demand.

In this paper we address the unit commitment problem, which is concerned with deciding what power units should be committed to satisfy a time variant demand in so as to minimize operating costs. These costs are defined on the basis of a quadratic criterion for power generation and start-up and shut-down costs for each unit. The problem is constrained by a load balance (the given demand over a time horizon), power limits for each generator, ramp limits, minimum up/down times, spinning reserve and carbon dioxide emission constraints. In some studies also power loss constraints and fuel constraints are added to the problem. Typical problem sizes are networks of 10-100 units that should be committed over a time horizon: 4-24 hours.

The above sketched unit commitment problem has been formulated in the literature as a non-convex MINLP, due to multiplications of discrete and continuous variables in cost function, and different formulations of the constraints for the up/down times (see for example Niknam et al. 2009). Also, the solution approaches reported in the literature often use meta-heuristic approaches e.g. GA, SA, FA (see Sasaki et al., 1992; Kazarlis et al.; 1996; Zhuang and Galiana, 1990). In this paper we show that the problem can be

formulated as a <u>convex MIQCP</u> that can be solved effectively with deterministic mixed-integer optimization tools.

2. Problem statement and proposed model

Given a network of power generators, the economic dispatch problem is concerned with finding how much power each unit should generate for a given demand, while minimizing the total operational costs, which are generally expressed in nonlinear form. By extending the problem over a multi-period time and including the decisions of switching units on or off at a given time period, this gives rise to the unit commitment problem. In this paper we solve the problem for a deterministic case, where the demand load over time is known.

In the following the deterministic security constrained unit commitment case study is formulated as a convex MIQCP model. The cost function, containing a convex quadratic production cost term and linear shutdown and startup cost terms is as follows:

$$\min C = \sum_{i=1}^{I} \sum_{t=1}^{T} \left[\left(u_{i,t} a_i + b_i p_{i,t} + c_i p_{i,t}^2 \right) + \left(y_{i,t} SD_i \right) + \left(z_{i,t} SU_i \right) \right] \tag{1}$$

where p_{it} is the power produced at time t at unit i. u_{it} is the 0-1 decision whether unit i is turned on or off at time t, and y_{it} and z_{it} are 0-1 decision variables for the addition of shut-down costs (SDi) and/ or start-up costs (SU_i). In the nonconvex model reported in the literature the quadratic term in the summation is multiplied by the variable $u_{it.}$ The following constraints apply:

a) Demand (D_i) load balance:

$$\sum_{i=1}^{I} p_{i,t} = D_t \forall t \tag{2}$$

b) Spinning reserve (R_t) constraint:

$$\sum_{i=1}^{I} u_{i,t} p_i^{U} = R_t + D_t \forall t \tag{3}$$

c) Output limitations:

$$u_{i,t} p_i^{L} \leq p_{i,t} \leq u_{i,t} p_i^{U} \forall i,t \tag{4}$$

d) Logic constraints for the shutdown cost term:

$$u_{i,t-1} - u_{i,t} \leq y_{i,t} \forall i,t \tag{5}$$

e) Logic constraint for the startup term:

$$u_{i,t} - u_{i,t-1} \leq z_{i,t} \forall i,t \tag{6}$$

Constraints (5) and (6) can be derived from logic conditions (see Raman and Grossmann, 1991), for start-up holds: "if $u_{i,t} = 1$ and $u_{i,t-1} = 0$ then start-up costs should be added to the objective function", assigning the Boolean literals to each action $P_A =$ unit i at time t is on ($u_{i,t} = 1$), $\neg P_B =$ unit i at time t-1 is turned off ($u_{i,t-1} = 0$), and $P_C = y_{it}$, where $y_{i,t} \in \{0,1\}$ (do not add costs, add costs), the logic expression is given by:

$P_A \wedge \neg P_B \Rightarrow P_C$

Removing the implication:

$\neg (P_A \wedge \neg P_B) \vee P_C$

Applying De Morgan's theorem leads to:

$\neg P_A \vee P_B \vee P_C$

It is noted that $P_A = u_{i,t}$ implies $\neg P_A = 1 - u_{i,t}$. Assigning the corresponding 0-1 variables to each term of the above conjunction leads to:

$1 - u_{i,t} + u_{i,t-1} + y_{i,t} \geq 1$

which can be rearranged to the following inequality

$u_{i,t} - u_{i,t-1} \leq y_{i,t}$

which is equal to the inequality (5) of the startup constraints. Generator i has a minimum down time TD_i, once it has been shut down. This constraint can be derived from the logical condition: "if $y_{i,t} = 1$ then $u_{i,t+1} = 0$ and $u_{i,t+2} = 0$,..., and

$u_{i,t+DT_i} = 0$", $y_{i,t} \Rightarrow \underset{j \in D_i}{\wedge} \neg u_{i,t+j}$

where $D_i = \{1,2,...,TD_i\}$. The logic expression above leads to the following set of constraints:

f) Minimum downtime generator

$$y_{i,t} + u_{i,t+j} \leq 1 \quad j = 1,...,TD_i \tag{7}$$

Generator i also has a minimum up time TU_i, once it has started up. This constraint can be derived from logical conditions: "if $z_{i,t} = 1$ then $u_{i,t+1} = 1$ and $u_{i,t+2} = 1$,..., and

$u_{i,t+ST_i} = 1$", the logic expression is given as:

$z_{i,t} \Rightarrow \underset{j \in U_i}{\wedge} u_{i,t+j}$

where $U_i = \{1,2,...,TU_i\}$. The above expression leads to the constraints:

Maximum up time generator

$$g) \quad z_{i,t} \leq u_{i,t+j} \quad j = 1,...,TU_i \tag{8}$$

h) Ramp rate limits for the continuous variables $p_{i,t}$:

$$p_{i,t} \leq p_{i,t-1} + RU_i \forall i,t$$
$$p_{i,t} \geq p_{i,t-1} - RD_i \forall i,t \tag{9}$$

i) System emission limit:

$$\sum_{i-1}^{I} \sum_{t=1}^{T} EC_i p_{i,t} \leq E^U \tag{10}$$

j) Binary variables:

$$u_{i,t}, y_{i,t}, z_{i,t} \in \{0,1\} \tag{11}$$

3. Numerical example

The data for the numerical example considered was taken from a security constrained unit commitment problem (SCUC) from (Niknam et al., 2009). The problem consists of a network with 10 generators (2 oil fired and 8 gas fired generators) that should be committed over a time horizon of 24 hours. The spinning reserve is fixed at 10% of the scaled demand load. Table 1 and Table 2 give the demand and the data of the generator units, respectively (Peak load). The MIQP problem contains 5330 Equations, 961 Variables (240 Nonlinear and 720 Discrete).

Hour	D_t (MW)
1	700
2	750
3	850
4	950
5	1000
6	1100
7	1150
8	1200
9	1300
10	1400
11	1450
12	1500
13	1400
14	1300
15	1200
16	1050
17	1000
18	1100
19	1200
20	1400
21	1300
22	1100
23	900
24	800

Table 1: Load demand

Table 2: System data

Unit	a_i ($/h)	b_i ($/MWh)	c_i ($/MW^{2}h)	EC_i (g/kWh)	p_i^U (MW)	p_i^L (MW)
1	1000	16.19	0.00048	500	455	150
2	970	17.26	0.00031	500	455	150
3	700	16.60	0.00200	500	130	20
4	680	16.50	0.00211	500	130	20
5	450	19.70	000398	500	162	25
6	370	22.26	0.00712	500	80	20
7	480	27.74	0.00079	780	85	25
8	660	25.92	0.00413	780	55	10
9	665	27.27	0.00222	780	55	10
10	670	27.79	0.00173	780	55	10

Unit	RU_i (MW/h)	RD_i (MW/h)	SU_i ($/h)	SD_i ($/h)	TU_i (h)	TD_i (h)
1	200	200	10	10	8	8
2	200	200	10	10	8	8
3	100	100	8	8	5	5
4	100	100	8	8	5	5
5	100	100	8	8	6	6
6	50	50	10	10	3	3
7	50	50	10	10	3	3
8	50	50	8	8	1	1
9	50	50	8	8	1	1
10	50	50	8	8	1	1
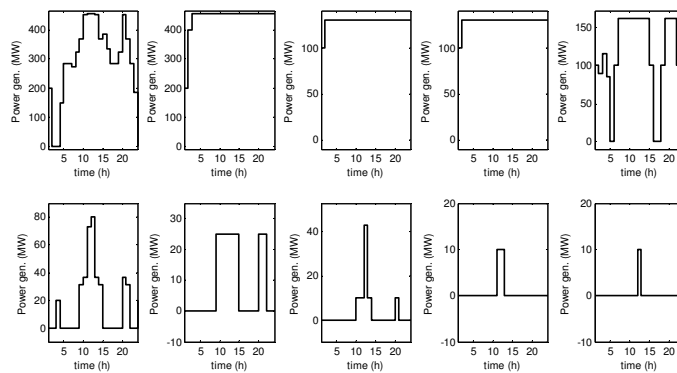
Figure 1: Optimized power production profiles for each generator

Energy optimization in the process industries: Unit Commitment at systems level

In Figure 1 the optimized power profiles for the global optimum solution with objective 578176 are shown. As can be seen the heavier generators are scheduled to their maximum capacity, while the smaller units are used to deal with the fluctuations in the demand. This result in fact corresponds to a commonly used scheduling strategy called priority list method (PLM) (Wood and Wollenberg, 1996).

4. Computational Results

We have compared the numerical results for two formulations with three MINLP solvers: DICOPT, SBB and BARON. The first model corresponds to the convex MIQCP presented in section 3. In the second model the objective function contains a product of a discrete and a continuous variable, resulting in the traditional non-convex problem.   Table 3 shows the results for the two models.

MODEL	SBB		DICOPT		BARON	
	CPU time*	*Nodes*	*CPU time**	*Maj. It.*	*CPU time**	*BaR Nodes*
Convex MINLP	3600	30566	2.7	3	3600	505
	Obj. val. 586979 (found at node 28801)		578176		Obj. val. 581427	
Non-convex MINLP	252	10340	50.4	4	3600	3289
	Obj. val. 595419 (found at node 266)		Obj. val. 598902		Obj. val. 637832	

Table 3: Comparison of convex and non-convex models with different solvers ** CPU time in seconds*

As can be seen in Table 3, for the convex case, DICOPT is the fastest in solving the requiring only 2.7 sec and finding the global optimum of 578176. In contrast, SBB and BARON are unable to find the global optimum for the convex case as they reach the time limit of 3600 sec. The best possible solution found for SBB produces a higher value of 586979, which was found in node 28801 out of a total of 30566 nodes in the branch and bound search tree. BARON obtains a slightly lower value of 581427. For the non-convex formulation, none of the three solvers finds the global optimum. SBB yields the best value of 595419, while DICOPT yields a slightly higher value of 598902. SBB requires 252 sec while DICOPT requires 50.4 sec. SBB presumably performs faster than in the convex case since the bounds in the branch and bound search are not rigorous for the non-convex case. BARON again cannot terminate the search after 3600 seconds and obtains a significantly higher value of 637832. From the results it is clear that the non-convex formulation is inferior to the convex MINLP, which is of course not surprising.

Instead of using general MINLP solvers, the convex MIQCP can also be solved with CPLEX. This results in a further reduction of computational time, as shown in Table 4. While DICOPT requires a CPU time of 2.8 sec, CPLEX solves the problem in 0.7 sec. This can in part be attributed to the small relaxation gap of the MIQP problem (1.6%). Compared to any of the SBB and BARON solvers, it is clear that solving the MIQCP with CPLEX can be solved orders of magnitudes faster. Compared to DICOPT the reduction is less dramatic. We should also note that CPLEX cannot be applied to the non-convex MIQCPs since in CPLEX the Hessian of the MIQCP is assumed to be positive definite (i.e. convex objective function).

DICOPT (MINLP)		CPLEX (MIQCP)	
CPU time*	*Iterations*	*CPU time**	*Iterations*
2.8	1287	0.7	1670
	3 Maj. It.		5 Nodes
Obj. value		Obj. value	
578176		578176	
Relaxation gap		Relaxation gap	
1.6%		1.6%	

Table 4: Comparison of the nonlinear and quadratic form of the unit commitment problem
* CPU time in seconds

5. Conclusions

We have developed a convex MIQCP model for the unit commitment problem, which traditionally has been formulated as a non-convex MINLP problem. Numerical results have been reported for a security constrained unit commitment problem consisting of a network of 10 power generators that must satisfy a time-variant demand over a horizon of 24 hours. Comparisons were presented between the convex MIQCP and the non-convex MINLP clearly showing the superiority of the former. As for the solvers, it is clear that the MIQCP capability in CPLEX performs considerably faster than general MINLP solvers SBB and BARON, while DICOPT is somewhat slower but still competitive.

References

Dochain, D. , W. Marquardt, et al. (2006). *"Monitoring and control of process and power systems: Towards new paradigms: Status report prepared by the IFAC Coordinating committee on Process and Power Systems."* Annual Reviews in Control **30(1):** 69.

Iyer, R. R. and I. E. Grossmann (1998). *"Synthesis and operational planning of utility systems for multiperiod operation."* Computers & Chemical Engineering **22(7-8):** 979.

Kazarlis, S. A., A. G. Bakirtzis, et al. (1996*). "A genetic algorithm solution to the unit commitment problem."* IEEE Transactions on Power Systems **11(1):** 83.

El Mahgary, Y. and E. Tamminen (1993). *"Relevant methods for planning and optimization of power and water production systems of the Gulf countries."* Desalination **92(1-3):** 149.

Niknam, T., A. Khodaei, et al. (2009). *"A new decomposition approach for the thermal unit commitment problem."* Applied Energy, In press.

Sasaki, H., M. Watanabe, et al. (1992). *"A solution method of unit commitment by artificial neural networks."* IEEE Transactions on Power Systems **7(3):** 974.

Raman, R. and I.E. Grossmann, "Relation Between MILP Modelling and Logical Inference for Chemical Process Synthesis," *Computers and Chemical Engineering* **15**, 73 (1991)

A.J. Wood and B.F. Wollenberg (1996), 2nd Ed. Power Generation, Operation and Control, John Wiley and Sons

Zhuang, F. and F. D. Galiana (1990). "Unit commitment by simulated annealing.**"** IEEE Transactions on Power Systems **5**(1): 311.

Fuzzy-like Optimization Approach for Design and Scheduling of Multipurpose Non-Periodic Facilities

Tânia Pinto-Varela[§1,2], Ana Paula F. D. Barbósa-Póvoa[2] and Augusto Q. Novais[1]

[1]*Unidade de Modelação e Optimização de Sistemas Energéticos (former Departamento 11de Modelação e Optimização of INETI), LNEG, Lisboa, Portugal*
[2]*Centro de Estudos de Gestão do IST, UTL, Lisboa, Portugal*

Abstract

This paper deals with the design and scheduling optimization of multipurpose non-periodic facilities, considering three different approaches to solve the conflicting objectives often present in this type of problems. Like in most facilities, this problem involves the maximization of the total revenue as well as the minimization of the total cost. The best way to deal with these two goals simultaneously is either to combine them into a single criterion (e.g., profit) or to apply a multi-objective optimization approach, although the latter may become very lengthy with the increase of the problem complexity. In this work we present an alternative optimization approach, adapted from symmetric fuzzy linear programming (SFLP), which is applied to the design and scheduling of this kind of facilities. An example is used to show the methodology application and results are compared to those located on the efficient frontier, which were obtained with the mono-objective model through the ε-constraint approach.

Keywords: Scheduling, Design, Multi-objective, Fuzzy.

1. Introduction

In multipurpose batch facilities, a wide variety of products can be produced via different processing recipes through the sharing of available resources, such as equipment, raw and intermediate materials, and utilities. The design of multipurpose batch facilities, real-world problems alike, involves multiples objectives, but most of the existing literature on their design has been centred on mono-criterion objectives (Barbosa-Povoa, 2007). There is therefore considerable scope for further investigation into the multi-objective optimization of such facilities. Dietz *et al*. (2006) presented a multi-objective cost-environment design of multiproduct batch plants. The approach used consists in coupling a stochastic algorithm, defined as a genetic algorithm, with a discrete event simulator. A multi-objective genetic algorithm was developed with a Pareto optimal ranking method. The same author proposed the problem of the optimal design of batch plants with imprecise demands using fuzzy concepts (Dietz, 2008) where the multi-objective problem was extended to take into account simultaneously maximization of the net value and two performance criteria, i.e., the production delay/advance and flexibility. Aguilar-Lasserre *et al.* (2009) addressed the problem of

---

[§]Author to whom all correspondence should be addressed. E-mail: tania.pinto@ineti.pt

batch design with imprecise demands on products amounts. The model presents an alternative of the imprecision demands treatment by introducing fuzzy concepts, embedded in a multi-objective generic algorithm. Despite the work already presented in the literature, the application in multipurpose batch facilities still remains unexplored.

In previous work the authors Pinto *el at.*, (2008b) undertook the bi-objective optimization of the design and scheduling of non-periodic multipurpose batch facilities. Contrarily to common practice, production costs and revenue were treated as two separate criteria, rather than being merged into a single one such as profit. The rationale was to generate a Pareto-optimum surface, which reflected the trade-off between these two objectives, thus allowing the decision maker to exercise his/her judgment in selecting the desired solution, while being aware of the characteristics of its neighbors. This optimization approach, while effective can become very lengthy with an increase in the problem size. For this reason, in this paper we propose to explore the application of a linear fuzzy approach to the design of multipurpose batch plants and in order to establish its relative merits, it is applied to an example reported in the authors' previous work and the results obtained are evaluated and compared to those previously identified on the Pareto surface. In addition and in order to allow a clear-cut comparison between approaches, optimization results for profit will also be estimated, which will be employed as a more generally accepted reference term.

2. Modelling framework

The design and scheduling of multipurpose non-periodic facilities is obtained through the solution of a MILP formulation. This was based on previous work (Pinto *et al.* 2008a). A discretization of the time interval is assumed, where a fixed time interval is defined. The following objective functions are considered in the model:

Revenue objective function

$$f_1 = \frac{H}{T} \sum_{r \in C_P} R_{rt} p_r \tag{1}$$

Cost objective function

$$f_2 = \left(\sum_t \sum_{k \in T_p} (\alpha_k^0 N_{kt} + \alpha_k^1 \xi_{kt}) + \sum_{r \in C_f} \left[R_{r_0} - R_{rt} \right] v_r \right) \frac{H}{T}$$

$$+ \left(\begin{array}{l} \sum_{r \in D_p} \sum_s \left(\Delta_r CC_{rs}^0 + V_r CC_{rs}^1 \right) + \sum_{r \in D_c} \sum_s \left(Ec_r CC_{rs}^0 + V_r CC_{rs}^1 \right) \\ + \sum_{r \in D_v} \sum_s \left(Et_r CC_{rs}^0 + V_r CC_{rs}^1 \right) \end{array} \right) \times CCF \tag{2}$$

The above formulation is explored using three different approaches. The first one is mono-objective, where the profit maximization is assumed, while the others correspond to two different multi-objective approaches. A first one where the ε- constraint method is applied, whereby one of the objectives functions is optimized, while the other is defined as a constraint (Pinto *et al.* 2008b) leading to a Pareto solution. The second one explores fuzzy linear programming (FLP) concepts, which might offer computational efficient alternatives to standard optimization approaches. This latter will be from now on the focus of this paper, with the aim to conclude on its effectiveness and computational cost. Such approach borrows the concept of degree of feasibility from a fuzzy approach that makes use of Zimmermann's (1978) symmetric fuzzy linear

programming (SFLP) method. Constraints are made flexible with the introduction of the concept of degree of feasibility, thus allowing multiple objective functions to be treated as fuzzy constraints and reducing the optimization problem to the maximization of the degree of feasibility of all objectives simultaneously. The key features of SFLP are (Tan 2005): (1) Crisp or non-fuzzy constraints are converted into fuzzy constraints by introducing tolerances. These modifications introduce the concept of degree of satisfaction of a constraint bounded to the interval [0, 1]; (2) An aspiration level is identified for each objective function, such that optimization entails maximizing the degree to which the objective is satisfied. This step involves identifying the best and worst values for each objective. The degree to which an objective is satisfied is also bounded to the interval [0, 1].

Objectives and constraints are treated in the same manner in SLPF, hence the use of the term symmetric. A new variable, λ, is introduced in the model serving to simultaneously modulate the degree of satisfaction of all the constraints or objectives. The SFLP is then formulated to maximize λ, which in effect expresses the global degree of satisfaction in the model.

Let us depart from the classic assumptions that all the coefficients are crisp numbers and the maximization is strictly imperative. The model presented in the previous section can be stated for the multi-objective approach as:

$$
\begin{aligned}
&Maximize\ f_m(x) && m = 1, 2; \\
s.t.\quad &g_j(x) \leq 0 && j = 1, 2, \ldots J; \\
&h_k(x) = 0 && k = 1, 2, \ldots, K; \\
&x_i^{(L)} \leq x_i \leq x_i^{(U)} && i = 1, 2, \ldots, n.
\end{aligned}
\tag{3}
$$

We now assume that the decision maker can establish an aspiration level, z, to be achieved for each objective function. The objective functions are then modeled as fuzzy sets, as follows, where f_1 and f_2 stand, respectively, for revenue and cost:

$$
\begin{aligned}
&f_1(x) \gtrsim z_1 \\
&f_2(x) \lesssim z_2 \\
s.t.\quad &g_j(x) \leq 0 && j = 1, 2, \ldots J; \\
&h_k(x) = 0 && k = 1, 2, \ldots, K; \\
&x_i^{(L)} \leq x_i \leq x_i^{(U)} && i = 1, 2, \ldots, n
\end{aligned}
\tag{4}
$$

The fuzzy multi-objective approach will be applied to construct as many fuzzy sets as there are objective functions, with the next step being the specification of the membership functions $\mu_{f1}(x)$, $\mu_{f2}(x)$ for each of them, where individual min and max preset bounds are used as calibration points. These bounds can be obtained either from known characteristics of the system and its operating environment (market, available investment, etc) or, in case these are not obvious, from preliminary exploratory tests. The membership functions:

$$
\mu_{f1}(x) = \begin{cases}
1 & if\ f_1(x) > max \\
\in [0,1] & if\ min \leq f_1(x) \leq max \\
0 & if\ f_1(x) < min
\end{cases}
\tag{5}
$$

$$\mu_{f_2}(x) = \begin{cases} 1 & \text{if } f_2(x) < min \\ \in [0,1] & \text{if } \min \leq f_2(x) \leq max \\ 0 & \text{if } f_2(x) > max \end{cases} \qquad (6)$$

The $\mu_{f_1}(x)$ should equal 0 if the lower bound for the respective objective function is strongly violated and 1 if it is very well satisfied (in the crisp sense), and should increase monotonically over the interval [*min; max= min+pi*], where p_i defines the tolerance spread. With proper adaptations an equivalent statement applies to $\mu_{f_2}(x)$, noting that now the lower bound is replaced by the upper bound and that $\mu_{f_2}(x)$ should decrease, rather than increase, over the corresponding interval. This means that we assume that the level of satisfaction for the objective function revenue f_1 rises from 0 (no production) to 1 (maximum value derived from its expected market share); and for the objective function cost the satisfaction level rises from 0 (no facilities installed) to a maximum allowed investment.

So far the objective functions and the constraints were considered crisp, but with the introduction of one new variable, λ (fuzzy degree), the crisp model becomes:

$$\text{Maximize } \lambda$$
$$s.t. \quad -\lambda p_1 + f_1(x) \leq \min$$
$$\lambda p_2 + f_2(x) \geq \max$$
$$g_j(x) \leq 0 \qquad j = 1,2,...J;$$
$$h_k(x) = 0 \qquad k = 1,2,...,K;$$
$$x_i^{(L)} \leq x_i \leq x_i^{(U)} \quad i = 1,2,...,n$$
$$(7)$$

that defines the fuzzy-like version of the previous model. The variable λ serves to modulate the objectives, where each objective must be satisfied up to a degree of at least λ. Thus, maximizing this variable effectively pushes the model towards the best compromised solution. The relative importance of the various objectives can be reflected in the model through a careful choice of the upper and lower bounds values.

3. Example

The proposed method is applied to the design of a multipurpose batch plant as presented in (Pinto *et al.* 2008b) where five products are to be produced within certain quantities in tons (S5: [0; 170]; S9 and S10: [0; 166]; S6: [0; 270] and S11: [0; 143]). Three raw materials, S1, S2 and S7, are used over the horizon of 24 h. The products S5 and S6 are both intermediate and final products. There are six main reactors (R1 to R6) available, and nine dedicated vessels. Reactors R1 and R2 may carry out two processing tasks, T1 and T2, while each storage vessel and reactors R3, R4, R5 and R6 are only dedicated to a single state/task. The problem described is solved using the three approaches mentioned above: symmetric fuzzy linear programming, ε-constraint and profit maximization.

Results for the fuzzy-like and profit maximization approaches are shown in Figure 1 (where in eq.(7) the min and max are respectively 0 and 1982), superimposed on the efficient frontier obtained with the ε- constraint approach, and identified with an arrow and point F, respectively. The points A, B, C, D and E represent points where a topology change occurs, caused by the addition of one or more main equipment units. Table 1 presents for each approach, the final products and their amounts. For simplicity

Fuzzy-like Optimization Approach for Design and Scheduling of Multipurpose Non-Periodic Facilities

for the ε-constraint approach only points C and D are shown. Analyzing the results and having been assumed that enough resources are available to satisfy the maximum demand, the point that defines the profit maximization, F, as expected corresponds to the maximum production for all products. For the new approach (fuzzy-like) the result is located between points C and D and reflect the best compromise solution, taking into account the relative importance of revenue and cost (there is an inflection in the marginal revenue), with a fuzzy degree of 0.58. Looking at the final products amounts and optimal design (Tables 1 and 2), it is possible to conclude that the topology obtained at point C (which stands through to D), is also adopted for the fuzzy solution, but with an improvement in the facility utilization, since production is increased (at the expense of S6 and also of S5) by adding further capacity to R1, R3 and R4.
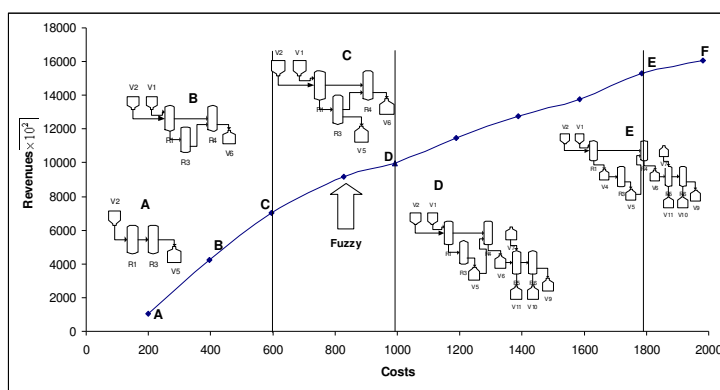
Figure 1- Optimal solution representation for the three approaches.

Table 1 – Quantities produced for each final product.

	C	Fuzzy	D	F -mono
S5	-	135	170	170
S6	258.6	270	270	270
S9	-	-	7.6	166
S10	-	-	7.6	166
S11	-	-	6.5	143

In Table 2 are presented the optimal design values for the main equipment obtained for each approach, in terms of capacities. The facility topology from the fuzzy-like and ε-constraint is the same but the fuzzy-like topology presents an improved design, since for the profit maximization presents the higher capacity able to satisfy the whole demand.

Table 2 – The optimal design for the main equipment.

	R1	R3	R4	R5	R6	V4	V5	V6	V9/V10	V11
C	103	103	129	-	-	-	52	259	-	-
Fuzzy	122	122	135	-	-	-	135	270	-	-
D	141	141	141	22	15	-	170	270	7.6	6.5
F- mono	124	124	169	119	83	169	170	449	166	143

Table 3 shows computational results obtained for the three approaches. Comparing the CPU time used, it is possible to conclude that the fuzzy approach is a very efficient approach, consuming only 52 s to obtain the best compromise solution between the two objectives.

Table 3: Computational results.

F01	F02	LP	CPU	Point
1606933	1982	120	2.17	F- mono objective
996543	991	789607	1843.45	D
919260	828	2985	52.39	Fuzzy
704237	594.6	62987	415.0	C

4. Conclusions

In this paper the design of multipurpose batch plants is presented, where different aspects are considered: plant topology, equipment design, scheduling and storage policies of multipurpose batch plants. A multi-objective problem is explored where simultaneous production maximization and cost minimization are considered. For its solution a fuzzy-like optimization approach is applied, which is compared to a multi-objective optimization using the ε-constraint method and to a mono-objective approach based on profit maximization. While the application of the ε-constraint method involves repetitive and time consuming calculations, by comparison the SFLP yields, in a very efficient manner, an optimal solution that embodies a compromise between the two conflicting objectives. A computational time reduction is observed leading to a very "economical" formulation to deal with multi-objective problems.

From a mathematical point of view, the main advantage of the explored approach, compared to the crisp problem formulation, is that it can handle objective functions with some imprecision associated to them, as reflected in the values of the tolerances. It also provides the decision maker with added insight into non-apparent trade-off solutions, thus enriching his/her decision scope. From a mere operational standpoint, this fuzzy-like approach can also be expedient to reveal, at the preliminary optimization stages, the existence of viable optimal solutions within the preset bounds.

The proposed approach can be expected to become increasingly effective as the number of simultaneous objectives increases, a line of work being currently under investigation.

References

A. A. Aguilar-Lasserre, M. A. B. Bautista, A. Ponsich and M. A. G. Huerta (2009). "An Ahp-Based Decision-Making Tool for the Solution of Multiproduct Batch Plant Design Problem under Imprecise Demand." Computers & Operations Research **36**(3): 711-736.

A. P. Barbosa-Povoa (2007). "A Critical Review on the Design and Retrofit of Batch Plants." Computers & Chemical Engineering **31**(7): 833-855.

A. Dietz, A. Aguilar-Lasserre, C. Azzaro-Pantel, L. Pibouleau, S. Domenech (2008). "A Fuzzy Multiobjective Algorithm for Multiproduct Batch Plant: Application to Protein Production." Comp. Chem, Eng **32**: 292-306.

A. Dietz, C. Azzaro-Pantel, L. Pibouleau and S. Domenech (2006). "Multiobjective Optimization for Multiproduct Batch Plant Design under Economic and Environmental Considerations." Computers & Chemical Engineering **30**(4): 599-613.

T. Pinto, A. Barbosa-Povoa and A. Q. Novais (2008a). "Design of Multipurpose Batch Plants: A Comparative Analysis between the Stn, M-Stn, and Rtn Representations and Formulations." Industrial & Engineering Chemistry Research **47**(16): 6025-6044.

T. R. Pinto, A. Barbosa-Povoa and A. Q. Novais (2008b). Multi-Objective Design of Multipurpose Batch Facilities Using Economic Assesments. 18th European Symposium on Computer Aided Process Engineering (ESCAPE-18), Lyon, France.

H. J. Zimmermann (1978). "Fuzzy Programming and Linear Programming with Several Objective Functions " Fuzzy Sets and Systems **1**: 45-55.

Superstructure optimization of Lignocellulosic Bioethanol plants

Mariano Martín, Ignacio E. Grossmann*

Chemical Engineering Department. Carnergie Mellon University.
 5000 Forbes Avenue Pittsburgh PA 15213
mmariano@andrew.cmu.edu; grossmann@cmu.edu

Abstract

In this paper we present the superstructure optimization for the production of bioethanol via gasification of lignocellulosic material, switchgrass. We describe the alternatives considered for the superstructure, which mainly include two types of gasification, different clean up processes and two different synthetic paths, catalytic and fermentation. We optimize the superstructure using a special decomposition technique. The results obtained are promising for the profitability of lignocellulosic ethanol since the energy consumed (20MW) and the production cost (0.66$/gal) are both lower than the optimized values for corn-based bioethanol plants.

Keywords: Lignocellulosic Bioethanol, MINLP model, Superstructure optimization, Energy.

1. Introduction

Bioethanol production has become the major alternative for renewable fuels given its compatibility with the current automobiles and the available infrastructure. However, the concerns regarding the production of the so called 1st generation of bioethanol from raw materials like corn or sugars [1,2], such as the extension of land needed to grow the crops, its competition with food production and supply, and the high water and energy consumption in the production process, have lead to place high expectations upon the 2nd generation of bioethanol which uses as feedstocks residues or non food dedicated crops. Nowadays, no industrial production process of this type is yet in operation. Thus, there is a great motivation to design optimized processes for producing bioethanol using lignocellulosic raw materials. In this paper, we focus on the production of bioethanol via gasification of lignocellulosic materials like switchgrass.

2. Description of supestructure

We consider a process superstructure that consists of four different parts. The first one is the gasification of the raw material. Two different technologies have been considered: (1) Indirect low pressure gasification, where the combustion of char provides the energy

for the gasification of the biomass by heating sand, which is transferred back to the gasifier, and (2) Direct high pressure gasification of the raw material with steam and oxygen, to avoid the dilution of the gas.

The gas generated must be cleaned from solids as well as other compounds like hydrocarbons, NH_3, CO_2 or H_2S also generated in the gasification. The hydrocarbons are partially removed in the tar where they are reformed with steam or are partially oxidized [3]. In the case of gasification at high pressure, solids are removed in a ceramic filter and later the gas is expanded generating energy. If the indirect lower pressure method of gasification is used, the solids are removed together with NH_3 in a wet scrubber. In both cases, the last traces of hydrocarbons are removed in a PSA system with a bed of Silica gel.

At this point the composition is adjusted to a molar ratio $CO : H_2$ of 1 [4]. In order to accomplish this objective a Water Gas Shift reactor, a bypass and a PSA for H_2 (with a bed of oxides) are considered. The split fraction selected will depend on the performance of the gasifier and the tar reformer. Reverse water shift has not been considered since the high price of hydrogen and the extreme conditions needed for the reverse water shift (T > 650 C, and conversions around 50%) makes it more profitable to sell any excess of hydrogen.

After the composition adjustment, CO_2 and H_2S are removed. The three technologies considered for this task are: (1) the absorption of the sour gases in Monoethylamine (MEA), (2) a PSA system with a bed of Zeolite 5A, and (3) the use of a membrane permeable to CO_2 using MEA as carrier. In case of using catalytic synthesis, H_2S must be completely removed from the gas due to its poisoning effect on the catalysts. In contrast, fermentation with bacteria can handle up to 2.5% in volume of H_2S.

Once the gas is purified, two synthetic paths are considered: The first is the fermentation path where the syngas is fermented in a stirred tank reactor in water [5].

The unreacted gases are recycled to the gas cleanup section of the process. The water from the water-ethanol solution must be removed to obtain fuel quality ethanol. A beer column reduces a large amount of the water in the ethanol solution. Next, four technologies are evaluated to dehydrate the ethanol to fuel grade: (1) azeotropic distillation, (2) adsorption in corn grits (3) use of molecular sieves with a bed of Zeolite 13 X [6], and (4) pervaporation. These processes can operate in parallel or sequentially to obtain fuel grade ethanol. The second possible path for the synthesis of ethanol is the high alcohols synthesis production using the optimal operating conditions by Philips et al. 2007 [4]. A catalyst based on the one used for the production of methanol is used. The light hydrocarbons and the unreacted gases are recycled back to the cleanup section of the process. The purification of ethanol is carried out using a sequence of distillation columns, either direct or indirect, for separating methanol and propanol. Figure 1 shows the superstructure.

The process is modelled as an MINLP using short-cut methods for each unit operation. We optimize the superstructure for minimum energy required for the different, composition, sour gases removal and purification methods. Later, economic evaluation of the options is used to select the most profitable process.

Superetructure optimization of lignocellulosic bioethanol plants
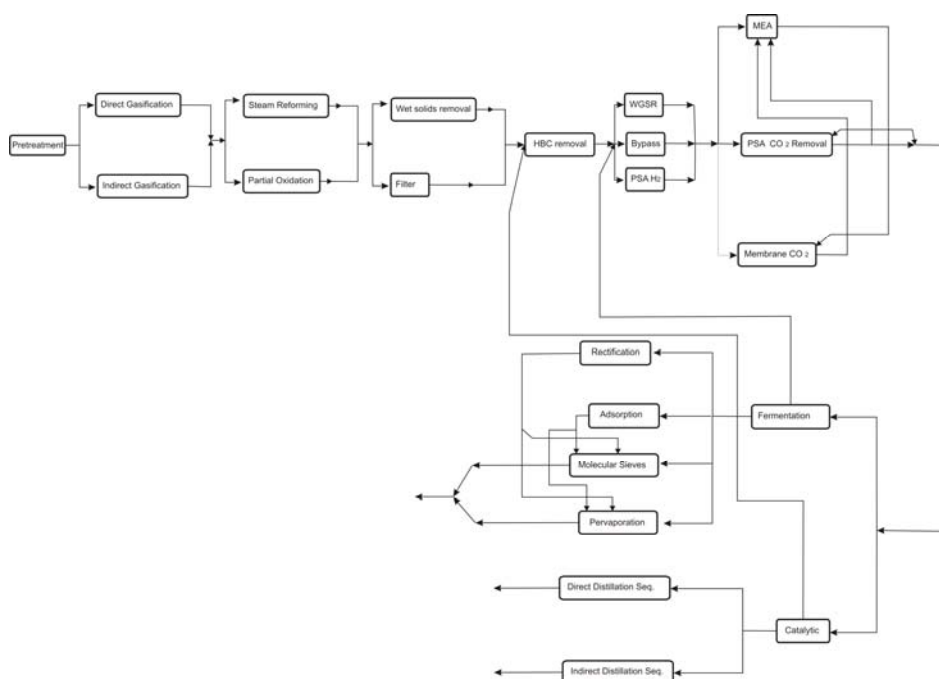
Figure 1.- Superstructure for the production of Lignocellulosic ethanol

3. Solution procedure

The sy nthesis and heat i ntegration f or t he pr oduction o f et hanol fr om li gnocellulosic raw m aterials is p erformed in two stag es. First we op timize th e st ructure of th e flowsheet based on the minimum consumption of energy. Second, we carry out the heat integration.

For the structural optimization of the problem, the MINLP m odel of the superstructure is decomposed using a hybrid search method. We first decompose the superstructure by means of a partial enumeration of alternatives in terms of gasification, reforming mode and synthetic path, creating a p artial tree. More sp ecifically, we fix the g asification technology (l ow p ressure or hi gh p ressure), t he reforming m ode (st eam or pa rtial oxidation) and the sy nthetic path (fermentation or cat alytic), which gives ri se to ei ght subproblems of the superstructure.

Each of the subproblems con sists of t hree subsystems: (1) composition adjustment for CO and H_2, (2) sour gases (CO_2 and H_2S) removal, and (3) ethanol purification, which depends on the synthetic path under consideration. The subsystems are not independent but they are linked by concentration, composition and operating pressure requirements. If not met, the next s ubsystem cannot operate. Thus, we first optimize the structure of each of the subsystems sequentially (1) to (3) based on minim um energy consumption to m eet th e composition, concen tration and o perating pressure con straints to en ter th e next subsystem. Several sol vers (MINOS, KN ITRO, CONOPT3/DICOPT) are used to

initialize the solution of each subsystem. The solution of the subsystems in sequence provides a good initial guess for the solution of the subproblem.

It turns out that the solution of the subsystems in sequence is the same as the solution to the subproblem because the operating conditions at the reactors were fixed to the current optimum, the recycles are also fixed and the operating conditions of the next subsystem (composition ratios, purity or inlet pressure) are part of the constraints of the previous subsystem.

At the end of the solution of the subproblems, we end up with eight alternatives (low or high pressure gasifiers, steam or partial oxidation reforming modes and finally fermentation or catalytic synthetic paths) where each of the subproblems yields a flowsheet with an optimal configuration of purification technologies in terms of energy consumption.

Once the flowsheets of each of the eight subproblems is determined, heat integration is performed in each of them. The heat integration consists of two steps. First, we replace the distillation columns by multieffect columns with three effects [6]. Next, we integrate the hot and the cold streams. A combined method using Synheat® and heuristics is used to match the cold and hot streams available in the process since several hot streams in the process are above the steam with the highest pressure which is the hottest utility.

Finally, an economic evaluation is performed for each of the eight subproblems accounting for the consumption of raw material, electricity, labour, chemicals and equipment. The alternative with lowest production cost is then selected.

4 Results and discussion

For a plant for producing 60 M gal/yr of ethanol from lignocellulosic raw material switchgrass, the economic evaluation, including the contribution of hydrogen as byproduct, yields a flowsheet with the lowest production cost of 81.7 MM\$/yr with an income of 41.5 MM\$/yr for sales of hydrogen, as follows:

The swichgrass is grinded to obtain a particle size of about 10 mm suitable for gasification [7]. Then, the biomass is gasified with steam and oxygen using the high pressure Renugas gasifier working at 23 bar and 893 ºC [8]. The gases generated are reformed with steam to eliminate hydrocarbons. The reforming takes place adiabatically. The gases exiting the tar reformer are cleaned up from solids (Ash, Char) in a ceramic filter at 500 ºC, expanded (where energy is obtained) and treated to remove the traces of hydrocarbons left in a PSA system at 25 ºC and 4.5 bar [9].Then, the composition of the gas is adjusted to a ratio $CO:H_2$ of 1 using a bypass combined with PSA which operates at 4.5 bar and 25 ºC [9] so that it is possible to remove the surplus of hydrogen, which can be sold to improve the economics of the process.

The gas with the correct composition of CO and H_2, is then treated to remove the sour gases, CO_2 and H_2S. The optimal solution found is a process in which the gas is treated sequentially in a PSA system operating at 4.5 bar and 25 º C, followed by the absorption of the gases in MEA at 29 º C and 29 bar [4]. The MEA solution is regenerated in a

distillation column operating at 1.7 bar and reused in the process adding a small amount of fresh MEA as makeup.

The synthetic path is the catalytic one. Thus, the gas exiting the absorption column at 29 bar and 49 ºC of the ME A system is compressed, heated and fed to the reactor. The reactor ope rates at 300 ºC and 68 bar [4]. In t he re actor, a number of products is obtained together with ethanol such as l ight hydrocarbons and al cohols. The u nreacted gases and the hydrocarbons of low molecular weight are separated from the water and the alcohols in a flash. The gases are rec ycled back to the clean up stag es wh ile th e alcohols will be condensed and separated to obtain fuel grade ethanol. In order to purify the eth anol, t he d irect d istillation sequ ence is th e op timum. Th e m ethanol is recycled back to the reactor to increase the yield

The overall yield of eth anol pr oduction is 20 % (kg eth anol/kg biomass) w ith a production cos t of 0.66 $/ gal of etha nol accounting for t he inc ome from the sales o f hydrogen. T he ener gy c onsumption i s 20MW, l ess t han half t he e nergy required t o produce ethanol from corn (46MW Karuppiah et al 2008). Figure 2 sh ows the optimal flowsheet.
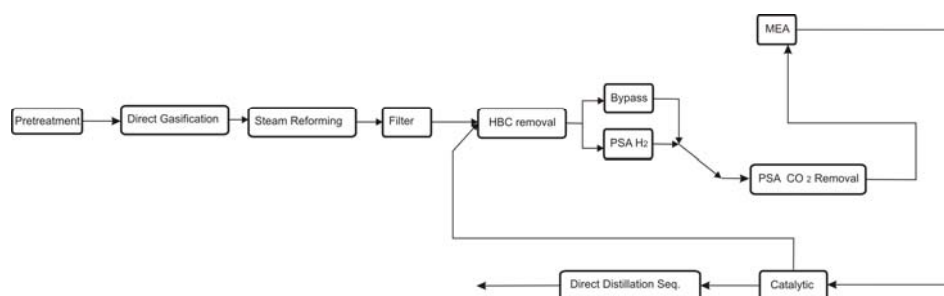
Figure 2.- Optimized superstructure for the production of lignocellulosic ethanol

5. Conclusions

We ha ve p roposed a su perstructure optimization ap proach f or t he sy nthesis o f t he production process o f et hanol fr om lignocellulosic ra w m aterials. The o ptimization of the superstructure involves the selection of different gasification technologies (low and high p ressure), ref orming m odes (steam reform ing o r partial oxi dation), com position adjustment, so ur gases removal, sy nthesis (catalytic or fe rmentation) an d p urification processes. Heat in tegration and multieffect columns are also con sidered. Th e prob lem was formulated as a n M INLP an d a new h ybrid s earch method was used t o solve the problem. Pro mising results ar e obt ained compared to corn-et hanol p roduction. For the lowest p roduction cost , t he fl owsheet i nvolves t he R enugas Gasifier wi th st eam reforming of hydrocarbons and catalytic synthesis. The energy consumption is 20MW. The production of H_2 as by product is an asset since it not only helps in the economics

of t he process re ducing t he p roduction c ost down t o 0.66$/gallon, but i t provides flexibility to th e p lant in the sen se th at t he h ydrogen can b e so ld or u sed in t he production of other fuels within the chemical complex.

Acknowledgments

References

1. Shapouri, H., Duffield, J. A., and Wang, M., 2002, USDA, Rept. No. 813. 14 p.
2. Pimentel, D., 2003, Natural Resources Research, v. 12, no. 2, p. 127–134
3. Rand, D.A.J. , Dell, R.M. (20 08) The Ro yal Society of Che mistry, Thom as Graham House, Science Park, Milton Road, Cambridge CB4 0WF, UK
4. Phillips, S. , Ad en, A ., Jech ura, J. ,and Eggeman D. T .(2007) *Technical Report,* NREL/TP-510-41168, April 2007
5. Klasson, K. T.; Ackerson, M. D., Clausen , E. C. and Ga day, j. L. (1991) Fuel, 70, 5 , 605-614
6. Karuppiah, R., Peschel,A. Grossmann, I.E., Martín, M., Martinson, W., Zullo, L. (2008) AICHE Journal. 54, 6, 1499-1525
7. Mani, S. , Tabil, L.G. ;, Sokhansanj, S. (2004) Biomass and Bioenergy, 27, 339 – 352
8. Gissy, J. Knigh t, R.A., Onisch ak, M., Cart y, R.H., Babu , S.P. (1992) U.S. Finland Biofuels Workshop II. Espoo August 24-30
9. Olofsson, I., Nordin, A., and Söderlind, U. (200 5) ISSN 1653-0 551 ETPC Repo rt 05- 02

A model for the multi-objective optimisation of a polymer electrolyte fuel cell micro-combined heat and power system

Sheila Mae C Ang[a,b], Daniel JL Brett[a] and Eric S Fraga[a]

[a]Department of Chemical Engineering, University College London (UCL), Torrington Place, London WC1E 7JE, United Kingdom, e.fraga@ucl.ac.uk

[b]Department of Chemical Engineering, University of the Philippines, Diliman, Quezon City 1101, Philippines

Abstract

This paper presents a system-level mathematical model for a polymer electrolyte fuel cell micro-combined heat and power (PEFC /μCHP) system suitable for multi-objective optimisation. The fuel cell stack model based on the previous work of the authors is integrated with the necessary sub-systems, namely fuel processing, thermal and power management, in order for it to operate as a residential heat and power generator. There is a trade-off between the net electric power output and the fuel consumption when the system is operated in a heat-led manner to deliver a particular thermal demand. The two criteria are vital in assessing the economic benefits of the technology. The aim is to determine a set of non-dominated trade-off optimal solutions, called the Pareto set, which simultaneously maximises the net power output and minimises the fuel consumption. In this work, the weighting method is used to generate the Pareto sets at different thermal demands.

Keywords: fuel cell, combined heat and power, multi-objective optimisation, modelling, design

1. Introduction

Micro-combined heat and power (μCHP) is a promising technology for residential applications. By simultaneously generating electricity and heat locally, losses due to electrical transmission and inefficiencies in distribution are diminished. μCHP technology presents a potential decrease in demand for grid electricity and heating systems based on fossil fuels, possible reduction in carbon emissions, and cost savings in the long run. Amongst the candidates for μCHP applications, fuel cells have the highest electrical efficiency and lowest emissions. However, choosing between candidates requires their evaluation *a priori* which motivates the development of rigorous mathematical models for use in computer aided design frameworks.

This work presents a system-level mathematical model for a polymer electrolyte fuel cell (PEFC) /μCHP system suitable for multi-objective optimisation. The operating strategy is heat-led where electricity is only generated when there is a demand for hot water or space heating. This approach results to a maximum utilisation of heat as it avoids generating useful heat in excess of the site requirements. The electricity can be imported from or exported to the grid at times of shortfalls or excess, respectively, in meeting the demands.

For a PEFC /μCHP system operating in a heat-led manner, there is a trade-off between the net power output and the fuel consumption. Intuitively, more fuel is needed to produce

additional power. In some cases, electricity in excess of the site requirements is generated which can be sold to the grid. However, exported electricity has a lower value than the electricity used on site. Deciding which of the two objectives – net power output or fuel consumption – is more important depends on the costs of fuel (i.e. natural gas) and electricity and the buy-back rate of electricity exported to the grid. The inherent variability in the costs introduces obscurity in deciding which operating point is most beneficial economically. Thus, information that shows the compromise between the net power output and the fuel consumption is a valuable tool in identifying the most suitable operating design for a given thermal and electrical demands.

The objective of this work is to determine a set of trade-off optimal solutions, called the non-dominated or Pareto set, that maximises the net power output and minimises the fuel consumption, with respect to design parameters and operating conditions, subject to physical constraints such as those imposed by the complex nonlinear mass and energy balances, equilibrium relations, fuel cell's electrochemical model, amongst others.

2. A model for a PEFC /μCHP system

Figure 1 shows the schematic of the modelled PEFC/μCHP system consisting of four sub-systems, namely fuel processing, fuel cell, power management and thermal management.
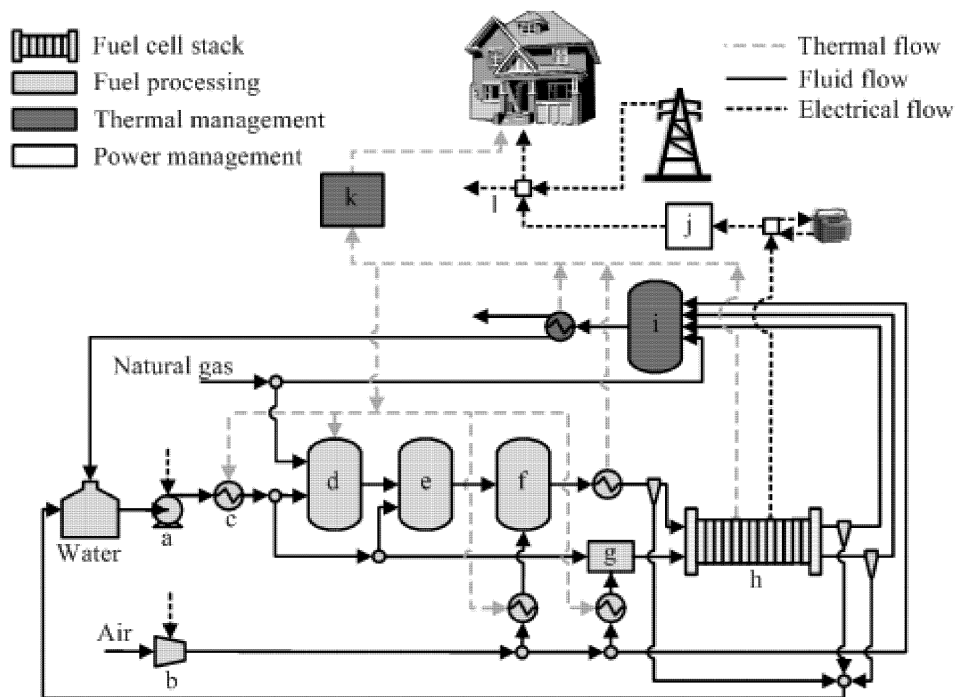
Figure 1: Schematic of the modelled PEFC /μCHP system. The letters in the diagram refer to: (a) water pump, (b) air compressor, (c) heat exchanger, (d) steam reformer, (e) water gas shift reactor, (f) PROX, (g) humidifier, (h) fuel cell stack, (i) afterburner, (j) DC/AC inverter, (k) water and space heating, and (l) to parasitic load.

A Model For The Multi-Objective Optimisation Of A Polymer Electrolyte Fuel Cell
Micro-Combined Heat And Power System

2.1. Fuel processing sub-system

The fuel processing sub-system generates a hydrogen rich gas mixture from the natural gas via steam reforming. The two main reactions taking place in the steam reformer are steam reforming and slight water gas shift reaction. The model for the steam reformer is based on the model presented by Jahn and Schroer [1], in which the steam reformer is considered as an equilibrium reactor. The correlations for the extent of reaction are only valid within a certain range of temperature and steam-to-carbon ratio, specifically [773 K, 1073 K] and [2, 5], respectively. The steam reformer model has been shown to be in good agreement with the experimental data of a demonstration plant.

As the platinum catalyst in the fuel cell is extremely prone to CO poisoning, the reformate gases are taken to the water gas shift reactor where CO reacts with additional steam producing more H_2 and CO_2. The shift reactor is also modelled as an equilibrium reactor. The equilibrium composition and temperature of the product gases are calculated from mass and energy balances and temperature-dependent correlation for the equilibrium constant for the shift reaction [2].

The CO content is further reduced in the preferential oxidation (PROX) reactor where CO is catalytically oxidised with oxygen from air. It is assumed that the PROX reactor is operating perfectly, i.e. all CO is preferentially converted to CO_2. This might be an oversimplification because depending on the type of catalyst and the operating conditions, combustion of H_2 present in the reformate gases may occur. The PROX model will be updated once a suitable data for selectivity becomes available in the literature.

2.2. Fuel cell sub-system

The fuel cell stack is the heart of the PEFC/μCHP system. It transforms the chemical potential in the hydrogen gas mixture from the fuel processing sub-system into direct current (DC) electricity. The fuel cell stack model is based on the previous work of the authors, where it has been used to investigate the trade-offs between the efficiency and the size of the system. The model takes into account the electrochemical reaction, the vapour-liquid equilibrium of water, the electro-osmotic drag transport of water, the back diffusion of water, and the chemical component balances. The detailed description and formulation of the fuel cell stack model can be found in [3].

2.3. Thermal management sub-system

The major sources of heat are the fuel cell stack and the heat exchangers used to cool down the hot reformate gases and afterburner exhaust gases. At times of high thermal and low electricity demands, the fuel processor and the fuel cell stack may be bypassed and the fuel may be fed directly into the afterburner. Part of the recovered heat is returned to the system to provide the heat needed by the steam reformer, steam generation, and pre-heating of the reactant gases. The net thermal output is used for water and space heating.

2.4. Power management sub-system

The power management sub-system coordinates the electric power produced by the fuel cell with that drawn from or delivered to the grid. The output of the fuel cell stack is a direct current (DC). A DC/AC inverter converts the DC electric power into alternating current (AC) appropriate for electrical appliances and for export to the grid. In this work, the efficiency of the inverter is assumed to be 95% [4]. Part of the generated AC electric power is used for the parasitic loads such as the air compressor and the water pump.

3. Multi-objective optimisation

The model for the PEFC /μCHP system is composed of over 500 equations (mostly non-linear and complex) representing the mass and energy balances, the equilibrium relations, the electrochemical model, amongst others. The model is used within an optimisation-based design framework with the aim of identifying the trade-offs between the net output power and the fuel consumption for any given thermal demand.

Multi-objective optimisation involves repeated evaluation of a large number of design alternatives with correspondingly high computational requirements. Our proposed model is simple enough to be solved in a reasonable time with acceptable accuracy, whilst complex enough to differentiate between alternative designs.

There are various methods for solving multi-objective optimisation problems. In this work, the weighting method is used to approximate the Pareto set. This method transforms the multi-objective optimisation problem into a single objective optimisation problem by associating each objective function with a weighting coefficient. The weighted sum of the objectives is then minimised resulting to a single solution that correspond to the weighting coefficient used. A Pareto set can be generated by evaluating a series of single-objective optimisation problems at different values of the weighting factor to avoid having to, *a priori*, select a particular weighting between objectives [3].

The multi-objective optimisation problem is formulated using the weighting method as

$$\min z = -\omega W_{\text{elec}} + (1 - \omega) W_{\text{fuel}}$$

$$\text{w.r.t.} \quad \text{design parameters and operating conditions}$$

$$\text{subject to} \quad \text{mass and energy balances}$$

$$\text{electrochemical model}$$

$$\text{equilibrium relations}$$

$$\vdots$$

$$\text{physical constraints}$$

$$\text{bounds on some of the design variables}$$

where W_{elec} is the net power output in kW, W_{fuel} is the fuel consumption in kW, z is the weighted sum of the objectives, and $\omega \in [0, 1]$ represents the weighting factors. The negative sign preceding the net power output objective denotes a maximisation problem. Single-objective optimisation problems, i.e. minimisation of the fuel consumption and maximisation of the net power output, are represented at the extreme points $\omega = 0$ and $\omega = 1$, respectively. Evaluating the optimisation problem for any $\omega \in (0, 1)$ will produce

A Model For The Multi-Objective Optimisation Of A Polymer Electrolyte Fuel Cell
Micro-Combined Heat And Power System

solutions between these extremes where both objectives are simultaneously considered. The value of ω gives the relative importance of each objective.

The optimisation model was written in GAMS modelling language and was solved using LINDOGlobal. LINDOGlobal employs branch-and-cut method to break a nonlinear programming (NLP) model down into a list of subproblems. For a given weighting factor, an optimisation run usually converges to a solution with a relative tolerance of 0.01% after approximately 300 seconds on a desktop computer with a 2.66 GHz Intel Core Duo CPU and with 2GB RAM.

4. Results and discussion
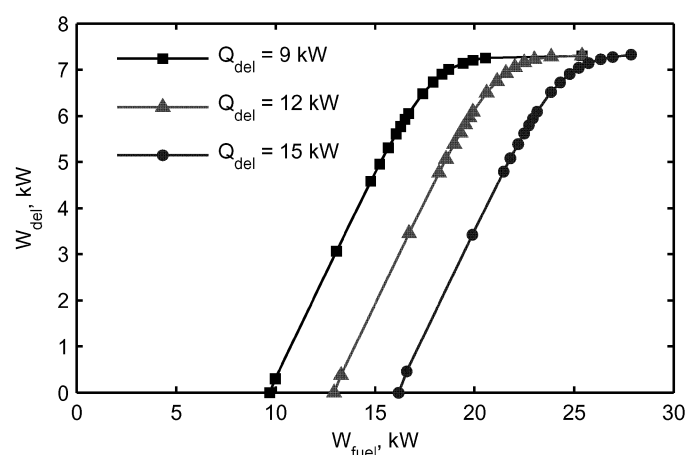
Figure 2: Pareto sets showing the trade-offs between the net power output (W_{del}) and fuel consumption (W_{fuel}) at different thermal demands (Q_{del}).

The electrical and thermal demands of a typical household are in the range of 1-3 kW and 9-15 kW, respectively [5]. In this work, the Pareto sets at different thermal demands, namely, 9, 12 and 15 kW are generated and presented in Figure 2. The highest point (top right) in each curve represents the optimal solution at $\omega = 1$, which corresponds to the single objective optimisation problem of maximising the net power output without taking the fuel consumption into account. The limit on the maximum net power output attainable can be attributed to the fixed size of the fuel cell. In this study the total active area of the membrane electrode assembly is considered to be 1.6 m². Conversely, the lowest point (bottom left) in each curve corresponds to the optimal solution at $\omega = 0$, which is the minimisation of fuel consumption regardless of the net power output. The result indicates that the lowest fuel consumption occurs when the system is operating in "boiler only" mode, i.e. the net output power is zero. In this case, the system is still generating some electric power but all of this is used to meet the parasitic demands such as the air compressor and the water pump. For the net power output below 7 kW, the net power output trades almost linearly with the fuel consumption. In this region, ~ 0.85 Watt additional net power output is produced for every Watt of extra fuel. At the net power output above 7 kW, there are no significant gains in power output with

increases in fuel consumption. Furthermore, increase in the thermal demand increases the fuel consumption, but has no substantial effect on the net power output. Finally, for a particular thermal demand, a surplus or a shortage in the net power output may result for some values of the weighting factors. Deciding whether or not this amount of power is exported to or imported from the grid will be determined by economic factors such as the costs of natural gas and electricity and the buy-back rate of electricity exported to the grid.

5. Conclusion and future directions

A system-level mathematical model for a PEFC /μCHP system suitable for multi-objective optimisation is described. Sub-systems, such as the fuel cell stack, fuel processing, thermal and power management, necessary to operate the system as a residential heat and power generator are modelled. There is a trade-off between the net power output and the fuel consumption when the system is operated in a heat-led manner to deliver a particular thermal demand. For the net power output below 7 kW, the net power output trades almost linearly with the fuel consumption, specifically ~ 0.85 Watt additional net power output is produced for every Watt of extra fuel. For some values of the weighting factors, a surplus or a shortage in the net power output may result. This may be exported to or imported from the grid, respectively.

A full economic analysis, including both operating and capital costs, will be necessary for the selection of the best trade-off. The model presented in this paper provides the base on which to develop a full economic model which would allow one to estimate the payback period for the equipment and its installation in a typical domestic scenario. This is currently under development.

6. Acknowledgement

References

[1] H. Jahn, W. Schroer, Dynamic simulation model of a steam reformer for a residential fuel cell power plant, Journal of Power Sources 150 (2005) 101–109.

[2] Y. Choi, H. Stenger, Water gas shift reaction kinetics and reactor modeling for fuel cell grade hydrogen, Journal of Power Sources 124 (2) (2003) 432 – 439.

[3] S. M. C. Ang, D. J. L. Brett, E. S. Fraga, A multi-objective optimisation model for a general PEM fuel cell system, Journal of Power Sources 195 (2010) 2754–2763.

[4] A. Hawkes, I. Staffell, D. J. L. Brett, N. Brandon, Fuel cells for micro-combined heat and power generation, Energy and Environmental Science 2 (2009) 729 – 744.

[5] Carbon Trust, Micro-CHP Accelerator: Interim Report, Tech. rep., Carbon Trust, London (November 2007).

Design, Simulation and Optimization of Polymerization Processes Using Advanced Open Architecture Software Tools

Apostolos Krallis[a], Prokopis Pladis[b], Vassilis Kanellopoulos[b], Vassilis Saliakas[b], Vassilis Touloupides [c] and Costas Kiparissides[a,b,c]

[a]PolymerS Ltd, 6th km Charilaou-Thermi Road, Thessaloniki 57001, Greece, info@polymersimulation.com
[b] Chemical Process Engineering Research Institute, 6th km Charilaou-Thermi Road, Thessaloniki 57001, Greece
[c]Department of Chemical Engineering, Aristotle University of Thessaloniki, Thessaloniki 54006, Greece

Abstract

As the polymer industry becomes more global and competitive pressures are intensifying, polymer manufacturers recognize the need for the development of advanced process simulators for polymer plants. The overall goal is to utilize powerful, flexible, adaptive design and predictive simulation tools that can follow and predict the behaviour of polymer production processes in an accurate, prompt and comprehensive way. In response to the current needs, a new generation of software packages has been developed for the simulation, design, parameter and state estimation, optimization and control of specific polymerization processes aiming at increasing plant efficiency, improving product quality and reducing the impact to environment. The new software tools provide a user-friendly interface, including an object-oriented design environment that can be accessed from the engineer's windows-based desktop environment and provide full graphical interaction and expert system guidance on how to use the program or making engineering decisions (such as selection of unit operation or physical property method). Moreover, an open-system architecture is adopted and applied to the process modeling components (PMCs) in order to be transparent to any other compatible process modeling environment (PME). Finally, recent advances regarding the development of software applications for specific polymerization systems (i.e., an LDPE high-pressure tubular reactor process and a PVC batch suspension process) are presented.

Keywords: Computer Aided Design, Polymerization Processes, Software Packages, Object Oriented Design

1. Introduction

Product quality, plant efficiency and safety can be significantly improved by the use of process models. A mathematical model that can reliably predict the behavior of a specific unit or/and process becomes a valuable tool that can be applied to all tasks of process operation. Software modularity, user friendly interfaces and computing power have increasingly opened up new opportunities for the application of advanced mathematical models in process operations. This growing computational potential has made possible the use of the same mathematical model for solving different problems of process operation hierarchy across the plant's life-cycle. Indeed, the broader use of

these models in process design, simulation, optimization and control, promises to have a profound commercial impact on the chemical and biochemical industry.

Regarding the polymer manufacturing industry both the challenges and the rewards are distinctively amplified via the application of advanced mathematical models and computer-aided process simulators. However, in contrast to the advances in CAD of chemical processes, the polymer engineer cannot always find help in the established general-purpose software packages either because the pertinent process modules are lacking completely or they are quite simplistic. The fact that each polymerization process involves a number of unique physical and chemical phenomena (e.g., reaction kinetics, physical and transport phenomena, thermodynamics, etc.) increases the scope for the development of custom-made CAD software tools for specific polymerization processes. Thus, the adoption of an open-system architecture should be a major feature of a custom-made simulator for a specific polymerization process so that the user can select or/and combine available in-house software or/and specific routines obtained from different software vendors in order to build his own process simulator.

In the present work, recent advances regarding the development of CAD software tools for two specific polymerization systems (i.e., an LDPE high-pressure tubular reactor process and a PVC batch suspension process), are presented. Our new PMCs can be used for the simulation, design, parameter and state estimation, optimization and control of polymerization processes. An object-oriented approach has been utilized so that the developed PMCs (i.e., polymerization reactor models) can be connected with other upstream or downstream equipment models (e.g., separators, heat exchangers, etc.) obtained from other sources (e.g., ASPEN). The adopted open-system architecture for PMCs largely facilitates the development of reusable components and offers additional advantages for standardization in process modeling technology.

2. Software Functionality and User Interface

In the present work, a series of advanced software packages have been developed to simulate the dynamic operation of a great number of polymerization systems. These tools provide an easy-to-use environment with a variety of interfaces including a fully interactive process flow diagram (PFD) (see Figure 1). Reaction materials, polymerization kinetics, different reactor configurations and other unit operations can be readily selected from the simulator's menu and combined together to build a desired process flow-sheet. Full access to kinetic parameters, species properties and concentrations, as well as process conditions allows the faithful description and simulation of a polymerization process.

The adopted object-oriented design and open-system architecture make it possible to simulate polymerization processes in a simple and straightforward way. Such easy-to-use validated process models can bring significant economic benefits to a polymer producer, including throughput and yield optimization, process trouble-shooting and analysis, grade transition strategies, development of new grades, process debottlenecking and design of advanced control systems.

The developed software tools support the application of model-based approaches to the design, operation, optimization and control of polymerization processes. Consequently, the model of a specific polymerization process becomes the central repository of much of the available process knowledge. This may encompass both detailed understanding of fundamental physical phenomena and empirical knowledge gained from practice that, in general, leads to a 'hybrid mechanistic model'. Ideally, such a model can predict the steady-state and dynamic behavior of a process over a

Design, Simulation and Optimization of Polymerization Processes Using Advanced
Open Architecture Software Tools

wide range of operating conditions to an acceptable degree of accuracy. The developed
polymerization models (marketed by PolymerS Ltd) are easy to use and have been
extensively validated against laboratory as well as pilot- and industrial-scale plant
measurements.

An 'open-system architecture' has been adopted that allows the export and, thus,
integration of the developed PMCs with other compliant PMEs. In Figure 1, a schematic
representation of the conceptual implementation of compatible PMCs of PolymerS Ltd,
simulating a high-pressure LDPE tubular reactor and the operation of high- and low-
pressure flash separators, within a commercial process simulator (e.g., ASPEN) is
depicted. Notice that models for the compressors can be provided by the host PME,
while the operation of the extruder can be simulated by a co-compliant PMC obtained
by a third party.

The heart of a process simulator is its user-interface environment that provides the
means for simulation of a polymerization process, ranging from a single unit to an entire
plant. This can be achieved by combining various PMCs from a database of available
unit operation models. For this purpose, a standard flow-sheeting user interface has been
developed comprising a palette of icons, each representing a different unit operation
model in the library database, and a 'white space area' used for constructing and editing
a new process flow-sheet (see Figure 1). Thus, the user can create an 'instance' of a
process unit model by dragging an icon from the palette and dropping it in the white
space. A 'unit' placed on the white space can be configured by the user via dialog
windows that allow the specification of materials, unit design parameters (e.g., reactor
and jacket geometric characteristics, controller parameters, etc.), operating conditions
(e.g., temperature, pressure, materials flow rates, etc.), kinetic, transport and
thermodynamic properties, etc. Each unit has one or more ports that allow its 'instance'
to be connected to instances of other models in the flow-sheet. Standard graphical
means are provided for effecting such connections. Once the process flow-sheet is built,
then various simulation, optimization and parameter estimation runs may be initiated
from it.
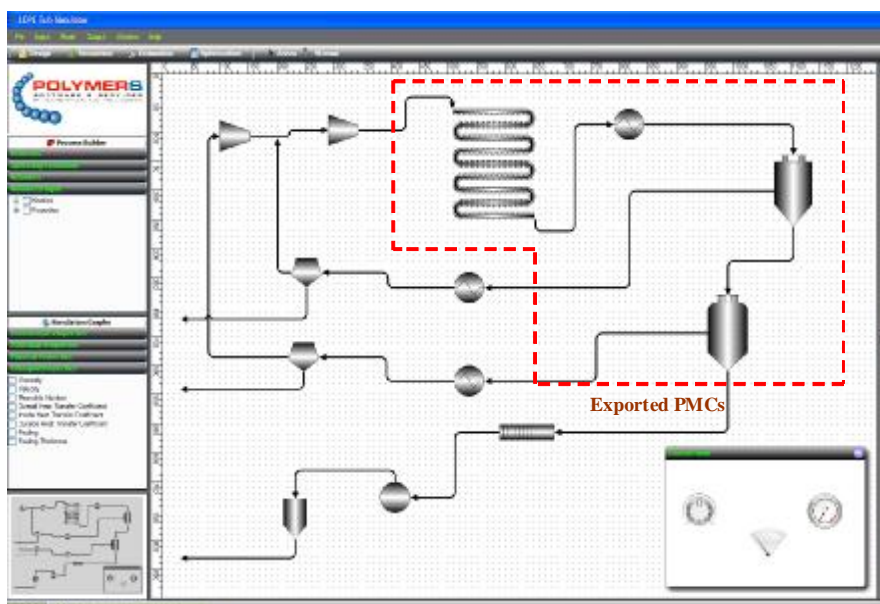
Figure 1: Screenshot of PolymerS software package.

3. Polymerization Process Models

A process simulation study involves a sequence of stages starting with the definition of the goals and is usually carried out in an iterative manner (Carido, 2009). In particular, the simulation study includes the conceptual phase (i.e., definition of the problem, possible collection of process data and definition of the conceptual model), the development phase (i.e., development of the model, simulation and model verification) and the post development actions (i.e., process optimization and control). The model implementation is usually carried out using a high-level programming language or an integrated software simulation environment.

Process modeling and simulation, process optimization and control can have a significant impact on polymer plant's operability and economics. Polymer manufacturers face increasing pressures for production cost reductions and more stringent quality requirements. However, product quality in polymer manufacturing is a much more complex issue than in more conventional simple chemical systems. Thus, a major objective of polymerization process modeling and simulation is to understand how the reaction mechanism, the physical transport properties (e.g., mass and heat transfer), reactor type and operating conditions affect the 'polymer quality' (Kiparissides, 2006). The last term includes all the polymer molecular properties (e.g., molecular weight distribution (MWD), copolymer composition distribution (CCD), long-chain branching (LCB), bivariate molecular weight – long chain branching distribution (MW-LCBD), topological properties of polymer chains, etc.) as well as the morphological properties of the product (i.e., particle size distribution (PSD), pore size distribution, bulk density, etc.). Since the end-use properties (i.e., physical, chemical, mechanical, rheological, etc.) of the polymers are directly linked with the molecular characteristics of the polymer chains, control of the polymer chain microstructure is of profound interest to the polymer manufacturing industry. This presupposes a thorough knowledge of the polymerization kinetics and the availability of advanced mathematical models to quantify the effects of process operating conditions on the molecular and morphological properties.

Over the past thirty years, a great number of mathematical models to simulate the dynamic or steady-state operation of various polymerization processes have been developed in our laboratory. Our process-specific software packages include comprehensive kinetic, thermodynamic and mass-transfer models coupled with macroscopic energy and mass balances as well as with population balance models that are numerically integrated to provide unique information on the 'polymer quality', polymer productivity, morphological polymer properties, polymer melt rheological behavior, etc. (Meimaroglou et. al., 2007; Kiparissides et. al., 2005, Krallis et. al., 2004). Linear and non-linear programming algorithms are used for the off-line and on-line parameter and state estimation in the various polymerization models. Finally, optimization schemes are implemented to determine the optimal control policies to improve polymer quality, maximize reactor throughput, minimize energy consumption or/and the amount of off-spec polymer during a grade transition.

4. Polymer Process Simulators

LDPE Tub Simulator™: The LDPE Tub Simulator™ is a powerful software tool developed for the simulation, parameter estimation, optimization and control of industrial high-pressure LDPE tubular reactor processes. It can be used either to predict the molecular properties of the ethylene homopolymer or copolymer grades (i.e., LDPE, EVA, EMA, EEA, EBA, EAA, etc.), to simulate the control moves of key operating

parameters (e.g. initiator and chain transfer agent flow rates), or even to predict the operational and product characteristics of alternative design options.

The overall goal of the software package is to provide adaptive design and predictive simulation tools that can follow and predict the operating conditions of a given high pressure polymer reactor process in an accurate, prompt and comprehensive way. Its range of use includes the prediction of molecular properties of the polymer produced, the estimation of key process variables as well as the prediction of the operational and product characteristics of alternative design options. Major points of consideration during the program development are the user friendliness of the input/output and the execution speed enabling its online use as a predictive tool (Kiparissides et. al., 2005).

The design of the reactor module is flexible enough to allow the incorporation of alternative reactor configurations, multiple injection points for monomers, initiators and solvents as well as multiple coolant streams. The simulator's output include reactor and coolant temperature profiles, reactor pressure, monomer conversion, polymer molecular properties (e.g. number, weight and Z-average molecular weights, polydispersity, long chain branching, short chain branching, vinyl and vinylidene groups, polymer density, melt index, copolymer composition), heat transfer variables (e.g. overall, inside and coolant heat transfer coefficients, fouling coefficient), transport properties (solution viscosity, mixture velocity, Reynolds number) etc. The simulation results can be compared with available experimental measurements of the LDPE plant. Figure 2 depicts a comparison between experimental measurements and simulation results on polymerization temperature. The red line represents the variation of monomer conversion with respect to the tubular reactor length.

The LDPE Tub Simulator™ incorporates as well software modules for the dynamic simulation of high-and low-pressure separation units, compressors, heat exchangers, etc., creating a process simulation tool. The process simulation studies can be used to optimize a grade or to perform grade transitions policies to minimize the undesired (off-spec) polymer product. Finally, the optimization modules are employed to find the optimum operating conditions (e.g. initiator flowrates, coolant flowrates) that maximize the reactor productivity at the desired polymer quality. Moreover, the on-line implementation modules provide the necessary tools for the on-line real-time use of the simulator package.

PVC Simulator™: Various kinetic, thermodynamic and mass-transfer models coupled with macroscopic energy and mass balances as well as with population balance models are numerically integrated to provide unique information concerning the polymer productivity, molecular and morphological polymer properties, energy process requirements, etc. A multi-phase kinetic model is employed to predict monomer conversion, polymerization rate, reaction heat, time evolution of the reactor pressure, thermodynamic properties of the reaction mixture as well as the monomer distribution in the monomer-rich, polymer-rich, gas and aqueous phases.

The model also predicts the dynamic evolution of polymer molecular properties (e.g., number and weight average molecular weights, number of short chain branches, etc). Based on macroscopic mass and energy balances, the simulator provides important information on the species concentrations and polymer temperature profiles in the reactor, the jacket and the overhead condenser as well. A detailed dynamic population balance model is used to predict the time evolution of the transient droplet size / particle size distribution. A dynamic population balance model is utilized to predict the evolution of the internal particle morphology (e.g., primary particle size distribution, average porosity, etc.) (see Figure 3). Moreover, dynamic optimization modules are

utilized in order to provide the optimal quantities of initiators, initiator addition strategies or reactor temperature profiles that produce a square polymerization rate profile capable of minimizing batch time.
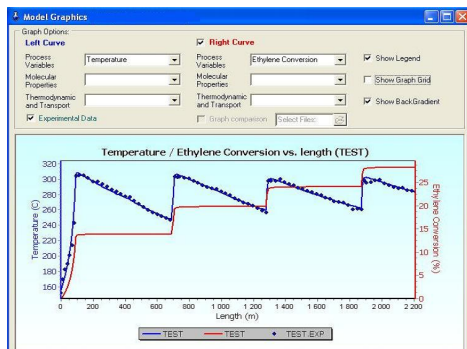
Figure 2: Comparison of model predictions with experimental data in a high-pressure LDPE tubular reactor (LDPE Tub Simulator™)
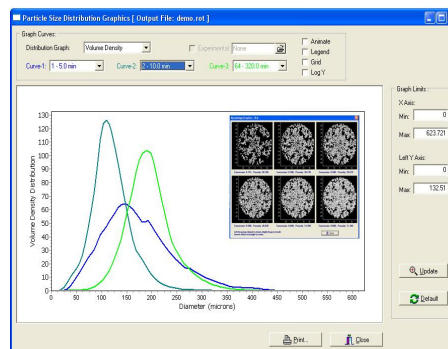
Figure 3: Prediction of PVC Internal Particle Morphology of PVC Polymer Particles using the PVC Simulator®

5. Conclusions

The present paper describes the development of an open-system platform for the computer aided design, simulation, parameter estimation and optimization of industrial polymer production processes. The utilization of an object-oriented programming environment is a major feature providing the user with the capability of selecting equipment relevant to the specific process, connecting the unit operations with other upstream and downstream equipment and, thus, building the process flow-sheet in an easy and comprehensive way. State-of-the-art mathematical models (e.g., kinetic models, thermodynamic models, mass-transfer models, population balance models, end-use properties models, etc.) capturing the peculiarities of specific polymerization processes can provide accurate information regarding polymer productivity, molecular, morphological, end-use polymer properties, etc. Finally, on-line and off-line optimization algorithms can be utilized to calculate the optimal operating conditions that can drive a polymerization process to some desired final properties specifications and, thus, gaining significant benefits on the polymer plant operability and economics.

References

J.M. Carrido (2009). Object Oriented Simulation. A Modeling and Programming Perspective, Springer Science & Business Media, New York, USA.

C. Kiparissides (2006). Chalenges in Polymerization Reactor Modeling and Optimization: A Population Balance Perspective, *Journal of Process Control*, 16, 205-224.

D. Meimaroglou, A. Krallis, V. Saliakas, C. Kiparissides (2007). Prediction of the Bivariate Molecular Weight – Long Chain Branching Distribution in Highly Branched Polymerization Systems Using Monte Carlo and Sectional Grid Methods, *Macromolecules*, 40, 2224-2234.

C. Kiparissides, C. Baltsas, E. Papadopoulos, J.P. Congalidis, J.R. Richards, M.B. Kelly, Ye Yi (2005). Mathematical Modeling of Free-Radical Ethylene Copolymerization in High-Pressure Tubular Reactors, Industrial Engineering and Chemistry Research, 44, 2592-2609.

A. Krallis, C. Kotoulas, S. Papadopoulos, C. Kiparissides, J. Bousquet, C. Bonardi (2004). A Comprehensive Kinetic Model for the Free-Radical Polymerization of Vinyl Chloride in the Presence of Monofunctional and Bifunctional Initiators, *Industrial Engineering and Chemistry Research,* 43, 6382-6399.

Optimisation of heterogeneous batch extractive distillation

Alien Arias Barreto[a], Ivonne Rodriguez Donis[a], Vincent Gerbaud[b], Xavier Joulia[b]

[a] *Instituto Superior de Tecnologías y Ciencias Aplicadas (InSTEC). Ave. Salvador Allende y Luaces. Ciudad Habana AP 6163. Cuba, Email: ivonne.rdguez@infomed.sld.cu*
[b] *Université de Toulouse, Laboratoire de Génie Chimique (LGC), CNRS, INP, UPS, 4 Allée Emile Monso, Toulouse, F-31 31432. France, Email: Vincent.gerbaud@ensiacet.fr*

Abstract

This paper considers the optimisation of batch extractive distillation, using heterogeneous entrainers for the first time. The objective function includes the maximum of overall profit and the optimisation variables are the entrainer flowrate and the reflux ratio that is an optimal combination of both decanted phases. Simulation and optimization is performed within MATLAB, by using a genetic algorithm coupled to a short-cut model of the distillation column. The performance of the optimisation scheme is illustrated through the separation of chloroform – methanol mixture with water considering either a constant or a piecewise constant policy for both optimization variables.

Keywords: batch extractive distillation, heterogeneous entrainer, genetic algorithm

1. Introduction

Batch distillation becomes irreplaceable when it is necessary to treat small quantities of materials with a great diversity in composition. Azeotropic and extractive distillation processes are the most used processes for separating azeotropic or close boiling mixture, always involving the addition of an auxiliary entrainer. Although heterogeneous entrainer has been widely used in batch azeotropic distillation, it has only been recently considered for batch heterogeneous extractive distillation (BHED) of the chloroform – methanol mixture with water [1]. Continuous feeding of the heterogeneous entrainer allows the residue curve map saddle binary heteroazeotrope to be drawn at the top of the column. Besides, unlike to the homogeneous entrainer, the heterogeneous entrainer can be fed at the column top and the process takes place with an extractive section only. The following operating steps for BHED were considered in this work: (T1) start-up of the column at infinite reflux, (T2) filling the top decanter along with continuous feeding of the entrainer F_E (T3) total reflux operation keeping the continuous feeding of the entrainer (F_E) until the unstable node is replaced by the saddle binary heterogeneous azeotropic mixture (T4) operation at a given R' together with F_E until the still is depleted of the immiscible key component (T5) separation between the remain homogeneous component and the entrainer at R'' without F_E. The top decanter is considered as a total condenser with a significant liquid hold-up. Both decanted phases can be refluxed and also supplement the entrainer feeding at the column top.

A stochastic optimisation method, genetic algorithm real-coded in MATLAB, is used along with the simulation of the BHED considering the short-cut modelling with the typical assumptions: theoretical plates, negligible pressure drop and liquid hold-up on the trays and constant molar overflow. The global optimization problem is decomposed into a series of independent single optimizations, each one related to an operating task and considering the same objective function. Preliminary parametric studies have demonstrated that the reflux ratio (R') and the entrainer flowrate (F_E) are the variables having a key incidence over the overall profit of the BHED. Mujtaba pointed that the optimal overall profit in homogeneous BED is mainly determined by assuring optimal values for F_E and R' in the task (T4) [2]. If unlimited capacity in the boiler is taken into account, optimization of task (T3) is not necessary because the operating time for setting the binary heterogeneous mixture chloroform - water at the top of the column logically decreases when F_E increases. Hence, this work is devoted to optimization of task (T4) related to the withdrawal of the heterogeneous key component to improve an overall profit function. A constant value and, also, a piecewise constant policy considering two intervals of time for F_E and R' will be considered. We solve the optimisation problem considering all combinations for F_E and R'.

2. Optimization problem formulation

2.1 Case of study: Separation of chloroform – methanol azeotropic mixture

Optimisation problem formulation concerns to the separation of azeotropic mixture chloroform – methanol which is widely used for separating bioactives substances from biological sources. Water was shown to be an effective heterogeneous entrainer [1]. Thermodynamic and topological features of the resulting ternary system are shown in Figure 1 including the univolatility curve chloroform – methanol (α_{12}). Thermodynamic calculations were done by using Simulis®Thermodynamics, a thermodynamic property server available in Microsoft Excel [3]. NRTL was chosen as thermodynamic model with literature binary coefficients [1]. As explained [1], because the univolatility line

α_{12}=1 ends at the chloroform – water edge, the saddle binary hetero-azeotrope chloroform – water can be drawn as a vapour overhead at the column if water is fed continuously at the column top generating two liquid – liquid phases into the decanter after condensation. The heavy chloroform-rich phase ($x^{II} = x_D$) can be drawn as distillate product whereas the water – rich phase (x^J) or a mixture composed by both decanted liquid phases can be refluxed toward the column top.
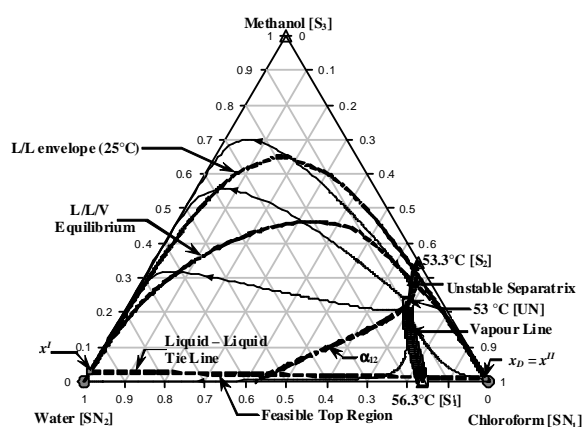
Figure 1. Chloroform – methanol - water residue curve map

Optimisation of heterogeneous batch extractive distillation

2.2 Modelling of heterogeneous BHED by short-cut model

Dynamic behaviour of BHED is described by the following differential and algebraic equations:

- Boiler mass balances: $\dfrac{dB}{dt} = F_E - D$ (1)

$$\frac{dx_B}{dt} = \frac{F_E}{B}(x_{FE} - x_B) + \frac{D}{B}(x_B - x^{II}) \qquad \text{where} \quad D = (1-\alpha)(1-\omega)V \tag{2}$$

- Extractive liquid profile inside the extractive column section (differential mode proposed by Lelkes et al. [4]; column with significant height)

$$\frac{dx_i}{dh} = \frac{V}{L}(y_i - y_i^*) \tag{3}$$

y_i^* is determined applying the modified Rachford-Rice procedure for a three-phase mixture as the vapour phase selected as a reference phase.

$$\sum_1^{c=3} \frac{z_i(K_{ij}-1)}{1 + \sum_1^{Phases=3} \beta_j(K_{ij}-1)} = 0 \tag{4}$$

β_i is the molar amount of each phase, z_i is the molar phase composition and K_{ij} is the equilibrium coefficient. The operating vapour composition y_i can be expressed in terms of R' and F_E/V as:

$$y_i = \left(\frac{R'}{R'+1} + \frac{F_E}{V}\right)x_i + \frac{1}{R'+1}x^{II} - \frac{F_E}{V}x_{FE} \tag{5}$$

Usually, BHED involves a reflux combination of the whole water-rich entrainer phase L^I along with a portion "α" of the chloroform rich-phase L^{II}. The proportion of L^I and L^{II} inside the decanter is determined by the liquid – liquid splitting ratio ω. Therefore, the total reflux liquid $L_R = L^I + \alpha L^{II}$. The reflux ratio can be determines as:

$$R' = \frac{\omega + \alpha(1-\omega)}{(1-\alpha)(1-\omega)} \tag{6}$$

2.2 Objective function and constraints

Optimization problem will deal with the maximization of the overall profit of the process and it can be formulated as following:

$$\min\{-OP\} = C_1 * D_1 + C_2 * D_2 - C_S * S - C_f *(t_{total}) \tag{7}$$

$$F_E(T_2, T_3, T_4), \alpha(T_4), R''(T_5)$$

St. $\quad x_{D\,chloroform}(T_4) \geq 0.99$

$\qquad x_{D\,methanol}(T_5) \geq 0.99$

$\qquad Re_i > 0.90$

Where Re_i is the recovery yield for component i. If one phenotype of the population doesn't fulfil any constraint, the associated OP takes the mandatory value of 10^6. C_1 (3.012 \$/mol), C_2 (0.5085 \$/mol) and C_S (0.001 \$/mol) are the prices of the products chloroform; methanol and the make-up of water, respectively. C_f (0.0027 \$/min) is the total operating cost of a real bench column. Optimisation using some market product

prices revealed a non sensitive effect for a given C_f. It was also assumed that the possible off-cut product (ternary heteroazeotrope) is not a commercial product and it can be recycled to the next batch. The optimization variables are the ratio (F_E/V) for tasks T2 T3 and T4 and the reflux policies *R'* for task T4 and *R"* for task T5. The aim of the optimization is searching an optimal value of (F_E/V) along with a comparison between a constant reflux ratio and a piecewise constant reflux policy considering two time intervals. If off-cut operation is required, reflux ratio is set at unity in order to reduce the chloroform molar content into the boiler lower than 0.01.

Genetic algorithm real-coded in MATLAB is used as optimization method. The initial population was set at 50, the selection rate is 0.8 and the mutation rate is 0.01. The optimization is stopped if no improvement of the objective function is achieved after 10 generations. Bounds for the optimization variables are: $1.4 \leq (F_E/V) \leq 2$ and $0.4 \leq \alpha \leq 0.9$ for chloroform recovery and $1 \leq R" \leq 10$ for the separation of methanol – water. Those bounds for (F_E/V), α and *R"* were taken from previous study according to the purity product constraints [1]. Optimisation by genetic algorithm comprises a large evaluation of the objective function including the solution of the dynamic model. Use of short-cut model allows a good approximation of the optimal operation conditions with less computational effort and time. The optimal results obtained by simplified method is validated by rigorous simulation using ProSim Batch for all optimal cases.

3. Results and discussion

Optimization of the BHED was performed considering several cases:
- Case I: (F_E/V), α and *R"* constant;
- Case II: constant (F_E/V) and piecewise constant α for two time intervals
- Case III: piecewise constant (F_E/V) and α for two time intervals.

Operating conditions for simulation of the separation of chloroform (1) – methanol (2) with water (3) are: initial charge (20 mol), composition charge (x_1=0.2704/ x_2=0.6714/ x_3=0.0582), decanter holdup (1 mol), vapour flow (0.016 kmol/hr) and column pressure (1.013 bar). Real mixture to be separated contains a little amount of water.

Table 1 displays the optimal values for (F_E/V), α and the respective *R'* for cases I, II and III along with the objective function for simplified model OP_{SM} and rigorous simulation OP_{RS} and the total operating time. Profitability for cases II and III are similar and they are about 24 % higher compared to case I. Case III, implementation of piecewise constant policy for F_E together with α during the withdrawal of chloroform, is the best option because chloroform can be drawn with high purity without increasing the time compared to case II and maintaining the recovery yield. But case II will be preferred for a simpler control of the BHED.

case	F_E/V	α	*R"*	time (h)	OP_{SM}	OP_{RS}
I	1.74	0.823	8.2	9.15	20.74	19.1
II	1.78	0.5751, 0.8468	7.8	8.16	25.23	21.9
III	1.67, 1.85	0.6159, 0.8866	7.7	8.3	25.86	22.04

Table 1. Summary of the optimization results

Simulation results using the short-cut model are presented in table 2, setting the constraint purities as ending criteria for the concerned task. The optimized variables provided the specified purity and recovery requirements, in particular enabling to achieve significant recovery of the products while maintaining their purity, thus

Optimisation of heterogeneous batch extractive distillation

demonstrating the interest of the BHED process. They were in agreement with the experimental validation done in ref. [1].

For all three cases, better than 0.99 molar fraction of chloroform is achieved in the heavy liquid phase into the decanter at the end of task T2 by filling the decanter while feeding the entrainer. Therefore, typical operation at total reflux and $F_E > 0$ (task T3) was not necessary. On the other hand, the off-cut operation task is almost negligible for case I. In our study, the cost of the raw material has not been included. It would lower the OP and improve the case I performance and lower the cases II and III.

Table 2. Simulation Results of each operating tasks (short-cut model)						
	case I	Rigorous simulation	case II	Rigorous simulation	case III	Rigorous simulation
time decanter filling (h)	0.13		0.13		0.13	
x_{D1} in decanter (task T2)	0.9945		0.9967		0.9981	
time of task T3 (h)	0		0		0	
switch time in task T4 (h)	-	-	0.55	*0.55*	0.685	*0.685*
time of task T4 (h)	2.02	*1.93*	1.3	*1.33*	1.57	*1.52*
x_{D1} in chloroform tank I	0.9911	*0.99*	0.9931	*0.99*	0.9981	*0.99*
chloroform recovery (%)	91.5	*99.6*	93.7	*99.8*	93.0	*99.81*
off-cut time (h)	0.02	*0*	0.25	*0*	0.18	*0*
time of task T5 (h)	7.0	*7.06*	6.7	*6.88*	6.6	*6.3*
x_{D2} in methanol tank II	0.9901	*0.99*	0.9901	*0.99*	0.99	*0.9901*
methanol recovery (%)	94.0	*95.1*	96.8	*96.0*	96.3	*96.6*

Table 2 also presents the rigorous simulation using BatchColumn® [3] with column features taken from literature [1]: assumptions of negligible liquid hold-up and pressure drop inside the column were considered. Distillation column was considered having 45 theoretical plates with the entrainer fed at the column top. Figure 2 displays the trajectory of the composition into the still for all studied cases involving steps T2, T3 and T4. The break point for changing α and F_E/V is pointed for cases II and III as well.

Start-up of the batch column is done at total reflux without F_E (task T1). The filling-up of decanter (task T2) shows identical still path for all reflux policies. The filling of the decanter (cross line in Figure 2) was performed until the boiler compositions were similar to those computed by the short-cut method taking about 0.13 hr for all cases. Then the chloroform-rich phase was drawn as distillate until its average purity was 0.99 (task T4). The operating total time for task (T4) agree well to those determined by short-cut method. In all three cases, the recovery yield of chloroform was higher 99.5% and no off-cut step was necessary. Due to this significant chloroform recovery yield, less difference of OP (approximately 15%) was obtained by rigorous simulation between case I and the other cases than for the short-cut model optimization (Table 1).

Figure 2 displays the trajectory of the still during the withdrawal of chloroform-rich phase (task T4). Good agreement was obtained between short-cut model results (continuous lines) and rigorous simulation results (symbols). Rigorous simulation of the separation of methanol – water (task T5) led to similar results of R", time, methanol purity and methanol recovery for all three cases (Table 2).

Finally, the entrainer water recovery yield was higher than 98%, with a molar purity around 0.988 retained into the still at the end of the process. These results indicating that optimal operation of BHED is mostly determined by operating variables associated to the separation of chloroform, F_E/V and the portion of chloroform rich phase (α) in the top reflux ratio R'.
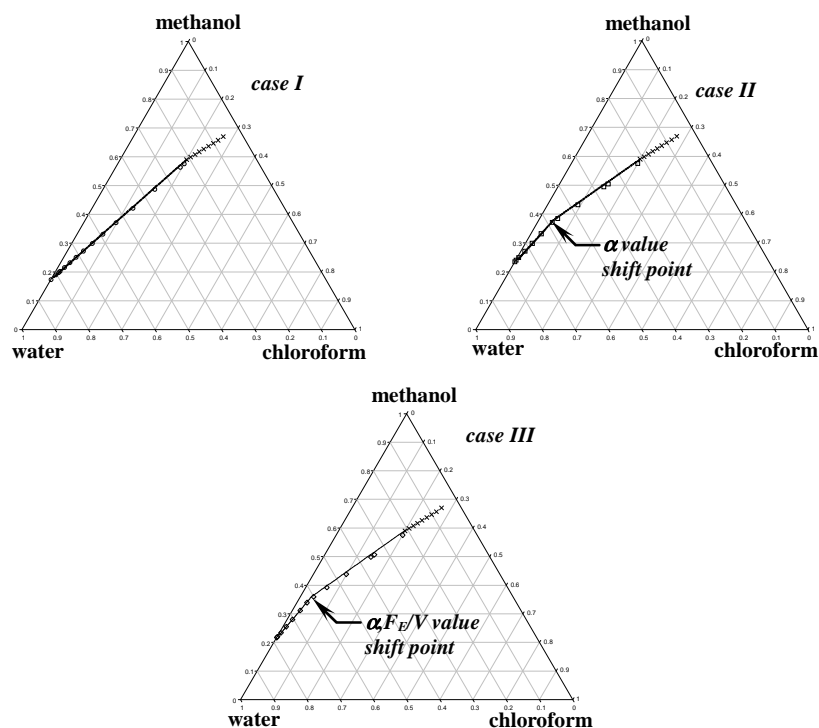
Figure 2. Comparison of still path computed by short-cut model and rigorous simulation.

4. Conclusions

Optimal operating values for reflux ratio and entrainer flowrate have been determined using the genetic algorithm implemented in MATLAB and connected to a short-cut model for simulation of batch heterogeneous extractive distillation. Three operating alternatives have been considered, combining either a constant or a piecewise variation with two time intervals for both optimization variables; the entrainer flowrate F_E/V and the amount (α) of the chloroform-rich phase returned to the column top from decanter along with the entrainer phase reflux.

The best profitability were achieved with piecewise constant F_E/V and α, with a 25% better OP than for constant F_E/V and α. Constraints of purity and recovery for both distillate products were met for all operating alternatives. Good agreement of rigorous simulation results taking the optimal conditions demonstrated the convenience of using the short-cut model along with genetic algorithm in order to accelerate the preliminary optimization analysis. Perspectives concern the incorporation of the energy balance to study the effect of the entrainer and decanter temperature.

References

[1] R.Van Kaam, I.Rodriguez-Donis, V. Gerbaud. Chem. Eng. Science, 63 (2008) 78
[2] I.M. Mujtaba. Trans IChemE, Vol. 77, Part A (1999) 588
[3] ProSim SA, 2001, www.prosim.fr
[4] Z. Lelkes, P. Lang, B. Benadda, and P. Moszkowicz, AIChE J., 44 (1998) 810

Energy Efficiency Optimization of Wastewater Treatment – Study of ATAD

Jaime Rojas*[a], Toshko Zhelev[a] and Moises Graells[b]

[a]*Charles Parsons Initiative on Energy and Sustainable Environment, University of Limerick, National Technological Park, Plassey. Co. Limerick, Ireland*
[b]*Chemical Engineering Department, Universitat Politècnica de Catalunya, Av. Diagonal 647, E-08028, Barcelona, Spain   *jaime.rojas@ul.ie*

Abstract

Autothermal thermophilic aerobic digestion (ATAD) is an activated sludge process used in wastewater treatment to stabilize and pasteurize the sludge. Due to the high oxygen uptake rates of thermophilic microorganisms, there is a relatively high energy requirement regarding the aeration of the reactors. As a result, ATAD is an energy intensive process. Given this background, several authors agree on the need to identify the optimum operating conditions (OCs) of ATAD. Nonetheless, there is no such study in the available literature. The aim of this ongoing work is to optimize the energy efficiency of ATAD systems by altering the OCs while complying with treatment goals. The selected methodology to achieve this is dynamic optimization. We have chosen two case studies: the single reactor design (2 kWh/kg) and two-reactor-in-series design (0.3-0.5 kWh/kg). Preliminary optimization results show a reduction in the energy requirement of up to 57% for the single reactor system.

Keywords: Wastewater treatment, ATAD, Energy efficiency, Dynamic optimization.

1. Introduction

Autothermal thermophilic aerobic digestion (ATAD) is an activated sludge process used in wastewater treatment to stabilize and pasteurize the sludge. In this context, stabilization refers to the reduction of the organic matter or volatile solids (VS) concentration in the sludge, while pasteurization refers to pathogen elimination via heat treatment. Several review papers are available on ATAD development, design, and operation (LaPara & Alleman, 1999; Layden et al., 2007a; Layden et al., 2007b).

The principle of the ATAD reaction can be described very simply: Raw sludge containing large amounts of organic matter and pathogens is fed into a well insulated reactor, where it is aerated and mixed for a certain time. The thermophilic microorganisms present in the sludge start to feed and grow at the expense of oxygen and organic matter, thus contributing to sludge stabilization. During their digestion, the thermophiles release vast amounts of metabolic energy, hence rising reactors' temperatures to the thermophilic range. The high temperatures are lethal for pathogens, thus contributing to sludge pasteurization. The end-product is, then, a stabilized, pasteurized sludge (also called Class A Biosolids) that can be used on agricultural land as a fertilizer without restrictions.

Due to the high oxygen uptake rates of thermophilic microorganisms, there is a relatively high energy requirement regarding the aeration of the reactors. Consequently, ATAD is energy intensive with 9-15 kWh/m^3 of treated sludge or 0.3-0.5 kWh/kg of VS treated (USEPA, 1990).

1.1. Motivation

There are experimental indications that the operating conditions (OCs) of ATAD can be significantly improved: (i) LaPara and Alleman (1999) concluded that further work is needed to determine the best way to accommodate the enormous OURs of ATAD systems. (ii) Temperature control of conventional ATAD systems is often very poor sometimes requiring heating and cooling loops (Scisson, 2003). No study has yet precisely calculated the potential economic and energy savings that could be attained by pre-heating the influent sludge. (iii) The volume change frequency has been found to affect the specific energy requirements for the removal of defined organic matter quantities (Ponti et al., 1995b). However, conventional ATAD systems make use of one single volume change per day, thus not allowing a complete exploitation of the thermophiles' efficiency (Ponti et al., 1995a).

In light of the previous considerations, several authors agree on the need to identify the optimum OCs of ATAD (LaPara & Alleman, 1999; Layden et al., 2007a).

1.2. Aim

The aim of this ongoing work is to develop optimization strategies leading to substantial reductions of the energy requirement of conventional ATAD systems by altering the OCs while complying with treatment goals.

2. Methodology

It is clear that such a problem falls within the realm of optimization. Given the discontinuous (semi-batch) and, therefore, inherently dynamic nature of ATAD, dynamic optimization is the methodology needed to solve this problem. To our knowledge there is no study in the available literature devoted to this specific problem. There are some recent review papers on dynamic optimization in the context of chemical engineering (Banga et al., 2003; Bonvin et al., 2003). The optimization problem in question can be formulated as follows:

$$\min_{\vec{u}(t),t_f} E_m[\vec{x},\vec{u}] = \frac{1}{m_{in}} \int_{t_0}^{t_f} P(\vec{u}(t)) \cdot dt \tag{1}$$

subject to

$$\vec{f}(\dot{\vec{x}},\vec{x},\vec{u}) = \vec{0} \tag{2}$$

$$\vec{x}(t_0) = \vec{x}_0 \tag{3}$$

$$0.38 - r_{VS}(t_f) \le 0 \tag{4}$$

$$1 - L_P(t_f) \le 0 \tag{5}$$

$$\vec{u}_L \le \vec{u}(t) \le \vec{u}_U \tag{6}$$

Where E_m is the gravimetric energy requirement (kWh/kg of VS treated), m_{in} the total mass of VS in the influent (kg), t_0 and t_f the initial and final time (days), respectively, P

the power of the aeration equipment (kW), $x(t)$ and $u(t)$ are the state and optimization variables, respectively, r_{VS} the VS reduction (%), and L_P the pasteurization lethality (%). In our problem, the objective function to be minimized is E_m (see Eq. 1). Sought are the optimum trajectories of the optimization variables $u(t)$ that minimize E_m while satisfying the stabilization and pasteurization constraints. Eq. 2 represents the set of differential equations describing the dynamics of the ATAD reaction which is subject to the initial value problem in Eq. 3. In other words, Eq. 2 is a model of the reaction kinetics, and for this purpose we will make use of the model presented in our previous work (Rojas et al., 2009). Eqs. 4 and 5 express the so-called path constraints and they represent the stabilization and pasteurization constraints, respectively. They ensure that a minimum required level of stabilization and pasteurization is achieved by the end of the reaction. Eq. 6 sets the lower and upper boundaries of the optimization variables.

The problem represented by Eqns. 1-6 represents a nonlinear programming problem with differential and algebraic constraints.

2.1. Case studies

Here we will consider the dynamic optimization of the most widely used ATAD designs: the single-reactor design and the two-reactor-in-series design.

2.1.1. Case study 1: Single-reactor design

This case study and its corresponding design parameters (such as reactor volume, aeration power, etc.) were extracted from Gomez et al. (2007). This design has an energy requirement of 2 kWh/kg of VS treated. The optimization variables considered in this case are the aeration flowrate, the sludge flowrate, the influent temperature, and the final time (see Eq. 7). The choice of these variables is based on our previous work (Rojas et al., 2009).

$$\vec{u}(t) = [t_l, t_f, T_{in}, q_a(t), q(t)] \tag{7}$$

Where t_l represents the loading time (days), t_f the final time (days), T_{in} the temperature of the influent sludge (°C), $q_a(t)$ the aeration flowrate (vvh), and $q(t)$ the sludge flowrate (m³/day). Figure 1 illustrates case study 1.
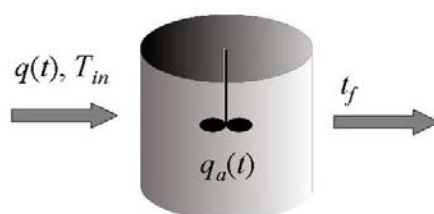
Figure 1: Single (300 m³) reactor design with selected optimization variables.

2.1.2. Case study 2: Two-reactor-in-series design

This case study (perhaps the most widespread), consisting of two equally sized 100 m³ reactors operated in series and its corresponding design parameters were extracted from USEPA (1990). This design has an energy requirement of 0.3-0.7 kWh/kg of VS treated. The choice of the optimization variables includes the sludge flowrate $q(t)$ (from the feed tank to the first stage reactor), the aeration flowrate of both reactors $q_a^i(t)$, the

influent temperature T_{in} (of the feed from the feed tank to the first reactor), and the final time of both reactors t_f (see Eq. 8). Figure 2 illustrates case study 2.

$$\vec{u}(t) = [q_a^1(t), q_a^2(t), q(t), T_{in}, t_f] \tag{8}$$
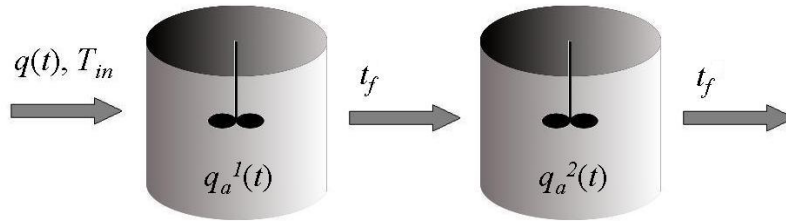
Figure 2: Two-reactor-in-series design with selected optimization variables.

2.2. Details on the Implementation

The optimization problem formulated in Eqs. 1-6 was implemented in the MATLAB® platform and it was solved using the *fmincon* function which is a deterministic method that uses the sequential quadratic programming algorithm to find a local minimum.
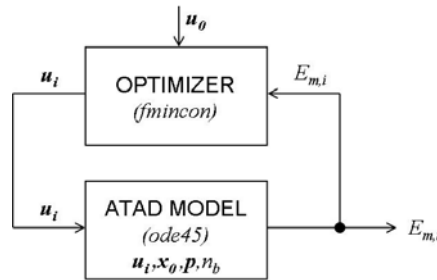
Figure 3: Optimization diagram showing different steps during each iteration.

Figure 3 describes the way how the optimization routine operates: a starting point u_0 is fed into the optimizer and later into the ATAD model which also requires an initial condition x_0 and a certain set of parameters p to produce the output E_m. This output is then used to calculate (via gradients and Hessians) the next value of the optimization variables u_i. This procedure continues until a certain tolerance criterion is satisfied.

At this point, it is relevant to say that after each batch of the ATAD reaction, only a certain reactor volume (between 10 to 20% of the total) is discharged so as to keep the high reactors' temperatures. Thus, the initial condition of the *j*-th batch depends on the final condition of the *(j-1)*-th batch. The implication of this is that it takes a number of batches to reach a cyclic or periodic operation, so that the value of the objective function and the constraints do not depend on the initial condition of the first batch. Hence, for each iteration within the optimization loop (see Fig. 2) and for a given initial condition x_0 and control vector u, the ATAD model simulates a total of n_b consecutive batches. In so doing, we ensure that the optimal solutions obtained correspond to cyclic operation and not to non-periodic states. In other words, the solutions no longer depend on the initial condition x_0 of the first batch; they only depend on the characteristics of the influent sludge and the OCs. The *i*-th value of the objective function and the path constraints are calculated based on the n_b-th batch.

This characteristic makes our problem unique and different to other dynamic optimization problems found in the literature where the optimal values depend on the

initial condition of the first batch. This is due to the fact that in other problems the whole reactor volume is discharged after each batch.

According to our experience, to obtain cyclic solutions n_b has to be higher than 10. However, due to "out of memory" problems the value 7 was used in the preliminary optimizations.

3. Preliminary Results and Discussion

Figure 4 shows the optimal trajectories of the state variables for the optimal control vector u = [0.234, 1.114, 25, 1.4, 240000] (the initial control vector was u_0 = [0.083, 1, 18, 1, 240000]). The corresponding value of the objective function is 0.86 kWh/kg, which represents a reduction of 57% regarding the reference value of 2 kWh/kg. Note that neither aeration flowrate nor sludge flowrate were discretized for this optimization (they are constant along the reaction).

Several optimization trials were run using different starting points u_0. It was found that the optimal values of the objective function differed significantly (up to 40%). This indicates that the objective function is multimodal. In other words, the objective function has several minima, and, generally, the optimization algorithm is caught in the local minimum that is closest to u_0.
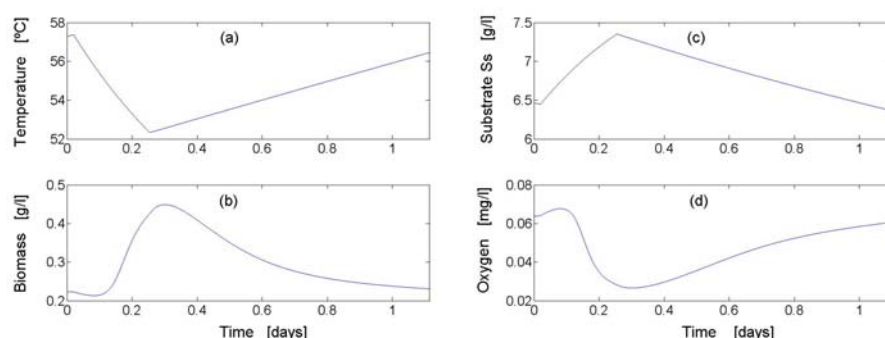
Figure 4: Optimal trajectories of state variables for case study 1: (a) reactor temperature, (b) biomass concentration, (c) substrate concentration, and (d) oxygen concentration.

4. Conclusion

The aim of this ongoing work is to develop strategies to minimize the energy requirement of ATAD systems by altering the operating conditions while complying with treatment goals. We first explained the need to optimize the ATAD reaction. Then, the dynamic optimization problem in question was formulated. The selection of the optimization variables was based on our previous work (Rojas et al., 2009). We selected two case studies: the single-reactor design and the two-reactor-in-series design, which are the most widely used designs. Preliminary optimization results show reductions in the energy requirements of 57% for case study 1.

Two important difficulties were found during the implementation of the optimization problem:

1. It was found that when optimizing the dynamics of a single batch the solutions would correspond to non-periodic states. This is due to the only partial discharge of the reactor content after each batch. Therefore, and to ensure the obtaining of

cyclic solutions, the objective function and constraints are evaluated based on the last of n_b batches. To our knowledge, this aspect of the optimization is new and different to other implementations found in the literature where the dynamics of one single batch is optimized. In our specific problem, the value of n_b should be higher than 10, which is computationally expensive and will require the use a 64-bit operating system to allow a greater allocation of RAM memory for future work.

2. After performing several optimization trials with different starting points, it was found that the optimal values of the objective function varied significantly (up to 40%). It was therefore concluded that the objective function is multimodal. Therefore, future work will be dedicated to solving the problem with stochastic algorithms or global optimization techniques to explore a wider region of the control space in search of better solutions.

5. Acknowledgements

References

J.R. Banga, E. Balsa-Canto, C.G. Moles, 2003, Dynamic optimization of bioreactors – a review, *Proc. Ind. Natl. Sci. Acad.* 69A, pp. 257-265.

D. Bonvin, S. Palanki and, B. Srinivasan, 2003, Dynamic Optimization of Batch Processes: I. Characterization of the nominal solution, *Computers and Chemical Engineering*, **27**, pp. 1–26.

J. M. Gomez, M. de Gracia, E. Ayesa, and J. L. Garcia-Heras, 2007, 'Mathematical modelling of autothermal thermophilic aerobic digesters'. *Water Res.*, vol. 41(5), pp. 959–968.

T. M. LaPara and J. E. Alleman, 1999, 'Thermophilic aerobic biological wastewater treatment.' *Water Research*, vol. 33, no. 4, pp. 895-908.

N.M. Layden, H.G. Kelly, D.S. Mavinic, R. Moles, and J. Barlet, 2007, Autothermal thermophilic aerobic digestion (ATAD) – Part I: Review of origins, design, and process operation. *Journal of Environmental Engineering and Science*, vol. 6, no. 6. pp. 665-678.

N.M. Layden, H.G. Kelly, D.S. Mavinic, R. Moles, and J. Barlet, 2007, Autothermal thermophilic aerobic digestion (ATAD) – Part II: Review of research and full-scale operating experiences. *Journal of Environmental Engineering and Science*, vol. 6, no. 6, pp. 679-690.

C. Ponti, B. Sonnleitner, and A. Fiechter, 1995, Aerobic thermophilic treatment of sewage sludge at pilot plant scale 1: operating conditions, *J. Biotechnol.* vol. 38, no. 2, pp. 173–182.

C. Ponti, B. Sonnleitner, and A. Fiechter, 1995, Aerobic thermophilic treatment of sewage sludge at pilot plant scale. 2: technical solutions and process design, *J. Biotechnol.*, vol. 38, no. 2, pp. 183-192.

J. Rojas, T. Zhelev, A. Bojarski, 2009, Modelling and sensitivity analysis of ATAD, *Computers and Chemical Engineering*, doi:10.1016/j.compchemeng.2009.11.019

J.P. Scisson, 2003, ATAD, the next generation: Design, construction, start-up and operation of the 1rst municipal 2nd generation ATAD, In *WEF/AWWA/CWEA Joint Residuals and Biosolids Management Conference and Exhibition 2003*, Baltimore, MD, February 2003.

USEPA, 1990, Environmental regulations and technology: Autothermal thermophilic aerobic digestion of municipal wastewater sludge, Technical report, United States Environmental Protection Agency, Office of Research and Development, Washington, D.C.

On the prediction of properties for diesel / biodiesel mixtures featuring new environmental considerations

Konstantinos Kalogeras,[a] Stella Bezergianni,[b] Vasiliki Kazantzi,[c] and Petros A. Pilavachi[a]

[a]*University of Western Macedonia, Department of Mechanical Engineering, 50100 Kozani, Greece*
[b]*Center for Research and Technology Hellas, Chemical Engineering Research Institute, 6th km Harilaou-Thermi Rd, 57001 Thermi-Thessaloniki, Greece,*
sbezerg@cperi.certh.gr
[c]*Technological Educational Institute of Larissa, Department of Project Management, 41110 Larissa, Greece*

Abstract

A model predicting 12 properties of diesel – biodiesel mixtures was developed. This was based on existing correlations capable of providing quality characteristics for the mixtures. The model was also used to maximize the biodiesel fraction in the diesel-biodiesel mixtures, while taking into consideration all product quality specifications as they are defined by Greek Legislation. The properties examined were density, viscosity, cloud point, pour point, volatility at temperatures 250°C, 350°C and 360°C, cetane index, cetane number, sulfur, water, higher heating value, flash point and cold filter plugging point (CFPP). The model was developed in MATLAB and the corresponding biodiesel optimization studies were carried out with the MATLAB's optimization toolbox.

Keywords: Biodiesel, mixture properties, environmental regulations

1. Introduction

In order to promote the use of biodiesel, this bio-based diesel-like fuel is blended with conventional diesel at specific mixing ratios. Particularly in Europe, the 2003/30/EU directive requests that all Member States should blend biodiesel at a 2% ratio in 2005 which will gradually rise up to 5.75% in 2010. Biodiesel is the most dominant biofuel in Europe and is currently produced in most European countries [1]. Biodiesel-diesel mixture specifications in Greece are defined by the EN 590:2004 standard, which is the same for fossil diesel.

Biodiesel has several advantages over fossil diesel [2]. Firstly the cetane number of biodiesel is generally higher than that of the conventional diesel [3], which may improve its auto-ignition ability. Also, biodiesel viscosity is normally twice as high as that of conventional diesel fuel [4], which is important for the engine lubricity. Moreover biodiesel is biodegradable as well as free from sulfur and aromatics [2,5]. Regarding emissions, commercial biodiesel significantly reduces PM exhaust emissions, which are particularly high in large cities, but slightly increases NO_x emissions [2,6]. However, there are some drawbacks from using biodiesel as an alternative diesel-type fuel. Biodiesel has lower energy content over conventional diesel, causing power decrease and increased fuel consumption [2]. A significant

problem of biodiesel is that its usage exhibits cold start problems due to the long chains that increase cloud point, pour point, cold-flow plugging point (CFPP) etc [7].

Biodiesel quality depends on the feedstock utilized (raw vegetable oil) and process conditions, rendering it a varying component for the final biodiesel-diesel blend. Blending biodiesel at increasing ratios imposes a need to estimate the quality of the final mixture based on the quality of the two blending components (i.e. biodiesel and fossil-based diesel). In this context, there are several efforts which aim to determine individual properties of biodiesel – diesel blends based on the quality of the two blending components [8-14].

This work focuses on combining all existing blending models of individual properties into a single model. This model enables the prediction of all diesel-biodiesel mixture properties for different mixing ratios as well as different diesel and biodiesel qualities. Furthermore, this model is used for estimating the maximum mixing ratio while ensuring that the final product is within specifications.

2. Methodology

MATLAB was utilized to develop the model which predicts the diesel – biodiesel mixture properties. The fossil diesel and biodiesel properties as well as the biodiesel fraction x in the diesel-biodiesel mixture are the model inputs. The blending models f_1, f_2, ... utilized involved 12 properties. Some of the blending models are given in the literature, while some other models depend on the following simple blending equation:

$$prop_A^m = (1-x) \cdot prop_A^f + x \cdot prop_A^b \qquad (1)$$

where $prop_A^f$ is the value of property A for fossil diesel, $prop_A^b$ is the value of property A for biodiesel, x is the mixing ratio of biodiesel to diesel, and $prop_A^m$ is the predicted value of property A for the diesel-biodiesel mixture. The sources of the blending models employed for each property are given in Table 1.

Table 1. Source of the blending model used for each property

Property	Blending model used
Density	Eq. (1)
Viscosity	[11]
Cloud Point	[12]
Pour Point	[12]
Distillation/volatility	[15]
Cetane Index	Eq. (1)
Cetane Number	Eq. (1)
Sulfur	Eq. (1)
Flash Point	[8]
HHV	[8]
Water	Eq. (1)
CFPP	[12]

Inserting the diesel – biodiesel mixture specifications as constraints, this new model is able to predict the maximum biodiesel fraction which does not violate the specifications. This procedure is described in Figure 1.
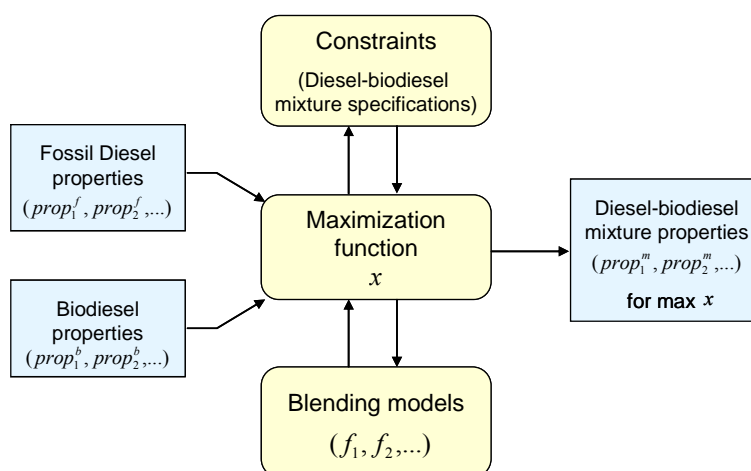
Figure 1. Model predicting the maximum biodiesel fraction in the mixture

3. Results

The above model is employed to predict the properties of diesel – biodiesel mixtures of different mixing ratios and to identify the maximum biodiesel mixing ratio that does not violate the product specifications of the produced mixture. Here, typical diesel and biodiesel fuels were considered and their properties are given in Table 2.

3.1. Predicting properties of diesel-biodiesel mixtures for different mixing ratios

For this study, four mixtures were considered, where normal diesel was blended with a typical biodiesel at the mixing ratios of 4, 6, 10 and 20%v/v. The mixing models described in section 2 were utilized to determine all associated properties of the different diesel-biodiesel mixtures. The properties of these four mixtures, denoted as B4, B6, B10 and B20, are given in Table 2.

The predicted properties of the four mixtures gave different results in terms of violations of properties' specifications. The diesel-biodiesel mixture density for the first three mixing ratios was within limits while it exceeded the upper limit in the case of the highest mixing ratio (20% v/v). This is expected as biodiesel density is significantly higher than diesel density and higher than the upper specification limit, causing problems at higher mixing ratios.

Volatility is another important property indicating the boiling point range and indirectly the size of molecules. In some cases, this falls outside the specification limits. Volatility at 360⁰C is lower than the minimum specification limit for all four mixing ratios as the biodiesel volatility at 360⁰C is significantly lower than the minimum specification limit, as a large proportion of biodiesel molecules have a boiling point which exceeds 360⁰C. Another problem is observed for the volatility at 350⁰C and particularly at higher biodiesel mixing ratios. This is also due to the fact that biodiesel molecules are bigger and therefore have a high boiling point. As the volatility at 350⁰C is 92.81% for diesel and 35.00% for biodiesel, the mixture volatility is between 92.81% and 35.00%, and is particularly closer to the higher end at small biodiesel mixing ratios, while it is closer to the lower end at the high biodiesel mixing ratios. As the biodiesel mixing ratio increases, the volatility at 350⁰C will eventually go below the threshold. However, no

problems appear for the volatility at 250⁰C as the upper specification limit is much higher than that of both diesel and biodiesel.

Table 2. Properties' prediction of diesel – biodiesel mixtures for different mixing ratios

	Properties	Units	Normal diesel	Bio diesel	Diesel-Biodiesel Blends				EN 590:2004	
					B4	B6	B10	B20	Min	Max
1	Density	g/cm^3	0,8368	0,8833	0,8387	0,8396	0,8414	0,8461	0,82	0,845
2	Viscosity	mm^2/s	2,7109	4,4700	2,7813	2,8164	2,8868	3,0627	2	4,5
3	Cloud Point	$°C$	-5,00	-6,00	-5,04	-5,06	-5,10	-5,20		
4	Pour Point	$°C$	-21,00	-7,00	-20,31	-19,96	-19,26	-17,58		
5	Volatility 250$°C$	%v/v	22,14	1,15	21,30	20,88	20,04	17,95		65
	Volatility 350$°C$	%v/v	92,81	35,00	90,50	89,34	87,02	81,23	85	
	Volatility 360$°C$	%v/v	96,07	41,67	93,90	92,81	90,64	85,21	95	
6	Cetane Index		54,57	46,17	54,31	54,19	53,94	53,27	46	
7	Cetane Number		50,00	53,00	50,12	50,18	50,30	50,60	51	
8	Sulfur	ppm	12,00	1,00	11,56	11,34	10,90	9,80		10
9	Flash Point	$°C$	135,94	176,37	137,56	138,36	139,98	144,02	55	
10	HHV	MJ/kg	34,97	35,82	35,01	35,03	35,06	35,14		
11	Water	ppm	0,50	281,00	11,72	17,33	28,55	56,60		200
12	CFPP	$°C$	-6,00	-13,00	-6,32	-6,47	-6,76	-7,48		-5/+5

*Grey cells indicate the properties that are out of the specification limits

Cetane number is another property which is outside the specification limits for all mixing ratios. This is due to the conventional diesel cetane number being below the minimum specification limit, rendering the cetane number of the mixture lower as well. The concentration of sulfur is also problematic due to the diesel sulfur concentration being above the maximum specification limit. However, as the biodiesel is mixed at higher ratios, it lowers the total sulfur concentration and at the highest ratio (20% v/v) brings it under the maximum level.

All the other properties are within the specification levels. This is due to the fact that for all these properties, both diesel and biodiesel values are well within the designated specification levels, rendering the associated properties of the mixtures also within specification.

3.2. Maximimizing biodiesel mixing ratio

For this study, the same diesel and biodiesel fuels were used and now every property was examined separately in an effort to determine the maximum ratio of biodiesel which maintains this property within specification limits, while estimating the remaining properties.

The optimization methodology described in section 2 was employed and the results are summarized in Table 3. The optimization aimed to identify the effect of maximizing the biodiesel mixing ratio considering a single specification bound each time. Each column in Table 3 represents a different maximization run conducted, considering only a single limitation, while providing the remaining properties for that maximum mixing ratio. For example, in the case of the first optimization run, the biodiesel mixing ratio was maximized until the density reached the corresponding upper bound (0.845 g/cm^3) as shown in the column "Density". For this case, the maximum mixing ratio achieved was 17,634% of biodiesel. For this mixing ratio, three other properties violated their specifications: volatility at 350⁰C (82.60% v/v); volatility at 360⁰C (86.49% v/v); and

On the prediction of properties for diesel / biodiesel mixtures

cetane number (50.53). The fact that as density reached its maximum limit, the volatilities at 350ºC and 360ºC were already outside specifications is due to the fact that biodiesel is the heavier compound with very low volatilities (see Table 2). Thus, at high mixing ratios biodiesel renders volatility below its minimum levels. Cetane number violation, however, is attributed to the cetane number of diesel being well below the minimum specification limit, not allowing the overall cetane number of the mixture to be within specification.

Table 3. Diesel-biodiesel property predictions when maximizing biodiesel mixing ratio while considering individual property specification constraints

Properties		Properties individually examined										
		Density	Viscosity	Volatility 250°C	Volatility 350°C	Volatility 360°C	Cetane Index	Cetane Number	Sulfur	Flash Point	Water	CFPP
1	Density	0.845	0.8833	0.8833	0.8431	0.8377	0.8833	0.8523	0.8833	0.8833	0.8699	0.8833
2	Viscosity	3.0211	4.4700	4.4700	2.9482	2.7456	4.4700	3.2973	4.4700	4.4700	3.962	4.4700
3	Cloud Point	-5.17	-6.00	-6.00	-5.13	-5.02	-6.00	-5.33	-6.00	-6.00	-5.71	-6.00
4	Pour Point	-17.97	-7.00	-7.00	-18.66	-20.66	-7.00	-15.49	-7.00	-7.00	-10.35	-7.00
5	Volatility 250°C	18.44	1.15	1.15	19.31	21.73	1.15	15.15	1.15	1.15	7.21	1.15
	Volatility 350°C	82.60	35.00	35.00	85.00	91.67	35.00	73.51	35.00	35.00	51.62	35.00
	Volatility 360°C	86.49	41.67	41.67	88.74	95.00	41.67	77.96	41.67	41.67	57.44	41.67
6	Cetane Index	53.41	46.17	46.17	53.69	54.44	46.17	52.60	46.17	46.17	50.07	46.17
7	Cetane Number	50.53	53.00	53.00	50.40	50.06	53.00	51.00	53.00	53.00	52.13	53.00
8	Sulfur	0.75	1.00	1.00	0.74	0.71	1.00	0.80	1.00	1.00	0.91	1.00
9	Flash Point	143.07	176.37	176.37	141.39	136.74	176.37	149.41	176.37	176.37	164.69	176.37
10	HHV	35.12	35.82	35.82	35.09	34.99	35.82	35.26	35.82	35.82	35.58	35.82
11	Water	49.96	281.00	281.00	38.34	6.03	281.00	94.00	281.00	281.00	200.00	281.00
12	CFPP	-7.31	-13.00	-13.00	-7.01	-6.16	-13.00	-8.43	-13.00	-13.00	-11.12	-13.00
Biodiesel ratio (%)		17.63	100.00	100.00	13.49	1.97	100.00	33.33	100.00	100.00	71.12	100.00

* Grey cells indicate the properties that are out of the specification limits
** The underlined properties are the properties which are examined each time

For volatilities at 350ºC and at 360ºC, the maximum biodiesel ratios reached were 13.49% and 1.97% respectively. In both cases, the cetane number was below the minimum specification limit as diesel has inherently a low cetane number. However, in the case of the volatility at 350ºC, another property that violated its bound was the volatility at 360ºC, as the corresponding value of biodiesel was too low.

When considering only cetane number limitations, the maximum ratio obtained was 33.33%v/v. At this mixing ratio, however, three other properties were violated: density (0.8523 *g/cm^3*); volatility at 350ºC (73.51% v/v); and volatility at 360ºC (77.96% v/v). These three violations are due to the fact that biodiesel is the heavier compound and in such high mixing ratios this will cause the violation of all properties associated with it (such as density and volatility).

Furthermore, the water content upper limit is reached at a biodiesel mixing ratio of 71.12% v/v. At that high mixing ratio, the same three properties (density and volatilities at 350ºC and 360ºC) violate their specifications due to the same reasons as described for the case of the cetane number.

When all the other properties (viscosity, volatility at 250ºC, cetane index, sulfur, flash point and CFPP) are considered individually, biodiesel can be mixed at any ratio, as these properties are well within the specifications for both diesel and biodiesel. However, in all cases some other properties are violated, which are the three properties described above (density and volatilities at 350ºC and 360ºC) as well as water content.

If cetane number violations are ignored, as cetane number additives are always upgrading diesel fuel, and considering all property specification limitations, the

maximum mixing ratio of biodiesel allowed is 1.97% v/v, at which the volatility at 360°C lower limit is reached.

4. Conclusions

A model was developed for predicting 12 properties of diesel-biodiesel mixtures based on the properties of each compound and the mixing ratio. The model was developed in MATLAB and it utilized blending correlations found in literature. This was demonstrated for a normal diesel and a conventional biodiesel. The model was employed for predicting the properties of mixtures with different mixing ratios. The property predictions were consistent with the expected quality of the mixtures and with the effect of increasing the biodiesel mixing ratio. The same model was also utilized as the basis for the constrained optimization in maximizing the biodiesel mixing ratio. This optimization methodology was able to identify the maximum biodiesel mixing ratio of 1.97% v/v for which all property specifications are maintained within limits.

References

[1] EurObserv'ER. Biofuels Barometer. Le journal des énergies renouvelables. No. 185. June 2008

[2] Demirbas. A.. Progress and recent trends in biodiesel fuels. Energy Conversion and Management. 50(1). pp. 14-34. 2009

[3] Bala. B.K.. Studies on biodiesels from transformation of vegetable oils for diesel engines. Energy Educ Sci Technol. 15. pp. 1–43. 2005

[4] Balat. M.. Bio-diesel from vegetable oils via transesterification in supercritical ethanol. Energy Edu Sci Technol. 16. pp. 45–52. 2005

[5] Shay. E.G.. Diesel fuel from vegetable oils: status and opportunities. Biomass Bioenergy 4. pp. 227–242. 1993

[6] Knothe. G.. Sharp. C.A. and Ryan. T.W.. Exhaust emissions of biodiesel. petrodiesel. neat methyl esters. and alkanes in a new technology engine. Energy Fuels 20. pp. 403–408. 2006

[7] Demirbas. A.. Importance of biodiesel as transportation fuel. Energy Policy. 35. pp. 4661–4670. 2007

[8] Demirbas. A.. Relationships derived from physical properties of vegetable oil and biodiesel fuels. Fuel. 87(8-9). pp. 1743-1748. 2008

[9] Demirbas. A.. Mathematical Relationships Derived from Biodiesel Fuels. Energy Sources. Part A. 30. pp.56–69. 2008

[10] http://awakeatthewheel.net. Cold weather biodiesel: Royal Turf Toe. Cloud Point and CFPP. 22/12/2006

[11] Alptekin. E.. Canakci. M.. Determination of the density and the viscosities of biodiesel-diesel fuel blends. Renewable Energy. 33(12). pp. 2623-2630. 2008

[12] Semwal. P. B.. Varshney. R. G.. Predictions of pour, cloud and cold filter plugging point for future diesel fuels with application to diesel blending models. Fuel. 74(3). pp. 437-444. 1995

[13] Sharp. C.A.. Emissions and lubricity evaluation of rapeseed derived biodiesel fuels-Final report. Prepared for Montana department of environmental quality. Nov. 1996

[14] Pasadakis. N.. Sourligas. S.. Foteinopoulos. Ch.. Prediction of the distillation profile and cold properties of diesel fuels using mid-IR spectroscopy and neural networks. Fuel. 85(7-8). pp. 1131-1137. 2006

Factors affecting the acid pretreatment of lignocellulosic biomass: Batch and continuous process

Teresa López-Arenas,[a] Punit Rathi,[b] Edgar Ramírez-Jiménez,[c] Mauricio Sales-Cruz [a]

[a]*Departamento de Procesos y Tecnología, Universidad Autónoma Metropolitana-Cuajimalapa, Artificios 40, 01120 Mexico D.F., Mexico, E-mail: mtlopez@correo.cua.uam.mx, asales@correo.cua.uam.mx*
[b]*Department of Chemical Engineering, Indian Institute of Technology Bombay, Powai, 400076 Mumbai, India, E-mail: punitrathi@iitb.ac.in*
[c]*Departamento de Ingeniería Química Petrolera, ESIQIE-Instituto Politécnico Nacional, Zacatenco, 07738 Mexico D.F., Mexico, E-mail: eramirezj@ipn.mx*

Abstract

In the acid pretreatment of lignocellulosic biomass, a first drawback that must be solved is the removal of lignin and hemicellulose through hydrolysis reactions. In this work, the influence of several operating conditions, as well as the reactor design (batch or continuous), for the dilute acid pretreatment is analyzed. Rigorous kinetic modeling is considered and simulation tools are employed. The results give guidelines for improving the process efficiency (i.e. high conversions and low reaction times).

Keywords: Pretreatment, lignocellulosic biomass, steady state, dynamic simulation

1. Introduction

Lignocellulosic biomass, such as corn stover and sugarcane bagasse, is a domestic feedstock that has potential to produce considerable quantities of bioethanol and other bioenergy and biobased products. Processing of lignocellulosic biomass to ethanol consists of four major operations: pretreatment, enzymatic hydrolysis, fermentation and ethanol separation/purification. To implement successfully the bioethanol production process, the first drawback that must be solved is the efficient removal of lignin and hemicellulose through a pretreatment process, but considering a low cost. Recently it has been demonstrated that the dilute acid pre-hydrolysis can achieve high reactions rates in short time and significantly improve cellulose hydrolysis [1]. However, pretreatment operating conditions must be tailored to the specific chemical and structural composition of the various sources of biomass.

Despite continuing interest in the kinetic mechanism of solid-phase acid-catalyzed hydrolysis for several type of biomass, little attention has been given to incorporating those kinetic models into the plant process modeling. For the one side, most of the techno-economic studies based on process simulations for bioethanol production from lignocellulosic biomass (e.g. [2-4]) take into account the hydrolysis reactions of hemicellulose (polymers) to produce mainly sugar monomers (glucose, xylose, arabinose, mannose) and acid-soluble lignin, but using conversion fractions at fixed operating conditions. On the other side, from batch kinetic studies [5-7], it has been revealed that the main factors affecting the acid pretreatment are the type of biomass,

the type of acid, the feed acid concentration, the reaction time and the reaction temperature. So that kinetic modeling and the operating conditions of the pretreatment unit play an important role in the design, development, and operation of the complete process of bioethanol production.

In this work, a systematic study of the reactor design and of the selection of the operating conditions for the dilute acid pretreatment of lignocellulosic biomass is presented, using modeling and simulation tools to improve the process efficiency. As case study, the pretreatment of corn stover in dilute sulfuric acid is considered. Numerical simulations are performed to analyze the influence of several factors: (a) the reactor design: batch or continuous, (b) the feed concentration of the dilute sulfuric acid, (c) the reaction temperature, (d) the reaction time, (e) the fraction of solids in the feed stream, and (e) the kinetic model. Static and dynamic simulations were performed using Aspen Plus and Matlab. The results are presented in terms of yields of fermentable sugars and reaction times, defining operating ranges for an efficient (batch or continuous) process. The methodology also gives guidelines to: improve experimental design, optimize the operating conditions and control the pretreatment process; reducing costs and effort of research and development.

2. Problem Statement

2.1. Process description

A (batch/continuous) reactor for the lignocellulosic biomass pretreatment using dilute sulfuric acid and high temperature is considered, where hydrolysis reactions are carried out. Corn stover is selected as lignocellulosic biomass, whose composition is (%w/w): 37.4% glucan, 21.1% xylan, 18% Lignin, 2.9 % arabinan, 2.0% galactan, 1.6% mannan, 15% moisture, 5.2% ash, 2.9% acetate, 3.1% protein, 4.7% extractives, and 1.1% unknown soluble solids. The reactor design basis was taken according the proposal of Aden et al. (2002): acid concentration = 1.1%, residence time = 2 min, temperature = 190°C, solids concentration = 42%.

Table 1. Reactions and conversion fractions (X).

No.	Reaction				X
1	$(Xylan)_n$	+	$n\ H_2O$	$\rightarrow$ n Xylose	0.90
2	$(Xylan)_n$	+	$m\ H_2O$	$\rightarrow$ m Xylose oligomer	0.025
3	$(Xylan)_n$			$\rightarrow$ n Furfural + $2n\ H_2O$	0.05
4	$(Glucan)_n$	+	$n\ H_2O$	$\rightarrow$ n Glucose	0.07
5	$(Glucan)_n$	+	$m\ H_2O$	$\rightarrow$ n Glucose oligomer	0.007
6	$(Glucan)_n$	+ ½	$n\ H_2O$	$\rightarrow$ ½ n Cellobiose	0.007
7	$(Arabinan)_n$	+	$n\ H_2O$	$\rightarrow$ n Arabinose	0.90
8	$(Arabinan)_n$	+	$m\ H_2O$	$\rightarrow$ m Arabinose oligomer	0.025
9	$(Arabinan)_n$			$\rightarrow$ n Furfural+ $2n\ H_2O$	0.05
10	Acetate			$\rightarrow$ Acetic acid	1.0
11	$(Lignin)_n$			$\rightarrow$ n soluble Lignin	0.05

The main reactions [2] are given in Table 1, where mannan and galactan have not included but they are assumed to have the same reactions and conversions as arabinan. It can be seen that most of the hemicellulose portion is converted to soluble sugars (primarily xylose, mannose, arabinose, and galactose). Glucan in the hemicellulose and a small portion of the cellulose are converted to glucose. Moreover the reactor operating

Factors affecting the acid pretreatment of lignocellulosic biomass: Batch and continuous process

conditions also solubilize some of the lignin in the feedstock. In addition, acetic acid is liberated from the hemicellulose hydrolysis. Degradation products of pentose sugars and hexose sugars (primarily furfural) are also formed.

2.2. Kinetic mechanism

Several models for acid hydrolysis in batch reactors have been proposed in the literature [5-8]. Due to the difficulty in finding a strict mechanism for hydrolysis reactions, it is usual to use simplified models to determine the kinetics of the hydrolysis of lignocellulosic materials. The simplest and widely used model involves a series of pseudo-homogeneous irreversible first-order reactions from solid polymer (i.e. A = xylan, glucan, arabinan, mannan, galactan) to aqueous monomer (i.e. B = xylose, glucose, arabinose, mannose, and galactose), and then onto decomposition products, Eq. (1a), where the kinetic rate constants k_i (i =1, 2) are given according Eq. (1b):

$$A_{polymer\,(biomass)} \xrightarrow{k_1} B_{monomer} \xrightarrow{k_2} D_{Decomposition\,products}, \qquad k_i = C_{acid}^n A_i e^{-E_i/RT} \qquad (1a, b)$$
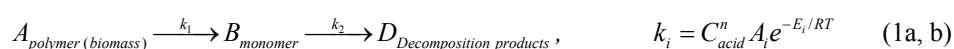
where A_i is the pre-exponential constant for reaction i, C_{acid} is the sulfuric acid concentration (%w/w of liquid), E_i is the activation energy for reaction i, n is the order of reaction w.r.t. acid concentration, T is the temperature, ϕ is the ratio of solid material to liquid (w/w).

Regarding the acetic acid, it has been reported [8] that generation of acetic acid does not follow the mechanism of Eq. (1a). Its kinetic mechanism is according Eq. (2), which depends slightly of the temperature and of the acid concentration.

$$A_{Acetic\,groups} \xrightarrow{k_1} B_{Acetic\,acid} \qquad (2)$$

For reactions reported in Table 1, kinetic parameters are given in Table 2. It is important to note that reactions with conversion fractions less than 0.025 are not considered in this work, since their contribution do not affect meaningfully the general results.

Table 2. Kinetic parameters for the hydrolysis reactions

Reaction No.	A_1 (s^{-1})	E_1 (J mol^{-1})	A_2 (s^{-1})	E_2 (J mol^{-1})	n_2	Eq.	Ref.
1	2.16×10^7	82.8	2.66×10^{12}	118.9	0.82	1	[5]
3	1.00×10^8	103.1	4.12×10^4	60.3	0.72	1	[5]
4	1.80×10^{10}	107.3	2.66×10^{12}	125.5	1.01	1	[5]
7	1.71×10^7	84.1	1.51×10^8	87.9	0.49	1	[5]
9	1.00×10^8	103.1	4.12×10^4	60.3	0.72	1	[5]
10	8.4×10^{-4}	-	-	-	-	2	[8]
11	2.16×10^6	85.2	1.23×10^9	95.7	0.39	1	[5]

2.3. Reactor modeling

For a CSTR, the following model predicts the concentration of monomers:

$$\frac{d[A]}{dt} = \frac{q_{in}[A]_0 - q_{out}[A]}{V} - k_1[A], \qquad \frac{d[B]}{dt} = \frac{q_{in}[B]_0 - q_{out}[B]}{V} + k_1[A] - k_2\phi[B] \qquad (3)$$

which, for batch reactor, can be simplified (i.e. $q_{in} = q_{out} = 0$) and solved, obtaining:

$$[B] = [A_0]\left(e^{-k_1 t} - e^{-\phi k_2 t}\right) k_1 / (\phi k_2 - k_1) \qquad (4)$$

3. Case studies: Results and discussion

In order to study the influence of several parameters on the design and operability of the process, the following variables were analyzed: (a) the reactor design: batch reactor or CSTR, (b) the feed concentration of the dilute sulfuric acid (from 1.1 to 6% w/w), (c) the reactor temperature (from 100°C to 190°C, this latter in close proximity to the calculated boiling point), (d) the reaction time to achieve maximum conversion, (e) the fraction of solids in the feed stream (from 0.05 to 0.6), and (f) the kinetic model. Regarding this one, two scenarios are compared: model A that considers only the generation reactions (i.e., k_1 given in Table 2 and $k_2 = 0$ in Eq. (1a)), and model B that considers both generation and decomposition reactions (i.e., k_1 and k_2 given in Table 2). Aspen Plus simulator was used to evaluate the boiling point of the reaction mixture, as well as the steady state of the CSTR. Matlab software was used to solve dynamic states of the CSTR and batch reactor. The results are presented next in terms of reaction times and yields of some fermentable sugars and ASL (acid-soluble lignin).

3.1. Batch Reactor

Fig. 1 shows the influence of reaction temperature and of the kinetic model for xylose and ASL. Temperature was varied from 373 to 463 K, since the boiling points was found to be at 467 K. For other monomers (glucose, arabinose, mannose, galactose) same behavior as xylose was obtained. Some highlights from these plots are: (a) as it can be expected, the higher the temperature the rate of reaction is increased, however due to this fact along with the reactions decomposition, the maximum in concentration is reached faster but is one of the lowest values in comparison with mild temperatures (423 – 443 K); (b) only ASL concentration has same behavior, but maximum values for Model B are reached with longer reaction time (not shown in Fig. 1b); (c) lower concentrations are obtained when Model B (realistic case) is used; and (d) there is a strong dependence on the temperature, high temperatures values are recommended in order to achieve higher concentrations in low reaction times.
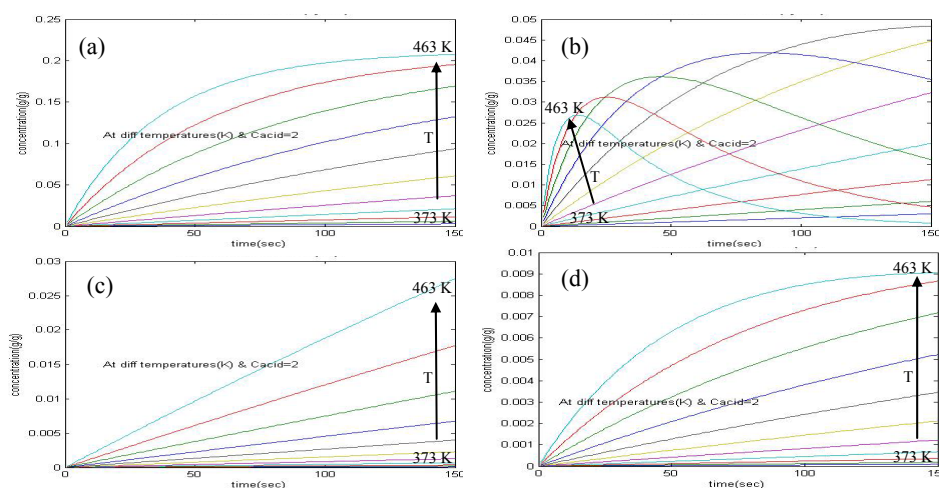
Figure 1. Influence of temperature in a batch reactor ($C_{acid} = 2\%$, $\phi = 0.42$): (a) Xylose, model A; (b) Xylose, model B; (c) ASL, model A, (d) ASL, model B.

To study the influence of acid concentration, this variable was varied from 1.1% (design basis) to 6%. Results for $C_{acid} = 2\%$ were shown in Fig. 1, while results for the two

Factors affecting the acid pretreatment of lignocellulosic biomass: Batch and continuous process

extreme values (1.1% and 6%) are shown in Fig. 2. It can be seen that the use of higher acid concentrations increase slightly the product concentration, however the reaction temperature becomes again a stronger variable to increase reaction conversions.

Lastly the influence of the solid fraction was studied varying its value around the design basis (42%), from 0.05 to 0.6. The solid fraction can be seen as the availability of material to be reacted, in other words, a limiting reactant. Results are shown in Fig. 3 for xylose at two specific temperatures. It is observed that at low ϕ-values and low temperature, concentrations are higher but reaction time is slower. Once again there is a strong influence of temperature, higher temperatures lower reaction times.
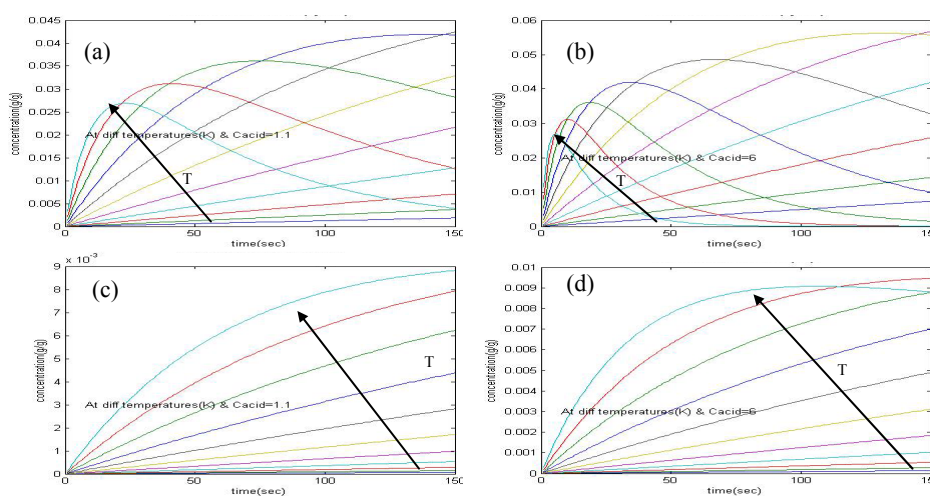
Figure 2. Influence of the acid concentration in a batch reactor ($\phi = 0.42$, model B): (a) Xylose, $C_{acid} = 1.1\%$; (b) Xylose, $C_{acid} = 6\%$; (c) ASL, $C_{acid} = 1.1\%$; (d) ASL, $C_{acid} = 6\%$.
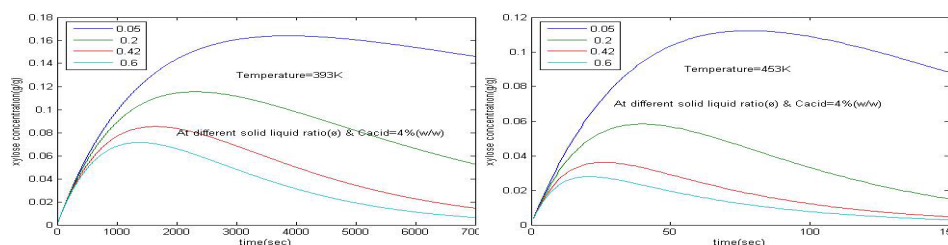
Figure 3. Influence of the solid fraction in a batch reactor ($C_{acid} = 4\%$, Model b): (a) Xylose, T = 393 K; (b) Xylose, $T = 463$ K.

3.2. CSTR

For a CSTR, similar studies as batch reactor were done. Firstly, for the steady state, results are summarized in Fig. 4. As in the batch reactor, the main highlights are: (a) Model B gives a non-monotonic behavior with lower conversions in comparison with Model A, (b) monomer concentration increases as the temperature is increased, and (c) higher conversions are obtained when C_{acid} is reduced. Secondly, for dynamic state, several analyses were done. In particular, it was corroborated the influence of Model A and B, as shown in Fig. 5. Finally the dynamic behavior of all monomers is presented in Fig. 6, where the specific operating conditions where define to keep a trade-off between high conversion and low reaction times: $T = 160°C$, $C_{acid} = 4\%$, and $\phi = 0.42$.
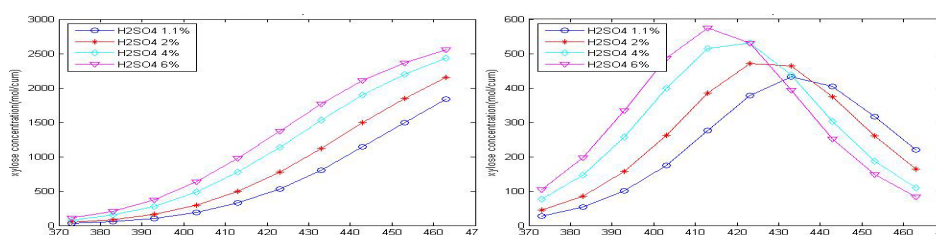
Figure 4. Steady state of a CSTR. Influence of temperature and acid concentration in a CSTR reactor ($\phi = 0.42$): (a) Xylose, model A; (b) Xylose, model B.
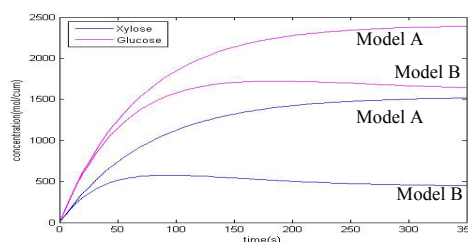
Figure 5: Dynamic behavior of a CSTR: Influence of the kinetic model.
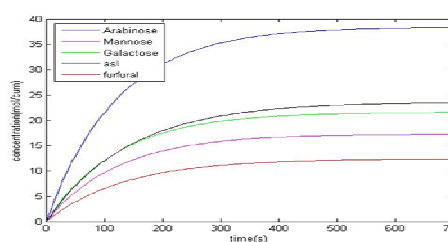
Figure 6: Dynamic behavior of a CSTR: Model B.

4. Conclusions

It has been shown that process modeling and simulation is critical and decisive for the well design of the pretreatment process of lignocellulosic biomass. Through rigorous kinetic mechanisms, together with numerical simulation in steady and dynamical states, it is demonstrated that the rigorous kinetics (Model B, which exhibits a high nonlinear behavior) should be used to simulate a realistic case. It is also concluded that efficient reaction conversions (with short reaction times) are strongly affected mainly by temperature, solid fractions and reactor-type, and there is slight dependency on the acid concentration. Currently, we are working on finding the optimum conditions by solving a strict optimization problem. From an industrial point of view, these results are relevant for the process design, operation and control.

5. Acknowledgements

References

[1] T. Eggeman, R.T. Elander, 2005, 96, 2019-2025.
[2] A. Aden, M. Ruth, K. Ibsen, J. Jechura, K. Neeves, J. Sheehan, B. Wallace, 2002 , National Renewable Energy Laboratory, NREL/TP-510-32438.
[3] J.R. Kwiatkowski, A.J. McAloon, F. Taylor, D.B. Johnstonet, 2006, Ind. Crops & Prod., 23, 288-296.
[4] C.A. Cardona, O.J. Sanchez, 2006, Energy, 31, 2447-2459.
[5] B.P. Lavarack, G.J. Griffin, D. Rodman, 2002, Biomass and Bioenergy (23) 367-380.
[6] S.H.A. Rahman, J.P. Choudhury, A.L. Ahmad, 2006, Biochem. Engng., 30, 97–103.
[7] A.H. Conner, B.F. Wood, C. G. Hill, J.F. Harris, 1985, J. Wood Chem. & Tech., 5, 461-489.
[8] R. Aguilar, J.A. Ramírez, G. Garrote, M. Vázquez, 2002, J. Food Engng., 55, 309-318.

Simulation and Control of Pollutant Propagation in Someş River Using COMSOL Multiphysics

Vasile M. Cristea, Elena D. Bâgiu, Paul Ş. Agachi

Babes-Bolyai University, 11 Arany Janos Street, Cluj-Napoca 400028, Romania, mcristea@chem.ubbcluj.ro

Abstract

The paper presents a dynamic simulator tool designed for predicting effects of the pollutant accidental discharge in a segment of the river Someş. The first approach considers the pollutant source distributed along the river bank and the second one the pointwise discharged pollutant in the river stream. The accidental release of the pollutant is considered either constant or varying with time. The complex time and space distribution of the pollutant concentration is revealed by the simulator. The k-ε turbulence model of the flow has been used and the advection-dispersion-reaction processes have been considered. Further, the effects of adding a neutralizing agent in certain points of the river segment are investigated with the aim of reducing the downstream pollutant concentration. The influence of the neutralizing agent discharge position and concentration are investigated in order to counteract the pollutant spreading effect in the river.

Keywords: pollutant transport and transfer, advection-dispersion-reaction, neutralizing agent

1. Introduction

The worldwide need for clean water has continuously increased with the population growth, being associated to the development of economic activities, urbanization and need for comfortable life. Water has become a global problem of the modern society as over 1.5 billon inhabitants do not have access to potable water. As rivers represented a necessary condition for establishing inland urban or rural human settlements, but also prerequisites for industrial activities, they have been always subject to undesired discharge. River pollution is responsible for more than 10% of diseases in the developing countries. As a result, it is most important to carefully manage the river water quality. One of the not yet solved problems consists in counteracting pollution accidents and completely removing their negative effects.

Prediction of the pollutant transport and transformation is providing valuable information for the management of the river quality. Modelling and simulation of the river system is not a trivial task due to the complex physical, chemical and biological processes involved. CFD software tools allow both the description of the hydrodynamics of the water flow together with the (bio)chemical transformation of the pollutant along and across the river stream. Both cases of common discharges and accidental contamination with pollutants may benefit on the simulation results. Simulators may become useful tools not only for preventing pollution accidents but also for designing the counteracting actions succeeding to the emergence of the polluting event.

Commercial river quality software simulators are not very rich in customizing detailed specification of the way pollutant is released into the river and usually act as a black

box. Continuous pollution may have more ravaging effects compared to instantaneous pollution as more important quantities of pollutant are released and for a larger period of time. The present developed simulator is able to cope with this challenge in a very flexible way, as both spatially and pointwise distributed pollutant sources may be simulated. Furthermore, time varying spill of the pollutant may be specified in a straightforward way.

This results in a useful tool for water quality management because it enables: forecasting impact of the pollutant in the river, linking data on pollution load with data on water quality, providing information for water quality policy analysis and testing, predicting pollutant peak propagation for early warning purposes, enhancing of supervising and alarming network design and revealing useful information for taking counteracting measures.

2. Modeling approach

Someş is an important trans-frontier river which springs from the Carpathians Mountains and stretches for 376 km in Romania and for 51 km in Hungary, where it merges with the Tisa River. The main pollutant sources of the Someş River consist in the chemical and mining companies, associated to the animal breeding farms, which may accidentally discharge both inorganic and organic compounds.

The implemented modeling approach is based on the advection-dispersion-reaction processes describing the transport and transformation of chemical species in a segment of the Someş River. Although analytical solutions of the partial differential equations are highly desired, the cases that may be dealt with the available particular developed explicit solutions are of very limited use, especially for the real river conditions [1]. This fact leads to the alternative of the numerical solutions [2]. As the geometry of the river is usually complex, obtaining realistic simulation results implies the use of software tools able to embed the irregular river profile and the complex intrinsic transport and transformation phenomena [3, 4].

CMSOL Multiphysics is a versatile CFD software package that solves coupled PDEs based on the Finite Element Method [5]. It consists in a large variety of application modules that enable, in a straightforward way, the coupling of PDEs from different application fields. Furthermore, the capability of exporting COMSOL application to the Simulink graphical user interface extension of MATLAB makes possible to benefit of modelling, simulation and analysing instruments that MATLAB with its Toolboxes may offer.

The k-ε model implemented in COMSOL for the fluid flow is one of the most used turbulence models for engineering applications:

$$\rho \frac{\partial \mathbf{U}}{\partial t} - \nabla \cdot \left[\left(\eta + \rho \mathbf{C}_\mu \frac{k^2}{\varepsilon} \right) \cdot (\nabla \mathbf{U} + (\nabla \mathbf{U})^T) \right] + \rho \mathbf{U} \cdot \nabla \mathbf{U} + \nabla P = \mathbf{F} \qquad (1)$$

In this model, two extra transport equations are solved for two introduced variables: the turbulence kinetic energy, k, and the dissipation rate of turbulence energy, ε.

The Chemical Engineering Module of COMSOL Multiphysics is appropriate for describing the pollutant space and time distribution in advection-dispersion-reaction driven processes and for providing truthful predictions of the pollutant distribution. It proves to be appropriate for the Someş River polluting simulator. Export of the COMSOL model into the Simulink environment has been achieved on the basis of a novel developed application.

Simulation and Control of Pollutant Propagation in Someş River Using COMSOL Multiphysics

Basic information was needed for the simulator development of the Someş River, such as: stream and geometry data (segment length, variation of channel width with depth, bottom slope, variation of wetted perimeter), hydraulic data (velocities, flows, water depths); meteorological data and water quality data (specific chemical species and compounds) [6, 7].

3. Simulation Results

The geometry of the river segment for which the simulator has been developed consists in a segment of the Someş River downstream the Cluj-Napoca city, in a region of potential pollutants accidental discharge. It has a complex configuration, as it includes a bifurcation of the main water stream, making the pollutant prediction a challenging task. The first investigated case, as an example, considers the pollutant source distributed on the river bank (lateral source). Staring with the value of 2 mg/L, the lateral polluting source concentration is decaying in time according to the function $1/(t+1)$, for a period of 30 minutes. In figure 1 are shown the river geometry, the pollutant source and the three particular points (Points 1–3) in which the pollutant concentration change in time is investigated.

a)   b)
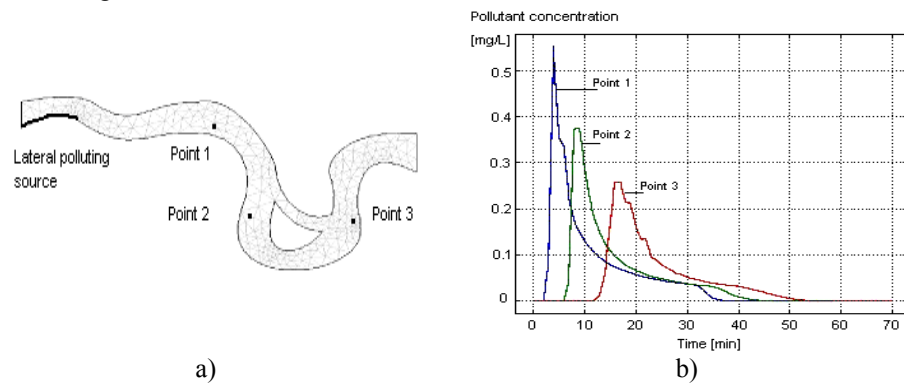
Figure 1. a) Geometry of the river segment for the case of the lateral pollutant source; b) Pollutant concentration in the three marked points 1–3

The simulation results predicting the spatial distribution of the pollutant concentration wave, at two different time moments, are presented in figure 2.

a)   b)
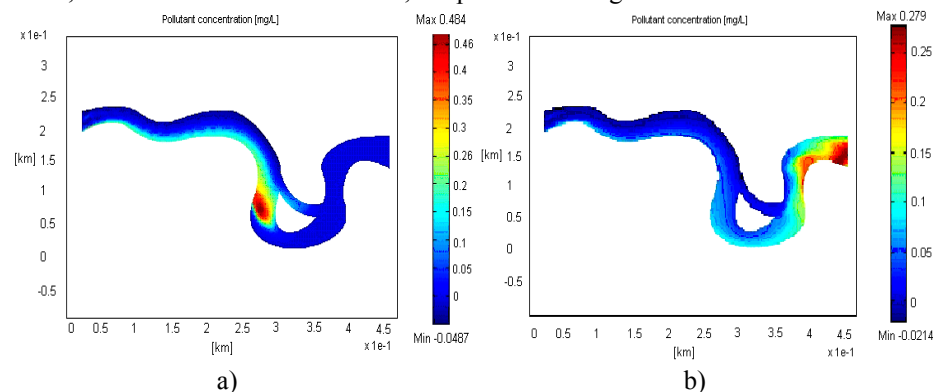
Figure 2. Pollutant concentration distribution at: a) t=8 min and b) t=20 min, after the lateral pollutant source release

The second investigated case considers the accidental release of the pollutant emerging from three sources situated in three points located in the cross section of the river. For this case, different hydrodynamics of the river flow have been set in order to better reveal the polluting effect in case of reduced flow rate. For the point sources the release is constant in time for a time period of 5 minutes and vanishes afterwards. The position of the three point sources on the river segment is presented in figure 3.
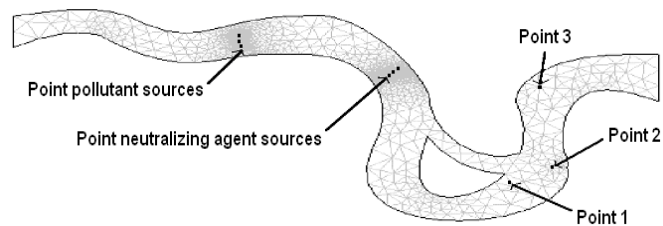
Figure 3. Localization of the point pollutant sources, neutralizing agent sources and pollutant concentration measurement three marked points 1–3

The polluting simulation scenario with the three point sources is comparatively performed for two circumstances. One is presenting the prediction of the pollutant discharge effect originating from the three point sources, for which advection and diffusion of the pollutant have been considered. The second one predicts the pollutant discharge effect, having the same origin, but considering that downstream to the polluting sources the neutralizing agent is added by three neutralizing agent point sources, figure 3. This second simulation circumstances account for the advection-diffusion-reaction processes, the latter process aiming to the pollutant consumption.

Comparative simulation results consisting in snapshots of the pollutant concentration distribution along the river segment, succeeding 30 minutes and 50 minutes to the pollutant discharge, are presented in figures 4 and 5.
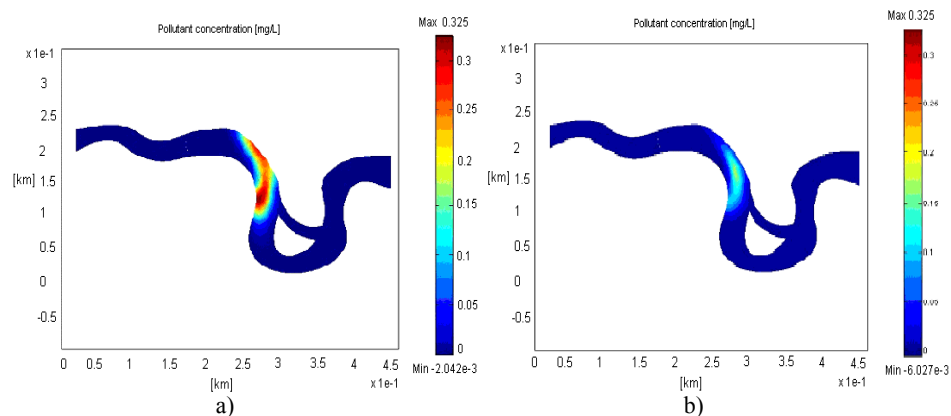
Figure 4. Pollutant concentration distribution at t=30 min after the point pollutant sources release for: a) case without neutralizing agent and b) case with neutralizing agent

*Simulation and Control of Pollutant Propagation in Someş River Using COMSOL
Multiphysics*

a)   b)
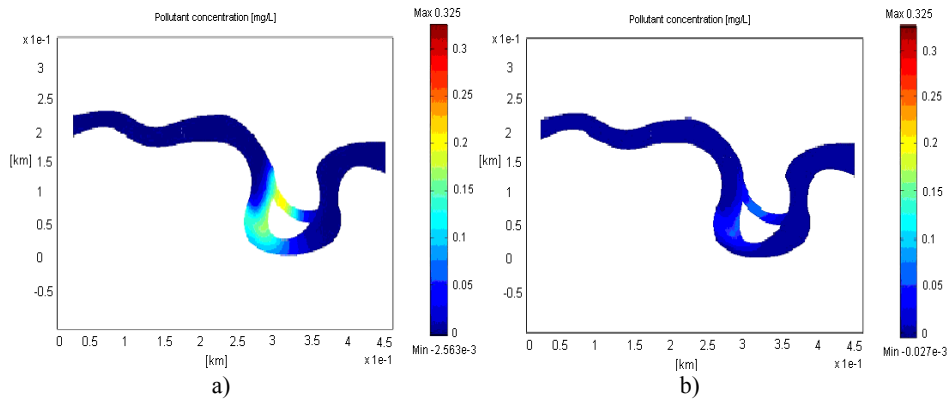
Figure 5. Pollutant concentration distribution at t=50 min after the point pollutant sources release
for: a) case without neutralizing agent and b) case with neutralizing agent

The simulation results presented in figures 4 and 5 reveal the counteracting effect of the
neutralizing agent, as the concentration of the pollutant is consistently diminished.
The effect of the concentration of point neutralizing agent sources on the pollutant
concentration value has been also comparatively investigated. Pollutant concentration
changes in time for the three points situated downstream the bifurcation region, and
marked by Points 1–3 in figure 3, are shown in figure 6.

a)

b)

c)
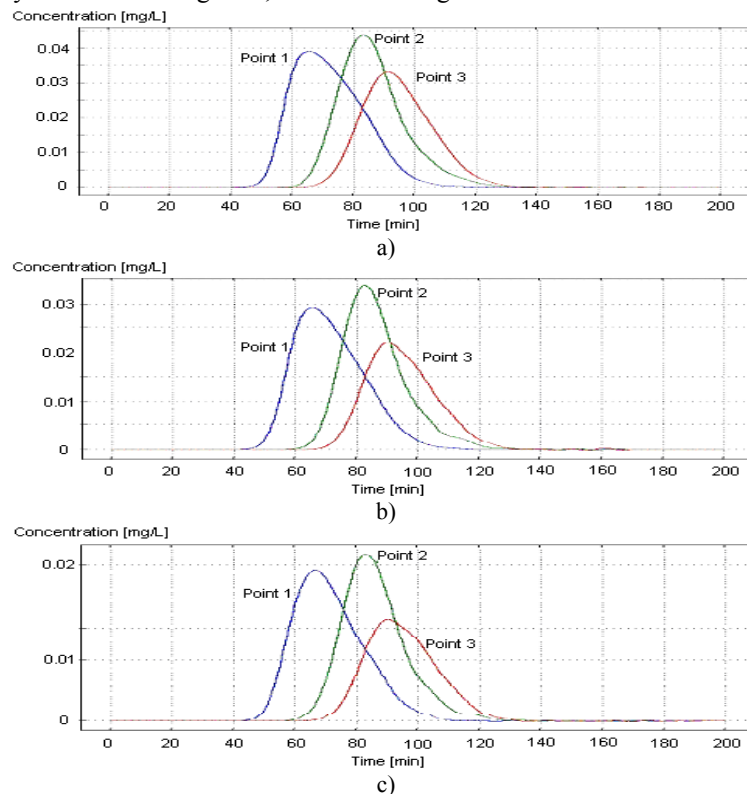
Figure 6. Pollutant concentration time change in the three considered Points 1–3, for the different
neutralizing agent source concentration values of: a) 10 mg/L, b) 25 mg/L and c) 40 mg/L

As expected, increase in the concentration of the neutralizing agent leads to the reduction of the downstream pollutant concentration. The simulator allows the investigation of the best conditions for obtaining the most efficient application of counteracting actions emerged from the pollutant accidental discharge, i.e. position of the neutralizing agent sources, requested concentration of the neutralizing agent, position of the measurement points and the way the neutralizing agent concentration or flow have to be changed in time.

4. Conclusions

The paper presents the development of a dynamic simulator for predicting the pollutant space and time distribution in a segment of the Someş River having a complex geometry, illustrated by a bifurcation configuration of the river flow. Both lateral and point polluting sources effects have been investigated, with constant or time varying concentration of the pollutant discharge. Complex hydrological parameters of the river may be embedded in the application and, associated to the benefits of the COMSOL Multyphysics CFD software, allowed the development of a powerful tool for river quality assessment in case of polluting accidents. Details revealing the spatial distribution of the pollutant wave along and across the river segment provide valuable information for risk assessment and for designing measures aimed to limit the undesired pollution effects. Application of potential counteracting measures has been also studied by the introduction of point sources releasing neutralizing agent. Effect of the concentration and position of the neutralizing agent sources have been considered. The simulator opens challenging perspectives for the further study of optimal conditions the neutralizing agent may be spread out, in order to efficiently reduce the pollution consequences. Coupled with Geographical Information System data and applications, the simulator may become a versatile and useful tool for river quality management.

5. Acknowledgements

References

[1] S.A. Socolofsky, G.H Jirka, 2005, Special Topics in Mixing and Transport Processes in the Environment, Engineering – Lectures, fifth ed. Coastal and Ocean Engineering Division, Texas A&M University.

[2] A.D. Rubio, A .Zalts, C.D. El Hasi, 2008, Numerical solution of the advection–reaction–diffusion equation at different scales. Environmental Modelling &Software 23, (1), 90–95.

[3] J.B. Boxall, I. Guymer, 2007, Longitudinal mixing in meandering channels: new experimental data set and verification of a predictive technique, Water Research 41, (2), 341–354.

[4] W. Czernuszenko, P. Rowinski, 2005, Water Quality Hazards and Dispersion of Pllutants, Springer Science+Business Media, Inc., New-York.

[5] COMSOL Multyphysics User's Guide, version 3.3, 2006.

[6] E. Roberts Alley, 2000, Part II Water Quality Management and Part IV Water Pollution Control, Water Quality Control Handbook, McGraw-Hill Company, Inc., New-York.

[7] D. Kavetski, G. Kuczera, S.W. Franks, 2006, Calibration of conceptual hydrological models revisited: 1. Overcoming numerical artefacts. Journal of Hydrology, 320, (1-2), 173–186.

Clean Coal: A new power generation process with high efficiency, carbon capture and zero emissions

Thomas A. Adams II[a], Paul I Barton[b]

Department of Chemical Engineering, Massachusetts Institute of Technology, 77 Massachusetts Ave, Cambridge, MA USA 02139. [a] adamsta@mit.edu, [b] pib@mit.edu

Abstract

An environmentally friendly electricity generation process using coal gasification and solid oxide fuel cells (SOFCs) produces electricity with high efficiency, 99.95% carbon capture and essentially zero atmospheric emissions. Coal is gasified into syngas, cleaned and shifted to hydrogen gas to fuel SOFCs. The primary waste products, CO_2 and H_2O, are separated with a very small energy penalty. The carbon dioxide purity is high enough to meet most specifications for geological sequestration. Even with carbon capture capability, the power plant has a higher efficiency (4-10 percentage points) than standard pulverized coal or integrated gasification combined cycle processes without carbon capture and consumes significantly less fresh water. If cooling towers are replaced with dry-cooling technology, net water can be produced and recovered, rather than consumed. Moreover, under a cap-and-trade scenario, the process has the lowest cost-of-electricity, even with carbon capture, for carbon prices above \$5-10/tonne.

Keywords: coal, gasification, power, carbon capture, sequestration

1. Introduction

Coal, with high reserves and a relatively low cost, will continue to be one of the most important energy sources for the United States and elsewhere. In the US, coal accounts for roughly one-third of all CO_2 emissions arising from fossil fuel use [1]. Worldwide, coal provides for about one-third of all electricity production, one-quarter of transportation use, and is projected to account for 22% of all energy needs by 2030 [2].

Although alternative sources are being improved, it is likely that alternatives alone will never be sufficient to meet the entire energy demand. For example, it is estimated that if all land in the US were devoted entirely to biomass production, it would only be able to meet 58% of its current energy needs for the transportation sector [3]. Therefore, demand for environmentally friendly coal technologies continues through the development of carbon capture and sequestration (CCS) technologies [4].

However, capturing CO_2 from traditional combustion exhaust gases is expensive, both from a capital and energy perspective, because the CO_2 in the exhaust is greatly diluted with N_2. For example, adding post-combustion capture capability to a traditional pulverized coal plant increases the levelized-cost-of-electricity (LCOE) by as much as 85% and reduces the efficiency by 12 percentage points [5]. This means that one would have to burn 45% more coal just to power the CO_2 capture process.

The integrated gasification combined cycle (IGCC) approach has been proposed to alleviate this problem. Here, coal is gasified into syngas, which can be converted to a mixture of H_2 and CO_2 in roughly equal quantities. Using appropriate solvents, the CO_2

can be recovered from this stream before combustion, but at the expense of a 6 percentage point drop in efficiency and a cost of electricity increase of up to 30% [5,6]. Nevertheless, it is preferable to traditional pulverized coal in the context of CO_2 capture.

A novel process is proposed that replaces the combustion step with solid oxide fuel cells (SOFC). A SOFC oxidizes H_2 electrochemically, producing electricity without the thermodynamic limitations of heat engines. The SOFC is designed to have separate anode (fuel) and cathode (air) sections, so air can be used for oxidation without diluting the fuel stream with N_2. This permits easy separation of the anode exhaust (H_2O and CO_2) with a very small energy penalty and no solvent recovery step [7]. The spent air stream, being heated by the SOFC, can provide additional power through the Brayton cycle. Together, these innovations provide a higher plant efficiency, significantly reduce the energy penalty of CO_2 capture, and facilitate recycle of water in the process.

2. Process Description

The general process is shown in Fig. 1. Details on each unit are discussed below.
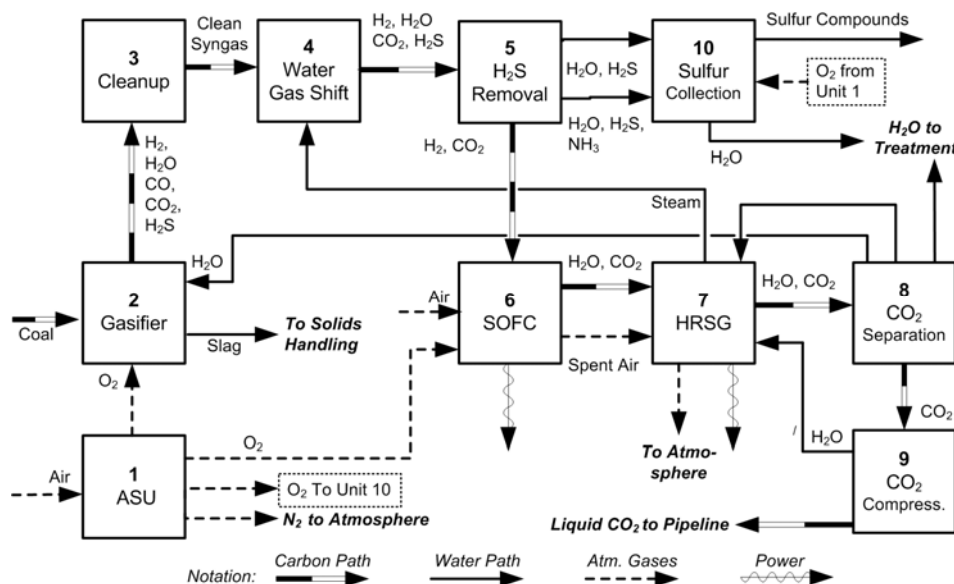
Figure 1: Simplified flowsheet of the SOFC-based coal-to-electricity process with 100% CCS.

2.1. Air Separation Unit

The Air Separation Unit (ASU) provides oxygen by cryogenic separation at 95% purity. Approximately 97% of the recovered O_2 is used for gasification.

2.2. Gasification

In the gasifier section, milled-coal slurry is fed to the gasifier with O_2 and water. The water is provided through recycle from downstream units. The gasifier operates near 56 bar pressure and 1300°C, and produces syngas (a mixture of H_2, CO and wastes such as CO_2 and H_2O). It is equipped with a cooling mechanism where radiant heat is recovered through steam in Unit 7. Slag (ash, oxidized wastes, etc.) is collected through a solids-recovery system. The remaining wastes present in the syngas stream consist primarily of N_2, Ar, HCl, H_2S, COS, NH_3, and Hg, which are handled downstream.

Clean Coal: High efficiency power plant with zero emissions.

2.3. Gas Cleanup

A scrubbing process removes HCl from the syngas stream in Unit 3, since HCl can degrade the performance of an SOFC at concentrations as low as 20 ppm [8].

2.4. Water Gas Shift

The energy-bearing components of the syngas are CO and H_2. However, CO can cause degradation of the SOFCs through carbon deposition, leading to power loss, lower efficiency, and a shorter lifetime [9]. Therefore the water gas shift (WGS) reaction is used to convert the CO into additional H_2:

$$CO + H_2O \leftrightarrow CO_2 + H_2 \tag{1}$$

2.5. Sulfur Removal

The presence of H_2S can cause poisoning of the SOFC anodes [9], and so it is removed by absorption using a solvent (e.g., Selexol). NH_3, Hg, and residual H_2O are also removed in Unit 5 as well. The cleaned stream contains approximately equimolar amounts of H_2 and CO_2, which is sent to the SOFCs. The recovered H_2S is converted into sulfur compounds via the Claus process in Unit 10 for use in other processes.

2.6. SOFCs

The SOFCs provide the bulk of the power generation capability for the plant. As shown in Fig. 2, the fuel stream is expanded to 20 bar (the maximum safe operating pressure for the SOFCs [10]) and fed to the anode side. Air is compressed and fed to the cathode side. Oxygen ions from the cathode are conducted through the solid wall and react with the fuel in the anode, producing DC power and heat according to:

$$H_2 + O^= \rightarrow H_2O + 2e^- \tag{2}$$

Stacks of SOFCs are arranged in modules, with intercooling stages between, to prevent overheating above 1000°C. To maximize fuel utilization, unreacted H_2 in the anode exhaust is oxidized with O_2 from the ASU, producing additional heat. The heated cathode exhaust is expanded to atmospheric pressure, producing electricity.
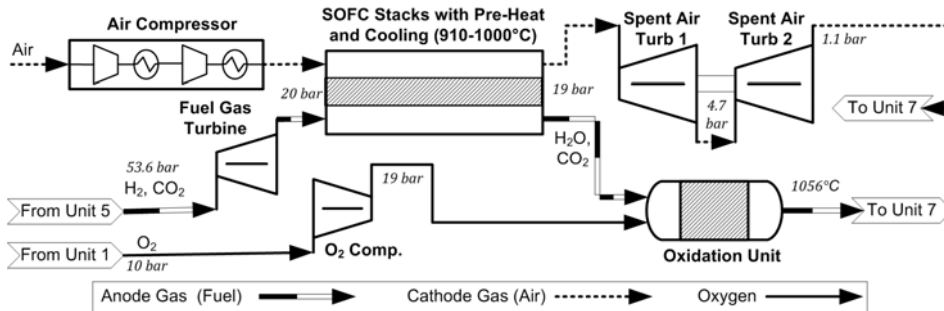
Figure 2: Detail of the SOFC section of the plant.

2.7. HRSG

The HRSG section of the plant generates steam and electricity using heat recovered from the other process units. The power produced is about half of the SOFC section.

2.8. CO_2 Recovery

In Unit 8, the cooled, depleted fuel stream (H_2O and CO_2) is partially condensed and flashed in a series of cascading flash drums, each at a lower pressure. The vapor phases are recompressed and recycled to higher pressure stages. The vapor product of the

highest pressure stage contains about 96% CO_2 by mole (the balance mostly N_2 and Ar). The liquid phase from the last stage is water at >99.9 mol% purity, which can be recycled for use in cooling towers or steam makeup as necessary.

2.9. CO_2 Compression

The 96% CO_2 stream is compressed close to the critical pressure, and then partially condensed, recovering most of the residual water in the liquid phase. The vapor stream is then totally condensed and pumped to supercritical pressures for pipeline transport and sequestration.

2.10. CAPE Tools

Aspen Plus 2006.5 was used for flowsheet simulations using built-in models for reactors, compressors, expanders, turbines, pumps, staged separation columns, and heat exchangers. More complex models for the gasifier and SOFCs are described in [7]. Physical properties were modeled with the Peng-Robinson/Boston-Mathias EOS, except for pure water streams (NBC/NRC steam tables), and CO_2/H_2O mixtures (Electrolyte NRTL method with Henry coefficients near the critical point, and Redlich-Kwong-Soave with Holderbaum mixing away from the critical point.)

3. Results

3.1. Electrical Efficiency

Even with CCS capability, the resulting SOFC-based power plant as a whole is significantly more efficient than the IGCC base case or traditional pulverized coal (PC) plants. As seen in Fig. 3, the SOFC-based plant with once-through cooling achieves about 44.8% efficiency, which is significantly higher than IGCC at 38.2%, and PC at 36.8% [5]. When CCS capability is included, the SOFC plant efficiency is reduced by only 1 percentage point. However, CCS causes the IGCC plant to drop by about 4 percentage points and as much as 12 percentage points for PC. Based on surveys of existing coal plants [11], using cooling towers instead of once-through cooling causes a 0.8-1.5% drop in the net power production, while using dry cooling (using air cooling instead of water) results in a 4-9% drop in power output. These are reflected in Fig. 3.
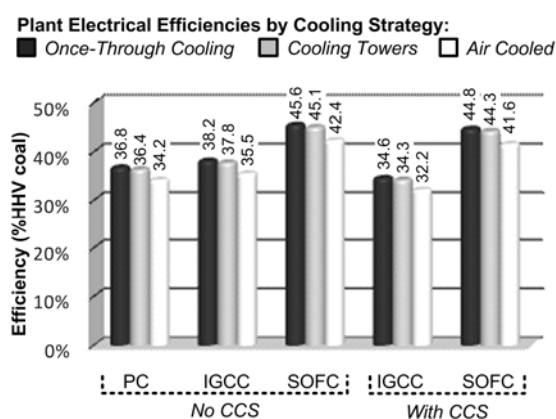
Figure 3: Summary of electrical efficiencies of various coal-to-electricity processes.

3.2. Environmental Issues

In Fig. 4A, the approximate water consumption of PC, IGCC, and SOFC-based power plants are shown. IGCC with cooling towers requires 30% less fresh water than the

Clean Coal: High efficiency power plant with zero emissions.

average PC plant, but the SOFC-based plant requires even less. This advantage arises from the recovery of the water in the CO_2/H_2O exhaust which can be achieved at modest temperatures with little power. This water can be recycled to completely account for all water input needs to the gasifier, leaving a surplus of about 1 billion L/yr. If cooling towers are used, the surplus can be used toward the evaporative cooling requirement. If dry cooling is used, the surplus can be treated in a wastewater treatment plant, providing a net positive output of water from the process. The surplus water generated in Unit 8 has very high purity, and so the costs of treatment are minimal.
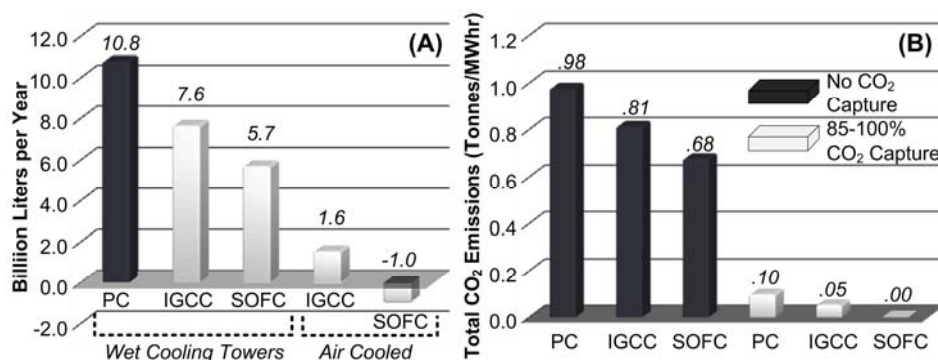
Figure 4: (A) Water consumption and (B) CO_2 emissions of various coal-to-electricity processes, for 227 tonne/day coal input.

In Fig. 4B, the CO_2 emissions of each plant type, with and without CCS, are shown. Without CO_2 capture, the emissions on a per MW-hr basis of the IGCC and SOFC plant are lower than PC by virtue of their improved efficiency. However, post-combustion capture by absorption (used by PC) achieves only about 85% CO_2 capture, pre-combustion capture (used by IGCC) likewise achieves about 90% capture, while the SOFC process is able to achieve 100% capture.

3.3. Levelized Costs of Electricity (LCOE)

Capital costs, fuel prices, and manufacturing costs for each process were scaled from quotations given in [5] or computed with Aspen Icarus 2006.5 where appropriate. Two SOFC prices are considered: $500 and $1000/kW, representing expected and conservative estimates, respectively. Also, the proposed American Clean Energy and Security Act (ACES) [12] is taken into account through an appropriate model, which considers a range of market prices for CO_2 emissions credits, a gradually diminishing supply of credits, free credits given to capture-enabled facilities during the early years of its effect, and other factors.

The resulting LCOEs for each process are shown in Fig. 5 for a range of average market prices of CO_2 credits. The SOFC-based processes, with carbon capture, have the lowest LCOEs than any of the capture-enabled processes for any carbon price. Furthermore, compared to a PC plant without carbon capture, the SOFC-based plants have lower LCOEs above $5-10/tonne CO_2 emitted, thus setting a lower bound on the carbon price needed to incentivize carbon capture. This is much lower than the $40-50/tonne price commonly discussed as the necessary minimum to incentivize carbon capture; note from Fig. 5 that approximately $47/tonne is required for one to choose to add post-combustion capture to a PC plant (as opposed to a traditional PC plant).
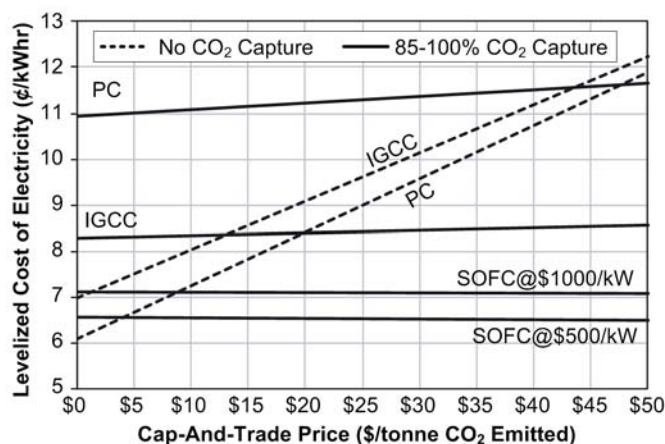
Figure 5: Levelized cost of electricity of various processes using 227 tonne/day coal input.

4. Conclusions

The use of SOFCs in a coal-to-electricity process enables inherently easy carbon capture, since CO_2 exhaust is kept free of N_2 dilution by maintaining passive separation of air and fuel sources. As a result, the process is more efficient and less costly than other CCS-enabled processes, and has a lower cost than even traditional PC at relatively low carbon tax rates. This represents a significant improvement over existing concepts. The same approach can be applied to natural gas, biomass, petcoke, and other carbonaceous fuel sources, and is a subject of future work.

5. Acknowledgements

References

[1]   T.J. Blasing, C.T. Broniak, and G. Marland, 2005, Tellus B, Chem. Phys. Meteorol., 57B (2005) 107-115.
[2]   A.J. Minchener and J.T. McMullan, J. Energy Inst., 81 (2008) 38-44.
[3]   R. Agrawal et al., Proc. Natl. Acad. Sci., 104 (2007) 4828-4833.
[4]   C.J. Cleveland (ed.), Encyclopedia of Energy, Elsevier Science, New York, 2004.
[5]   M.C. Woods et al., Cost and Performance Baseline for Fossil Energy Plants. Volume 1: Bituminous Coal and Natural Gas to Electricity Final Report. DOE/NETL-2007/1281.
[6]   R.E. Anderson, H. Brandt, and F. Viteri, US Patent 2003/0131582 A1 (2003).
[7]   T.A. Adams, and P.I. Barton, High efficiency power production from coal with carbon capture. AIChE J., in press 2010.
[8]   J.P. Trembly, R.S. Gemmen, and D.J. Bayless, J. Power Sources, 169 (2007) 347-354.
[9]   A.I. Marquez et al., J. Power Sources, 164 (2007) 659-667.
[10]   EG&G Technical Services, Inc., Fuel Cell Handbook 7th Ed., 2004 , pp 7.28-7.29.
[11]   R.W. Gross, Energy penalty analysis of possible cooling water intake structure requirements on existing coal-fired power plants. NETL, DOE/OFE, Argonne NL (2002).
[12]   H.A. Waxman, E.J. Markey. The American Clean Energy and Security Act of 2009. Proposal to US House of Representatives, March 31, 2009.

Exergoeconomic Analysis of Post-Combustion CO_2 Capture Processes

M.-O. Schach[a], R. Schneider[b], H. Schramm[b], J.-U. Repke[a]

[a]*Chair of Process Dynamics and Operation, TU Berlin, Sekr. KWT 9, Strasse des 17. Juni 135, 10623 Berlin, Germany*
[b]*Siemens AG Energy Sector, Fossil Power Generation, Industriepark Höchst, 65926 Frankfurt am Main*

Abstract

One of the instruments against the global warming is capture and sequestration of carbon dioxide from the flue gas of coal-fired power plants. Different concepts of capture are being pursued. The advantage of post-combustion processes, like processes based on absorption and stripping, is the possibility of retrofitting a state-of-the-art power plant with a capture facility under reasonable effort. Capturing CO_2 by using an absorption/stripping process requires energy in form of electricity and steam both supplied by the power plant. The capture process thereby reduces the overall efficiency of the power plant by up to 13%-pts. A way to lower these energy requirements is the development of new process configurations. In this study three different configurations were investigated by performing an exergoeconomic analysis. Revealing the major cost sources the processes could be improved by assessing and adjusting the cost effectiveness of each component. Furthermore a new configuration could be derived from the results of the analysis.

Keywords: CO_2 capture, MEA, exergoeconomic analysis, process configurations

1. Introduction

CO_2 is one of the greenhouse gases which are made responsible for the current period of global warming. Among the biggest emission sources for this gas are coal fired power plants. To reduce these emissions processes are being developed to capture the CO_2 from the flue gas. Basically there are three main categories of capture processes: post-combustion, pre-combustion and oxyfuel processes. The advantage of the post-combustion processes is that they are applicable to already existing power plants with only reasonable changes. A drawback of all capture processes is the reduction of the power plant efficiency. For the conventional absorption/stripping cycle used for post combustion CO_2 capture the efficiency will be reduced by up to 13%-pts. In order to lower these losses the processes are being improved in different manners. Big efforts have been put into the development of new solvents. Lower energy requirements for the regeneration and a high loading capacity are the main objectives.

Furthermore the development of new process configurations is a second field of investigation. For the comparison and improvement of different absorber/stripper configurations a suitable evaluation scheme must be developed and applied. In this work an exergoeconomic analysis was performed. It is a well known tool for optimizing process designs when a mathematical optimization cannot be used as in this case due to the complexity of the processes. An economic analysis based on the cash flow analysis method considering all capital, operating and maintenance cost was already performed in a previous work (Schach et al. 2010). For the current study this analysis was extended

by an exergy analysis. Thus it was possible to identify the location, the magnitude, and the sources of thermodynamic inefficiencies of the processes. With the combination of both analyses the cost effectiveness of different process configurations was improved by iterative evaluation and optimization. A reference process represented by a conventional absorption/stripping cycle, a configuration with an absorber intercooler and a matrix stripper configuration were analyzed.

2. Process Simulation

The objective of this study was to analyze, compare and improve different process configurations by applying an exergoeconomic analysis. To demonstrate the performance of this method processes were chosen which represent different technical innovations and different cost structures like variable investment cost. Hence, the influence of the different parameters on the separation costs could be observed. In addition to the typical absorption and stripping configuration, two other process alternatives were selected. The concept based on an absorber intercooler represents an improvement, which has only a small influence on the investment cost, whereas the matrix configuration has a stronger impact on investments.

The different process configurations were simulated with Aspen Plus 2006.5 using the amine package MEA-REA. This includes a reaction model considering both kinetically controlled and equilibrium reactions. For the absorption and stripper columns the RadFrac model was used with rate-sep calculation.

A 30 wt-% monoethanolamine (MEA) solution was used as a solvent. All processes separated 90% of the CO_2 in the flue gas, which was then compressed up to 110bar. The treated flue gas had a mass flux of 779.5 kg/s and the following composition: $x_{N2} = 0.7$, $x_{CO2} = 0.14$, $x_{H2O} = 0.13$, $x_{O2} = 0.03$.

2.1. Baseline Process

All processes were compared to a baseline process represented by the standard absorption/stripping cycle (Figure 1). The flue gas enters the absorber after passing a blower and a water cooler. In the absorber the CO_2 is absorbed by the solvent. The treated gas is vented to the atmosphere after passing a water scrubber to remove MEA traces. The loaded MEA solution is pumped through a cross heat exchanger to the top of the stripper. In this column the rich solvent is regenerated by providing heat in form of heating steam of the power plant. The vapours of the column are condensed in a partial condenser at 40°C. As gaseous product CO_2 with a purity of >95 mol-% is obtained, which is liquefied in a multistage compressor by pressurizing the CO_2 up to 110 bar. During the compression water condenses and the resulting liquefied CO_2 has a purity of >99.5 mol-%. The lean solvent is routed back to the absorber.
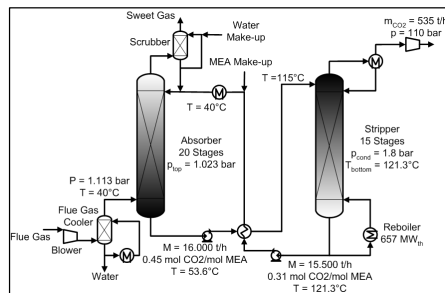
Figure 1: Baseline Process
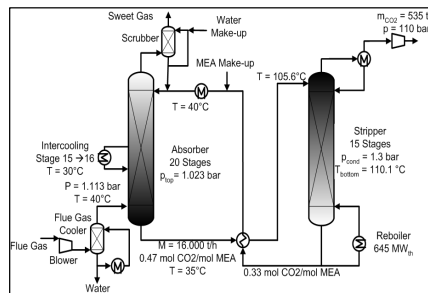
Figure 2: Absorber Intercooler

Exergoeconomic Analysis of Post-Combustion CO$_2$ Capture Processes

2.2. Absorber Intercooler

The configuration shown in Figure 2 is an extension of the baseline process. In this configuration an intercooler is applied to the absorber. Several authors, like Thompson and King (1987), performed studies dealing with this configuration. The whole liquid stream from stage 15 is cooled down to 30°C and returned to the subjacent stage 16. The location of the intercooler was a result of parametric studies. The advantage of this process option is that the heat released during the chemical reaction between CO$_2$ and MEA can be removed. Hence, the temperature profile in the absorber can be smoothed and more CO$_2$ can be absorbed. By the application of lower temperatures it is possible to increase the CO$_2$-loading, while the mass flow of the solvent remains constant. Due to this, the lean solvent can enter the absorber column at higher loadings in comparison to the reference case. Thus energy can be saved in the stripper.

2.3. Matrix Stripper

Figure 3 shows a matrix stripper configuration. This process option was originally mentioned by Oyenekan and Rochelle (2007). The rich solvent is split into two streams with a ratio of 50/50. One split stream is directed to stripper 1, where a part of the solvent is regenerated. The bottom product is forwarded to the middle section of a second stripper, which operates at a lower pressure. The other split stream of the rich solvent is fed to the top of this stripper. From the middle section of the second stripper a semi-lean solvent is directed to the middle of the absorber, whereas the bottom products are fed to the top of the absorber. Heat is supplied in the form of steam to both strippers. Since the first stripper operates at a higher pressure (1.8bar) but has the same bottom temperature as the second one, only a small amount of CO$_2$ is obtained in this column. According to Oyenekan and Rochelle (2007), the advantage of this configuration is the smoothed temperature profile throughout the second stripper resulting in a lower energy demand

3. Exergoeconomic analysis

The exergoeconomic analysis consists of two parts, an exergy analysis and an economic analysis. Both were performed according to Bejan et al. (1996). In the following chapters both analysis and the combination of them will be described.

3.1. Economic Analysis

In order to calculate the costs of the different configurations a cost model was developed. Table 1 summarizes the main assumptions and boundary conditions for the economic evaluation.

Table 1: Assumptions for the economic analysis

Project life	25 years
Plant operating	7500 h/y
Cost of CO$_2$ certificate	17,68 €/t
Interest rate	8%
Inflation rate	3%
Rate of price increase of apparatuses	10%
Rate of price increase of OMC	5%
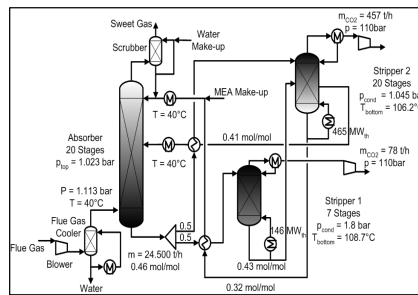
Figure 3: Matrix Stripper

Based on the results of the simulation of the process the component costs (e.g. for columns, heat exchangers or pumps) have been calculated with a scaling factor and reference costs.

The sum of the component costs represents the purchased equipment cost (PEC). Based on this value the capital costs (CC) and the operating and maintenance costs (OMC) could be assessed. The compositions of these costs were estimated using Peters et al. (2002). Both costs were then converted into a constant series of payments for every year of project life. This value together with the annual production of electricity and the CO_2 emissions led to the important cost of CO_2-avoided. These cost are a characteristic measure of the process performance. They are defined as follow

$$\text{cost of } CO_2\text{-avoided [€/t]} = \frac{\text{cost of electr.}_{capture} \text{ [€/kWh]} - \text{cost of electr.}_{ref} \text{ [€/kWh]}}{CO_2\text{emission}_{ref} \text{ [t/kWh]} - CO_2\text{emission}_{capture} \text{ [t/kWh]}} \tag{1}$$

3.2. Exergy Analysis

Exergy is the maximum theoretical useful work obtainable from a thermal system as it is brought into thermodynamic equilibrium with the environment. The exergy of a system consists of different components. The components considered in this analysis were the physical and the chemical exergy. For calculating the chemical exergy the exergy-reference environment of Szargut et al. (1988) was used. For each apparatus an exergy balance was performed from which the exergy destruction could be calculated. This is together with the exergy loss a measurement for the thermodynamic inefficiencies of a system. Another important value resulting from the exergy analysis is the exergetic efficiency ε which is defined as the ratio between the exergy of the product and the exergy of the fuel ($\varepsilon = E_P/E_F$). With the exergy destruction and the exergetic efficiency it was possible to evaluate the processes from the thermodynamic point of view.

3.3. Exergoeconomic Analysis

The exergoeconomic analysis combines the results of the economic and the exergy analysis. It gives an idea about the cost of the exergy destruction. Each apparatus was separately analyzed. The most important value for this evaluation was the exergoeconomic factor f_k which expresses the contribution of the capital cost to the sum of capital cost and cost of exergy destruction of an apparatus k. It is defined as

$$f_k = \frac{\dot{Z}_k^{CI}}{\dot{Z}_k^{CI} + \dot{C}_{D,k}}, \tag{2}$$

where Z_k^{CI} is the cost rate associated with capital investment and $C_{D,k}$ is the cost rate associated with the exergy destruction. With this factor it is possible to identify the major cost source of an apparatus. This can be the capital investment (high f_k) or the cost of exergy destruction (low f_k).

4. Results

The results of the exergy analysis and the economic analysis are shown in Table 2. The shown values, except the exergetic efficiency, are normalized. The reference case represents the benchmark. The results of the other processes reflect the performance in comparison to the benchmark. Both alternative configurations show better results than the reference case. The matrix stripper configuration has the best exergetic efficiency and also the lowest energy demand. However, the cost of CO_2-avoided are higher than

Exergoeconomic Analysis of Post-Combustion CO₂ Capture Processes

the cost of the configuration with intercooler. This is due to the higher investment cost for the matrix stripper. An additional stripper column and a second cross heat exchanger are needed, whereas for the intercooler configuration only an additional heat exchanger is required. From the energetic and exergetic point of view the matrix stripper with high exergetic efficiency and lower energy requirement would be the best process. However, the additional economic analysis revealed that not the process with the lowest energy demand or highest exergetic efficiency has the best overall performance. In this case the simple configuration with an additional intercooler has the lowest cost of CO_2-avoided.

Table 2: Results of the exergetic and economic analysis

	Reference	Intercooling	Matrix
exergetic efficiency [%]	30	34	35
normalized cost of CO2-avoided [€/t]	1	0,95	0,97
normalized equivalent power demand [MW]	1	0,97	0,94

Based on the results of the exergoeconomic analysis it was possible to improve the processes and to derive design modifications. Table 3 shows the results of the analysis. The apparatuses with a very high or very low f_k value are interesting for further improvement. The processes consist of three different kinds of apparatuses: heat exchanger, pumps and columns. The f_k value of the pumps can only be changed by using different pumps with different efficiencies. However, due to data limitations this analysis was not included. The columns have relative low f_k values. This signifies that an increase of the efficiency at the expense of investment cost would improve the cost effectiveness. However, the major part of the exergy destruction in these components is caused by the chemical reactions. Since the reactions cannot be avoided it is difficult to improve these apparatuses only by adjusting the process parameters. A more proper method is the synthesis of new configurations. As can be seen in Table 3 the columns in the matrix configuration have a higher f value which is a result of the improved process design.

Table 3: Results of the exergoeconomic factor f_k for the different apparatuses

Base Case Apparatus	f[%]	Intercooling Apparatus	f[%]	Matrix Apparatus	f[%]
Flue gas cooler	99,46	Flue gas cooler	99,46	Flue gas cooler	99,45
Compressor	21,85	Compressor	23,33	Compressor 1	36,87
Absorber	13,06	Absorber	11,60	Compressor 2	19,88
LS Pump	11,59	Blower	10,21	Absorber	14,32
Blower	10,10	LS Pump	9,77	Desorber 1	11,02
Cross HX	8,90	Cross HX	8,61	Blower	10,16
Desorber	3,06	Desorber	2,92	LS Pump	9,70
RS Pump	2,61	RS Pump	1,32	Desorber 2	4,31
LS Cooler	0,34	LS Cooler	0,04	RS Pump	1,54
				Cross HX 1	1,29
				Cross HX 2	0,17
				LS Cooler	0,01
				Semi-LS Pump	0,01
				Semi-LS Cooler	0,01

Apparatuses which seem to be adequate candidates for further improvements are the lean solvent cooler, the flue gas cooler and the cross heat exchanger. As the f_k value of

the lean solvent cooler is very low, an improvement at the expense of investment is reasonable. For a heat exchanger this would mean a reduction of the logarithmic temperature approach and an increase of the heat exchange area. There is a similar relation for the cross heat exchanger. The indicator of the process overall performance are the cost of CO_2-avoided. Figure 4 shows the normalized cost of CO_2-avoided and the f value for different logarithmic temperature approaches in the cross heat exchanger. A maximum of the f value means a minimum for the cost of CO_2-avoided.
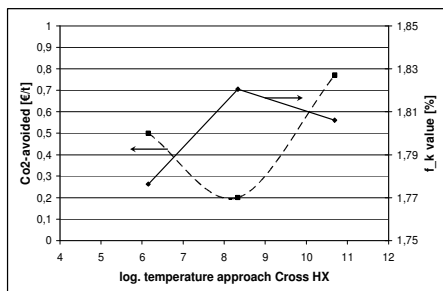
Figure 4: Optimization of Cross HX   **Figure 5:** Configuration without flue gas cooler
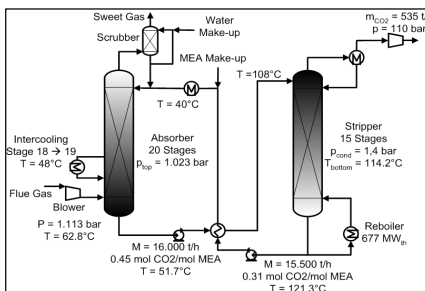
The high f_k value of the flue gas cooler implies that a reduction of the investment cost could improve the process performance. The highest possible reduction of the investment cost in this case would be the elimination of the apparatus. Therefore a new flowsheet was designed without flue gas cooler (Figure 5). The purpose of the cooler is to enhance the absorption due to the lower temperature. In order to keep the absorption performance constant, an intercooler was integrated in the column. Thereby the investment cost of the process could be lowered while the performance of the process was kept constant. Normalized cost of CO_2-avoided of 0.96 in comparison to the reference case could be attained. With an exergoeconomic analysis it was not only possible to improve the processes but also new configurations could be synthesized.

5. Conclusion

An exergoeconomic analysis was performed to improve the cost effectiveness of different process configurations of post-combustion processes for CO_2 capture. The results revealed the components which can be improved at the expense of investment cost or efficiency. Besides an example for process improvement a new process configuration derived from the results was shown.

References

Bejan, A., Tsatsaronis, G., Moran, M., 1996, Thermal Design and Optimization, Wiley.

Oyenekan, B. A., Rochelle, G. T., 2007, Alternative Stripper Configurations for CO_2 Capture by Aqueous Amines, AIChe Journal, 53, 23, 3144 – 3154.

Peters, M. S., Timmerhaus, K. D., West, R. E., 2002, Plant Design and Economics for Chemical Engineers, McGraw-Hill Professional.

Schach, M.O., Schneider, R., Schramm, H, Repke, J.U., 2009, Techno-economic analysis of post-combustion processes for the capture of carbon dioxide from power plant flue gas, Ind.Eng.Chem.Res., 2010 (accepted).

Szargut, J., Morris, D. R., Steward, F. B., 1988, Exergy Analysis of Thermal, Chemical, and Metallurgical Processes, Springer.

Thompson, R. E., King, C. J., 1987, Energy Conservation in Regenerated Chemical Absorption Processes, Chem. Eng. Process, 21, 115 – 129.

Intensification of Sulfur Dioxide Absorption: Environmental and Economic Optimization

Patricia Luis, Aurora Garea, Ángel Irabien

Universidad de Cantabria, Avda. Los Castros s/n, Santander 39005, Spain,
luisp@unican.es

Abstract

Process intensification offers significant improvements in chemical manufacturing and processing, leading to cheaper, safer and sustainable technologies. Recovery of sulfur dioxide from gas emissions using an intensified process instead of the absorption by means of scrubbers is in the spotlight of many investigations. The substitution of the equipment by a membrane device intensifies the process from an environmental point of view, increasing process efficiency and reducing solvent losses, but the economic impact needs to be studied.

This work considers a ceramic hollow fibre membrane contactor as membrane device and the modeling of mass transfer in the membrane contactor is performed in order to establish the influence of the operation conditions on the process efficiency and to carry out an environmental and cost evaluation-optimization study.

Keywords: process intensification, sulfur dioxide recovery, membrane contactor, mass transfer, cost reduction-environmental optimization.

1. Introduction

Removal of sulfur dioxide from gas emissions by selective absorption is a common method to separate and concentrate sulfur dioxide and to reduce air pollution and environmental risks. Organic solvents (e.g. N,N-dimethylaniline) are typically used as absorption solvents due to their affinity and reversible interaction with sulfur dioxide leading to a regenerative process [1]. However, the use of scrubbers and other systems where a direct contact between the gas stream and the absorption liquid occurs produces some economic and environmental drawbacks due to solvent losses. The substitution of the absorption equipment by a membrane device intensifies the process [2, 3], increasing process efficiency and reducing solvent losses.

Membrane processes have great interest in recovery of target compounds from a gas stream due to widespread advantages [4, 5]: controlled interfacial area, independent control of gas and liquid flow rates and it avoids solvent losses due to drops dragging, which is a key factor because of environmental and economic considerations. However, the membrane introduces a new resistance to mass transfer and it must be considered in the process design and optimization.

In this work, the modeling of mass transfer in a ceramic hollow fibre membrane contactor using N,N-dimethylaniline as the absorption liquid, experimentally studied in previous works [6-8], allows establishing the operating conditions (gas and liquid flow

rates) and the process efficiency to fulfil a specified environmental target of sulfur dioxide recovery. A cost evaluation is performed in terms of investment costs, related to the membrane area, and operating costs, related to energy consumption.

2. Mass transfer device

The main features of the hollow fibre membrane contactor used in this study (Hyflux Ceparation BV, The Netherlands) are show in Table 1. Specific details of the experimental system can be found elsewhere [7].

Table 1. Hollow fibre membrane module.

Fibre material	α-Al$_2$O$_3$
Housing material	316 stainless steel
Potting material	Epoxy
Fibre o.d. (d_o), m	4×10^{-3}
Fibre i.d. (d_i), m	3×10^{-3}
Fibre length (L), m	0.44
Number of fibres (n)	280
Effective membrane area, m^2	0.8
Pore size of the fibre, nm	100

3. Mathematical modeling

In order to describe the mass transfer in the hollow fibre membrane contactor as a function of the operation conditions, the mass balance has been applied in the shell and tube sides, using the Happel's free surface model [9] in the shell side and the laminar regime model in the tube side, according to the flow configuration. Gas flows through the shell side countercurrently with the liquid flowing through the tube side. Mass transfer takes place through the membrane pores without mixing between phases and it is supposed that the membrane pores are filled with gas. The coordinates of a fibre are shown in Figure 1.
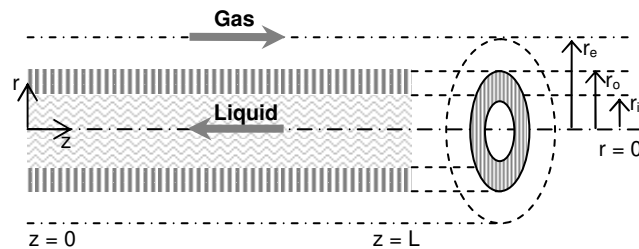
Figure 1. Axial and radial coordinates of a fibre.

The partial differential equations of mass balance for cylindrical coordinates are obtained using Fick's law of diffusion and the following assumptions have been considered: (1) steady state and isothermal condition; (2) no axial diffusion; (3) Happel's free surface model [9] to characterize the velocity profile in the shell side and laminar regime model in the tube side; (4) the physical properties of the fluid were constant; (5) constant shell-side pressures.

Intensification of Sulfur Dioxide Absorption Process: Environmental and Economic Evaluation

3.1. Mass transfer in the shell side

According to the Happel's free surface model [9], the following dimensionless equations have to be solved in order to describe the mass transfer behavior in the shell side:

$$\frac{Gz_{ext}}{2} \cdot f(\bar{r}) \cdot \frac{\partial \overline{C}_{SO_2,g}}{\partial \bar{z}} = \frac{1}{\theta} \frac{\partial}{\partial \theta} \left(\theta \frac{\partial \overline{C}_{SO_2,g}}{\partial \theta} \right) \tag{1}$$

$$f(\bar{r}) = \left[1 - \left(\frac{r_o}{r_e} \right)^2 \right] \cdot \left[\frac{\left(\frac{\bar{r} \cdot r_o}{r_e} \right)^2 - \left(\frac{r_o}{r_e} \right)^2 + 2 \cdot \ln\left(\frac{1}{\bar{r}} \right)}{3 + \left(\frac{r_o}{r_e} \right)^4 - 4 \cdot \left(\frac{r_o}{r_e} \right)^2 + 4 \cdot \ln\left(\frac{r_o}{r_e} \right)} \right] \tag{2}$$

$$\theta = \frac{r_e}{r_o}, \quad \frac{\partial \overline{C}_{SO_2,g}}{\partial \theta} = 0 \tag{3a}$$

$$\theta = 1, \quad \frac{\partial \overline{C}_{SO_2,g}}{\partial \theta} = Sh_m \cdot (\overline{C}_{SO_2,g} - \overline{C}_{SO_2,l}) \tag{3b}$$

$$\bar{z} = 0, \quad \overline{C}_{SO_2,g} = 1 \tag{3c}$$

where r_e is the free surface radius defined as:

$$r_e = \left(\frac{1}{\phi} \right)^{0.5} \cdot r_o \tag{4}$$

and φ is the packing density of the module, calculated as:

$$\phi = \frac{n \cdot r_o^2}{r_{cont}^2} \tag{5}$$

where n is the number of fibres and r_{cont} is the radius of the hollow fibre contactor. The following dimensionless parameters have been considered:

$$\theta = \frac{r}{r_o}, \qquad \bar{z} = \frac{z}{L}, \qquad \overline{C}_{SO_2,g} = \frac{C_{SO_2,g}}{C_{SO_2,in}} \tag{6a}$$

Sherwood number:

$$Sh_m = \frac{K_m S \cdot r_o}{D_{SO_2,g}} \tag{6b}$$

Graetz number in the shell side, with $d_o = 2r_o$:

$$Gz_{ext} = \frac{u_{m,g} d_o^2}{D_{SO_2,g} L} \tag{6c}$$

3.2. Mass transfer in the tube side

The model equations can be written in the dimensionless form as:

$$\frac{Gz_{int}}{2} \left[1 - \bar{r}^2 \right] \frac{\partial \overline{C}_{SO_2,l}}{\partial \bar{z}} = \frac{1}{\bar{r}} \frac{\partial}{\partial \bar{r}} \left(\bar{r} \frac{\partial \overline{C}_{SO_2,l}}{\partial \bar{r}} \right) \tag{7}$$

with the boundary conditions

$$\bar{r} = 0, \quad \frac{\partial \overline{C}_{SO_2,l}}{\partial \bar{r}} = 0 \tag{8a}$$

$$\bar{r} = 1, \quad \frac{\partial \overline{C}_{SO_2,l}}{\partial \bar{r}} = \frac{\partial \overline{C}_{SO_2,g}}{\partial \bar{r}} \cdot \frac{D_{SO_2,g}}{D_{SO_2,l}} \cdot H \tag{8b}$$

$$\bar{z} = 1, \quad \overline{C}_{SO_2,l} = 0 \tag{8c}$$

where the dimensionless variables are defined as

$$\bar{r} = \frac{r}{r_i}, \qquad \bar{z} = \frac{z}{L}, \qquad \overline{C}_{SO_2} = \frac{C_{SO_2}}{C_{SO_2,sat}} \tag{9a}$$

Graetz number in the tube side, with $d_i = 2r_i$

$$Gz_{int} = \frac{u_{m,l} d_i^2}{D_{SO_2,l} L} \tag{9b}$$

and the dimensionless mixing cup is calculated as:

$$\overline{C}_{SO_2,l,z=0} = 4 \int_0^1 \overline{C}_{SO_2,l} \left[1 - \bar{r}^2 \right] \bar{r} \, d\bar{r} \tag{10}$$

The numerical solutions of Equations (1) to (10) are obtained using the commercial software Aspen Custom Modeler (Aspen Technology Inc., Cambridge, MA.). The discretization was carried out in the radial and axial directions, considering two radial directions: r and θ and countercurrent flow. The 4th Order Central Finite Difference (CDF4) was applied for both axial and radial directions. The parameters to solve the model are shown in Table 2.

Table 2. Model parameters.

$D_{SO_2,DMA}$ (m^2·s^{-1})^a	2.10×10^{-9}
$D_{SO_2,g}$ (m^2·s^{-1})^b	1.26×10^{-5}
H^c	1.31×10^{-3}
$Sh_m{}^d$	1×10^{-3}

a Diffusion coefficient of SO_2 in N,N-dimethylaniline. Estimated from the literature [10].
b Diffusion coefficient of SO_2 in air. Estimated from the literature [10].
c Calculated from the industrial process for a gas stream with 5 vol.%SO_2.
d Experimentally obtained from previous works [7].

4. Results and Discussion

4.1. Influence of operation conditions on the process efficiency
Figure 2 shows the process efficiency as a function of the Graetz number referred to the shell side (Gz_{ext}), i.e. gas phase, and to the tube side (Gz_{int}), i.e. liquid phase. It can be observed that the higher the Gz_{ext}, the lower the process efficiency. The residence time of the gas phase in the contactor decreases when the gas flow rate increases, leading to a poorer mass transfer in the contactor. It can be also observed that to achieve specific process efficiency, an increase in the liquid flow rate involves an increase in the gas treatment capacity but a limit is found. When $Gz_{int} > 1$, the treatment capacity is not improved because the concentration of sulfur dioxide in the liquid phase is very far from the saturation value and it does not have any influence on the absorption process.

Intensification of Sulfur Dioxide Absorption Process: Environmental and Economic Evaluation
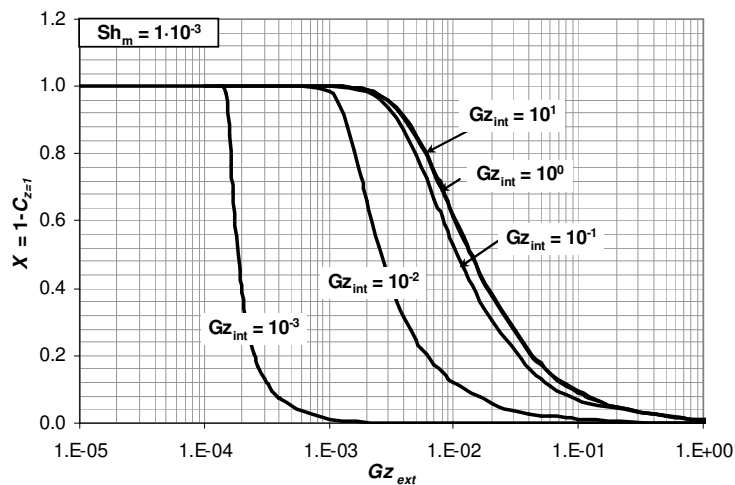
Figure 2. Influence of operation conditions on the process efficiency (Gz_{ext} refers to gas phase; Gz_{int} refers to liquid phase).

4.2. Environmental and economic optimization

The evaluation of the operation conditions on sulfur dioxide recovery is important to fulfil technical requirements but an evaluation in terms of environmental and economic considerations is needed to determine the application of the studied system. Figure 3 shows the total cost, calculated as the summation of the operation costs (60 €/(m^3/h) [11]) and the investment cost based on the membrane area required (613 €/m^2). An operation time of 20 years has been considered in the study.
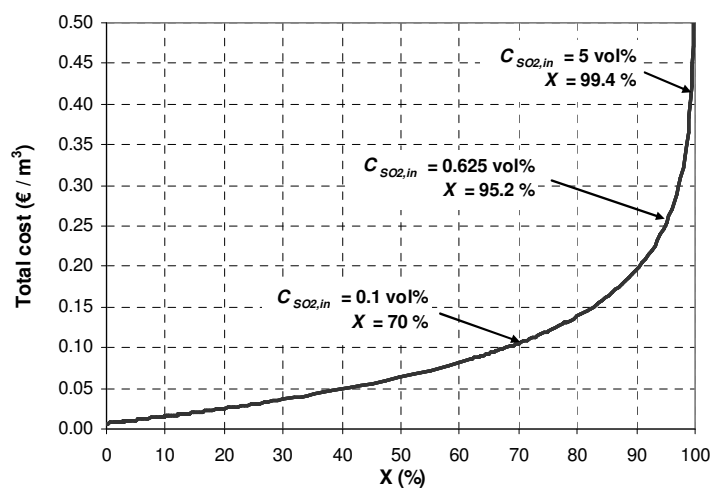
Figure 3. Total cost as a function of the process efficiency (X). Examples are calculated when $C_{SO2, out}$ = 0.03 vol% is required.

The emission limits determine the process efficiency that has to be achieved, which also depends on the concentration of sulfur dioxide at the inlet of the membrane contactor. As a first look, the higher the inlet concentration, the higher the process efficiency

required. Figure 3 shows three cases for different inlet concentrations but with the same environmental target: achieving a concentration of sulfur dioxide at the outlet of the contactor of 0.03 vol%. It can be observed that the higher the inlet concentration, the higher the cost treatment per volume unit. However, the sulfur dioxide produced is superior for streams with elevated load of sulfur dioxide, which could increase the interest of membrane systems for highly concentrated streams depending on the market. A deeper analysis is currently being performed to compare the trade-off solutions.

5. Conclusions

An optimization procedure based on a diffusion-controlled mass transport has been applied to the sulfur dioxide absorption process using a hollow fibre membrane module and N,N-dimethylaniline as the absorption liquid. The model allows the selection of the best operation conditions, expressed in terms of the Graetz numbers in the shell and tube sides, according to requirements.

The evaluation of the membrane process shows that the inlet concentration of sulfur dioxide is an important variable in the decision making from the point of view of environmental and economic considerations, mainly depending on the environmental restrictions.

6. Acknowledgements

References

[1] S. Bhattacharya, BK. Dutta, M. Shyamal, RK. Basu, 1996, Absorption of sulfur dioxide in aqueous dispersions of dimethyl aniline, The Canadian Journal of Chemical Engineering, 74, 339 – 346.

[2] A. Criscuoli, E. Drioli, 2007, New metrics for evaluating the performance of membrane operations in the logic of process intensification, Industrial and Engineering Chemistry Research, 46, 8, 2268 – 2271.

[3] J.A. Moulijn, A. Stankiewicz, J. Grievink, A. Gorak, 2008, Process intensification and process systems engineering: A friendly symbiosis, Computers and Chemical Engineering, 32, 3 – 11.

[4] A. Gabelman, S. Hwang, 1999, Hollow fiber membrane contactors, Journal of Membrane Science, 159, 61 – 106.

[5] A. Mansourizadeh, AF. Ismail, 2009, Hollow fiber gas-liquid membrane contactors for acid gas capture: A review, Journal of Hazardous Materials, 171, 38 – 53.

[6] P. Luis, I. Ortiz, R. Aldaco, A. Garea, A. Irabien, 2007. Recovery of sulfur dioxide using non-dispersive absorption, International Journal of Chemical Reactor Engineering 5, art. no. A52.

[7] P. Luis, A. Garea, A. Irabien, 2008. Sulfur dioxide non-dispersive absorption in N,N-dimethylaniline using a ceramic membrane contactor, Journal of Chemical Technology and Biotechnology, 83, 11, 1570 – 1577.

[8] P. Luis, A. Garea, A. Irabien, 2009. Zero solvent emission process for sulfur dioxide recovery using a membrane contactor and ionic liquids, Journal of Membrane Science, 330, 1 – 2, 80 – 89.

[9] J. Happel, 1959, Viscous flow relative to arrays of cylinders. AIChE Journal, 5, 2, 174 – 177.

[10] RH. Perry RH, DW. Green, 1997, Perry's Chemical Engineers' Handbook, 7th ed. McGraw-Hill New York, pp. 5–48.

[11] BREF (02.03), February 2003, Integrated Pollution Prevention and Control, Reference Document on Best Available Techniques in Common Waster Water and Waste Gas Treatment / Management Systems in the Chemical Sector, pp. 176.

Integrated Framework For Enterprise Management—A Synergistic Approach Towards Sustainable Biorefineries

P. Sharma[a], B.R. Sarker[b], J.A. Romagnoli[a]

[a]*Department of Chemical Engineering, LSU, Baton Rouge ,LA 70803, jose@lsu.edu*
[b]*Department of Industrial Engineering, LSU, Baton Rouge, LA 70803, bsarker@lsu.edu*

Abstract

The objective of the proposed formulation is to investigate the cross-functional links between different levels of a process enterprise, and how these links can be exploited in order to benefit a prototypical biorefining enterprise. A biorefining enterprise is characterized as an entity having three major layers in its functional hierarchy—the strategic planning layer, the tactical planning layer, and the production layer. The emphasis of this paper is on the strategic and the tactical planning layers. Each layer is modelled with an optimization model. The corporate layer is formulated as a mixed integer capital budgeting and network design problem while the supply chain layer is formulated as a resource allocation non-linear problem. A biorefining enterprise is used as a case study to highlight the utility of the proposed framework.

Keywords: Supply Chain Management, Enterprise-wide optimization, Cellulosic Ethanol

1. Introduction

A highly competitive environment in the process industry has compelled companies to explore cost-cutting measures in order to make their enterprise profitable. A nascent area concerning such an endeavour is enterprise-wide optimization. Modern process enterprises function as cohesive entities involving several degrees of cross-functional co-ordination across enterprise planning and process functions [1]. The complex organizational structures underlying horizontally and vertically integrated process enterprises challenges our understanding of cross-functional coordination and its impact on business. Provision of robust decisions support tools throughout the enterprise hierarchy can hence have a pronounced effect on its profitability.

A sustainable enterprise is often defined as an enterprise that does not have a negative socio-environmental impact on the society [2]. We further refine this definition to encompass not only the ability to positively impact the environment, but also maintaining such an impact through value creation and profitability. An enterprise is defined as being sustainable if it produces goods and services that benefit our environment and is able to preserve such an influence through continued growth. The National Renewable Energy Laboratory (NREL) has identified biorefineries to be the most promising route towards creation of a sustainable energy portfolio. While recent government initiatives and private undertakings have focused on developing process technologies to make a biorefinery more profitable, little emphasis has been laid on developing a robust supply chain for any biorefining enterprise. Recent studies [3] also indicate that supply chain design will play a key role in determining the commercial viability of cellulosic ethanol.

In this paper, for the first time, a model for a typical biorefining enterprise is formulated and implemented towards enterprise wide management. In the proposed model the

enterprise is represented with three interdependent, functional layers—the *corporate (strategic) layer, the supply chain (tactical) layer,* and *the production layer (operations).* Each layer is functionally dependant on the others for information for complete optimization of the model. Special care has been taken to define pertinent cross-functional linkages between these three planning layers while imparting as much realism to emulate an actual serviceable enterprise.

2. Problem Statement

Our work adopts a decentralized approach for supply chain management in order to mirror actual enterprise architecture and to make the decision support tools function autonomously. What sets our work apart from other similar decision support frameworks is the formulation of the optimization models for the three planning layers that aim to exploit the synergy between different levels of planning in order to satisfy different enterprise objectives for varying time horizons. A representation of the architecture is presented in Figure 1.

Long-Term (strategic) Financial Pan	Network topology, plant capacities, quarterly production and sales targets	Medium-Term (tactical) Plan	Production targets, Raw materials allocation	Short-Term (operations) Production Plan
Time-horizon 5 years		**Time Horizon** 3 Months		**Time Horizon** 2 weeks
Objective Maximize Enterprise Value		**Objective** Maximize Profit		**Objective** Minimize Production costs
Outputs Network Topology; Raw Materials portfolio; Financing of assets and debt obligations; Capacity planning;	Earnings before interest and taxes, inventory, production and sales profiles	**Outputs** Resource allocation; Inventory profiles; Profit ; Production targets; Raw materials and sales profiles	Production levels, operating costs	**Outputs** Plant set-points; Production levels

Figure 1: Interaction between different plans-long, medium and short

As a preliminary formulation, we will analyse the strategic and the tactical planning layers. We have reduced the solution space by devising our problem as follows: the enterprise is dedicated to the production of a single product, cellulosic ethanol (P1); there are three potential feedstocks that yield different amounts of biomass (R1, R2, R3); there are two potential technologies (TA, TB); product demand and prices are stochastic and divided into low, medium, and high values; there are three potential production facilities (S1,S2,S3); there are two potential blending facilities that the finished product can be transported to (B1, B2); there are five potential markets served by the enterprise (M2, M2, M3, M4, M5).

3. Model Formulation

As mentioned in the problem statement, the three levels of planning and optimization have different objectives, time horizons, and time steps. This may lead to inconsistent results as plans from one layer may contradict plans generated by other layers. To accommodate these differences, we suggest the following implementation: (a) Implement the strategic optimization module every quarter with updated input parameters in order to take into account the realization of the of the tactical and production plans, (b) Implement the tactical optimization module every two weeks but with a shrinking time horizon, and updated input parameters in order to take into

Integrated Framework For Enterprise Management—A Synergistic Approach Towards Sustainable Biorefineries

account the realization of the first time step of the previous tactical plan. Additionally, the constraint formulation for the supply chain layer assumes a high level of importance in our framework. Although profit maximization is the most important short-term goal, we want to keep perspective of our long-term goal of continued value creation. Hence, the constraints in the tactical plan are formulated such that the information contained in outputs generated by the strategic planning model is included in the tactical model in the form of constraints.

3.1. Strategic Planning

Strategic level planning involves deciding the configuration of the network, i.e., the number, location, capacity, and technology of the facilities [4]. In general, a biorefining supply chain network will consist of the following entities—lands for producing feedstocks for the biorefinery, transportation fleets to move raw materials and finished products, inventories for raw materials and finished products, production facilities, blending facilities, and markets. We derive our corporate valuation model following the principles proposed by Grant [5, 6]. The objective function defined in our valuation model is to maximize the shareholder's value and the method used to calculate the objective value is called discounted-free-cash-flow (DFCF) method. The major decision variables that the strategic planning model yields on a quarterly basis are as follows: raw materials portfolio and amounts; production and blending facility locations and capacities; inventory levels for raw materials and finished product; raw materials and finished product transportation amounts; capacity increments for production and blending facilities; securities transactions; and loan repayments.

Due to space limitations only a sample of constraints related to raw materials, debt obligations, material balances, and capacity increments, is provided.

$$RMHarvest_{lrt} - Crop_{lt} BiomassYield_{rt} (SeasonalAdjust_{rt} - Spoilage_{rt}) \leq 0$$

$$TotalDebt_t = TotalDebt_{t-1}(1 + irt) + AmountBorrowed_t - AmountRepaid_t$$

$$Productionm_{ist} - Inventory_{ist-1} = \sum_b ProductBlending_{isbt} + Inventroy_{ist} \qquad (1)$$

$$CapacityIncrement_{jst}^{UB} = \begin{cases} 37000, & t = 5,10,15,20 \\ 0 & , otherwise \end{cases}$$

3.2. Tactical Planning

The tactical plan assumes that the network topology, production and transportation capacities, and supplier and customer portfolio are known from the strategic planning model. These are then used as constraints by the tactical planning layer. The objective of the tactical plan is to maximize profits with a planning horizon of three months with biweekly time steps. Constraints and parameters used in the model include material balances similar in structure to the strategic planning model, and the minimum inventory stock levels (safety stock) required at all times. As an initial estimate, the sum of the biweekly demand forecasts for each market is forced to equal the quarterly demand for the strategic planning model and the biweekly prices for the product are formulated such that their average equals the product price input to the strategic planning model. Decision variables yielded by the model are production and inventory profiles for each facility, transportation quantities for each transportation link, and sales profiles for each market.

Operational targets yielded by the strategic plan are used as constraints by the tactical planning model. We formulate the targets set by the strategic plan as "less than" or "greater than" constraints in order to provide a greater solution space for profit maximization. Keeping in mind certain physical limitations on the constraint development, the production and raw material harvest targets are expressed as "less than or equal to" constraints and the inventory and sales levels are expressed as "greater than or equal to" constraints. A sample of the constraints is provided below.

$$\sum_t BiweeklyProduction_{ist} \leq ProductionTargetStrategicLevel_{is}$$

$$\sum_t BiweeklySales_{ibmt} \geq SalesTargetStrategicLevel_{ibm}$$

(2)

3.3. Results and Discussion

Table 1 shows the breakdown of parameters associated with raw material selection while Table 2 is a summary of the costs associated with opening and operating a facility.

Table 1: Parameter breakdown for raw materials

Parameter / Raw Material	R1	R2	R3
biomass yield (kg / 1000kg)	607	689	600
Product yield (TA) (gal / kg)	0.0779	0.0885	0.0772
Product yield (TB) (gal.kg)	0.0809	0.0905	0.0797
Cost ($ / kg)	0.130	0.122	0.119

Table 2: Costs parameters for network design

Parameter	S1	S2	S3
Investment in (TA) ($/gal)	0.24	0.24	0.28
Investment in TB ($/gal)	0.32	0.30	0.32
Fixed Costs ($/gal)	0.26	0.31	0.15

Table 3 lists the optimal plant capacities, and the raw material usage for the biorefining network. As can be seen, plant S3 has the highest planned operating capacity while S2 has the lowest. The reasoning is evident when one analyzes the high investment and fixed costs associated with S2.

Table 3: Strategic planning outputs for network topology

Strategic Plan Enterprise Value: $81,783,408			
Sites	S1	S2	S3
Plant Capacities (1000 gallons)			
t=1	18529.500	11337.700	36371.500
Raw material usage (1000 kg)			
Raw Material	R1	R2	R3
t = 1	165220	286380	285000

What is not evident is capacity difference between S1 and S3 even though S3 requires a higher investment. This can be attributed to the *low fixed costs* associated with the operation of S3 as compared to S2. The raw materials portfolio has highest usage of R3 and low usage of R1. This can attributed to the high biomass and product yield associated with R2 which offsets its high cost of procurement. Despite mediocre

biomass and product yields, R1 has low usage since it has a high cost of procurement. R3 usage is higher as it has the lowest cost of procurement.

Based on the operating cost parameters of the technologies in question, the strategic optimization model prescribes using technology TA at sites S2 and S3, and using technology TB at site S1. This can be attributed to the fact that *product yield* for TB is higher and the investment costs for TA at site S3 are marginally higher than those at S1 and S2. In real life, this can happen due to different state tax structures that may support of even promote a certain type of technology more than the others. Despite its high investment costs it is found that it is optimal to maintain a diversified portfolio of technologies. The difference in investment costs between TA and TB is not sufficient enough in the case of S1 and S2 to make up for the higher product yields for TB. Another aspect considered in the model was the *time value of money*. Given that the value of money reduces by 1% every year, the strategic plan entails investing in the technologies at the beginning of the time horizon, even though product demands are much higher in later periods. Also, the optimizer rejects four of the five markets based on low demands and high costs of transportation. The sales volumes and production amounts as expected are low in low and medium demand scenarios, while they are high in the high demand scenario. Table 4 lists the forecasted sales volumes and production amounts for market M3, and site S3 respectively.

Table 4: Snapshot of operational decision variables yielded by the strategic planning model

Decision Variable	Scenario	High	Medium	Low
Production (1000gal/quarter)		18908	6364	6364
Sales (1000gal/quarter)		19454	6909	6909

As can be seen from the table, the majority of the sales in the time period are from site S3. Also noteworthy is the facts that even though product demands are low in the beginning, plants are built to a much higher capacity than the actual product demand. The plants did not run at full capacity in the beginning, but given the time-value of money, the optimizer prescribes adding all the capacity at the beginning of the time horizon in anticipation of elevated product demands late on. There are pitfalls to this approach; if the product demand does not materialize, the enterprise can suffer huge losses due to massive capacity investments up front. Therefore, it is imperative to have accurate long-term demand forecasts which ironically are fraught with large uncertainties. Consequently, a *real-options* analysis has been suggested in the *conclusions* section of this paper for portfolio evaluations and capacity planning to overcome such issues and is an area currently being investigated by the authors.

As an exercise to demonstrate the utility of the tactical plan, we simulated a feedback loop with a disturbance in production. It was assumed that the realization of the tactical plan followed the optimized results for the first three time steps. The resulting actions were fed back to the plan. A fault was then assumed to have occurred in the production facilities, S2 and S3, leading to a seven percent reduction in their capacity utilization rate. As expected, the production plans generated showed a reduction of seven percent but there was a reallocation of the sales plans in order to maximize the overall profit for the 3 period time horizon.

Also there is a seven percent reduction in the production and raw material consumption plans. The raw material allocation profiles, product shipping profiles, and the sales

profiles were reformulated to accommodate the disturbance while maximizing network profit for a reduced time horizon of three periods. Figure 3 compares the original and the reformulated profiles for market, M3, and production facility, S3.
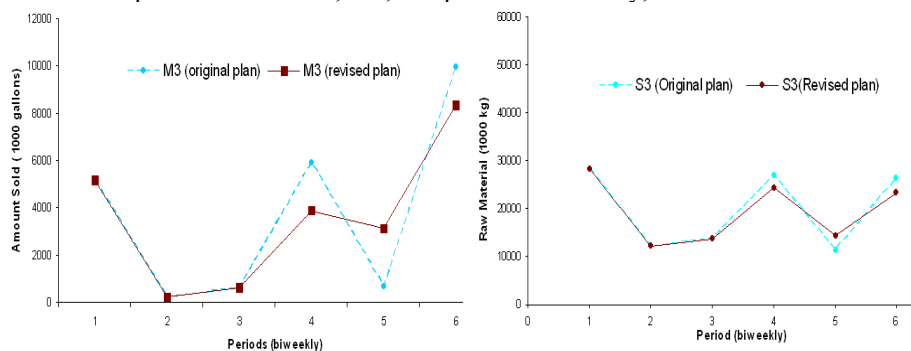
Figure 3: Sales plans for M3 and resource allocation for S3

The ethanol obtained in the process simulations was separated up to 95% purity by means of distillation. A set of two columns in series was used. 99% pure anhydrous ethanol was then obtained by means of dehydration.

4. Conclusions and future work

A preliminary framework of modeling a prototypical biorefining enterprise was considered. The framework proposed is modular in nature and can react efficiently to real-world disturbances. Results show a distinct shift in an enterprise's operational plans due to disturbances. Future work will concentrate on expanding the framework to a multi-product biorefinery network, using *real-options analysis* to evaluate enterprise portfolios and capacity decisions, adding more rigorous stochastic forecasting modules and modeling individual components of a supply chain such as inventories and transportation resource planning keeping in mind the complex characteristics of a bioreinfing supply chain.

References

[1] Varma V., Reklaitis G., Blau G., Pekny J. (2007), "Enterprise-wide modeling & optimization—an overview of emerging research challenges and opportunities," *Computers and Chemical Engineering*, **31**: 692–711.

[2] Petrini M., Pozzebon M. (2009), "Managing sustainability with the support of business intelligence: Integrating socio-environmental indicators and organizational context," *Journal of Strategic Information Systems*, doi:10.1016/j.jsis.2009.06.001.

[3] Slade R., Bauen A., Shah N. (2009), "The commercial performance of cellulosic ethanol supply-chains in Europe," *Biotechnology for Biofuels*, Online Journal, **2**(3)

[4] Naylor T. (1981), "How to Integrate Strategic Planning into your Management Process," *Long Range Planning,* **14**(5): 56-61.

[5] Grant J. (2002), "Enterprise Valuation I: Free Cash Flow Model," Chapter 6: 105-128; *Foundations of Economic Value Added*, 2nd Edition, New Hope: Frank & Fabozzi, Wiley & Sons.

[6] Puigjaner L., Lainez M. (2008), "Capturing dynamics in integrated supply chain management," *Computers and Chemical Engineering*, 32(11): 2582–2605.

Permeable reactive barrier for groundwater PCE remediation: the case study of a solid waste landfill pollution

Armando Di Nardo,[a] Michele Di Natale, [a] Alessandro Erto,[b]
Dino Musmarra,[a] Immacolata Bortonea

[a] *Centro Interdipartimentale di Ricerca in Ingegneria Ambientale, Seconda Università di Napoli, via Roma 29, Aversa (CE), 81031, Italy, armando.dinardo@unina2.it*
[b] *Dipartimento di Ingegneria chimica, Università di Napoli Federico II, P.le Tecchio, 80-80125 Naples, Italy, aleserto@unina.it*

Abstract

The remediation of a tetrachloroethylene (PCE) contaminated aquifer near a solid waste landfill, by an activated carbon Permeable Reactive Barrier (PRB) is presented as a case study. A 2D numerical model has been used to describe the pollutant transport within a groundwater and the pollutant adsorption on the barrier. The results show that the barrier has a good efficiency since the PCE concentration flowing out of the PRB is always lower than the limits provided for in the currently enforced Italian legislation.

Keywords: PCE removal, groundwater remediation, PRB.

1. Introduction

The presence of tetrachloroethylene (PCE) in groundwater is due to anthropogenic sources (e.g. industrial discharges from manufacturing activities, solid waste landfill leachate, etc.) and results in a severe alteration of water natural properties. PCE can be found in groundwater both in NAPL form (non-aqueous phase liquid) and dissolved in water. The physical and chemical properties, as well as the hydraulic characteristics of the surrounding media, greatly influence its mobility and persistence in groundwater. Adsorption treatments of contaminated groundwater are widely used for PCE pollution control; these treatments can be performed ex-situ, coupled with *pump and treat* techniques, or in situ with Permeable Reactive Barriers (PRB). In a PRB treatment, the barrier is commonly built with a reactive material having a higher hydraulic conductivity than the surrounding soils, so that the contaminated groundwater is forced to pass through the barrier itself, moving under natural hydraulic gradient. The mechanism of action of a PRB depends on the reactive material chosen to build the barrier. Usually the material used is zero-valent iron (EPA, 1998; Vogan et al., 1999; D'Andrea et al., 2005) in which the pollutants undergo a series of reduction reactions to achieve a complete degradation. These reduction reactions wear the barrier and may induce solid precipitation which can seriously affect the barrier efficiency, diminishing its porosity and conductivity. In this sense, the use of sorbent materials seems to be a valid alternative (Lorbeer et al., 2002, Di Natale et al., 2008). In fact, the removal of chlorinated organic compounds, such as PCE, from polluted water and wastewater can be efficiently achieved with an adsorption process that combines process efficiency and configuration simplicity (Suzuki, 1990). As adsorption phenomena take place, the pollutant is immobilized into the barrier, avoiding precipitation phenomena. In PRB

design and optimization, the hydrological and geotechnical properties of the polluted aquifer must be considered and a thorough characterization of the site is required.

In this work, the remediation of a PCE-contaminated aquifer by an activated carbon PRB has been studied. A polluted aquifer near a solid waste landfill in Giugliano in Campania (Italy) is presented as a case-study. The CFD (Computational Fluid Dynamics) approach permits to predict barrier performances in a wide range of working conditions, providing the spatial and temporal distribution of the contaminant plume.

2. PRB design

Before designing a barrier it is necessary to properly characterize the site, to assess the contaminant properties, distribution and tracking; to describe the fluid dynamics within the aquifer; to determine the chemical-physical phenomena involved in the adsorption process and to meaningfully represent the results Hence, mathematical modelling becomes an essential tool to predict barrier performance in different working conditions. The design of a barrier consists in the definition of location, orientation and dimensions; in particular its width allows to calculate the total GAC required. The residence times of the contaminated flow travelling through the barrier should be long enough for adsorption process to take place. Therefore, the barrier width (W) must satisfy the following inequality:

$$\frac{W}{u_b} > \left(k_c a\right)^{-1} \tag{1}$$

In the above equation u_b represents groundwater flow velocity through the barrier, k_c the overall mass transfer coefficient for adsorption reactions and a represents the external specific surface of the adsorbent particles.

It must be considered that in-flowing concentrations may vary during the barrier working period and that desorption phenomena may occur within the barrier when the PCE in-flowing concentration is lower than the equilibrium values corresponding to the concentration of the carbon-adsorbed PCE. When these conditions occur, a wider barrier ensures a gradual release of the PCE adsorbed, avoiding critical out-flowing concentrations. Therefore, the barrier must be designed both to retain intense concentration peaks and to guarantee long term performances.

2.1. Modeling equations

The contaminant migration in a porous medium is due to advection–dispersion processes; therefore, considering a two-dimension system, the dissolved PCE mass balance equation may be written as follows:

$$\frac{\partial C}{\partial t} + \frac{\rho_b}{n_b}\frac{\partial \omega}{\partial t} + \frac{\bar{u}\nabla C}{n_b} = \nabla\left(D_h \nabla C\right) \tag{2}$$

In (2) C represents PCE concentration in fluid, $\bar{u}$ the unit flux vector, ω the PCE concentration on solid, ρ_b the dry adsorbing material bulk density, n_b the soil porosity. The hydrodynamic dispersion coefficient D_h can be shown to be a second-rank tensor expressed as:

$$D_h = D + D_d^* \tag{3}$$

In eq. (3), D is the tensor of mechanical dispersion and D_d^* is the coefficient of molecular diffusion (a scalar), respectively.

Permeable reactive barrier for groundwater PCE remediation: the case study of a solid waste landfill pollution

The $\bar{u}$ the unit flux vector in eq. (2) can be determined by the application of the Darcy equation, written as:

$$\bar{u} = -K_{sat} \cdot \nabla h \tag{4}$$

where the hydraulic load can be calculated starting from the Laplace equation:

$$-\frac{\partial^2 h}{\partial x^2} - \frac{\partial^2 h}{\partial y^2} - \frac{\partial^2 h}{\partial z^2} = 0 \tag{5}$$

that can be integrated with appropriate boundary conditions. The second term on the left hand side of (2) reads:

$$\frac{\rho_b}{n_b} \frac{\partial \omega}{\partial t} = k_c a \left[C - C^*(\omega, C) \right] \tag{6}$$

In eq. (6), $C^*(\omega, C)$ derives from the adsorption isotherm and defines the mass transfer driving force in the transport model equations.

The initial PCE liquid concentrations are determined and the PCE solid concentrations are assumed to be zero, throughout the entire flow domain:

The boundary conditions are as follows:

$$C = 0 \qquad\qquad if\ x = 0\ \forall y\ \forall t\ OR\ y = 0\ \forall x\ \forall t\ OR\ y = Y\ \forall x\ \forall t$$

$$\frac{\partial C}{\partial t} + \frac{\bar{u} \nabla C}{n_b} - \nabla(D_h \nabla C) = 0 \quad x = X \quad \forall y\ \forall t \tag{7}$$

where X is the distance between the barrier and the west boundary of the domain, and Y is the extension of the domain in y direction.

The numerical integration of eqs. (2)–(6) with the boundary conditions (7) has been carried out by means of a first order finite difference implicit scheme using a commercial 2D model flow, PMWIN (by U.S. Geological Survey). Specifically PMWIN MODFLOW toolbox solves Laplace equation (5) and Darcy equation (4) and PWMIN MT3D toolbox solves transport equation (2).

2.2. Adsorption characterization

The solid used for the barrier set-up is a commercially available non impregnated granular activated carbon (GAC), Aquacarb 207EA^{TM} (Sutcliffe Carbon). A complete solid characterization has been carried out. This material has a BET surface area of 950m^2/g and an average pore diameter of around 26Å. The dry bulk density (ρ_b) is 500kg/m^3, the porosity (n_b) is 0.4m^3/m^3 and its hydraulic conductivity is about 0.001m/s. The PCE-Aquacarb 207EA adsorption isotherm at a temperature of 10 °C has been reported in a previous paper (Erto et al., 2009). Specifically the following Langmuir isotherm model has been used, based on experimental data with ω_{max}=913.9 (mg/g) and K=19,830 (l/mol):

$$\omega = \frac{\omega_{max} K \cdot C}{1 + K \cdot C} \tag{8}$$

A Visual Basic Application code (ADSORP-CODE) has been specifically developed to describe adsorption phenomena involving the pollutant.

2.3. PRB sizing

PRB sizing is obtained iteratively, as reported schematically in the flow chart of Figure 1. Specifically, after defining boundary conditions and all Input data of PMWIN such as Δx, Δy, Δz (computational grid), K_{sat} (hydraulic conductivity), n_b (soil porosity), h (hydraulic height), C_0 (initial pollutant concentrations), D_h (hydrodynamic dispersion coefficient), k_c (overall mass transfer coefficient for adsorption reactions) and a (external specific surface of the adsorbent particles), a first MODFLOW simulation is carried out, fixing X (PRB distance from pollutant plume) and L (PRB length) and choosing W (PRB depth) and m (PRB angle), in order to calculate $h(x,y,z,t)$ and $u(x,y,z,t)$. The barrier must be designed so as to be long enough to intercept the whole pollutant plume.

Then m is changed until velocity direction $\bar{u}$ is orthogonal to PRB less than angle tolerance Δm. Next step is MT3D simulation to compute pollutant concentration $C(x,y,z,t)$. Then $C_W(t)$ is computed by means of ADSORP-CODE and if this value is lower than a limit value (C_{lim}), PRB depth is correct, otherwise it must be increased until $C_W(t)<C_{lim}$.
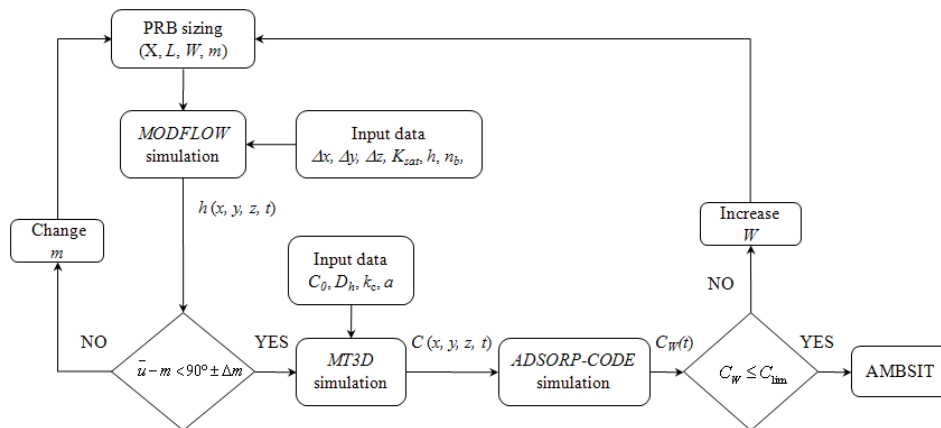
Figure 1. PRB design Flow chart

2.4. Numerical simulation results

Results are graphically represented show is obtained by means of AMBSIT$^©$ (by CIRIAM, Second University of Naples), a GIS (Geographic Information System) application specifically developed to improve contour plots of pollutant concentration.

In this way, it is possible to simplify results analysis with thematic spatial and time maps. AMBSIT$^©$ allows to draw concentration contour plots with different interpolation algorithms.

3. Case study

The case study refers to a large area (2.25km^2) in Giugliano in Campania in the metropolitan area North of Napoli (Italy), where various solid waste landfills exist. Over the past 20 years, about eight million tons of urban and special wastes were deposited, both legally and illegally, in these landfills.

The groundwater aquifer, located at a depth of 35-40m from the land surface and confined by an aquitard (-50m), is contaminated by a large number of pollutants, both inorganic and organic (ISOGEA, 2006). The soil composition can be approximated with a unique mineral type, whose hydraulic conductivity is $5 \cdot 10^{-5}$m/s (ISOGEA, 2006). The

Permeable reactive barrier for groundwater PCE remediation: the case study of a solid waste landfill pollution

groundwater flux lines are EW oriented under a piezometric gradient of 0.01m/m. In Figure 2a, the PCE isoconcentration in actual conditions have been reported, together with the piezometric lines of the aquifer.

The PCE concentration values are about 20 times higher than the Italian regulatory limits for groundwater quality, fixed at 1.1µg/l. Thus, a defined volume of contaminated groundwater can be identified and the PRB design can be approached. The barrier considered is a continuous trench penetrating the aquifer at full-depth (50m), up to the aquitard, perpendicular to the groundwater flow to ensure that the most efficient capture of the plume is obtained. To the same purpose, the distance between the barrier and the boundary of the pollutant plume has been established at 6m.

4. Results

A great number of numerical simulations, in different working conditions, have been necessary to determine the optimal barrier position and dimensions. The best results have been obtained for a barrier 3m wide and 900m long, as shown in Figure 2.

| C[µg/l] | 0 – 0.001 | 0.001 – 1.1 | 1.1 - 5 | 5 - 15 | 15 - 25 |
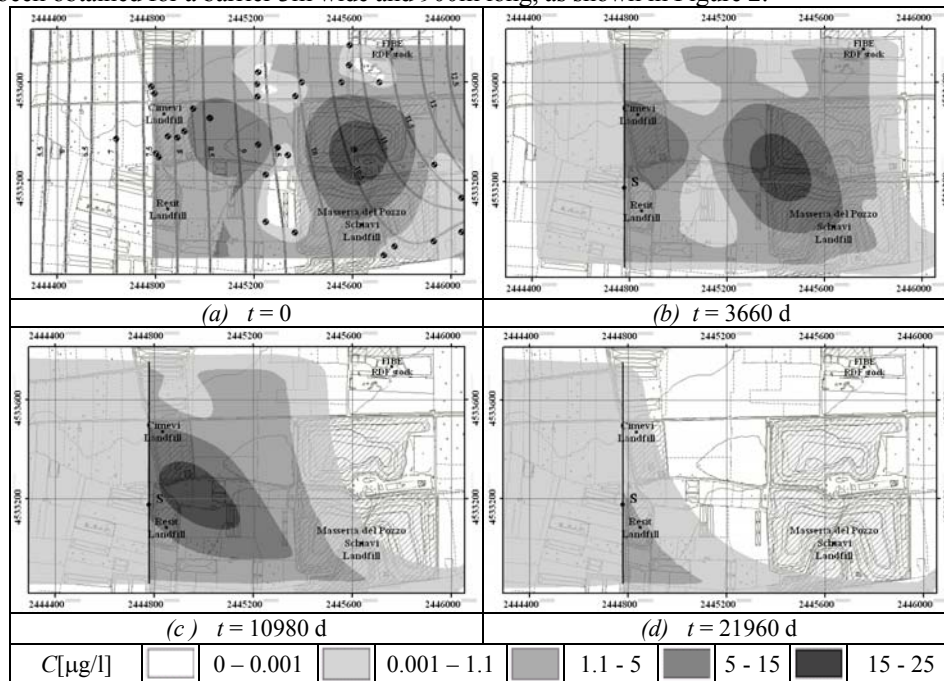
Figure 2. PCE iso-concentration and iso-piezometric lines for the case study

In the same figure, the numerical results, in terms of PCE polluted plume evolution as a function of run time, are reported. Figure 2 clearly shows that during the run period of about 60 years, the out-flowing concentrations are always lower than the regulatory limit. It is worth noticing that even when the adsorbing barrier complies with the standards the desorption of some previously captured PCE may occur.

This result is better illustrated in Figure 3 where the PCE concentration profile inside the barrier is reported in the form of a breakthrough curve.

In particular, Figure 3 reports, for point S in Figure 2, the PCE concentration at the inlet (C_{IN}) and at the outlet (C_W) of the barrier, as a function of time. Curves in Figure 4 clearly show that the PCE concentration flowing out of the barrier is always lower than

the standard limit, even when the inlet PCE concentration decreases (after a run time of about 20 years) and the desorption of previously captured PCE occurs.
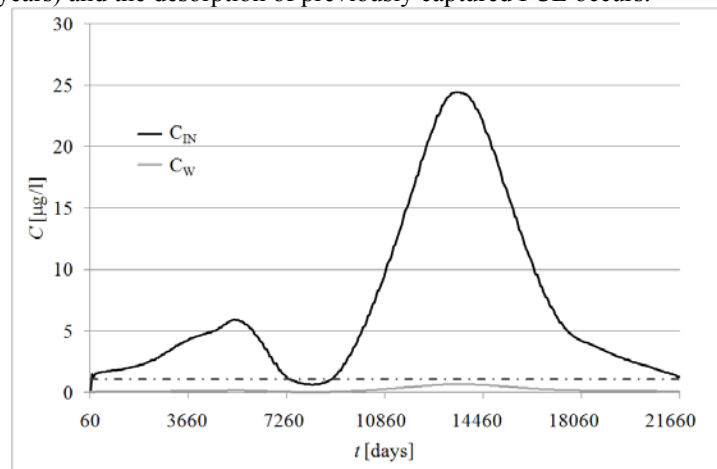
Figure 3. PCE concentrations in the barrier point S.

5. Conclusion

This work reports an approach to the design of an activated carbon Permeable Reactive Barrier (PRB). A case study, the remediation of a PCE polluted aquifer, has been considered in order to define the optimal barrier position and dimensions. During a working period of about 60 years in which first adsorption and then desorption take place, the out-flowing concentration is always lower than the regulatory limits. This result suggests that adsorption barriers can be considered for the remediation of the Giugliano in Campania (Napoli) site although some technological aspects, pollutant synergy and long term effects have yet to be taken into account.

References

EPA (U.S. Environmental Protection Agency), 1998, Updated Health Assessment Document for Tetrachloroethylene, EPA/600/8-82/005B.

EPA (U.S. Environmental Protection Agency), 1998, Permeable Reactive Barrier Technologies for Contaminant Remediation, EPA/600/R-98/125.

A. Erto, R. Andreozzi, F. Di Natale, A. Lancia and D. Musmarra, 2009, Experimental and isotherm model analysis on TCE and PCE adsorption from model water solutions – Chemical Engineering Transactions 17, 293-298.

P. D'Andrea, K.C. Lai, P. Kjeldsen and I.M. Lo, 2005, Effect of groundwater inorganics on the reductive dechlorination of TCE by zero-valent iron, Water Air Soil Pollution 162, 401-420.

F. Di Natale, M. Di Natale, R. Greco, A. Lancia, C. Laudante and D. Musmarra, 2008, Groundwater protection from cadmium contamination by permeable reactive barriers, Journal of Hazardous Materials 160, 428-434.

ISOGEA srl, 2006, Indagini preliminari sul sito di Masseria del Pozzo e Ampliamento di Giugliano in Campania, Rapporto interno del Comune di Giugliano in Campania.

H. Lorbeer, S. Starke, M. Gozan, A. Tiehm and P. Werner, 2002, Bioremediation of chlorobenzene-contaminated groundwater on granular activated carbon barriers, Water air Soil Pollution Focus 2,183-193.

M. Suzuki, 1990, Adsorption Engineering, Elsevier, Amsterdam.

J.L. Vogan, R.M. Focht, D.K. Clark and S.L. Graham, 1999, Performance evaluation of a permeable reactive barrier for remediation of dissolved chlorinated solvents in groundwater, Journal of Hazardous Materials 68, 97–108.

Hybrid semi-parametric modeling of biological systems: Application to spectroscopic data for the estimation of concentrations

von Stosch M.,a Oliveira R.,b Peres J.,a Feyo de Azevedo S.,a

a LEPAE, Departmento de Engenharia Quimica, Faculdade de Engenharia, Universidade do Porto, Rua Dr. Roberto Frias s/n, 4200–465 Porto, Portugal.
b REQUIMTE, Departamento de Quimica, Faculdade de Ciências e Tecnologia, Universidade Nova de Lisboa, 2829-516 Caparica, Portugal.

Abstract

In this work, bioprocess monitoring based on spectral data is improved when compared to commonly applied chemometric tools, by merging nonparametric modeling, biological and process *a priori* knowledge into a hybrid semi-parametric model. This particular semi-parametric structure comprises a nonparametric submodel inspired by a NPLS structure, as NPLS has been reported to be successful for dealing with massive numbers of highly correlated spectral data. The method was applied to Bordetella pertussis cultivations equipped with a Near-InfraRed (NIR) probe, showing that estimates of metabolite concentrations are improved when compared to those obtained through classical chemometric modeling, as expressed by lower mean square errors, better calibration properties and a higher statistical confidence.
Keywords: Dynamic modeling, Hybrid modeling, NIR, dynamic nonlinear PLS

1.Introduction

Many biopharmaceutical industries are implementing the new Process Analytic Technology (PAT) guidelines. The first steps therein aim at a better on-line characterization of the process state by implementing advanced monitoring techniques. Spectroscopy techniques such as Near-InfraRed, mid-InfraRed, Terahertz, Raman InfraRed, Fourier-Transform InfraRed or Fluorescence spectroscopy have been widely reported in the context of PAT since they are fast, non invasive, non destructive, and amenable to chemically complex multiphase reaction media. The availability of such analytical devices provides extensive spectral data sets holding complex molecular scale information. To date, chemometric modeling tools such as Principal Component Analysis (PCA), Partial Least Square(PLS) and its nonlinear counterparts (nonlinear-PLS and Kernel-PLS) are applied to deconvolve the complex spectra and to correlate with target state variables. These chemometric techniques present, however, the limitation that they do not incorporate a priori knowledge about the target biological system, where in contrast it is envisaged that such information rich spectral data will in future be integrated with Systems Biology models and macroscopic process operation (Teixeira et al, 2007b).

The integration of spectral data and fundamental biological models is hindered by the fact that such data has no direct physical meaning. Hybrid semi-parametric systems can provide the ideal mathematical framework for bringing together biological and process mechanisms along with the data from such analytical devices. In general the knowledge

can be arranged in parallel or serial, the latter being particularly suitable for complex systems for which large data sets are available without direct physical interpretation (Teixeira et al. 2007b). Biological constraints such as metabolic reactions connectivity in the form of elementary modes (Teixeira et al., 2007a) can also be included in the hybrid structures, paving the way for on-line characterization of the fluxome.

In traditional hybrid models usually Artificial Neural Networks (ANN) found application (Oliveira 2004, Thomson & Kramer 1994, Psichogios & Ungar 1992). However when high-dimensional data are considered as inputs to the ANNs then the number of ANN parameters quickly exceeds the number of measured target state variables, which leads to a clearly underdetermined system of equations. On the other hand (N)PLS models cannot be directly integrated into the serial hybrid structure as for the training of such the kinetic rates would have to be known, and their estimation from noisy and sparse data is prone to errors (e.g., Oliveira 2004). In this article a NPLS like structure is integrated into a serial hybrid model structure and an algorithm for its parameter identification is developed. For the evaluation against a dynamic PLS model, the hybrid structure is applied to experimental data, namely NIR and concentration of target metabolites of a *Bordetella pertussis* cultivation.

2. The semi-parametric hybrid model

The framework for the serial semi-parametric hybrid model structure are the reactor material balances

$$\frac{dc}{dt} = f = r(L_x, w_A) - D \cdot c + u \ , \tag{1}$$

where c is the vector of concentrations, r is the vector of kinetic rate functions which depend on the inputs L_x and the parameters w_A, D is the dilution rate and u is the vector of volumetric control inputs. The vector of the kinetic rate functions, r, is the representative of the biological system, which is defined by the following semi-parametric model

$$r(L_x, w_A) = K \cdot \left\langle \phi_j(c) \times \rho_j(L_x, w_A) \right\rangle_{j=1,\dots,m} \ , \tag{2}$$

with K being a $n \times m$ matrix of yield coefficients, ϕ being m known kinetic functions and where $\rho(L_x, w_A)$ are unknown kinetic functions which are obtained from a nonparametric model comprising L_x and w_A. Many times these nonparametric models are ANNs (see Oliveira 2004, Thomson & Kramer 1994, Psichogios & Ungar 1992) but these cannot be applied along with spectral data as mentioned above. Due to the outstanding features of the (N)PLS for high-dimensions of redundant inputs, a structure equivalent to a NPLS structure which moreover restores the NPLS features is adopted here. Further a suitable parameter identification algorithm funding on the sensitivities approach is proposed, avoiding the estimation of the kinetic rates from sparse and noisy measurements. The structure which is proposed consists of o independent submodels such that

$$\rho_{1..m}(L_x, w_A) = \sum_{i=1}^{o} \rho_{1..m,i}(L_{i,1..k}, w_A) \ . \tag{3}$$

Each submodel $\rho_{1..m,i}(L_{i,1..k}, w_A)$ further can be divided into two parts, an outer and an inner model. The outer model linearly compresses the high number of dimensions of the

Hybrid semi-parametric modelling of biological systems: Combining Projection to Latent Structures (PLS) with fundamental material balances

inputs and outputs by the use of input loadings, $W_{x,i}$, and output loadings, $W_{y,i}$, to one inner and one outer latent variable, respectively. The inner model then links (non)linearly the input latent variable with the output latent variable, for details see (Qin & McAvoy 1992, Baffi et al 2000). In this study the inner models are chosen to be ANNs, which proved to be successful in Baffi et al (2000). The complete nonparametric model can then be written as

$$\rho_{1..m}(L_x, w_A) = \sum_{i=1}^{o} W_{y,i} \cdot \left(w_{2,i} \cdot g \left(w_{1,i} \cdot h (W_{x,i} \cdot L_{i,1..k}) + b_{1,i} \right) + b_{2,i} \right) , \qquad (4)$$

where the ANN of submodel i comprises the weights, $w_{1,i}$ and $w_{2,i}$, the biases, $b_{1,i}$ and $b_{2,i}$, and the transfer functions, h and g, which are linear and tangential, respectively. The vector of inputs $L_{i,1..k}$ with dimensions $1..k$ is for $i=1$ identical to L_x and can comprise the model estimates of concentrations and/or additional experimental data, such as spectral data. When $i>1$ then, equivalently to the PLS models the input is calculated as

$$L_{i,1..k} = L_{i-1,1..k} - W_{x,i-1} \cdot L_{i-1,1..k} \cdot W_{x,i-1} , \qquad (5)$$

such that information which is gathered in prior input latent variables is not processed again, avoiding redundancy of the latent variables. While the structure of the nonparametric model is the prerequisite, the success of the model depends on the parameter identification. The following approach restores the idea of the NIPALS algorithm by application of a twofold objective function. On the one hand this objective consists of the minimization of a least square error of the residual in the concentrations

$$\min_{w_A} \left(E_1 = \frac{1}{P \times n} \sum_{l=1}^{P} \sum_{j=1}^{n} \frac{(c_{experimental,j}(t) - c_j(t, w_A))^2}{c_{max,j}} \right) , \qquad (6)$$

where P is the number of time events, $c_{experimental,1..n}$ are the experimentally measured concentrations and $c_{max,1..n}$ are the respective standard variances, which account for statistic properties of the data.

On the other hand the objective consists of the maximization of the captured variance of the inputs to the model, which is analogous to minimizing the least square error of

$$\min_{w_A} \left(E_2 = \frac{1}{o \times k} \sum_{i=1}^{o} (W_{x,i,lin} - W_{x,i})^2 \right) . \qquad (7)$$

The first term in Eq. (7), $W_{x,i,lin}$, is the vector norm of $W_{x,i,lin,un}$, which is calculated as following

$$W_{x,i,lin,un} = \frac{L_{i,1..k} \cdot t_i}{t_i^T \cdot t_i} . \qquad (8)$$

Therein the input scores t_i are obtained from the inputs times the input loadings, i.e.

$$t_i = W_{x,i} \cdot L_{i,1..k} . \qquad (9)$$

This NIPALS inspired calculation of $W_{x,i,lin}$ restores some basic features of the PLS, such as independence of the latent variables and minimization of redundant information in the latent variables.

The parameter identification is accomplished by application of the sensitivities equation. These sensitivities equations are not presented here for the sake of briefnesses, but can be derived by building the derivative after all parameters w_A, where w_A comprises all input loadings, all output loadings and all inner model ANN parameters. However for the identification of the input and output loadings their normalization to unit length, which is a condition arising from PLS, needs to be taken into account. In any case, the sensitivity equations need to be integrated along with the reactor material balances, i.e. Eq. (1). In general the MATLAB® integration routines could be applied but they are rather time expensive when compared to a linear, Euler integration approximation schema, which therefor found application.

The two least square objectives, namely E_1 and E_2 are sought to be simultaneously minimized using the "lsqnonlin" MATLAB® function which uses a subspace trust region method and is based on the interior-reflective Newton method (MATLAB® Optimization toolbox). Depending on the inputs L_x a deficiency of the presented twofold objective might come into account, namely when model estimates are inputs to the nonparametric model then w_A might be optimized rather towards the second objective than the first. In order to circumvent this deficiency the parameter identification is carried out until the "best" parameters are obtained, which is accomplished as described below, and then a further parameter identification is carried out, in which only a subset of the parameters, namely all ANN parameters and all output loadings, along with only the first objective is applied.

Identification of the best parameters bares two well known challenges. One is due to the fact that gradient based nonlinear optimization does not necessarily identify the global minimum but rather a local minimum of the objective and that the optimization might get stuck there. To overcome this challenge in this study, as done by several other authors, Oliveira (2004), Peres et al. (2008), Teixeira et al. (2007a), at least four runs of the optimization with random initial values of all parameters are performed, where the consistency of the obtained solutions is taken as a measure whether or not to perform further runs. The other challenge is know as "over-training" of the nonparametric model and is due to the fact that after a certain threshold the optimization rather results in modeling the noise of the data then to further identify the underlying function. This challenge is usually overcome by the application of two data sets, one containing about 2/3 of the data for training the model, and the other containing 1/3 of the data referred to as validation data set. Then parameter identification is carried out on the training set and is then stopped when the residual of the validation set is the smallest.

The presented nonparametric structure admits two structural changes, when the number of inputs and outputs are fixed. One change can be made to the number of nodes used in the hidden layer of the ANNs, but in this study is fixed to be one because reported in Qin & McAvoy (1992) and Baffi et al (2000) to have only little influence on the quality of estimates. The other change that can be made is regarding the number of latent variables. These number of latent variables is however not like in PLS or NPLS consecutive increased till a certain amount of variance is captured but instead fixed at the beginning of the parameter identification. Therefore hybrid model structures with different numbers of latent variables are compared to each other in order to find the most "suitable" structure. For the comparison of the structures several facts need to be taken into account such as the model residual, the number of model parameters and the number of data the residual is build on. A criteria reported to be suitable for these model

Hybrid semi-parametric modelling of biological systems: Combining Projection to Latent Structures (PLS) with fundamental material balances

comparisons is the Bayesian Information Criteria (BIC), see Peres et al. (2008). In the sense of the BIC the model is the most suitable which exhibits the greatest BIC value.

3. Application to experimental data of a *Bordetella pertussis* cultivation

Cultivations of *Bordetella pertussis* are used for the production of a vaccine against whooping cough. The key for controlling the process is the online knowledge about biomass and the specific growth rate, Soons et al. (2008a). Such online knowledge can in principal be derived from the online NIR measurements applying a "suitable" model.

3.1.The process

The experimental data of *Bordetella pertussis* which find application in this study are the ones collected, described and used in Soons et al (2008a) and Soons et al (2008b). The processes were run in batch mode and variations to the process conditions, such as in pH, Temperature and dissolved oxygen, were made as reported in Soons et al (2008a). The recorded NIR data were pretreated as in Soons et al (2008a) by the application of a Savtisky-Golay smoothing with a 45-point window and a second-order polynomial in order to reduce noise and then shifted to zero-mean and scaled by the variance. Measurements of the concentrations of lactate, glutamate and biomass over time for eight batches were recorded Soons et al (2008b). However due to uncertainties in the measurements of the substrates in one of the batches, only the remaining batch data were used, i.e. split into a 5 batches comprising training data set and a 2 batches comprising validation set.

3.2.The *Bordetella pertussis* hybrid model

The hybrid model in this case contains mechanistic knowledge about the process, which was reported in Soons et al (2008b). The system of model equations reads,

$$\frac{d}{dt}\begin{bmatrix} Lac \\ Glu \\ X \end{bmatrix} = \begin{bmatrix} Lac \cdot X & 0 & 0 \\ 0 & Glu \cdot X & 0 \\ 0 & 0 & X \end{bmatrix} \cdot \begin{bmatrix} r_{Lac} \\ r_{Glu} \\ \mu \end{bmatrix} - D \cdot \begin{bmatrix} Lac \\ Glu \\ X \end{bmatrix} \qquad (10)$$

where *Lac*, *Glu* and *X* are the concentrations of lactate, glutamate and biomass, respectively and r_{Lac}, r_{Glu} and μ are the respective unknown kinetic functions which are obtained by the nonparametric model. The inputs L_x to the nonparametric model in this study contain the estimates of all concentrations, the pretreated NIR data and measured data of pH, Temperature and the percentage of dissolved oxygen. For the hybrid NPLS model the number of latent variables is varied in order to identify the best model structure and in addition for the (N)PLS models the inputs are varied, by the means of number and kind in the sense of AutoRegressive eXogenous (ARX) models.

3.3.Results &Discussion

The best obtained model performance criteria over model methodologies and structural parameters are shown in Tab.1. It therein can be seen that the hybrid model outperforms by far, in the order of one magnitude, the traditional dynamic ARX-PLS for all BIC as for all MSE values. These results are consistent with graphical observations and in-line with the expectations formulated in the introduction. It can be further noticed that the best identified hybrid model structure consists of only 2 latent variables which thus offers a drastically smaller number of involved parameters when compared to the best

identified ARX-PLS which comprises 3 latent variables and there due exhibit higher statistical confidence reflected in the higher BIC values shown in Tab. 1.

Table 1. Model performance criteria, namely the Bayesian Information Criteria (BIC) & the Mean Square Error (MSE) for training & validation sets over model methodologies & structural parameters. Lv stands for the number of latent variables & nt for the number of time series.

Model	Structure	BIC train	MSE train	BIC valid	MSE valid
ARX-PLS	[nt =1, lv =3]	-11615	0.2483	-8418	0.7810
Hybrid model	[lv =2]	-326	0.0991	-83	0.0753

4. Conclusions

The proposed semi-parametric hybrid model when applied to NIR data of *Bordetella pertussis* cultivations clearly improves the estimation of concentrations when compared to those obtained through classical chemometric modeling, namely ARX-PLS, as expressed by the mean square error, which is one order of magnitude lower for the hybrid model, and higher statistical confidence of the estimates. Further, as a result of applying this hybrid structure, the trajectories of the estimated fluxes are directly accessible on-line, opening the possibility for on-line metabolic flux control.

5.Acknowledgment

References

Baffi G., Martin E.B. and Morris A.J., (2000), Non-linear dynamic projection to latent structures modelling, Chemom. Intell. Lab. Syst., 52, 5–22

Oliveira R. (2004), Combining first principles modelling and artificial neural networks: a general framework, Computers & Chemical Engineering, 28, 755-766

Peres J, Oliveira R, Feyo de Azevedo S., (2008), Bioprocess hybrid parametric/nonparametric modelling based on the concept of mixture of experts, Biochemical Eng. J., 39, 190-206

Qin S.J. and McAvoy T.J., (1992), Non-linear PLS modelling using neural networks, Computers & Chemical Engineering, 23, 395–411

Psichogios D.D. Ungar L.H., (1992), Comparison of four neural net learning methods for dynamic system identification, IEEE Transactions on Neural Networks, 3, 1, 122-130

Soons Z.I.T.A., Streefland M., van Straten G., van Boxtel A.J.B., (2008a), Assessment of near infrared and "software sensor" for biomass monitoring and control, Chemometrics and Laboratory Systems, 94, 2, 166-174

Soons Z.I.T.A., Shi J,, Stigter J,D,, van der Pol L.A., van Straten G., van Boxtel A.J.B., (2008b), Observer design and tuning for biomass growth and k(L)a using online and offline measurements, J. of Processcontrol, 18, 7-8, 621-631

Teixeira A.P., Alves C., Alves P.M., Carrondo M.J.T., Oliveira R., (2007a), Hybrid elementary flux analysis/nonparametric modelling: application for bioprocess control, BMC Bioinformatics, 8, 30

Teixeira A.P., Carinha N., Dias J.M.L., Cruz P., Alves P.M., Carrondo M.J.T., Oliveira R., (2007b), Hybrid semi-parametric mathematical systems: Bridging the gap between systems biology and process engineering, J. of Biotechnology, 132, 418-425

Thompson M.L., Kramer M.A., (1994), Modelling chemical processes using prior knowledge and neural networks, AIChE Journal, 40,8, 1328-1340

Real-Time Inversion and Response Planning in Large-Scale Networks

Angelica V. Wong,[a] Sean A. McKenna,[b] William E. Hart,[c] Carl D. Laird[d*]

[a]*Graduate Student, Artie McFerrin Dept. of Chemical Engineering, Texas A&M Univ., College Station, TX 77843, ang_vanessa@tamu.edu*
[b]*Sandia National Laboratories, Albuquerque, NM 87109, samcken@sandia.gov*
[c]*Sandia National Laboratories, Albuquerque, NM 87109, wehart@sandia.gov*
[d]*Assistant Professor, Artie McFerrin Dept. of Chemical Engineering, Texas A&M Univ., College Station, TX 77843, carl.laird@tamu.edu; *corresponding author.*

Abstract

In this paper we present a mixed-integer linear programming formulation for performing source inversion in drinking water systems using discrete (yes/no) measurements available from manual grab samples. Given the large size of a typical water distribution system, standard water quality models are inappropriate for use in the mixed-integer framework. In this research, we demonstrate the use of an origin-tracking approach to develop the water quality model, and show how this model can be efficiently and exactly reduced prior to the formulation of the MILP, giving rise to a much smaller MILP. Furthermore, we demonstrate that this formulation can be efficiently solved in a real-time setting on large networks with over 10,000 nodes while considering over 100 time discretizations.
Keywords: Water distribution systems, Pollution sources, Optimization, Algorithms

1. Introduction

There is significant concern regarding water quality problems in drinking water distribution systems given their vulnerability to chemical and biological contamination. While security measures can be increased at water storage and treatment facilities to prevent accidental or intentional contamination, protecting the remainder of the distribution system poses a bigger challenge. Potential contamination locations include any one of the many access points in the network, including household water faucets or fire hydrants. One proposed method of protection is the installation of an early warning detection system composed of monitoring stations throughout the drinking water network. Due to the cost of installing and maintaining a fixed sensor grid, the number of sensors installed must be limited, and several researchers have investigated the problem of optimal sensor placement within drinking water networks [1-3]. However, on its own, an early warning detection system provides only a coarse measure of the time and location of the contamination event. During a contamination event, it is also important to determine the location, time, and/or time dependent magnitude of contamination injection. Once this information is available, the current extent of contamination within the system can be estimated, and an appropriate control and cleanup strategy can be devised to protect the population.

The source inversion problem has been addressed by a number of different researchers [4-10]. In previous work we introduced a large-scale nonlinear programming approach that used real-time concentration information from an installed sensor grid to accurately determine the time and location of the contamination event [4-7]. While this approach allowed systematic identification of contaminant sources in

real-time, it relied on concentration measurements of the contaminant from a fixed sensor grid. Given the cost and technical challenges, it is unlikely that utilities will have a sufficient sensor grid with the capability to measure concentrations of arbitrary contaminants. Instead, the majority of utilities will need to rely on standard water quality measures [11] (e.g. pH, salinity, residual chlorine) obtained through manual grab samples. Using fault detection approaches, these values can be used to estimate the presence or absence of contaminant (a yes/no discrete measurement).

With this in mind, the following integrated real-time process is proposed. Given an initial customer inquiry or detection of contamination event, several response teams are sent to gather manual grab samples near the location of the initial detection. Measurements from these grab samples are then used to invert for the best estimate of the source of the contamination event. Initially, with only a few measurements, it will generally not be possible to uniquely identify the contamination source location, however, a list of the possible locations can be identified. Using this list and simulation results, grab sample locations for the next measurement cycle are selected. Each cycle of taking additional measurements and performing the source inversion is repeated until the source of the contaminant is uniquely identified. As part of this real-time process, this paper specifically addresses the formulation of an effective mixed-integer linear programming problem for determining the source of a contamination event given discrete (yes/no) measurements from sparse manual grab samples.

2. Problem Formulation

Water distribution systems are usually represented as a network of links and nodes, where links represent pipes, pumps, and valves, and nodes represent tanks, junctions and water sources, like rivers and reservoirs. Here, plug flow is assumed for all pipes. Junctions are considered to be zero-volume, while tanks have volume, and complete mixing is assumed at each node. The hydraulic calculations are assumed to be independent of the water quality calculations [12-14]. Calculation of the hydraulic model for this system requires the network properties and demand patterns at each node. The flow patterns are then used as known inputs to the water quality model. Note that the time varying demand patterns and storage tank levels cause flow velocities to change in magnitude and even direction.

The contaminant spread through the network due to a mass injection term at one or more nodes is described by a set of algebraic mass balances for junctions, ordinary differential equations for tanks, and partial differential equations for the pipes. Discretizing the pipes spatially along their length would lead to an inverse problem that is far too large for current numerical tools. Instead, a time discretization is selected and an origin-tracking approach [6] is used to pre-calculate the time-dependent delays and express the time varying concentrations at pipe boundaries as a function of the time varying node concentrations. The origin-tracking approach is performed on a pipe-by-pipe basis, and has a computational complexity that is linear in the number of pipes, making it an efficient alternative for large networks. The differential equations for the tanks are approximated on the same time discretization using backward finite differences, and the entire water quality model can now be represented as a very large but sparse, linear system, $Gc = Mm$, where c is the complete vector of time discretized concentrations for pipe inlets, pipe outlets, and network nodes, m is the vector of time discretized injections for every node, and G and M describe the linear system resulting from the discretization and origin-tracking approach. To simulate a contamination event, it suffices to specify the injection vector m and solve the linear system for the concentration profiles. Simulation results using a time discretization of

15 minutes from our proposed water quality model compares well with the results produced by EPANET [15].

The goal of the source inversion formulation is to find the time profiles of the unknown injection variables m that result in calculated concentrations that best match the manual grab sample measurements. If contaminant concentration measurements were available, this could be formulated as a least-squares minimization. However, while the model calculates concentrations, the measurements are only discrete yes/no values. Therefore, to map the model output to the available measurement information, we select a threshold value for the contaminant concentration and assume that a concentration from the model above this threshold should yield a "yes" measurement and a concentration below this threshold should yield a "no" measurement. Here, let $\mathcal{T}$ refer to the set of time discretizations, while $\mathcal{P}$ and $\mathcal{N}$ refer to the complete set of network pipes and nodes, respectively. Also, let S be the set of node-time pairs representing the time and location where manual grab samples were taken. From the set S we can write two subsets. Set S_- contains the node-time pairs where the presence of contaminant was not detected (discrete measurement was "no"). Set S_+ contains the node-time pairs where contaminant was detected (discrete measurement was "yes"). A min-max objective is formulated that penalizes concentrations below the threshold (th) for node-pairs where contaminant was detected and penalizes concentrations above the threshold for node-pairs where contaminant was not detected:

$$\min \sum_{(i,j)\in S_-} \max(0, c_{ij} - th) + \sum_{(i,j)\in S_+} \max(0, th - c_{ij}) \tag{1}$$

Using a standard (and exact) reformulation for this objective, the complete contaminant source inversion problem is formulated as the following mixed integer linear program (MILP),

$$\min_{c,m,y,n,p} \sum_{(i,j)\in S_-} n_{i,j} + \sum_{(i,j)\in S_+} p_{i,j} \tag{2}$$

$$s.t. \quad Gc = Mm \tag{3}$$

$$0 \le m_{i,j} \le B y_i, \ \forall i \in \mathcal{N}, j \in \mathcal{T} \tag{4}$$

$$m_{i,j-1} \le m_{i,j}, \forall i \in \mathcal{N}, j \in \mathcal{T} : j \ne 0 \tag{5}$$

$$\sum_{i\in \mathcal{N}} y_i \le 1, \qquad y_i \in \{0,1\} \tag{6}$$

$$n_{i,j} \ge 0, \ \ n_{i,j} \ge c_{i,j} - th, \ \ \forall (i,j) \in S_- \tag{7}$$

$$p_{i,j} \ge 0, \ \ p_{i,j} \ge th - c_{i,j}, \ \ \forall (i,j) \in S_+ \tag{8}$$

where $c = [..c_{i,j}..]$ is the complete vector of time discretized concentrations for the inlet of every pipe, the outlet of every pipe, and every node. The vector $m = [..m_{i,j}..]$ is the vector of unknown time discretized mass injections for every node. Here, y_i is a 0-1 variable that allows injections at node i only if $y_i = 1$. The big-M parameter B is a reasonable upper bound on the injection terms. While the contamination event can start at any time, once the injection starts we assume that it will continue until the contamination source is found. This constraint is enforced in Eqn. 5. Using

measurement information from a particular sampling cycle, this formulation allows us to identify the best-fit estimate of the contamination source node. Due to the limited number of measurements, the source identification problem is almost guaranteed to have non-unique solutions. In order to find all possible solution (i.e. all potential contamination sources), the problem is solved repeatedly, cutting previous solutions until the measure of fit deteriorates.

Given the large number of nodes (>10,000) and the number of timesteps (~100) for a typical water network model, the above MILP is too large for efficient solution on a standard desktop computer. Noting that the objective function only requires concentrations at node-time pairs where measurements were taken, we describe an efficient procedure to reduce the size of the water quality model. We can rewrite the water quality model as $c = [G^{-1}M]m$ where, instead of the entire linear system, we are only interested in the rows corresponding to measurement node-time pairs. A particular row r_i of $G^{-1}M$ can be efficiently extracted by first solving for $\hat{r}_i$ from $G^T \hat{r}_i = e_i$, where e_i is a vector of zeros with a one in the ith index. Then the ith row is found by $r_i = M^T \hat{r}_i$. In this procedure, the matrix G is factored only once and the remaining calculations involve only a single backsolve and a single matrix-vector product for each row extracted. Eqn. 3 in the MILP is now replaced by a smaller linear system involving concentrations of the sampled nodes only,

$$c_R = Rm_R \tag{9}$$

Note that this is an exact reduction and not an approximation.

3. Numerical Results

The effectiveness of this formulation is tested on a real water distribution system model [16], shown in Fig. 1. This network contains 12523 nodes, two constant
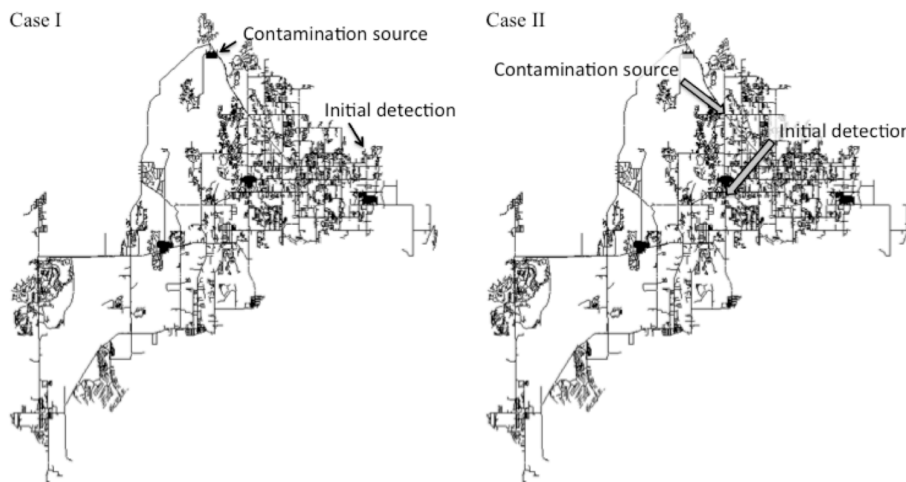
Figure 1: Municipal Water Network Model with Case Study Details

head sources, two tanks, 14822 pipes, four pumps, and five valves. Two contamination scenarios are considered, each over 48 hours with a time discretization of 15 minutes. Case I is a single injection scenario at Reservoir-12523 starting at hour 10. Case II is a single injection scenario at Junction-5901, also starting at hour 10. The measurement profiles for each contamination scenario are simulated using EPANET [15]. For each

sampling cycle, the mixed integer problems with the reduced water quality model are formulated in AMPL [17] and solved with CPLEX (www.ilog.com).

For the first sampling cycle in Case I, contaminant is initially detected at Junction-1980 at hour 41 by a customer. This location is then sampled by utility personnel, along with three other nearby locations. We continue the cycle of performing source inversion and retrieving additional grab samples until the source location is identified. A limit of 6 manual grab samples is considered for each additional sampling cycle, and each sampling cycle takes an hour for completion. In Case II, contaminant is detected first at Junction-4544 at hour 30, and the same procedure is performed.
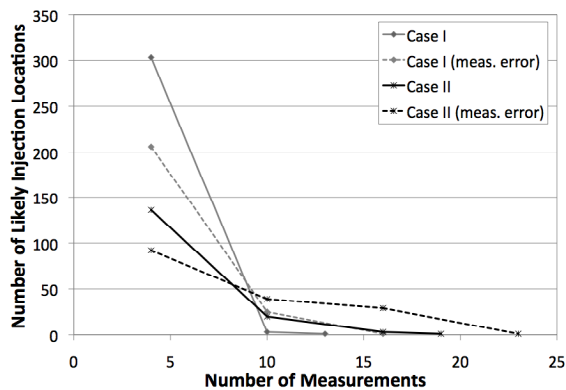
Figure 2: Case Study Progress

The formulation was effective in finding the contamination source for both case studies (and several others that are not included in this paper). Fig. 2 compares the effectiveness for exact measurements and when measurements contain errors. For the case where measurements contain errors, one of the grab samples from each cycle was randomly selected to return the incorrect yes/no value. For Case I with exact measurements only three sampling cycles and a total of 13 measurements were required to narrow the number of potential contamination sources from 303 locations to the single correct contamination location. When measurements errors were present, three sampling cycles and a total of 16 measurements were required. For Case II with exact measurements four sampling cycles and a total of 19 measurements were required to narrow the number of potential contamination sources from 136 locations down to one. With measurement error, four sampling cycles and a total of 22 measurements were required. For each of these case studies, computational time was about two minutes to create the water quality model, one minute to reduce the model, and 30 seconds to 4 minutes to solve the inversion problem on an Intel Xeon 3.0 GHz desktop machine.

4. Conclusions and Future Work

The results demonstrate that the presented formulation was effective in determining the source of a contamination using discrete (yes/no) measurements at sparse locations and times, even in the presence of measurement error. This paper also shows how the water quality model can be efficiently reduced prior to the formulation of the MILP giving rise to a much smaller MILP that can be solved efficiently in a real-time setting.

This paper only assumes a single contamination scenario, however the formulation allows for consideration of multiple events. This will be further investigated in future work. In addition, new grab sample locations during each cycle were selected manually. Mathematical programming can also be effectively used to determine optimal sampling locations. Formulations for grab sample placement will be combined with the inversion problem and included in future work. This work also

assumes accurate demand patterns and hydraulic information. The impact of demand uncertainty will be investigated in future work.

5. Acknowledgements

References

[1]   Ostfeld A. and Salomons E. (2004). "Optimal layout of early warning detection stations for water distribution systems security.", J. Water Resour. Plann. Manage. 130(5), 377-385.

[2]   Berry, J., Fleischer, L., Hart, W. and Phillips, C. (2005). "Sensor placement in municipal water networks.", J. Water Resour. Plann. Manage., 131(3), 237-243.

[3]   Murray, R., Baranowski, T., Berry, J. W., Hart, W. E., Janke, R., Phillips, C. A., Taxon, T. N., and Uber, J. G. (2009). "Sensor network design for drinking water contamination warning systems: a compendium of research results and case studies using TEVA-SPOT", EPA/600/R-09/141, U.S. Environmental Protection Agency, Office of Research and Development, National Homeland Security Research Center, Cincinnati, OH.

[4]   van Bloemen Waanders, B. G., Bartlett, R. A., Biegler, L. T., and Laird, C. D. (2003). "Nonlinear programming strategies for source detection of municipal water networks." Proc., EWRI Conf., Philadelphia.

[5]   Laird, C. D., Biegler, L. T., and van Bloemen Waanders, B. G., (2004). "Real-time, large scale optimization of water network systems using a subdomain approach.", Proc., 2nd CSRI Conf. on PDE-Constrained Optimization, Santa Fe, NM.

[6]   Laird, C. D., Biegler, L. T., van Bloemen Waanders, B. G., and Bartlett, R. A. (2005). "Contamination source determination for water networks.", J. Water Resour. Plann. Manage., 131(2), 125-143.

[7]   Laird, C. D., Biegler, L. T., and van Bloemen Waanders, B. G., (2006). "Mixed-integer approach for obtaining unique solutions in source inversion of water networks.", J. Water Resour. Plann. Manage., 132(4), 242-251.

[8]   Preis, A., and Ostfeld, A. (2006). "Contamination source identification in water systems: a hybrid model trees-linear programming scheme." J. Water Resour. Plann. Manage., 132(4), 263-273.

[9]   Guan, J., Aral, M. M., Maslia, M. L., and Grayman W. M. (2006). "Identification of contaminant sources in water distribution systems using simulation-optimization method: case study." J. Water Resour. Plann. Manage., 132(4), 252-262

[10] De Sanctis, A. E., Shang, F., and Uber, J. G. "Real-time identification of possible contamination sources using network backtracking methods." J. Water Resour. Plann. Manage., accepted (2009).

[11] Hart, D.B. and McKenna, S.A. (2009). "CANARY User's Manual Version 4.1." Tech. Rep. EPA/600/R-08/040A.

[12] Zierolf, M., Polycarpou, M., and Uber, J. (1998). "Development and autocalibration of an input-output model of chlorine transport in drinking water distribution systems." IEEE Trans. Control Syst. Technol., 6(4), 543-553.

[13] Rossman, L., and Boulos, P. (1996). "Numerical methods for modeling water quality in distribution systems: A comparison." J. Water Resour. Plann. Manage., 122(2), 137-146.

[14] Shang, F., Uber, J., and Polycarpou, M. (2002). "Particle backtracking algorithm for water distriution system analysis." J. Environ. Eng., 128(5), 441-450.

[15] Rossman, L. (2000) "Epanet 2." Tech. Rep. EPA/600/R-00/057, Washington, D.C.

[16] Ostfeld A., et al. (2008) "The battle of the water sensor networks (BWSN): a design challenge for engineers and algorithms." J. Water Resour. Plann. Manage., 134(6), 556-558.

[17] Fourer, R., Gay, D. M., and Kernighan, B. W. (1993). " AMPL: a modeling language for mathematical programmng, Scientific Press, Danvers, Mass.

Multi-objective optimization of solar assisted absorption cooling system

Berhane H. Gebreslassie[a], Mélanie Jimenez[a], Gonzalo Guillén-Gosálbez[b],
Laureano Jiménez[b] and Dieter Boer[a]

[a] *Department of Mechanical Engineering, University Rovira i Virgili*
[b] *Department of Chemical Engineering, University Rovira i Virgili*
Av. Països Catalans, 26, 43007-Tarragona, Spain, Dieter.Boer@urv.cat

Abstract

This work presents a multi-period and multi-objective optimization based on mathematical programming of solar assisted absorption cooling systems. Seven solar collector models combined with a gas fired heater and an absorption cooling cycle are considered. The optimization task is formulated as a multi-objective multi-period mixed-integer nonlinear programming (MINLP) problem that accounts for the minimization of the total cost of the cooling system and the associated environmental impact. The environmental performance is measured following the Life Cycle Assessment (LCA) principles. The capabilities of the proposed method are illustrated in a case study that addresses the design of a solar assisted ammonia-water absorption cooling system using the weather conditions of Tarragona (Spain).

Keywords: Solar assisted cooling, MINLP, multi-objective optimization, life cycle assessment (LCA), absorption cycle.

1. Introduction

Air conditioning and refrigeration have a significant contribution to the total energy consumption. The growing demand for air conditioning and refrigeration has caused a significant increase in the consumption of primary energy resources, to the end that it currently threatens the stability of electricity grids. Thus, a change in the energy structure should be made, promoting energy efficient technologies and renewable energies. A more sustainable concept, known as solar assisted refrigeration, is based on the use of absorption cooling cycles driven by thermal energy provided by solar collectors. The use of solar energy for cooling applications has a high potential to replace partially conventional cooling systems, given the fact that the cooling demand matches the availability of solar irradiation [1]. These technologies use renewable energy sources, thus decreasing the associated environmental impact. However, their cost is still higher compared to conventional cooling systems (*i.e.*, vapor compression cooling system).

The objective of this work is to provide a quantitative decision support tool for the optimal design of solar assisted absorption cycles. The model presented optimizes the operating and structural decisions of the absorption cycle taking into account simultaneously its environmental and economic performance. The environmental performance is measured based on the Eco-indicator 99, which follows the life cycle assessment (LCA) methodology.

2. Problem statement

The problem addressed is formally stated as follows. Given are the cooling capacity of the system, the inlet and outlet temperatures of the external fluids (chilled water and cooling water), capital cost data, monthly-averaged weather conditions, the performance characteristics of the solar collectors, and life cycle inventory data associated with the construction and operation of the cooling system. The goal is to determine the optimal design and associated operating conditions that simultaneously minimize the total cost and the environmental impact of the system.

3. Solar assisted absorption cycle

Figure 1 depicts the solar assisted absorption cooling system under study. The absorption cycle operation is discussed in detail in Berhane *et al.* [2]. The heat production subsystem includes two main units: a gas fired heater (GFH), and solar collectors (Col). GFH is a low pressure heater consuming natural gas. The solar collector panels used in this work are: evacuated tube collectors (ETC), flat plate collectors (FPC) and compound parabolic collectors (CPC). Within each collector type, we consider different possible alternatives and the associated models that describe their performance.
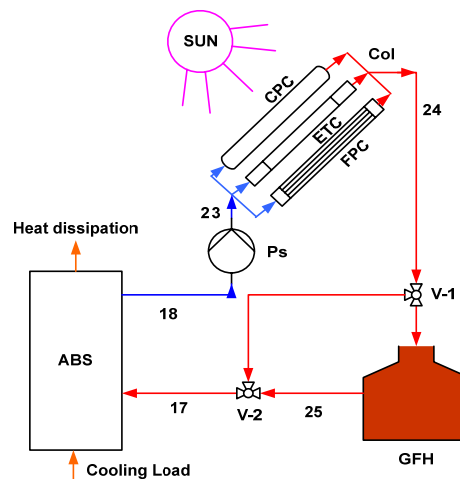
Fig. 1. Ammonia-water solar assisted absorption cooling system

4. Mathematical model

The mathematical model of the absorption cycle is based on the formulation introduced by the authors in previous works [2, 3]. The original formulation has been extended in order to integrate the heat production subsystem model.

4.1. General constraints

The model considers that the time horizon can be divided into *t* periods. The cycle can then operate in a different manner in each of these periods in order to get adapted to the specific solar radiation of that time interval. Without loss of generality, we consider that each of these periods corresponds to one month, although in general we could specify any other length.

The model is based on energy and materials balances applied to each unit of the system.

Multi-Objective Optimization of Solar Assisted Absorption cooling system

4.2. Heat production subsystem

The heat supplied to the absorption cycle is generated in the solar collectors and the gas fired heater.

4.2.1. Collector performance constraints

The selection of n solar collectors of type i is represented by the following disjunction [4]:

$$\bigvee_{i=1,\ldots I} \begin{bmatrix} Y_i \\ \bigvee_{i=1,\ldots I} \begin{bmatrix} Y_n \\ Q_t^{Col} = \eta_{i,t} n A_i^{Col} GB_t \end{bmatrix} \\ \eta_{i,t} = IAM(\Theta)C_{0,i} - C_{1,i}\dfrac{T_t^{av} - T_t^{amb}}{GB_t} - C_{2,i}\dfrac{\left(T_t^{av} - T_t^{amb}\right)^2}{GB_t} \end{bmatrix} \tag{1}$$

$$Y_i, Y_n \in \{True, False\} \qquad \forall i, n$$

where Y_i and Y_n are Boolean variables that decide whether the given disjunctive terms i and n inside the disjunctions are true or false. They are true if n collectors of type i are selected and zero otherwise. If the collector is chosen, then the equations inside the disjunction, which determine the heat supplied by the solar system, are active. If the collector is not selected, the corresponding equations are all set to zero. The useful heat collected from the solar system in each time-period t (Q_t^{Col}) is determined from the collector performance, which is calculated from the collector efficiency $(\eta_{i,t})$, its area and the global solar incident radiation on the collector surface in month t (GB_t) [1]. The area of the collectors is obtained by multiplying the size of the collector type i available in the market (A_i^{Col}) with the corresponding number of collectors (n) installed in the system. The second equation inside the disjunction allows to determine the efficiency of the collector i in month t [1].

4.2.2. Linking constraints

Eqn. (2) links the heat provided by the collectors and the gas fired heater with that consumed by the cycle. Hence, the heat consumed by the generator of the cycle in month t (Q_t^D) should be less than or equal to the sum of the heat collected from the collector (Q_t^{Col}) and the heat supplied by the gas fired heater (Q_t^{GFH}) t in the same month:

$$Q_t^D \leq Q_t^{Col} + Q_t^{GFH} \tag{2}$$

4.3. Objective functions

Our model includes two contradicting objective functions: 1) minimize the total cost; and 2) minimize the environmental impact of the cooling system.

4.3.1. Economic performance objective function

The total cost of the system (TC) accounts for the capital and operating costs (C^c and C^{op}, respectively) as shown in eqn. (3).

$$TC = C^c + C^{op} \tag{3}$$

4.3.2. Environmental performance objective function

The environmental performance of the system is measured based on the principles of Life Cycle Assessment (LCA) using the Eco-Indicator 99 framework. The total Eco-Indicator 99 (ECO_{99}^{tot}) is given by the sum of the Eco-Indicator 99 of the manufacturing part (ECO_{99}^{man}) and the operational part (ECO_{99}^{op}) shown in eqn. (4).

$$ECO_{99}^{tot} = ECO_{99}^{man} + ECO_{99}^{op} \qquad (4)$$

5. Solution method

The design problem can be finally formulated as a multi-objective mixed-integer nonlinear programming (MINLP) of the following form:

$$(P) \quad \min_{x,y} \quad U(x,y) = \left\{ TC(x,y), ECO_{99}^{tot}(x,y) \right\}$$

$$s.t. \quad h(x,y) = 0 \qquad (5)$$
$$g(x,y) \le 0$$
$$x \in R, \; y \in \{0,1\}$$

In this formulation, x represents the state or design variables such as thermodynamic properties, mass flows and equipment sizes. The discrete variables are denoted by y, and are used in the selection of a specific number of collectors n of type i. $TC(x, y)$ and ECO_{99}^{tot} denote the economic and environmental performances of the solar heat integrated absorption cooling system, respectively.

The solution to *(P)* is given by a set of Pareto optimal points that represent the trade off between the economic and environmental performance of the system. For the calculation of the optimal trade off solutions, the ε-constraint solution method is used. This method is coupled with a customized branch and bound algorithm that reduces the computational burden of the model. This strategy relies on branching on the Boolean variables that denote the type of collector. In every node of the tree, lower and upper bounds on the optimal solution are obtained by relaxing the integer variables that denote the number of collectors selected (lower bound) and then rounding them up and solving again the problem with fixed values of the integers (upper bound). The nodes of the tree for which the lower bound exceeds the current best upper bound are pruned, and the procedure is repeated until all the collector types are analyzed.

6. Case study

The capabilities of our approach are illustrated through a case study that addresses the design of a solar assisted absorption cooling system with a cooling capacity of 100 kW. The process data for the ammonia-water absorption cycle are taken from [2, 3]. Concerning the solar collectors, there are several types which can be used in solar air conditioning systems. We consider seven models of non-tracking collectors assuming the weather conditions of Tarragona (Spain).

7. Results and discussions

The non-convex bi-criteria mixed integer nonlinear programming (MINLP) and the associated solution procedure is implemented in GAMS [5]. Figure 3 shows for the extreme solutions the single damage categories that contribute to the total Eco-indicator 99. In all the cases, the main contributor to the overall environmental damage is the

Multi-Objective Optimization of Solar Assisted Absorption cooling system

depletion of natural resources. On the other hand, the eco-system quality damage is rather low, mainly because the cooling system does not use hydrochlorofluorocarbons (HCFC).
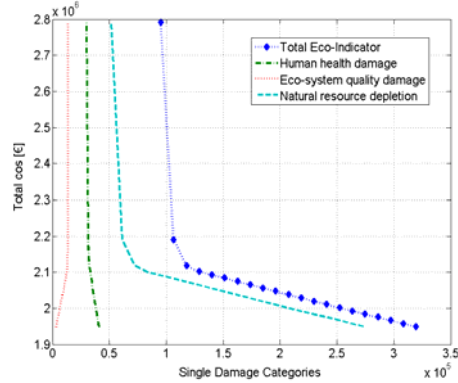
Fig. 2. Pareto set of solutions of the case study
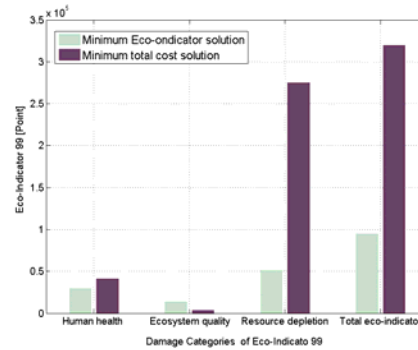
Fig. 3. Contribution of single damage categories to the total Eco-indicator at the extreme Pareto designs

The CPU time required to obtain the complete Pareto set of solutions is *88.33* seconds in a *2.29 GHz* machine. The Pareto points are depicted in Figure 2, which shows the single damage categories as well as the total Eco-indicator 99 value associated with each Pareto solution. Each point of the curve represents a different absorption cycle design. In the minimum total cost solution, the heat required by the cycle is exclusively supplied by the gas fired heater. However, as we move towards the minimum impact solution, the energy needs are gradually fulfilled using a larger number of collectors. Near to the minimum total cost Pareto design, flat plate collectors are selected. The reason is that they show performance improvement at lower temperatures more than the evacuated tube collectors. Moreover, they are cheaper than the evaluated tube collectors. On the other hand, near to the minimum environmental impact Pareto designs, the heat demand is covered by the evacuated tube solar collectors. The reason is at higher temperature the performance of evacuated tube collectors is better than the flat plate collectors.

Figures 4 and 5 depict the contribution that the manufacturing and operation of the subsystems of the cycle (i.e., absorption cycle itself, collectors and heater) have on its environmental performance for each of the extreme Pareto designs. In general, the construction of the equipment units has very little impact on the overall environmental damage. In fact, in the minimum total cost Pareto design, the main source of impact is the operating of the gas fired, whereas the contribution due to the manufacturing of the cooling system is negligible. However, in the minimum environmental impact solution, the main contributor to the total impact is the operation of the solar collectors. These are operated using pumps that consumes electricity, the impact of which is determined assuming the Spanish electricity grid that is mainly based on fuel oil and coal. In this latter case, the magnitude of the impact due to the manufacturing of the cooling system is larger, mainly because of the emissions of heavy metals that take place during the construction of the collectors. Hence, it is worthwhile to consider this part in the calculation of the life cycle impact of the solar assisted cooling system.

Note: The subscripts m, op and t represents the manufacturing, operation and total Eco-indicator respectively.
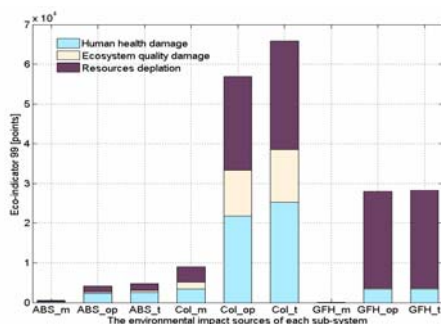
Fig. 4. Main sources of impact in the minimum environmental impact solution
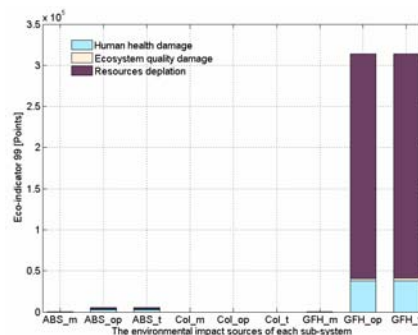
Fig. 5. Main sources of impact in the minimum cost solution

8. Conclusions

Based on mathematical programming, the design of more sustainable solar assisted absorption cooling systems has been presented. The method introduced relies on formulating a bi-criteria MINLP problem that accounts for the minimization of the total cost and the environmental impact of the cooling system. The capabilities of the proposed approach have been illustrated through its application to the design of a solar assisted ammonia-water absorption cooling system. Reductions up to 70 % in the environmental impact are feasible by replacing the consumption of primary energy resources by renewable solar energy sources. The tool presented in this work is intended to guide decision-makers towards more sustainable design alternatives for fulfilling the current cooling demand.

9. Acknowledgements

References

1. H. M. Henning. *Solar-assisted air-conditioning in buildings*. Springer Wien/New York; 2004.
2. B. H. Gebreslassie, G. Guillén-Gosálbez, L. Jiménez, and D. Boer. Design of environmentally friendly absorption cooling systems via multi-objective optimization and life cycle assessment. *Computer Aided Chemical Engineering*, 26:1099-1103, 2009.
3. B. H. Gebreslassie, G. Guillén-Gosálbez, L. Jiménez, and D. Boer. Economic performance optimization of an absorption cooling system under uncertainty. *Applied Thermal Engineering*, 29:3491-3500, 2009.
4. M. Türkay and I. E. Grossmann. Structural flow sheet optimization with complex investment cost functions. *Computers & Chemical Engineering*, 22:673-686, 1998.
5. A. Brooke, D. Kendrik, A. Meeraus, R. Raman, and R. E. Rosenthal. *GAMS - A User's Guide*. GAMS Development Corporation, Washington, 1998.

The Effects of Vehicle-to-Grid Systems on Wind Power Integration in California

Bri-Mathias S. Hodge, Shisheng Huang, Aviral Shukla, Joseph F. Pekny, Gintaras V. Reklaitis

School of Chemical Engineering, Purdue University, 480 Stadium Mall Dr., West Lafayette, IN 47907, [bhodge, huang47, shuklaa, pekny, reklaiti] @purdue.edu

Abstract

Renewable energy portfolio standards have already caused a large increase in the amount of electricity produced from renewable sources and this amount is expected to increase as the policy target dates draw closer. One technology that has benefited greatly from these standards is wind energy. The uncertainty inherent in wind electricity production dictates that nearly equal amounts of conventional generation resources be kept in reserve should wind electricity output should suddenly dip. The introduction of plug-in hybrid electric vehicles into the personal transportation fleet presents an interesting possible solution to this problem through the concept of vehicle-to-grid power. The ability of these vehicles to increase the wind production fraction for the California market when high levels of wind energy are present in the supply portfolio is examined.

Keywords: Energy systems modeling, Electricity systems, Plug-in hybrid electric vehicles, Vehicle-to-grid, Wind energy

1. Introduction

Renewable energy portfolio standards, mandates and goals set by individual states have helped to drive the recent increases in renewable energy electricity generating capacity within the United States. While a number of energy technologies are considered to qualify for these standards, including landfill gas, solar thermal and solar photovoltaic, the chief technology to benefit from these programs has been wind power. Many of the mandates dictate that a certain percentage of electricity sales within the state, up to 33%, must be generated from this renewable mixture, that thus far has been dominated by wind. As the amount of wind energy being added to the electricity grid increases the problem of wind energy intermittency becomes more pronounced. Intermittency in electricity generation from wind poses problems on the minutely, hourly and daily time scale. Due to the uncertain nature of wind energy supply larger than normal operating reserves must be kept on standby in areas where wind energy makes up a significant portion of the generating capacity.

Plug-in hybrid electric vehicles (PHEVs), a transportation technology on the cusp of commercial introduction, aim to reduce gasoline consumption by replacing some of the gasoline used as an energy source for personal transportation with electricity. PHEVs possess a battery for electricity storage that allows for an all-electric transportation range, here assumed to be 40 miles, before gasoline is used to power the vehicle. The widespread adoption of PHEVs into the electricity grid poses many challenges, but also some very promising possibilities.

One such possibility is the idea that the PHEV battery be allowed to be able to sell electricity back to the grid when needed. This idea, known as vehicle-to-grid (V2G), is significant for wind energy because it suggests the possibility that distributed automotive batteries could serve as a form of operating reserve for the electricity grid. When intermittent electricity sources are producing more than the current demand, excess electricity could be distributed to charge vehicles that are plugged in. When a reduction in wind energy production occurs the batteries could be tapped as a temporary reserve to balance electricity supply and demand until additional generating capacity could be brought online. It is thought that the ability of PHEVs to serve as an operating reserve for the electricity grid could lead to significant increases in the percentage of electricity production by renewable sources, as the reserve capacity of conventional electricity generators serving as stand-by power could be reduced and electricity produced from wind during periods when supply exceeds demand could be shifted to periods when demand exceeds instantaneous supply.

2. Background

The idea of using electric vehicles as a source of power, instead of strictly as a load, arose from the realization that the power capacity of the United States vehicle fleet greatly exceeds the capacity of United States electric utilities (Kempton and Letendre, 1997). A cost-benefit analysis performed for battery-based electric vehicles showed that the benefits accrued from performing electricity regulation services exceed the additional costs to the consumer that would be necessary to make a standard electric vehicle and home charging station V2G-ready (Kempton and Tomic, 2005a). A probabilistic reliability model that determines the level of operating reserves necessary for wind energy (Milligan, 2001) was then used as the basis for a calculation of the percentage of the national vehicle fleet that must be electric and V2G capable in order serve the function of an operating reserve (Kempton and Tomic, 2005b). Two different mathematical programming models have been used to estimate the amount of wind energy penetration possible with V2G in place and high levels of electric vehicle penetration. The EnergyPLAN model was used in a case study for Denmark, the results of which indicate that excess electricity production could be cut in half for 50% wind energy production with 100% electric vehicle penetration and V2G participation (Lund and Kempton, 2008). An earlier study (Short and Denholm, 2006) used an electricity capacity optimization model, WinDS, to estimate the 2050 total wind energy capacity in the United States would be 110% larger in the case of 50% penetration of PHEVs with an all-electric range of 60 miles than in the base case scenario, though still only 16% of total electricity generation.

3. Modeling Approach

A multi-paradigm modeling approach has been used in order to investigate the ability of PHEVs to act as operating reserves for the electricity grid in the case of high levels of wind power capacity in the generation portfolio. An agent-based framework is used to encapsulate models created using varied modeling paradigms and provide clear channels for communication and time synchronization (Hodge, et al., 2009). A discrete-event household demand model that incorporates the inclusion of PHEVs has been integrated with an agent-based generation assignment model for electricity supply (Huang et al., 2010). The household demand model has been modified to allow for the PHEVs to both charge with electricity from the grid, and discharge electricity to the grid

when connected to a home electrical outlet. This is accomplished by establishing a number of representative households where the PHEV availability and charge level can be continually communicated to the supply model. Each PHEV is represented as an agent within the discrete event household model with usage characteristics typical of a personal automobile. At any moment in time, the vehicle has a probability of being used for either personal trips or commuting to work. PHEV driving timing and distances are assigned probabilistically based on the results of a traffic simulation model (TRANSIMS, 2009). In this study, when the vehicle is not in use, and if the PHEV is parked at home with remaining battery capacity exceeding 80%, it is plugged in as an operating reserve; ready to supply electricity in the case of a shortfall or to absorb any surge in intermittent electricity production. Once the battery charge level drops below 80% of capacity, the vehicle is taken offline and charged again until full. This charging pattern, whereby any vehicle is charged as it returns home, is meant to better replicate anticipated consumer behavior in a flat-rate electricity tariff environment, instead of the utility controlled charging patterns commonly assumed in previous PHEV studies.

The supply model assigns availability to each generator for each hourly period using generator type availability percentages. In this model we do not consider the effects of unanticipated outages or availability forecasting errors, except in the case of wind power production. Within each hour timeframe spot markets are held at ten minute intervals to ensure that demand does not exceed supply at any particular instant. The electricity supply model consists of over 700 electricity generation units that individually determine their availability for every time step, as well as a scheduling agent that chooses from among the available units in order to ensure that electricity supply is sufficient to meet demand. The high levels of wind generation in the model dictates that the assignment agent must operate in a fashion so as to minimize the risk of supply shortages when wind output falls below forecasted levels. For this reason baseload generation, here defined as coal, nuclear and geothermal, is scheduled first as the long ramp-up times needed for these generators preclude their use as reserves assigned within the ten minute intra-hour periods. Forecasted wind capacity is scheduled next with the previous period's output serving as a crude estimate of the wind power available in the next time step. Subsequent generation technologies are added to the schedule based on estimated marginal cost until the forecast supply is equal to the forecast demand. This ordering is used as it would likely be the ordering in a generation pool bid system where wind power would have a near zero marginal cost bid for its forecasted capacity and is not suitable for peaking due to its unknown availability. The PHEVs with V2G capability act as short-term, or spinning, reserves in order to fulfill supply shortages, here caused by inaccuracies in wind forecasting, for one ten minute interval while replacement reserves such as hydroelectric or natural gas generation may be brought online in the spot market.

4. California Case Study

Scenario analysis has been used to examine the effect of V2G power systems on wind energy capacities needed to reach the California mandate. The fulfillment of California's renewable portfolio standard by substantial additions of wind generating capacity, both with and without V2G capability, is examined in order to gauge the technology's impact on integrating intermittent power into the electricity grid. The adoption of PHEVs is assumed in both cases so that electricity demand levels may be held constant between the two scenarios as PHEV usage could cause a significant

increase in electricity requirements that must be accounted for in both scenarios to ensure a fair comparison. Electricity usage from non-transportation sources is assumed to remain constant in California for the period from 2007 until 2020. Historically, overall electricity usage in the state has grown at approximately 1.13% per year for the period from 1990 until 2007 (EIA, 2007). However, the state's renewable portfolio standard will be costly and difficult to reach even without additional growth, so conservation efforts are assumed to offset growth in this scenario. Another simplifying assumption made is that all additional installed generation capacity added is in the form of wind power. This is justified by the fact that wind power is currently the cheapest of the qualifying renewable technologies. While it is acknowledged that a mixture of renewable generating technologies will be required to meet the standards set, wind power is likely to be the dominant technology and, thus, is the focus of this study. Retiring generation capacity from sources other than wind is assumed to be replaced by a new plant from the same source with equivalent capacity. Finally, while California currently imports approximately 20% of its electricity usage (EIA, 2007), in this study we assume that all generation will occur in-state by 2020, a feat that could currently be accomplished with existing generation capacity.

Wind data used in the analysis comes from the Wind Integration Dataset produced by the National Renewable Energy Laboratory (NREL, 2009). The dataset consists of simulated wind speed and power output for a particular location sampled every 10 minutes for the years 2004, 2005 and 2006. Each location is assumed to contain 10 Vestas V90 3 MW wind turbines with a hub height 100 meters above ground level. The datasets includes over 2,800 possible onshore sites in California that would allow the installation of over 86 GW of capacity. Since the full amount of potential capacity is not needed in the study, sites were chosen based on their simulated capacity factor. Those sites with higher capacity factors were assumed to be chosen before lower capacity factor sites. The 650 sites with the highest average capacity factors over the years 2004-2006 were used in the study, thus with 30 MW per site a total of 19.5 GW is assumed to be added to the system in the most extensive scenario.

Electricity supply and demand have been simulated over the course of a six month summer period in order to determine the amount of wind generation capacity that would need to be added in order to fulfill the California renewable energy mandate of 33% by 2020. The simulation was performed for the cases of PHEVs both with and without V2G capabilities for a number of wind and PHEV penetration rates. In both scenarios PHEVs are assumed to gain adoption to levels in line with current industry projections. The two different PHEV penetration levels used for 2020 are adoption of one PHEV in 10% and 20% of California households. Since California has approximately 1.8 cars per household these penetration rates correspond to approximately 5.5% and 11% of total California light-duty vehicles. While this may seem like a large fraction of the total vehicle stock, it must be considered that there is a ten year timeframe in which the vehicles may be absorbed into the mixture and that even those sold at the beginning of the timeframe should still be in use during the period under study. Additionally, California has historically been an early adopter of similar technologies, with approximately 25% of hybrid vehicles sales in the United States occurring in California. The NREL simulated wind data for the year 2006, using the months of May until the end of October, is used to represent the output of wind generation capacity additions made to fulfill the renewable energy mandate. This addition of wind generation capacity greatly increases the state's total generating capacity. The nameplate capacity

of all the generation units considered without the addition of new wind capacity is nearly 67 GW. Thus the addition of 19.5 GW of wind power capacity would represent a nearly 22% increase in the total generation capacity within the state. Twelve different cases of additional wind capacity additions, PHEV household penetration rates and V2G capabilities have been simulated in order to examine the effects of V2G on the integration of large amounts of wind energy into the California electricity grid.

Case	Additional Wind Capacity	PHEV Rate	V2G?	Renewable Fraction	Wind Fraction
1	7,500 MW	10%	No	19.7%	9.9%
2	7,500 MW	10%	Yes	19.9%	10.2%
3	7,500 MW	20%	No	19.6%	10.0%
4	7,500 MW	20%	Yes	20.1%	10.2%
5	15,000 MW	10%	No	27.5%	17.8%
6	15,000 MW	10%	Yes	27.9%	18.2%
7	15,000 MW	20%	No	28.0%	18.1%
8	15,000 MW	20%	Yes	28.0%	18.2%
9	19,500 MW	10%	No	31.0%	21.4%
10	19,500 MW	10%	Yes	30.7%	21.3%
11	19,500 MW	20%	No	31.1%	21.5%
12	19,500 MW	20%	Yes	30.9%	21.4%

Table 1: Results of the California case studies

At the lower levels of additional wind capacity a small increase in the wind and renewable electricity fractions is noticed when V2G capability is assumed for PHEVs. However, even this small advantage is not replicated as the amount of additional wind capacity is increased. It is also important to note that under none of the scenarios proposed would California reach its goal of producing 33% of the in-state electricity with renewable sources by 2020. This is despite the fact that the closest scenarios would require adding the equivalent of the current California wind generation capacity every year until the mandate deadline. These results appear to paint a rather more pessimistic view of the future for the V2G technology when compared to previous results reported in the literature. However, it must be noted that much of the previous work was done using rudimentary calculation techniques based upon optimistic assumptions, in order to set an upper-bound on the capabilities of the technology. In this study we have used advanced modeling and simulation techniques under a set of assumptions that we consider to be more realistic, most notably allowing charging when a vehicles return home instead of utility controlled vehicle charging, in an attempt to gauge the true usefulness of the technology. However, the period ahead wind output estimate is more likely to be accurate than the day or hour ahead estimates typically used in dispatching, which could lead to decreased V2G power supply. While the wind generation rates used in this study are much higher than has yet been seen in any large market in the United States, it is important to note that Denmark currently has similar

levels of wind power penetration, nearly 20% of electricity production (DEA, 2009) without subsequent electricity storage additions, and has not as yet experienced reliability issues due to its transmission connections to neighboring European nations. California is in a similar situation in that it currently imports electricity and could expect to continue electricity imports if necessary in the short-term. However, the scale of the problem must also be recognized. Denmark has a total installed wind capacity of just over 3GW while California currently has approximately 2.3 GW of wind capacity and would have over 20 GW of capacity with the additions discussed in this study. So if there were a sustained period of wind power downtime in the envisioned scenario the electricity shortfall would be much more difficult to replace with imports or peaking units than in the Danish example. So while battery storage techniques, or more specifically V2G technologies, are not yet critical for wind penetration levels on the order of 20%, they may still be required should even higher levels of intermittent electricity production be desired.

5. Conclusion

A multi-paradigm model of the California electricity system has been used to examine if the adoption of plug-in hybrid electric vehicles with vehicle-to-grid power system capability would aid in integrating large amounts of wind energy into the generating portfolio. The results of the simulation indicate that at wind generation capacity levels of approximately 20% and consumer determined vehicle charging patterns for realistic PHEV penetration rates the effect of V2G capabilities on wind usage rates is minimal. Further studies examining higher levels of wind power capacities and PHEV adoption rates are planned.

References

Danish Energy Agency, 2009, Energy Statistics 2008, www.ens.dk,

Energy Information Administration, 2007, Electric Power Annual 2007, DOE/EIA-0348

B-M. Hodge, J. Pekny, G. Reklaitis, 2009, A Multi-Paradigm Energy Model for Liquid Natural Gas Analysis, Proceedings of the 1st Annual Gas Processing Symposium - Qatar, H. Alfadala, G.V. Reklaitis and M.M. El-Hawagi (eds.)

S. Huang, B-M. Hodge, J. Pekny, G. Reklaitis, 2010, The Impact of PHEV Adoption on Natural Gas Demand in Electricity Generation, Proceedings of the 2nd Annual Gas Processing Symposium - Qatar, F.T. Eljack, G.V. Reklaitis and M.M. El-Hawagi (eds.)

W. Kempton and S. Letendre, 1997, Electric vehicles as a new power source for electric utilities, Transportation Research Part D, 2, 3, 157-175

W. Kempton and J. Tomic, 2005, Vehicle-to-grid power fundamentals: Calculating capacity and net revenue, Journal of Power Sources, 144, 1 268-279

W. Kempton and J. Tomic, 2005, Vehicle-to-grid power implementation: From stabilizing the grid to supporting large-scale renewable energy, Journal of Power Sources, 144, 1, 280-294

H. Lund and W. Kempton, 2008, Integration of renewable energy into the transport and electricity sectors through V2G, Energy Policy, 36, 9, 3578-3587

M. Milligan, 2001, A Chronological Reliability Model to Assess Operating Reserve Allocation to Wind Power Plants, Presented at the 2001 European Wind Energy Conference, Copenhagen, Denmark, July 2-6, NREL – CP-500030490

National Renewable Energy Laboratory, 2009, Wind Integration Datasets – Western Wind Dataset, http://www.nrel.gov/wind/integrationdatasets/western/data.html

W. Short and P. Denholm, 2006, A Preliminary Assessment of Plug-In Hybrid Electric Vehicles on Wind Energy Markets, NREL Technical Report, TP-620-39729

TRANSIMS Open Source, 2009, http://transims-opensource.org

Incorporating Negative Values in AHP Using Rule-Based Scoring Methodology for Ranking of Sustainable Chemical Process Design Options

Mohamad R. Othman,[a,b] Jens-Uwe Repke,[a] Günter Wozny[a]

[a]*Chair of Process Dynamics and Operation, Berlin Institute of Technology, Germany.*
guenter.wozny@tu-berlin.de
[b]*Faculty of Chem. Eng. & Nat. Res., Universiti Malaysia Pahang, Malaysia.*
rizza@ump.edu.my

Abstract

Adopting Analytical Hierarchy Process (AHP) for selection and ranking of sustainable chemical process design that considers the trade-off between economic feasibility, environmental friendliness and social advantages offers good decision methodology for multi criteria problem. Typical AHP problem however are limited to handle only positive preferences. The introduction of negative preferences into AHP often creates various contradicting scenarios that result in spurious and inconsistent decision. To overcome such contradiction a rule-based scoring methodology is proposed. The system works by initially define each indicator according to its value-desirability behaviour. Next, using specific conversion factors that act accordingly to that behaviour the indicators value is converted into a credit-penalize scoring concept. Positive preference is credited with positive value while negative preference is penalized with negative value that shows its undesirability. Using a rule-based approach the score which span over positive and negative value are treated to elicit the final selection and ranking solution. The functionality of the proposed methodology will be demonstrated to selection and ranking of several biodiesel case scenarios in presence of positive and negative preferences.

Keywords: Analytic hierarchy process (AHP), process design, multicriteria decision making,

1. Motivation

Feasibility of process design alternatives is usually influence by techno-economic criteria. With the increasing awareness of sustainability development (SD) it is important to take into account the triple bottom line of sustainability namely environment friendliness and social advantages along side with economic feasibility. These conflicting objectives pose a multi criteria problem in decision making. Some of the commonly used techniques for multi criteria decision making (MCDM) are the analytic hierarchy process (AHP), distance function method and the multi attribute utility theory (MAUT). Among the three techniques, AHP is the most suitable for MCDM problems (Narayanan et al., 2007). In order to evaluate sustainable design option, the three criteria have to be quantified using appropriate indicators. Various indicators have been proposed in the past by various researchers and organizations. A few examples are profitability and rate of return (ROR) to measure economic feasibility, waste reduction (WAR) algorithm and gas emission to measure environmental impacts and safety and hazard analysis for social related criteria. Normally, decision making
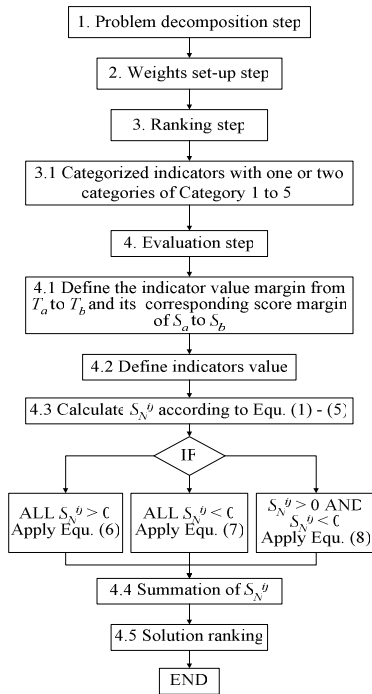
Figure 1. Flow diagram of the method

using AHP involve indicators with positive value without realizing few indicators are constrained to be non-positive such as debts, assets, loss and expression of undesirability (Millet & Schoner, 2005). Some of the indicators mentioned above despite having positive value may span over negative value. Calculating profitability for instance could result in negative value which indicates losses. Other than that, social indicators could also have negative preferences to express undesirability. In addition, the decision indicators may also have different or contradictable value-desirability behaviour. Arguably it can be categorized to either *higher-value-higher-desirability* (HVHD) or *lower-value-higher-desirability* (LVHD). While HVHD behaviour is obvious for example in measuring profitability, LVHD refers to decision indicators which prefers lower value for example environmental indicators i.e. potential environment indicator (PEI) value, CO_2 emission etc. Applying conventional AHP in presence of the above mention conditions will become challenging and could elicit spurious and inconsistent results unless some modifications are made to the AHP process. Millet & Schoner (2005) mentioned two typical approaches have to handle positive and negative values in the AHP. The first is to handle positive and negative value separately and to calculate a benefit to cost ratio. The second approach, which is a standard method, involves inverting negative values into positive preferences. These two approaches however, cause computational complexity and elicit inconsistent results. They then proposed a new approach called Bipolar AHP (BAHP) that introduces modifications to AHP software user interface and also its computational process. It provides a simple solution to accommodating negative preferences while maintaining a true zero reference point. The approach however is highly software dependable and it does not consider the value-desirability behaviour of the indicators. In this work a new approach is proposed using a rule-based scoring methodology. This approach requires several simple modifications to the ranking and evaluation step of AHP. The former step involves defining each indicator according to its value-desirability behaviour. Using specific conversion factor the indicators value is converted into a credit-penalize score reflecting desirability-undesirability. In the evaluation step, treating both positive and negative value simultaneously however may create scenarios which can pose numerical inconsistencies. Therefore, a rule-based approach is proposed to accordingly treat each scenario to elicit the final solution.

2. Overview of the methodology

Figure 1 shows the overall AHP process incorporating the proposed methodology. The modifications introduced are shown in the ranking and evaluation step (step 3 and 4 of the figure). As previously mention decision indicators may span to positive-negative

Incorporating Negative Values in AHP Using Rule-Based Scoring System for
Prioritization of Sustainable Chemical Process Design Options

values and in addition may also have contradictable value-desirability behaviour. The solution idea is to convert the initial indicators value to a scoring system with credit-penalize concept reflecting the accommodation of both HVHD and LVHD behaviour. In this concept any desirable value is credited with positive score while undesirability is penalized with negative score. Such concept will ensure that negative preferences are taken into consideration in the overall score instead of positively making it to lower preferable value. The conversion of the indicator value *v*, into its corresponding score *S*, depends on its behaviour which can be categorized as follows,

- Category 1: Credit score with HVHD behaviour for $v_{0 \to \infty}$ and $T_{a \to b} > 0$

$$S^{ij} = \frac{v^{ij}}{T_a^{ij}} \left(S_a^{ij} \right), \ S_a^{ij} > S_b^{ij}, \ T_a^{ij} > T_b^{ij} \tag{1}$$

- Category 2: Penalize score with HVHD behaviour for $v_{0 \to -\infty}$ and $T_{a \to b} < 0$

$$S^{ij} = \frac{v^{ij}}{T_b^{ij}} \left(S_b^{ij} \right), \ S_a^{ij} < S_b^{ij}, \ T_a^{ij} < T_b^{ij} \tag{2}$$

- Category 3: Credit score with LVHD behaviour for $v_{0 \to \infty}$ and $T_{a \to b} > 0$

$$S^{ij} = S_a^{ij} - \left[\frac{v^{ij}}{T_a^{ij}} \left(S_a^{ij} \right) \right], \ S_a^{ij} < S_b^{ij}, \ T_a^{ij} > T_b^{ij} \tag{3}$$

- Category 4: Penalize score with LVHD behaviour for $v_{0 \to \infty}$ and $T_{a \to b} > 0$

$$S^{ij} = \left[\frac{v^{ij}}{T_a^{ij}} \left(S_a^{ij} \right) \right], \ -S_a^{ij} < -S_b^{ij}, \ T_a^{ij} > T_b^{ij} \tag{4}$$

- Category 5: Credit score with LVHD behaviour for $v_{0 \to -\infty}$ and $T_{a \to b} < 0$

$$S^{ij} = \left[\frac{v^{ij}}{T_b^{ij}} \left(S_b^{ij} \right) \right], \ S_a^{ij} < S_b^{ij}, \ T_a^{ij} < T_b^{ij} \tag{5}$$

where v^{ij} is the original value of *i*-th indicator for *j*-th criteria. The indicators value margin, $T_{a \to b}$ is set from the highest to the lowest value that corresponds to score margin of $S_{a \to b}$. In the ranking stage (step 3.1) requires the assessor to define each indicator to either one or two of the Category 1 to 5 reflecting the indicator's value-desirability behaviour. This is important so that the indicators are evaluated accordingly that reflects its true behaviour. In the evaluation step (step 4) involves firstly determining the range of $T_{a \to b}$ and its corresponding score margin of $S_{a \to b}$. Note that setting the score margin $S_{a \to b}$ however depends on its value-desirability behaviour. For HVHD high value is assign with high score while LVHD assign low score to high value. It is also important to note that a negative score is given to negative preferences as penalization for its undesirability. After the indicators value has been defined (step 4.2) step 4.3 involve converting this value into its corresponding score using equation (1) to (5). The next step involves calculation of normalized score using the weighting vectors determined in the weights set-up step and the score determined previously. The calculation is however not straight forward since the score span over negative and positive value and this

creates contradicting scenarios that could affect the overall result. In order to correctly response to each scenario a rule-based approach is proposed. It consist sets of rules to handle specific scenario as follows,

- RULE 1: IF $S^{ij} > 0$ THEN

$$S_N^{ij} = \frac{S^{ij}}{\sum_1^i S^{ij}} \left(n^j \right) \tag{6}$$

- RULE 2: IF $S^{ij} < 0$ THEN

$$S_N^{ij} = \frac{S^{ij}}{\sum_1^i S^{ij}} \left(-n^j \right) \tag{7}$$

- RULE 3: IF $S^{ij} < 0$ AND $S^{ij} > 0$ THEN

$$S_N^{ij} = \frac{S^{ij}}{\sum_1^i S_{pos}^{ij} - \sum_1^i S_{neg}^{ij}} \left(n^j \right) \tag{8}$$

where S_N^{ij} is the normalized score for i-th alternative in the j-th criteria, n^j is the normalized value for criteria j. and S_{pos} and S_{neg} are the positive and negative score, respectively. Applying these rules helps to properly handles various situation of mix positive and negative scores for correct and consistent solution rankings. Adopting the rule-based scoring methodology offers the opportunity for automated decision support using computer programs i.e. spreadsheet, Visual Basic etc.

3. Application of the methodology

In 2003 Zhang et al. conducted a good techno-economic assessment on four simulated biodiesel processes namely alkali-catalyzed system using virgin oil (Case 1), alkali-catalyzed system using waste cooking oil (Case 2), acid-catalyzed process using waste cooking oil (Case 3) and acid-catalyzed system using hexane extraction (Case 4). Their simulation results concluded that all of these processes are proved to be technically and economically feasible but each had its limitations. Their conclusions however are limited to only techno-economic criteria and furthermore, no selection and ranking of alternatives methodology were applied. To test the functionality of the proposed methodology the four biodiesel processes will be used as case study in this work. As the work focuses only towards techno-economic assessment additional work on environment and social assessment have been performed utilizing the data and reviews included in their work. Note however, for the environmental assessment the effect of energy usage was not considered. The lists of indicators and its categorical behavior are shown in Table 1. This table also includes the range of indicators value and its corresponding score used in this work. A spreadsheet program has been developed embedding all the equations mention earlier to assist decision making. The user only needs to provide the data or parameters as in Table 1 into the spreadsheet. The results are shown in Figure 2. The proposed rule-based scoring methodology was able to recognize the positive-negative values and perform overall ranking of alternatives

Incorporating Negative Values in AHP Using Rule-Based Scoring System for Prioritization of Sustainable Chemical Process Design Options

according to the normalized score priorities. Figure 1a shows the segregation of the results according to the three sustainability criteria. In economic evaluation Case 3 and 4 are the most promising. Although all options have negative net annual profit but based on after-tax rate of return and break-even price of biodiesel, the acid-catalyzed processes (Case 3 and 4) were economically competitive alternatives compared to the alkali process systems. These results are consistent with the work from Zhang. Environment assessment shows a significant difference of environmental performance between different design systems whereby the alkali-catalyzed processes is environmentally friendlier than the acid-catalyzed systems. This result is contributed mostly by the large amount of methanol used by the acid-catalyzed system for transesterification reaction. High amount of calcium sulphate deposited also contribute to the high impact to the environment. Among cases in each system however the difference is small.

Table 1. Summary of the indicators' values and the score conversion specification

Indicators	Cat.	Case 1	Case 2	Case 3	Case 4	Score Conversion			
						Value, $T_{a \to b}$		Score, $S_{a \to b}$	
						T_a	T_b	S_a	S_b
Econ.									
Net annual profit, ($\times 10^{-6}$)	1.2	-2.06	-2.28	-0.35	-0.82	10	-2.28	10	-10
After-tax rate of return, %	1	-85.27	-51.18	-15.63	-21.48	100	≈ 0	10	1
Break-even price, $	3	857	884	644	702	885	0	1	10
Env. friendliness									
Total rate of PEI output	3	26.48	16.47	131.66	125.24	132	0	1	10
Total PEI output/product	3	0.02	0.01	0.12	0.11	0.13	0	1	10
Total rate PEI gen.	5	-550.1	-464.8	-663.7	-608.4	0	-665	1	10
Total PEI gen./product	5	-0.49	-0.42	-0.59	-0.54	0	-0.6	1	10
Societal									
Safety during operation	1.2	3	3	-1	-1	10	-10	10	-10
Operability of the plant	1.2	3	3	1	1	10	-10	10	-10
Safe start-up and shutdown	1.2	5	5	4	4	10	-10	10	-10
Fit for purpose	1.2	10	10	10	10	10	-10	10	-10
Design should meet location specific demands	1.2	10	10	10	10	10	-10	10	-10
Control of product quality and quantity	1.2	-1	-1	3	3	10	-10	10	-10
Maintenance	1.2	5	5	4.5	4.5	10	-10	10	-10
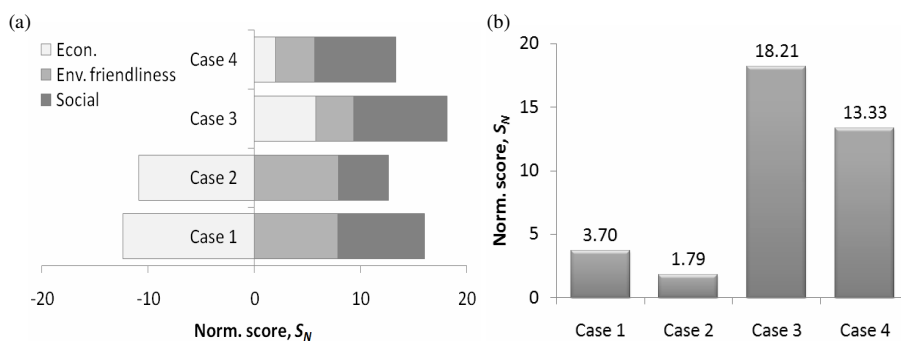
Figure 2. AHP results (a) normalized individual score (b) overall normalized ranking result

Social criteria show almost similar score for Case 1, 3 and 4. The fact that Case 2 was the most complex process with the greatest number of equipment because of the addition of a pretreatment unit for free fatty acid removal (Zhang et al., 2003) make it the least preferable. The overall ranking of the four biodiesel processes is shown in Figure 1b. Introduction of penalization concept makes alkali-catalyzed processes are less preferable than the acid-catalyzed systems mainly because of its huge economic disadvantages. The most sustainably feasible design option is Case 3. Although environmentally unattractive but taking into consideration the trade-off between the two other criteria it perform better the others. However, it is important to note that the decision results using AHP is very sensitive. Any modifications made either to the process models or weights in the AHP weights set-up step could significantly affect the decision outcome. Overall, from the results obtained it shows the proposed methodology is able to successfully consider both desirability-undesirability in design assessment, which cannot be done in the conventional AHP.

4. Conclusions & future work

A rule-based scoring methodology has been proposed to handle positive and negative preferences in process design selection and ranking. The functionality of the approach has been successfully demonstrated through the use of four biodiesel design options. It able to correctly evaluate the credit-penalize score and provide consistent result despite presence of various contradicting scenarios. The useful feature of this methodology is that the modification took place at the numerical calculation step of AHP specifically at the ranking and evaluation step thus enable decision makers to focus more on the real issues in process design development. Potentially, such approach offers the opportunity for automated decision support e.g. using spreadsheet.

5. Acknowledgements

References

[1] D. Narayanan, Y. Zhang, M.S. Mannan, 2007, Vol 85 (B5) 349-359.
[2] I. Millet, B. Schoner, 2005, Computers & Operations Research, Vol. 32, pp. 3163-3173.
[3] Y. Zhang, M.A. Dub, D.D. McLean, M. Kates, 2003, Bioresource Technology, 89, pp. 1–16.

Integration of CAPE and LCA Tools in Environmentally-Conscious Process Design: A Case Study on Biomass-Derived Resin

Yasunori Kikuchi[a], Kazuya Mayumi[a], and Masahiko Hirao[a]

a Department of Chemical System Engineering, The University of Tokyo, Eng. Bldg. 5, 7-3-1 Hongo, Bunkyo-ku, Tokyo 113-8656, Japan , kikuchi@pse.t.u-tokyo.ac.jp

Abstract

Life cycle assessment (LCA) is a strong tool for quantifying the environmental impact associated with a product life cycle. To implement LCA into environmentally-conscious process design, the lack of inventory data should be addressed by computer-aided process engineering (CAPE) tools. This study proposes the process design framework integrating CAPE and LCA tools. Missing inventory data is estimated by process simulation with design information extracted from literatures and design heuristics. At this time, the sensitivity of design settings to LCA results should be analyzed to generate design requirements for further process design stages. This study demonstrates the proposed procedure by a case study on biomass-derived polypropylene (PP). In the case study, it is demonstrated that the potential of biomass-derived resin to reduce environmental impact due to the production of PP. At the same time, it was also revealed that the environmental impact could be increased by implementation of biomass resource as raw material, because of the increase of GHG emission in biomass cultivation site.

Keywords: Process simulation, Life cycle assessment, Activity modeling

1. Introduction

Environmental impacts of emerging technologies such as production of resin derived from biomass have become an issue, that is, environmentally-conscious process design should be addressed in designing a process [1]. For sustainable chemical production, the life cycle of product should be taken into account. Life cycle assessment (LCA) has become a strong tool to quantify life cycle impacts [2, 3]. The evaluation results have a great impact on not only product and process design, but also social system design for sustainability. For appropriate and detail LCA in practical decision-making, various information, especially adequate inventory data, is required, which might be a hard obstacle to perform LCA at the early phase of process design. At that time, computer-aided approach can be a strong support to devise a solution to deal with and share a huge amount of information, alternative technologies, and constraints on design with multi-stakeholders in product life cycle. CAPE can play a role of enabling data acquisition and processing for LCA of social process system installing novel emerging technologies, because it has great abilities to combine available engineering knowledge and to estimate adequate data required for LCA.

In this paper, we propose a systematic method of process design integrating CAPE and LCA tools. As an actual case study, the production process of biomass-derived resin was designed. In this case study, process simulator and other CAPE databases were

utilized as data resources for LCA. Based on the case study, a procedure to perform LCA with CAPE tools was discussed.

2. Framework of Environmentally-Conscious Process Design Integrating CAPE and LCA Tools

2.1. Design procedure with LCA

Process design phase can be divided into several steps based on the detail of treated information and the number of alternative candidates, which are conceptual, basic, and detail design. Although there is the limitation of available information in early design phases such as conceptual and basic design, wide range of alternative candidates can be considered with simulation-oriented optimization [4]. LCA requires various kinds and much amount of information on the life cycle of product. This is why LCA has been regarded as the assessment methodology for detail design phase or existing ones. To consider potential alternatives reducing environmental load on life cycles, however, process designers should be able to perform LCA at early design phase. The applicability of CAPE tools should be enhanced to enable LCA at early design phase, where enough information for the quantification of environmental impacts cannot be obtained directly.

Design procedure with LCA includes activities required for the collaboration of CAPE and LCA tools. To visualize such complicated business model, conventional flow chart is not enough and the utilization of novel approach should be taken into account. The type-zero method of Integrated DEFinition language, or IDEF0 [5] is applicable to systematically describe the activities of environmentally-conscious process design integrating CAPE and LCA tools. IDEF0 can visualize the required activities complexly intertwined with information on process design and evaluation. It has been utilized for the modeling of the business model implementing the integrated process design [6, 7]. In business process reengineering field, IDEF0 are applied for the design of novel business systems implementing new business procedure [8]. IDEF0 was applied to visualize a design framework integrating CAPE and LCA tools.

2.2. Design framework integrating CAPE and LCA tools

Fig. 1 shows the activities required for environmentally-conscious process design. The design framework mainly consists of early design, detail design, and construction phases. Total management and resource providing activities are incorporated to facilitate these design phases [7]. In each design step, LCA can play a role of checking the environmental performance of design alternatives. Comparing LCA results and constraints about environmental impact, alternatives might be redesigned or eliminated. If design requirements can be specified for reduce the increase of environmental impacts, such information is outputted to downstream design phases, which is described in fig. 1 as outputted information from activity A2 "design requirement to meet reduction target".

Process inventories required for LCA are not available at early design phase. Process simulation is inevitable to perform LCA. In this process simulation, parameters might be designed not as specific values, but as possible range of values to remain design flexibility in detail design phase. For example, the approach temperature between hot and cold streams in heat exchangers or maximum compression ratio in compressors cannot be designed specifically at early design phase. These design flexibility can be a variability factor of total environmental impact. At the same time, there is the

Integration of CAPE and LCA Tools in Environmentally-Conscious Process Design: A Case Study on Biomass-Derived Resin
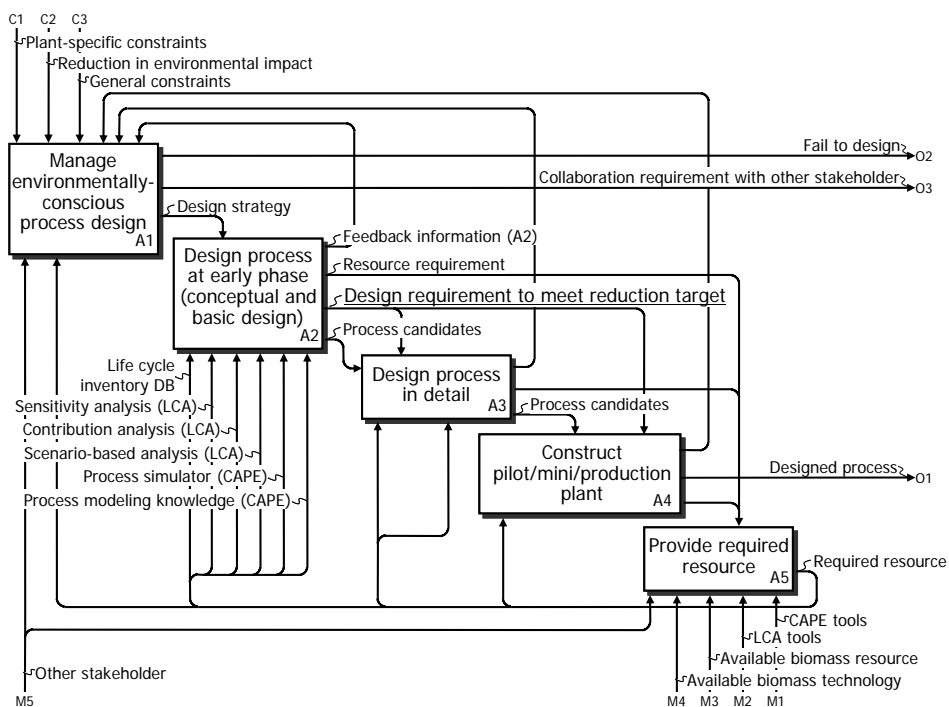
Fig. 1 Activities of environmentally-conscious process design

uncertainty factors of environmental impact due to the biomass-resource cultivation process, for example, transformation of land use [9], nitrogen fertilizer [10], and the adequacy of biomass to produce bio-ethanol [10]. These variability and uncertainty factors should be taken into account in early design phase with the integrated utilization of CAPE and LCA tools.

3. Case Study on Biomass-Derived Polypropylene

3.1. Synthetic route from biomass to polypropylene

Possible synthetic routes were investigated from literatures and patents. Fig. 2 schematically shows the life cycle boundary of biomass-derived polypropylene targeted in this case study. Two synthetic routes were considered as the production technologies of biomass-derived PP. The first one is bio-ethanol route. Bio-ethanol is available from biomass through fermentation [10]. Bio-ethanol can be easily converted into ethylene by dehydration [11]. Dimerization and metathesis reaction can produce butane from ethylene [12] and propylene from butane and ethylene [11]. The dehydration, dimerization, and metathesis reaction network can synthesize propylene from bio-ethanol. The second route is synthesis gas route. By the pyrolysis of biomass through gasification, synthesis gas can be obtained with a little of hydrocarbon [13]. Through the adjustment of composition of synthesis gas such as the ratio of H_2/CO, methanol can be synthesized [14]. Methanol-to-olefin reaction process can produce propylene [14].

To obtain process inventories, double lined processes in fig. 2 were simulated by Aspen Plus^{TM} or Aspen HYSYS^{TM}. As an example of process simulation, process flow
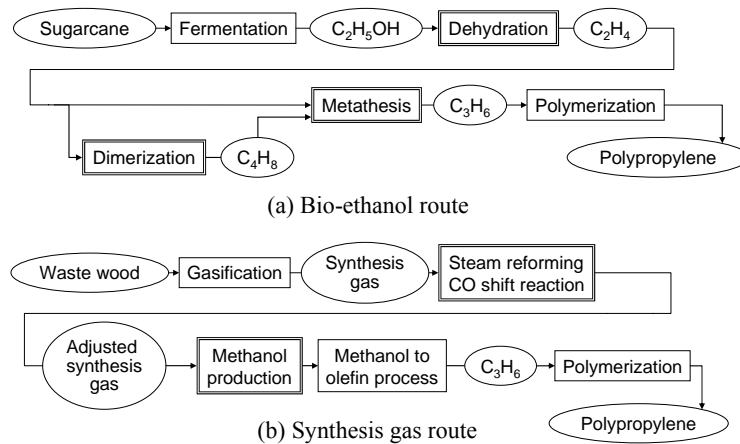
Fig. 2 Production processes of biomass-derived polypropylene
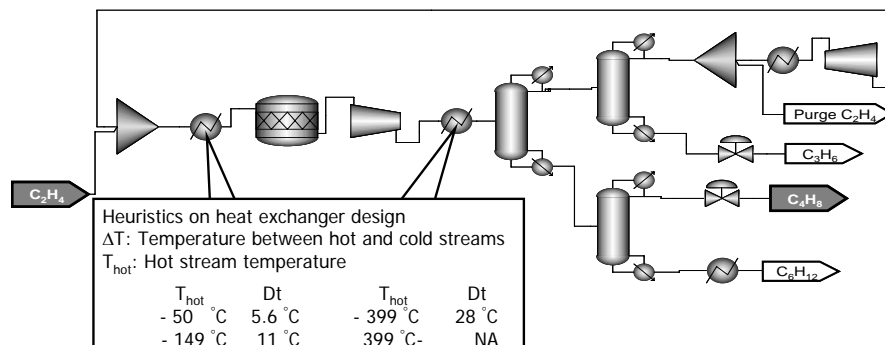
Fig. 3 Process flow diagram: dimerization process

diagram (PFD) of dimerization process in the bio-ethanol route is shown in fig. 3. This process includes the reactor for the dimerization and the separator network composed of three distillation columns for separating products and by-products. At the separation part in fig. 3, butene and by-products are separated, and unreacted ethylene is recycled to the reactor. Design parameters including reaction temperature, pressure and conversion ratio were obtained from literatures. Regarding heat exchanging, power, and separation systems, unit operations were designed based on available design heuristics [15]. For example, designing heat exchanging system, we followed pinch technology and set the minimum approach temperature between hot and cold streams (ΔT_{min}) as 50 C. Designing power system, the maximum compression ratio of each compressor was set at 2.5 and the maximum decompression ratio of each gas turbine at 0.4 [4]. Designing separation system, reflux ratios were designed as one and a half times of the minimum reflux ratios.

3.2. Scenario setting
LCA is performed for the production of the unit amount of PP from source to resin, that is the cradle-to-gate of the life cycle of PP. Table 1 organizes the scenario settings assessed in this case study. Cases were defined based on the condition of PP production: raw material, ΔT in heat exchanger, land transformation for cultivation, nitrogen

Integration of CAPE and LCA Tools in Environmentally-Conscious Process Design: A Case Study on Biomass-Derived Resin

fertilizer, and the location of PP production plant. PP production from fossil resource was set as case 1. Synthesis gas route from waste wood was set as case 2 and Bio-ethanol routes under different conditions were cases 3 and 4. In cases 3 and 4, the scenario settings were set based on the variability and uncertainty in the life cycle of PP to clarify the margin of error. When the conditions were set as the parenthetic ones, the LCA results showed the maximum or minimum environmental impacts.

3.3. Results and discussion

Fig. 4 shows the LCA result of scenarios in Table 1. This result demonstrates that biomass-derived PP has potential to reduce environmental impacts. In this regard, however, several points should be addressed at the early design phase for meaningful environmentally-conscious process design.

Table 1 Scenario settings assessed in case study. Parenthetic conditions are ones in the maximum or minimum environmental impact required to analyze the margin of errors.

	Case 1	Case 2	Case 3	Case 4
Raw material	Fossil resource: naphtha	Biomass: waste wood	Biomass: sugarcane	Biomass: sugarcane
ΔT in heat exchanger	-	50 C	50 C (0 C)	Heuristics (0 and 50 C)
Land transformation for cultivation	-	None	Forest to agricultural land (None)	None
Nitrogen fertilizer	-	None	Proper amount (twice proper amount)	Proper amount
Location of PP production	Japan	Japan	Brazil (Japan)	Brazil
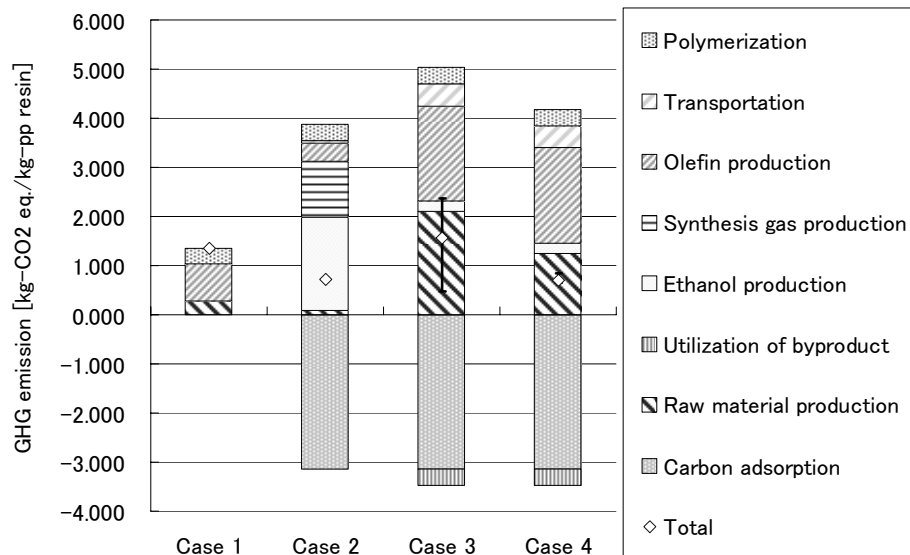
Fig. 4 LCA results of scenarios organized in table 1

Fig. 4 demonstrates that the biomass cultivation process has highly sensitive factors to total environmental impact. At least, the condition of case 3 results in the increase of GHG emission rather than PP production from fossil resource. The margin of error of the result of case 3 means that biomass-derived PP may increase environmental impact significantly. For example, if farmer transforms land use style to prepare crop land, the amount of fixed carbon in ground will vary over the year. At the case of transformation of land from forest to agricultural land for sugarcane, the fixed carbon in the ground decreases from 95.8 Mg/ha to 72.8 Mg/ha in fifty years [9]. Because decreased carbon in the ground is emitted to the air as CO_2, such transformation of land use results in the increase of GHG emission. Comparing the results of cases 3 and 4, the amount of GHG emission and the margin of error in case 4 result are much less than those in case 3. In case 4, the conditions of biomass cultivation were managed well, in other words, the biomass resources with law GHG emission in cultivation could be collected by a successful procurement from quality supplier. The small margin of error due to the design flexibility of heat exchanger means that the decision might be turned back by the conditions of biomass resources procurement.

4. Conclusion

Biomass-derived PP has great potential to reduce the environmental impact originating from the production of resin. In this regard, however, a large margin of error, that is uncertainty, must be addressed at the early design phase. The design flexibility of heat exchanger has a much less sensitivity to total environmental impact than the uncertainty of biomass cultivation condition. Sharing this kind of information can lead to effective process and life cycle design. Based on these results, we concluded that the integration of CAPE and LCA tools enables the effective assessment of environmental impact of novel technologies before their industrialization. The assessment results can reveal the considerations for environmentally-conscious process design.

References

[1]   J. A. Cano-Ruiz, and G. J. McRae. Ann. Rev. Ene. Environ. 23 (1998) 499-536.

[2]   A. Azapagic, and R. Clift. 1999. Comput. Chem. Eng. 23 (1999) 1509-1526.

[3]   H. Sugiyama, M. Hirao, R. Mendivil, U. Fischer, and K. Hungerbühler. Proc. Saf. Environ. Protec. 84 (2006) 63-74.

[4]   L. T. Bieglar, I. E. Grossmann, A. and W. Westerberg. Systematic Methods of Chemical Process Design. New Jersey: Prentice Hall PTR. (1997)

[5]   D. T. Ross. IEEE Trans. Soft. Eng. 3 (1977) 16-35.

[6]   Y. Kikuchi and M. Hirao, Proc.19th ESCAPE, (2009) 1123-1128

[7]   H. Sugiyama, M. Hirao, U. Fischer, K. Hungerbühler. J Chem. Eng. Jpn.. 41(9) (2008) 884-897.

[8]   Y. Naka (eds.), Introduction to Integration Engineering (original title in Japanese), Kogyo Chosakai, Tokyo, 2006

[9]   C. C. Rhoades, G. E. Eckert and D. C. Coleman, Ecol. Appl. 10, (2000) 497-505

[10] I. C. Macedo, E. A. J. E. A. Seabra and J. E. A. R. Silva. Biomass Bioenerg.. 32, (2008) 582-595

[11] WO2007/055361

[12] M. Iwamoto, PETROTECH, 27, 8, (2004) 628-632

[13] MEXT, Japan. Leading project: Project on the multi treatment and recycling of domestic and industrial wastes, and biomass (original title in Japanese). (2005)

[14] K. H. Kaggerud. Doctoral thesis, NTNU (2007)

[15] W. W. Seider, J. D. Seader and D. R. Lewin, "Product & Process Design Principles Synthesis, Analysis, and Evaluation", John Wiley & Sons, Inc. (2004)

Energy integration issues for hydrogen and electricity co-production based on gasification process with Carbon Capture and Storage (CCS)

Calin-Cristian Cormos,[a] Serban Agachi [a]

[a] *Babes – Bolyai University, Faculty of Chemistry and Chemical Engineering*
11 Arany Janos Street, RO-400028, Cluj – Napoca, Romania,
Emails: cormos@chem.ubbcluj.ro; sagachi@staff.ubbcluj.ro

Abstract

Integrated Gasification Combined Cycle (IGCC) is a technology for power production in which the feedstock is partially oxidized with oxygen and steam to produce syngas. In a conventional IGCC design without carbon capture, the syngas is purified for dust and hydrogen sulphide removal and then sent to a Combined Cycle Gas Turbine (CCGT) for power production. Carbon capture technology is expected to play a significant role in the coming decades for curbing the greenhouse gas emissions.

IGCC is one of the power generation technologies having the highest potential to capture carbon dioxide with the low penalties in term of efficiency and cost. In a modified IGCC design for carbon capture, the syngas is catalytically shifted to maximize the hydrogen level in the syngas and to concentrate the carbon species in the form of carbon dioxide that can be later capture in a pre-combustion arrangement. After carbon dioxide capture, the hydrogen-rich syngas can be either purified in a Pressure Swing Adsorption (PSA) unit and exported to the external customers (e.g. PEM fuel cells) or used in a CCGT for power generation.

This paper investigates the most important energy integration issues for hydrogen and electricity co-production scheme based on coal gasification process with carbon capture and storage (CCS). The coal-based IGCC case study produces around 400 MW net electricity and a flexible hydrogen output in the range of 0 to 200 MW (LHV) with 90 % carbon capture rate. The principal focus of the paper is concentrated on overall energy efficiency optimization by better heat and power integration of the main plant sub-systems (e.g. integration of steam generated in gasification island and syngas treatment line into combined cycle, integration of PSA tail gas in the power block, heat and power demand for Acid Gas Removal unit, ASU – GT integration etc.).

Keywords: IGCC; Hydrogen and electricity co-production; Energy integration aspects; Heat and power integration; Carbon capture and storage (CCS)

1. Introduction

The introduction of hydrogen in energy system as a new energy carrier is exciting much interest, as this offers significant advantages including reduction of greenhouse gas emissions at the point of end use, enhancement of the security of energy supply and improvement of economic competitiveness. Hydrogen can be produced from different feedstocks, such as natural gas, oil derived products, coal and water [1]. Usually it is used in the chemical and petrochemical sectors but in the future there is hope than it can be largely used in transport sector (e.g. Proton Exchange Membrane – PEM fuel cells).

It is well known that solid fossil fuels reserves (mainly coal and lignite) give a much bigger energy independence compared with liquid and gaseous fossil fuels [2] but coal utilization is regarded with concern because of bigger greenhouse gas emissions associated with it. In the future, solid fuel gasification (either fossil or fossil with addition with biomass or waste) is likely to play a key role in large-scale hydrogen and electricity production [3]. These processes will be based on entrained flow gasification as this type of gasifier maximizes hydrogen production and facilitates the capture of carbon as CO_2, whereby it can be stored in geological reservoirs or used for Enhanced Oil Recovery (EOR) [4].

Integrated Gasification Combined Cycle (IGCC) is one of the power generation technologies having the highest potential to capture CO_2 with low penalties in efficiency and cost [5]. In a modified IGCC design for carbon capture, syngas is catalytically shifted to maximize the hydrogen level in the syngas and to concentrate the carbon species in the form of CO_2 that can be later capture in a pre-combustion arrangement. After CO_2 and H_2S capture in a double stage Acid Gas Removal (AGR) system, the hydrogen-rich syngas is used in a CCGT for power generation of for production of purified hydrogen (using a Pressure Swing Adsorption unit) which can be used in (petro)chemical industry or for transport sector in hydrogen-fuelled fuel cells.

This paper investigates energy integration issues for co-generation of hydrogen and electricity from coal gasification with carbon capture. The focus of the paper is concentrated on analyzing and proposing practical ways to optimize the overall energy efficiency of the plant by better heat and power integration among the plant sub-systems e.g. steam integration between gasification island, syngas conditioning line and the steam cycle of the combined (Rankine) cycle, influence of heat and power demand for Acid Gas Removal unit, possible gains from air integration of the Air Separation Unit (ASU) and the gas turbine compressor, sensitivity analysis (e.g. influence of ambient condition like) for the power block, plant flexibility in term of varying hydrogen output, integration of the tail gas resulted from hydrogen purification step into the power block (compressing and burning in the gas turbine or HRSG duct burning), influence of hydrogen-fuelled gas turbine in comparison with syngas-fuelled gas turbine etc.

The case study investigated in the paper will produce a flexible ratio between power and hydrogen in the range of about 400 MW electricity and 0 – 200 MW hydrogen (based on lower heating value – LHV) with 90 % carbon capture rate.

2. Plant configuration and main design assumptions

In Integrated Gasification Combined Cycle scheme with carbon capture process, the syngas resulted from the coal gasification for maximizing the hydrogen content via water – gas shift conversion (WGS). Shift conversion stage has also the role to concentrate the carbon species present in the syngas in form of carbon dioxide that can be later captured in a pre-combustion arrangement (Acid Gas Removal unit). The hydrogen-rich gas resulted after AGR is split, part is further purified (>99.95 % vol.) and send to external customers and the rest is used for power generation in a combined cycle. The conceptual layout of a modified IGCC scheme for co-generation of hydrogen and electricity with simultaneous carbon capture is presented in Figure 1 [6-9].

The main differences of IGCC scheme without carbon capture and a similar plant with carbon capture are the following: introduction of carbon monoxide shift conversion stage, an AGR unit which capture in addition of H_2S also CO_2, introduction of captured carbon dioxide drying and compression stage, hydrogen purification unit and finally a hydrogen-fuelled gas turbine (compared with a syngas-fuelled gas turbine).

Energy integration issues for hydrogen and electricity co-production based on gasification process with Carbon Capture and Storage (CCS)
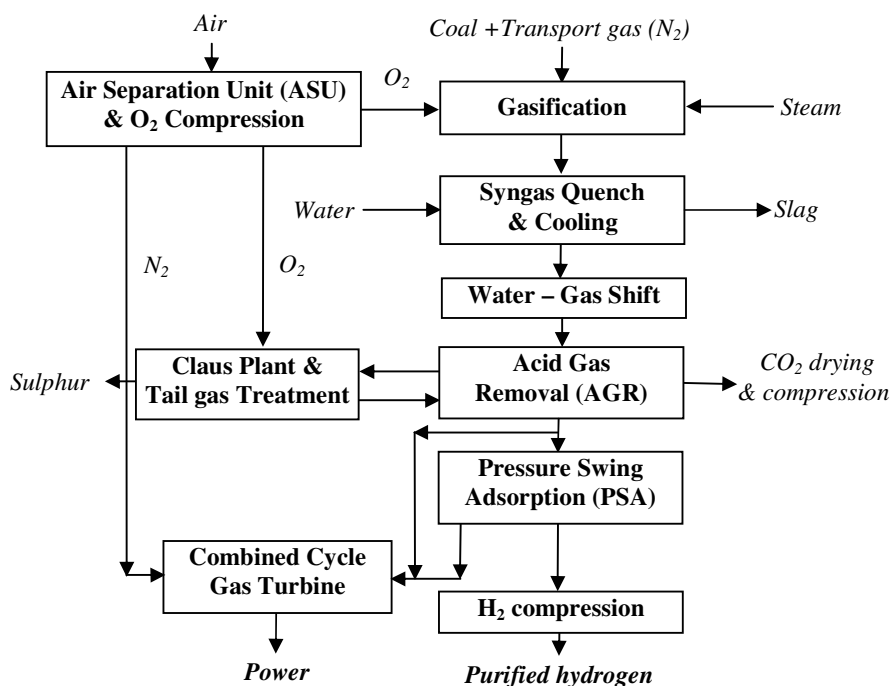
Figure 1. Layout of IGCC scheme for hydrogen and power co-generation of power with carbon capture

The main sub-systems of the plant for poly-generation of various energy vectors and theirs design assumptions used in the mathematical modeling and simulation are presented in Table 1 [6-8,10].

Table 1. Main design assumptions

Unit	Parameters
Air separation unit (ASU)	Oxygen purity: 95 % (vol.) Delivery pressure: 2.37 bar Power consumption: 225 kWh/ton O_2
Gasifier (Siemens)	Pressure: 40 bar Pressure drop: 1.5 bar Temperature: >1400°C Water quench
Shift conversion (WGS)	Sulphur tolerant catalyst Two adiabatic beds Pressure drop: 1 bar / bed
Acid gas removal (AGR)	Solvent: Selexol® Separate capture of CO_2 and H_2S Solvent regeneration: pressure flash
CO_2 compression and drying	Delivery pressure: 100 bar Compressor efficiency: 85 % Solvent used for drying: TEG (Tri-ethylene-glycol)
Claus plant & tail gas treatment	Oxygen-blown H_2S-rich gas composition: > 20 % (vol.) Tail gas recycled to H_2S absorption stage
Pressure Swing Adsorption (PSA)	Purified hydrogen: > 99.95 % (vol.) Purification yield: 85 % Tail gas pressure: 1.5 bar (recycled to the power island)

Gas turbine	Gas turbine type: M701G2 (Mitsubishi Heavy Industries)
	Net power output: 334 MW
	Electrical efficiency: 39.5 %
	Pressure ratio: 21
	Turbine outlet temperature (TOT): 588ºC
Heat recovery steam generation (HRSG) and steam cycle	Three pressure levels: 118 bar / 34 bar / 3 bar
	Reheat of MP steam
	Steam turbine isoentropic efficiency: 85 %
	Steam wetness ex. steam turbine: max. 10 %
Heat exchangers	$\Delta T_{min.} = 10$ºC
	Pressure drop: 1 % of inlet pressure

3. Evaluation of energy integration issues

The whole configuration of IGCC scheme with carbon capture was simulated with ChemCAD software. The case study was simulated in different situation (only electricity or various modes of hydrogen and electricity co-production in the range of 0 – 200 MW hydrogen). For electricity only mode, the gas turbine (GT) is running full load and for co-production modes on part load operation.

The simulation results were used to make a heat and power integration study of the Combined Cycle Gas Turbine (power island) for maximization of power generation. The steam generated in the gasification island and syngas conditioning (HP and LP steam) were integrated in the steam cycle of the CCGT. Also, the heat duties for various units in the plant (gasification, AGR, power block) were extracted from the steam cycle. The main steam cycle data for power generation only are presented in Table 2.

Table 2. Steam (Rankine) cycle parameters

Stream	Unit	Parameters
HP steam from process	t/h	128.00 @ 337.93ºC / 120 bar
HP steam to HP Steam Turbine	t/h	435.85 @ 573.35ºC / 118 bar
MP steam after MP reheat	t/h	469.55 @ 446.35ºC / 34 bar
MP steam to process units	t/h	35.3 @ 415.15ºC / 41 bar
MP steam to AGR (solvent reg.)	t/h	27.80 @ 250.63ºC / 6.5 bar
LP steam from process units	t/h	82.00 @ 201.62ºC / 3 bar
LP steam to LP Steam Turbine	t/h	606.70 @ 182.51ºC / 3 bar
Cooling water	t/h	31500 @ 15ºC / 2 bar
Hot condensate returned to HRSG	t/h	814.42 @ 115ºC / 2.8 bar
Flue gas at stack	t/h	2810.31 @ 108.47ºC / 1.01 bar

Hot and cold composite curves of CCGT are presented in Figure 2.
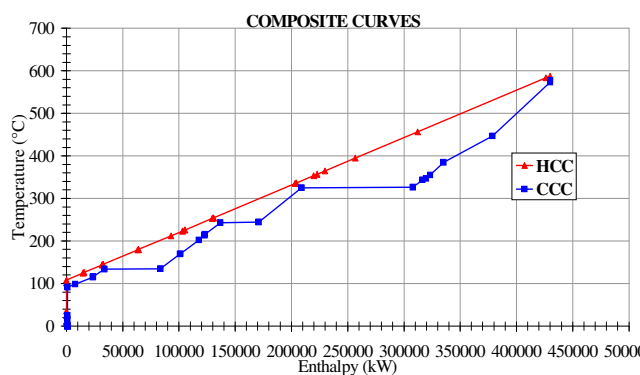
Figure 2. Composite curves for combined cycle gas turbine (CCGT)

Energy integration issues for hydrogen and electricity co-production based on gasification process with Carbon Capture and Storage (CCS)

Table 3 presents the overall plant performance indicators for the IGCC scheme with carbon capture operated in both scenarios: power only and hydrogen and power co-generation. For co-generation mode, the gas turbine is turned down in order to displace an energy stream (in form of hydrogen-rich gas) than can be purified. The tail gas resulted from the PSA unit is integrated back in the CCGT being compressed and mixed with main fuel line to GT.

Table 3.Overall plant performance indicators

Main Plant Data	Units	Power	Power + Hydrogen		
Coal flowrate (a.r.)	t/h	165.70	165.70	165.70	165.70
Coal LHV (a.r.)	MJ/kg	25.353	25.353	25.353	25.353
Feedstock thermal energy – LHV (A)	MW_{th}	1166.98	1166.98	1166.98	1166.98
Thermal energy of the syngas (B)	MW_{th}	934.75	934.75	934.75	934.75
Cold gas efficiency (B/A * 100)	%	80.10	80.10	80.10	80.10
Thermal energy of syngas exit AGR (C)	MW_{th}	830.70	830.70	830.70	830.70
Syngas treatment efficiency (C/B *100)	%	88.86	88.86	88.86	88.86
Gas turbine output (1 x M701G2)	MW_e	334.00	316.62	297.83	279.04
Steam turbine output (1 ST)	MW_e	197.50	187.44	178.60	169.19
Expander power output	MW_e	0.78	0.72	0.66	0.61
Gross electric power output (D)	MW_e	532.28	504.78	477.09	448.84
Hydrogen output (E)	MW_{th}	0.00	50.00	100.00	150.00
ASU consumption + O_2 compression	MW_e	44.72	44.72	44.72	44.72
Gasification island power consumption	MW_e	8.08	8.08	8.08	8.08
AGR + CO_2 drying and compression	MW_e	40.07	40.07	40.07	40.07
Hydrogen compression		0.00	0.66	1.34	2.01
Power island power consumption	MW_e	19.00	19.70	20.07	21.16
Total ancillary power consumption (F)	MW_e	111.87	113.23	114.28	116.04
Net electric power output (G = D - F)	MW_e	420.41	391.55	362.81	332.80
Gross electrical efficiency (D/A * 100)	%	45.61	43.25	40.88	38.46
Net electrical efficiency (G/A * 100)	%	36.02	33.55	31.08	28.51
Hydrogen efficiency (E/A * 100)	%	0.00	4.28	8.56	12.85
Cumulative efficiency (G+E/A * 100)	%	36.02	37.83	39.64	41.36
Carbon capture rate	%	92.35	92.35	92.35	92.35
CO_2 specific emissions (power)	kg/MWh	76.12	81.51	87.73	95.38
CO_2 specific emissions (power + H_2)	kg/MWh	76.12	72.08	68.77	65.74

As can be noticed from the Table 3, for power only case, overall plant efficiency is decreased with about 7 – 8 % compared with a conventional IGCC scheme without carbon capture [5,7]. This efficiency decrease is due to the penalty of capturing the carbon dioxide and it can be noticed compared the specific CO_2 emission figure compared with about 700-800 kg/MWh without capture [5].

Another fact that has to be mentioned is that for hydrogen and electricity co-production mode, the overall efficiency of the plant is increasing in the situation in which the ancillary power consumption is remaining virtually constant (see Table 3). This fact is very important and attractive for plant cycling (modification of the power generated by the plant according to the demand of the grid) considering that for low electricity demand the plant can produce mostly hydrogen which compared with power can be stored to be used either for covering the peak loads or for other applications (transport sector, petro-chemical sector etc.).

A way to increase the overall plant efficiency with about 1.5 – 2 % is to integrate Air Separation Unit (ASU) and the gas turbine compressor [11]. An air bleed from the compressor outlet of the GT can supply part or all of the air required by the

ASU. Typically, the total air required by the ASU amounts to around 20 - 25 % of the GT compressor air. The degree of (air) integration is usually defined as the percentage of the total ASU air required coming from the GT compressor. Figure 3 presents the variation of net electrical efficiency of the plant vs. ASU – GT air integration degree.
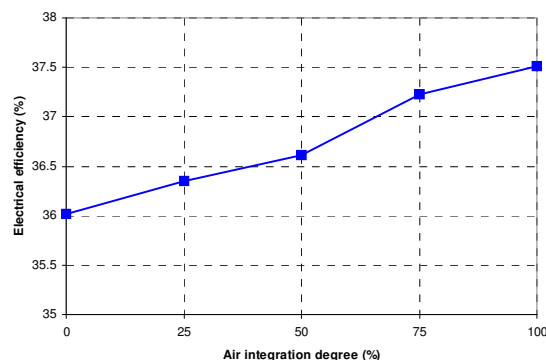
Figure 3. Net electrical efficiency gain from ASU integration with GT

4. Conclusions

This paper analyzes from energy integration point of view, the coal gasification for co-generation of hydrogen and electricity with CCS. Modeling and simulation techniques were used to evaluate different ways of improving main overall plant performance indicators (e.g. steam integration, ASU – GT integration, plant flexibility).

5. Acknowledgements

References

[1] F. Müller-Langer, E. Tzimas, M. Kaltschmidtt, S. Peteves, Int. J. Hydrogen Energy 32 (2007) 3797

[2] Statistical Review of World Energy BP 2008, www.bp.com

[3] European Hydrogen and Fuel Cell Technology Platform. Hydrogen Energy and Fuel Cells – A vision for our future. Report EUR 20719, https://www.hfpeurope.org/, 2005

[4] C. Cormos, F. Starr, E. Tzimas, S. Peteves, Brown A., Gasifier concepts for hydrogen and electricity co-production with CO_2 capture, Third International Conference on Clean Coal Technologies, Cagliari, Sardinia, Italy, 2007

[5] E. Tzimas, A. Mercier, C. Cormos, S. Peteves, Energy Policy 35 (2007) 3991

[6] P. Chiesa, S. Consonni, T. Kreutz, R. Williams, Int. J. Hydrogen Energy 30 (2005) 747

[7] International Energy Agency (IEA) – Greenhouse Gas Programe (GHG), Potential for improvement in gasification combined cycle power generation with CO_2 capture, Report PH4/19, 2003

[8] C. Cormos, F. Starr, E. Tzimas, S. Peteves, Int. J. Hydrogen Energy 33 (2008) 1286

[9] C. Higman, M. Van Der Burgt, Gasification, Elsevier Science, Second edition, 2008

[10] C. Cormos, Int. J. Hydrogen Energy 34 (2009) 6065

[11] O. Maurstad, An overview of coal based Integrated Gasification Combined Cycle (IGCC) technology, Massachusetts Institute of Technology, 2005

Life Cycle Assessment of a High Temperature Molten Salt Concentrated Solar Power Plant

Vincenzo Piemonte[1], Marcello De Falco[1], Pietro Tarquini[2], Alberto Giaconia[2]

[1]*Department of Chemical Engineering Materials & Environment, University of Rome "La Sapienza" via Eudossiana 18 00184 Rome-Italy*

[2]*ENEA (Italian National Agency for New Technologies, Energy and Sustainable Development), Casaccia Research Centre, via Anguillarese 310 00123 Rome-Italy*

Abstract

The well-known world energetic matter, mainly due to the worldwide growing energy consumption gone with a reduction of oil and gas availability, and to the environmental effects of the indiscriminate use of fossil fuels in our economy, is leading to the development of clean innovative technologies for the reduction of GHG emissions and the creation of a more sustainable economic structure worldwide. But, realizing and installing renewable energy plants have an environmental "footprint" that has to be evaluated to quantify the real impact of renewable technologies on the environment. Nowadays, the most important tool to evaluate this impact of a product is the Life Cycle Assessment (LCA). To this aim, several impact categories are defined; among these the most important are the Global Warming, the Abiotic Depletion, the Eutrophication, the Acidification, the Land Use and the Human toxicity.

The aim of this work is to present a Life Cycle Assessment of an innovative solar technology, the Molten Salt Concentrating Solar Power (CSP) plant, developed by Italian Research Centre ENEA and able to produce clean electricity by using solar energy. The Life Cycle Assessment was carried out by means of the SimaPro7 software, one of the most used LCA software in the world. It is worth assess that these types of software are an indispensable tools for leading LCA studies. In the second part of the study the environmental performance of the CSP plant was compared with these of conventional oil and gas power plants.

Keywords: LCA, Concentrated Solar Power Plant, Renewable Energy, Conventional Power Plant.

1. Introduction

The well-known energetic issue is stimulating the development of clean innovative technologies for the reduction of GHG emissions and the creation of a more sustainable economic structure worldwide. The exploitation of renewable energy sources for heat and energy production is commonly considered the most promising way to reduce the impact of human activities on environment, since clean energy technologies allow using no-fossil derived energy, without producing pollutants and GHGs. On the other hand, realizing and installing renewable energy plants have an environmental "footprint" due to the utilization of construction materials, transport, maintenance, final disposal, etc. It is extremely important evaluating this impact for renewable plants, and LCA is a crucial tool to understand how reducing technologies environmental footprint.

Among renewable ene rgy technologies, those exploiting solar energy seem to be t he most ap plicable th anks to t he h uge an d diffuse solar en ergy av ailability. Th e present work is focused on an innovative solar technology, the Molten Salt Concentrating Solar Power (CSP) plant, developed by Italian Research Centre ENEA.

CSP p lant (Fi gure 1) b asically co nsists o f a solar collector field, a re ceiver, a heat transfer flu id lo op; a suitable heat storage syste m is al so requ ired to maximize th e "capacity factor" (i.e . product ivity) of t he s olar plant, a nd to provide sol ar heat at the desired rate regardless the instantaneous solar radiation availability and fluctuations [1]. The mirrors of the solar field concentrate the direct solar radiation on the solar receiver set at the foca l point. T he heat transfe r fl uid rem oves the hi gh temperature s olar h eat from the receiver and it is afterwards collec ted into an insulated heat storage tank to be pumped, o n demand, t o the heat use rs where i t rel eases i ts sensi ble heat. Fi nally, t he heat carrier fluid is stored i nto a lowe r temperature tank ready to restart the solar heat collection l oop. A pr oper di mensioning o f the heat st orage sy stem all ows to dri ve t he process in continuous.

Recently, some molten nitrate mixtures at te mperatures up to 550°C have bee n positively tested as con venient, co st-effective an d en vironmental frien dly h eat tran sfer fluid and storage medium for CSP plants [2,3].

The high temperature molten salt sensible heat is used t o generate high pressure steam to be sent to a steam turbine Rankine cycle for the production of clean electrical energy.

The aim of this work is to evaluate the performance of the proposed CSP plant from an environmental poi nt of view by the use of a the Li fe C ycle Assessm ent M ethodology which i s base d o n cal culations a nd anal ysis of e ffects t o en vironment, h uman heal th, socioeconomic factors and climate change.
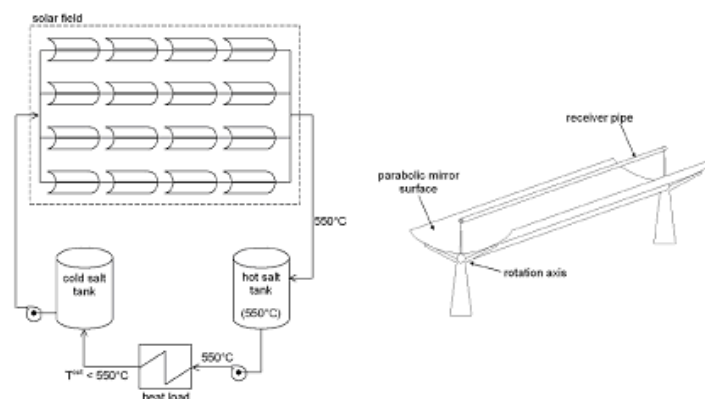
Figure 1: Simplified scheme of a single solar trough collector section (right) and a power plant with two-tank molten salt storage (left).

2. LCA methodology, software & Data base

The "Life Cycle Assessm ent", methodology allows to evaluate a set o f interactions that a pr oduct o r servi ce has with t he en vironment, consi dering its whole life cycle that includes t he preproduction p oints (extraction a nd p roduction of ra w m aterials), production, di stribution, use (i ncluding re -use a nd m aintenance), recy cling a nd fi nal disposal. So the objective s of the LCA are to evaluate the effects of th e in teractions between a p roduct an d t he envi ronment, hel ping t o understand t he envi ronmental impacts directly or indirectly caused by the use of a given product.

Life Cycle Assessment of a High Temperature Molten Salt Concentrated Solar Power Plant

In com pliance with IS O 14 040 and 1 4044, the Life C ycle Assessm ent is ac hieved through four distinct phases: first you need to make the description of the system under consideration, the evaluation methods used and the purpose of study (Goal and Sc ope). The second phase includes products manufacturing (including extraction and production of raw materials, production and distribution of the final product) and its final disposal system. For e ach process, the em issions into the environment (air, water, soil) or more generally the cost i n en vironmental t erms (including, i f p ossible, a need of wat er an d land) must be taken into account. This phase is called Life Cycle Inventory (LCI). The third phase is called Li fe Cycle Im pact Assessm ent (LC IA) an d pr ovides f or t he determination of a wi de range of cat egories of en vironmental impact (global warming, abiotic d epletion, Eu trophication, Acidification, et c..). T he l ast phase c onsists i n t he analysis o f the resu lts ob tained fro m the assessm ent o f th e in dividual categ ories of environmental impact; in this phase the processes of normalization and weighting must be carried out to obtain global results (overall impact indices), which are often used to compare the environmental impact of two products.

If th e an alysis is p erformed d irectly o n t he categories of environmental im pact, such methodology is called "Mid-point approach." A v iable and v alid altern ative is represented by the "End-point approach "or" Damage-oriented approach. "

In general, t he LC A can be conducted by assessing t he envi ronmental fo otprint of a product f rom raw m aterials to pr oduction (Cradle t o gat e), o r t o be e xtended t o t he whole product life cycle, including its disposal (Cradle to grave).

In this work the LCAs have been performed by using the Eco-Indicator 99 methodology which develops an "End-point" approach: the typical impact categories a forementioned are n ormalized an d grouped i n t hree dam age cat egories: Dam age t o Hum an Heal th, Damage to Eco system Qu ality an d Damage t o Reso urces [4]. Fu rthermore t he International Pan el for th e Cli mate Ch ange (IPCC) an d th e Cu mulative En ergy Demande (C ED) methods were also used to estimate the global warming and the total energy requirements.

The so ftware used fo r t he realizatio n o f t he LCA studies reported in th is wo rk is SimaPro7. The LCI of the CSP plant has been realized on the basis of the construction data directly provided by the ENEA centre, while the LCAs of the conventional oil and gas p ower pl ants hav e bee n per formed by using dat a i ncluded i n t he Ec oinvent v .2.0 database. The data relate to international scenarios, which cover the entire industrialized world.

From how reported above it is ev ident the contribution of co mputer simu lations to th e LCA realization a nd there fore to t he envi ronmental ass essment of the process plant considered in this study.

3. LCA Results

In th e tab le 1 are reported so me p reliminary ag gregated d ata relativ es to th e construction of a CSP pl ant for a continuous energy production of about 400 kWe for 300 days/year (i.e. 2880 MWhe/year). A bi omass furnace was also considered t o provide a pl ant back -up ene rgy of 1.58 M Wth, required to bal ance protracted cl oudy periods. The LCI data are directly provided by the Italian research centre ENEA.

The LC A was pe rformed co nsidering t he m ost i mportant m aterials use d for construction as well as th e energy requ irements fo r constru ction, wh ile tran sport was implicitly i ncluded i n e nergy cons umption. M aterials u sed wa s ex pressed per M J produced el ectricity over i ts life t ime, assum ed 2 0 y ears. In o rder t o p erform a LC A cradle to grave the disposal of the building was also considered.

The most important LCA results for the CSP plant are summarized in table 2. It is worth noting t hat th e h igh value assu med fro m t he im pact cate gories is m ainly d ue t o th e environmental cont ribution of t he b uilding m aterials pr oduction. In e ffect t he production of 1 kg of stainless steel is associated with an emission of 3.93 kg of $CO2_{eq}$.

Table 1: Raw materials, energy demand and land use for the CSP plant building

Raw material	Amount	Raw material	Amount
Concrete (kg)	334973	Mineral wool (kg)	6474
Steel (kg)	384000	Biomass (kg)	746048
Stainless steel (kg)	116284	**Energy Demand**	
Glass (kg)	62239	Solar radiation (MJ/year)	70006154
Plastics (polypropylene) (kg)	28000	Electricity (Italian mix) (MJ/year)	9027
Sodium nitrate (kg)	877241	**Land use**	
Potassium nitrate (kg)	584827	CSP plant (m^2)	36000
Zinc (kg)	116423	Biomass growth (m^2)	750000

Table 2: Main LCA's results for the CSP plant

Impact category	Value/year
Global warming 100a ($kgCO2_{eq}$)	141788.4
Ozone layer depletion 25a ($kgCFC-11_{eq}$)	0.012
Human toxicity 100a ($kg1,4-DB_{eq}$)	91302.9
Acidification ($kgSO2_{eq}$)	730.376
Eutrophication ($kgPO4^{3-}_{eq}$)	72.801
NOx ($kgNOx_{eq}$)	686.385

*Life Cycle Assessment of a High Temperature Molten Salt Concentrated Solar
Power Plant*

In the second part of the LCA study, the environmental performance of the CSP plant
was compared with respect to those of an oil power plant and a gas power plant
characterized by the same productivity of the CSP plant considered in this work.
The results of the LCAs comparison are reported in figures 2 and 3 in terms of
cumulative energy demand and climate change evaluated at 100 years, respectively.
From the pictures is evident that the CSP plant requires a high quantity of renewable
energy (from solar, biomass, water), while fossil energy duty is about 80% less that that
required from a gas power plant and 90% less that that required from an oil power plant.
These findings agree with a much lower emission in terms of $CO2_{eq}$ (measured as
Global Warming Potential) reported by the CPS plant with respect to these of the oil
and gas power plants (figure 3).
In the figure 4 it is also reported the comparison of the three power plants considered in
this work by using the Eco-indicator 99 methodology in terms of damage categories.
The figure clearly highlights the lower impact of the CSP plant both for the damage
category "Resources" and "Human Health", while its impact for the "Ecosystem
Quality" is substantially comparable with that of the oil power plant and slowly higher
that of the gas power plant.
From an overall point of view the figure 4 suggests that the CSP plant is always
preferable to the oil power plant (its impact is lower than that of the oil power plant for
all the three damage categories). On the contrary the comparison with respect to the gas
power plant must be further elaborated by means of the mixing triangle technique [5].
From figure 5 it is evident that the gas power plant is preferable to CSP plant only for
very high values of ecosystem quality macro-category, that is only if we assign a very
low importance to the resources depletion and to the human health.
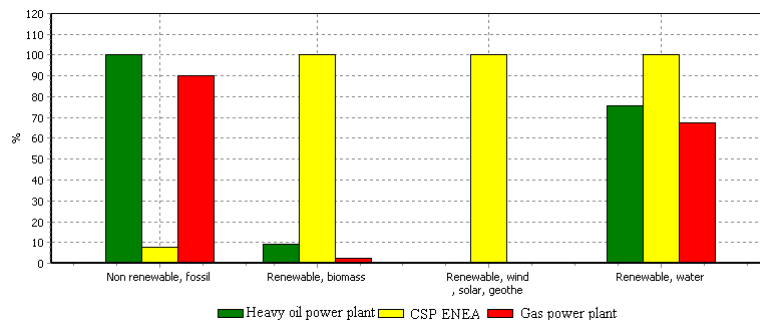
Figure 2: LCAs Comparison in terms of Cumulative Energy Demand (CED)
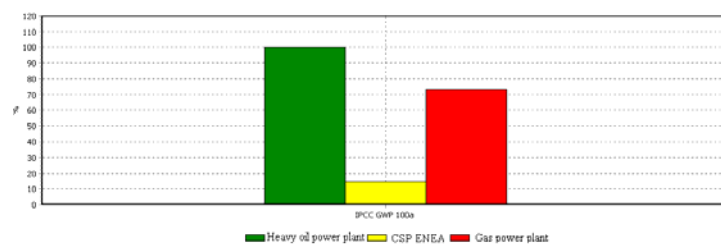
Figure 3: LCAs Comparison in terms of Global Warming 100a (IPCC)
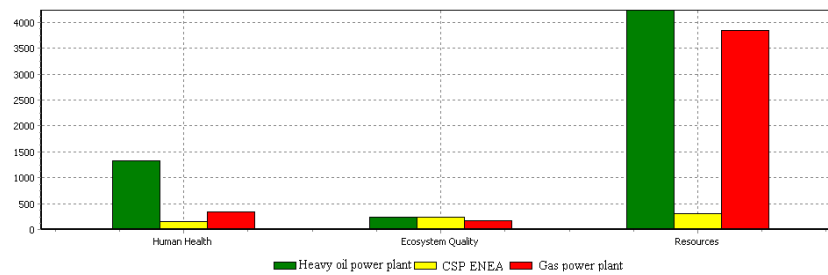
Figure 4: LCAs Comparison by means of the Eco-indicator 99 Methodology: damage oriented approach
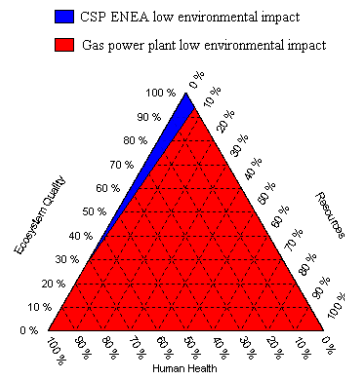
Figure 5: Comparison of the global impact of the CSP plant vs. a gas power plant by using the mixing triangle approach

4. Conclusions

In this work the CSP plant performance was assessed f rom an envi ronmental point of view by usi ng t he LC A methodology. The CSP plant was also com pared to conventional power plant (gas and oil) in order to evaluate its reliability. Even if this is only a preliminary study the results are very interesting: by assigning reasonable values to the three damage categories used in the eco-indicator 99 methodology, the CSP plant is always p referable with resp ect to th e co nventional p ower p lants. Th is find ing confirms the high potentials of this innovative plant technology.

References

[1] C.J. Winter, R.L. Sizmann, L.L. Vant-Hull, *Solar Power Plants*. Springer-Verlag Ed, New York, 1991.
[2] D. Mills, *Advances in solar thermal electricity technology.* Solar Energy. 76 (2004) 19-31.
[3] U. Herrmann, D.W. Kearney, *Survey of Thermal Energy Storage for Parabolic Trough Plants.* ASME J Sol Energy Eng. 124 (2002) 145-151.
[4] Goedkoop, M., Effting, F., and Collignon, M. 2000. The Eco-indicator 99, A damage oriented method for Life Cycle Impact Assessment, PRé Consultants B.V. second edition.
[5] P. Hofstetter , A. Braunschw eig, M. Me ttier, R.M. Wenk, O . Tietje, The Mixing Triang le: Correlation and Graphical Decision Support for LC A-based C omparison, Journal of Industr ial Ecology, 3(4) 2008, 97-115.

Sustainable Design of Different Seawater Reverse Osmosis Desalination Pretreatment Processes

Matan Beery[a], Günter Wozny, Jens-Uwe Repke

[a]*Institute of Process and Plant Technology, TU Berlin, Sekr. KWT 9, Strasse des 17. Juni 135, 10623 Berlin, Germany, matan.beery@mailbox.tu-berlin.de*

Abstract

Sustainable development is usually presented using the three pillar model: economy, environment and society. This research uses life cycle based assessments in conjunction with process system engineering methods for the design and analysis of sustainable processes with the example of pretreatment of seawater membrane desalination. At the first step, the sustainability potential of several state of the art process alternatives is analyzed (namely granular media- and membrane-based filtration). Conducting a broad environmental life cycle assessment (LCA, using GaBi4® software) together with a life cycle costs assessment (LCC, according to SETAC guidelines) forms the so called eco-efficiency analysis (based on a definition by Germany's institute for applied Ecology) which together with a qualitative process performance assessment gives a good idea about the sustainability potential of the different process alternatives. The results show that the current trend of using a membrane based pre-treatment is proven here to not always be a more sustainable process solution.

Keywords: Sustainability Assessment, Desalination, Life-cycle, Eco-efficiency

1. Introduction and Methods

Today, product and process sustainability are key issues considered in planning and designing industrial projects. This is especially true when concerning indispensable, broad-spectrum commodities such as water. The process of sea water reverse osmosis (SWRO) desalination is rapidly increasing in popularity to become the most commonly used method of producing potable water in arid, coastal areas. Producing sufficient amounts of desalinated water in a sustainable manner is one of humanity's most immediate challenges. It is clear today that if human beings are to survive on this planet, the continued development of our technological society must be done in a way which will meet the needs of the present, without compromising the prospects of future generations. It is often depicted using the 'three pillar' model, achieving the goal of sustainability when the three pillars: society, economy and environment complement one another (Fig. 1). Although some have already assessed seawater desalination and its pre-treatment methods according to environmental [1,2] or economical aspects [3], a combined approach taking into account all aspects of sustainable development is still missing. The common approach used today sets out to inspect and analyze each of the three pillars using life cycle methods. The product is examined over its entire life cycle, from the excavation of the natural resources, to the production processes, the use-phase by the client, and eventually to the product's disposal/recycling. This so called 'cradle-to-grave' approach is crucial in making sure all aspects are accounted for. The environmental life cycle assessment is done using the LCA (stands for 'life cycle assessment') method and is the only one of the three that is internationally standardized (ISO 14040 + 14044).
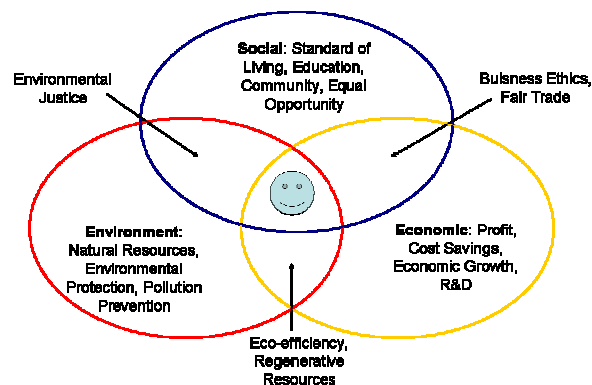
Figure 1: Three pillars of sustainability and their intercepting regions

It is composed of four steps: Goal definition, Inventory analysis, Impact assessment, and Interpretation. All calculations and measurements are scaled according to the 'functional unit', a pre-defined quantity which characterizes the process/product and makes it possible to compare alternatives. The popular commercial LCA software, GaBi4, was used due to its high acceptance by the industry. Two main drawbacks of an LCA are its inability to consistently integrate elements of risk assessment and to directly quantify the loss of biodiversity.

The economical life cycle assessment is done using the well known life cycle costing method (LCC). Here, all costs attributed to the life cycle of the product are taken into account. It is sometimes referred to as Environmental LCC, as it often has a more holistic nature than a normal cost assessment, accounting for all environment related costs (such as recycling) and standing side by side to an environmental LCA. It should be emphasized that LCC should always include only real monetary flows and that overlapping with LCA should be avoided in order not to double count for the same impacts. The aggregated and discounted results of the life cycle costing are usually expressed in the price of the functional unit. There is still no standard for conducting an environmental LCC however the guidelines of the Society of Environmental Toxicology and Chemistry (SETAC) are widely accepted by both scholars and industries. The current state of the art of societal life cycle assessment is still far from being ripe for international standardization and will therefore be only introduced here on a qualitative discussion level. The quantitative assessment of environmental and economical aspects is often referred to as eco-efficiency (see Fig. 1) as it inspects the trade-offs between better environmental performance and the necessary costs involved or vice versa. Such analyses have been documented for several other processes and products [4,5] but never before to seawater desalination. It should be made clear that all kinds of life cycle assessments are not ment for absolute scaling rather as tools for comparing alternatives and assisting in the process of decision making [5]. It is the goal of this work to conduct a similar comprehensive sustainability assessment to the pre-treatment process of SWRO desalination.

2. System Outline

Due to the very delicate nature of the reverse osmosis membranes, it is crucial that the water fed to them would be pre-treated to have a good quality. At the moment, the most common pre-treatment methods include coagulation/flocculation, media filtration and cartridge filtration, as well as the addition of different chemicals to the sea water (for pH

Sustainable Design of Different Seawater Reverse Osmosis Desalination Pretreatment Processes

control, disinfection, anti-scaling etc.). However, some newer pre-treatment methods have been suggested and examined in recent years, such as the use of membrane filtration (usually capillary dead-end UF). The two main process steps taken into consideration here are given in figure 2. Since the input parameters of the SWRO system as well as its surroundings are location-dependent, a narrowing down of the basic assumptions is needed if one wants to quantitatively assess the sustainability potential of different process designs.
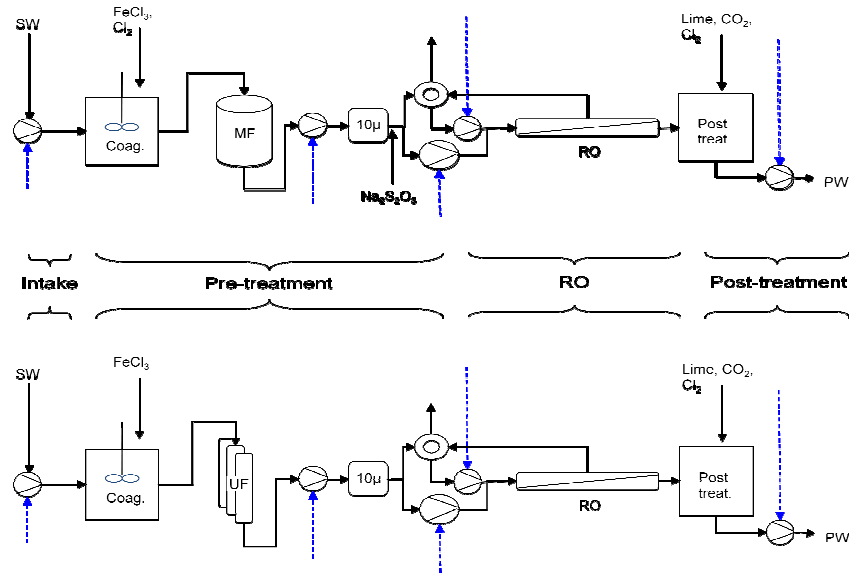
Figure 2: Basic steps of SWRO pre-treatment: conventional (top) and membrane (bottom)

The first step that had to be taken was deciding on the input water quality and some basic conditions that would be representative of state of the art SWRO plants. It was assumed that a 189 mega liters per day (MLD, or 50 MGD) SWRO plant was to be built on a coast with access to low to moderately fouled surface seawater. The following assumptions were made regarding the intake seawater quality: TDS of 35,000 ppm; Temperature of 25°C; pH of 7.8 – 8; NTU of 2.5; SDI of 4; The conventional set-up was based on coagulation followed by single stage multi-media gravity filters (1.5 m anthracite, 0.75 m sand, 0.6 m garnet) and 10 μ cartridge filters. The RO system is of a single-stage single-pass type with 7 elements per pressure vessels, using an isobaric energy recovery device. An over-all recovery rate of 45% with RO flux of 12.75 lmh (7.5 gfd) was used for all calculations. The loading rate of the filters was assumed to be 2.45 m³/m²/h. No anti-scaling additives were assumed necessary under these conditions. Disinfection by chlorination should also be avoided when possible as it is believed to be the cause of increased RO bio-fouling potential due to the break-down of organics in the water. However shock chlorination (and respectively de-chlorination with sodium bisulfite) might be inevitable from time to time and therefore intermittent chlorination was taken into account.

The UF alternative consists of a 120 micron strainer followed by inside-out capillary pressure driven membranes and 10 micron cartridge filters. The UF design calculations were based on an average filtration flux of 85 lmh and a backwash flux of 250 lmh (including one daily chemical backwash with 100 ppm free chlorine). The same 45% process recovery rate and 12.75 lmh RO flux were considered although a modification to these parameters is possible due to the better filtrate quality of the UF permeate (less

fouling potential). This was avoided in order to have a stronger basis of comparison between the alternatives as changing them requires changing of downstream processes. Both alternatives use iron (III) chloride as a coagulant agent (5 ppm for the conventional, 0.5 ppm for the membrane). Other operational data was based on literature averages. An important point that should be raised is the fact that in some cases an unanticipated increase of bio-fouling in UF based pre-treatment systems has been documented [6]. The addition of a UV disinfection unit after the UF membranes has been considered here as a third process design alternative.

3. Results

Environmental LCA

The system's boundaries were set to be the same as the plant's boundaries and the functional unit was defined as one cubic meter of product water. Due to lack of documented information, the impacts of production and transportation of the technical equipment (including membrane and UV lamp replacement) could not be taken into account but they are generally regarded to have a very small impact compared to the plant's operation. The information for the upstream chains was based on information from the software's internal data-base, the German ProBas project and previous publications in this area. The results of the impact assessment showed a clear dominance of the energy production over the impacts resulting from the other processes. The appraised climate change potential of the different alternatives is shown in figure 3. The green house gas emissions of the UF alternative are about 70 grams higher (in eq. CO_2 per m³ product water) than those resulting from the conventional media filtration. This alludes to a yearly additional emission of about 4600 ton CO_2 (equivalent to the average yearly emission of about 3300 cars). Not a very large amount, but it definitely disagrees with the desired global goal of massive CO_2 reduction in the near future.Other less important impacts proved to be marine ecotoxicity (mostly due to the use of chlorine) and ozone depletion (mostly due to the pre-production of coagulant).
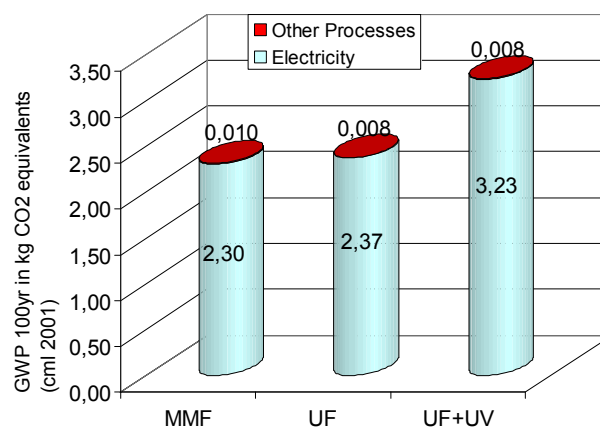
Figure 3: Climate change environmental impact of SWRO

Life Cycle Costing

The method of life cycle costing takes into account all money flows involved in a products' life cycle. The use-phase and end-of-life (EoL) costs are ignored since they are less meaningful in this case and have the same value for all of the pre-treatment alternatives. Input cost-flows are accounted for using the purchase price and not their

Sustainable Design of Different Seawater Reverse Osmosis Desalination Pretreatment Processes

"real" cost (excluding suppliers' profit margins) as recommended by SETAC. This is due to the fact that many of the input cost flows relate to commodities that hold very low profit margins and that other costs are hard to estimate as they relate to classified information, such as the production costs of membrane elements. The project life time was assumed to be 25 years with an interest rate of 6%. All estimations were based on common market prices in early 2009. The results can be seen in figure 4. An interesting conclusion of this analysis is that the membrane pre-treatment system, usually considered more expensive to purchase (requires about 5 million dollars more in capital investment), is actually irrelevant when amortizing the costs over the plants' entire life time, especially when assuming a slightly higher yearly availability factor in the UF case due to the ability of the membranes to better handle sudden changes in intake water quality. The savings resulting from lower chemical consumption (mostly coagulant) and the increased life-time of the RO membranes and cartridge filters are more than enough to cover for the extra costs associated with the UF system operation (such as energy and membrane replacements). Over-all the UF system offers a saving of 3 US Cents per cubic meter of product water: A significant saving of about 2 million Dollars per year.
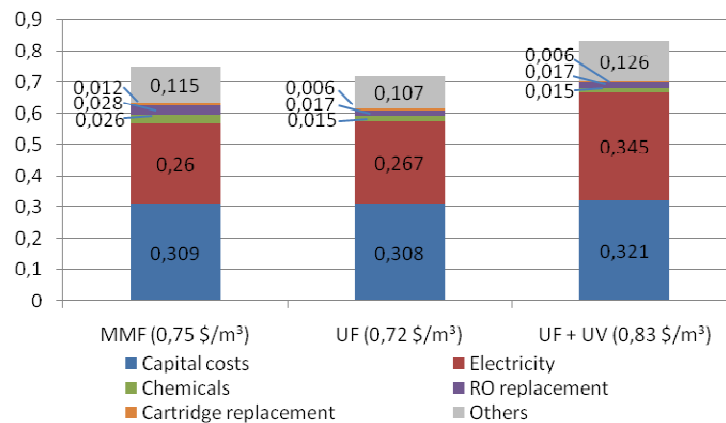
Figure 4: LCC of different SWRO pre-treatment configurations. All numbers are in $/m³

Eco-efficiency and Process Performance

Eco-efficiency is usually defined as the relation between environmental improvements to economical expenditure. At this point an aggregation of the environmental impacts to a single environmental performance score is usually unavoidable. The method of CML 2001 experts IKP was used to produce single, weighted, over-all environmental impact scores. These were 6.28e-13, 6.43e-13 and 8.76e-13, corresponding to conventional, UF and UF+UV accordingly. It is obvious that the UF + UV alternative is the less preferable of the three as it means more costs for less environmental performance. As to the sustainability based decision between a conventional and a UF pre-treatment, it is clear that one should proceed with caution. If the main goal is to reduce the total environmental impacts, then the conventional pre-treatment at the current state of the art still has an advantage over the membrane pre-treatment. However the relative difference in environmental performance (about 2.4 %) is smaller than the difference in LCC values (4 %), which gives an advantage to the membrane, especially in case the environmental impacts resulting from electricity production could be reduced. Such a reduction could be achieved by changing the electricity supply to a "greener" one such as a gas combined cycle (as in the Ashkelon plant) or wind turbines (as in Sydney and

Perth). Another promising option would be to optimize the design and operation of the membrane trains with the goal of minimizing the energy and chemical consumption, for example by working with lower filtration fluxes. However, a trade off with increased capital costs must be considered (lower flux means more membrane area is needed). Qualitative societal and operation performance issues might also be helpful in making a design decision. The UF employs a higher degree of automation which on one hand reduces the required human intervention but on the other hand includes many small moving technical parts which are more susceptible to malfunctions. Membrane pre-treatment systems are also limited in their flexibility as they are currently all proprietary systems, exclusively bonded to one installing manufacturer's membranes and services (unlike the RO system). Additionally, despite being more robust against sudden changes in intake water conditions (already taken into account in the plant availability factor), the membrane systems could be problematic in the case of increased bio-fouling. Even in sporadic events this would mean higher chemical and energy consumption as well as higher replacement rates of the RO membranes and cartridge filters, negatively affecting the eco-efficiency of the plant. In more persistent cases, the addition of a sterilization process (such as UV irradiation) might be inevitable. The membrane therefore stands at a slight disadvantage, mostly due to the issues of flexibility and robustness.

4. Conclusions

The goal of this work was to objectively assess the current sustainability of the most common seawater RO desalination pre-treatment methods. This analysis showed that when integrating all three aspects of sustainability, the membrane pre-treatment in this case, although being more economical, is somewhat less preferable in the environmental and societal views, mostly due to its higher energy demand (which shows a strict dominance over the chemical use) and to its lower degree of flexibility (proprietary systems) and robustness (bio-fouling risk). Therefore when working with non difficult waters (as was the case assessed here), the gravity media filter is currently still a more sustainable technological solution. In other cases, the needed extension of the conventional pre-treatment process train (to include flocculators, settlers, DAF, second stage filtration, etc.) will decrease its eco-efficiency, thus favoring with the membrane pre-treatment. The membrane based pre-treatment is never the less a more modern technology still at the beginning of its learning curve, with a large improvement potential at its disposal. Further optimization of its design and operation concerning the over-all process sustainability is called for.

References

[1] S. Latteman: Environmental Impact assessment of Desalination Plants, Lecture Notes, DME Seminar on Environmental Issues in Desalination, Berlin 2009.
[2] C. Fritzmann, C. Niewersch, A. Jansen, T. Wintgens, T. Melin: Comparative Life Cycle Assessment Study of Pretreatment Alternatives for RO Desalination, IDA World Congress-Maspalomas,Gran Canaria –Spain October 21-26, 2007.
[3] P. Glueckstern and M. Priel: Comparative Cost of UF vs. Conventional Pretreatment for SWRO Systems, Desalination & Water Reuse, 12(4) and 31(1) (2003) 34–39.
[4] R. Grießhammer, M. Buchert, C. Gensch, C. Hochfeld, A. Manhart, I. Rüdenauer: PROSA – Product Sustainability Assessment (Final Report), Institute for Applied Ecology, Freiburg 2007.
[5] W. Kloepffer: Life Cycle Sustainability Assessment of Products, Int J LCA 13 (2008) 89-95.
[6] M. Wilf: RO, NF, Membrane filtration and MBR Technology for Potable and Wastewater Reclamation Applications, Lecture Notes, EDS Course, L'Aquila 2009.

An Optimisation-based Approach for Integrated Water Resources Management

Songsong Liu,[a] Petros Gikas,[b,c] Lazaros G. Papageorgiou[a]

[a] *Centre for Process Systems Engineering, Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE, UK, E-mail: l.papageorgiou@ucl.ac.uk*
[b] *Department of Environmental Engineering, Technical University of Crete, Chania, 73100, Greece*
[c] *Hellenic Ministry of Environmental Planning and Public Works, (a) General Secretariat of Public Works,Special Service of Public Works for Greater Athens Sewerage and Sewage Treatment, Varvaki 12, Athens, 11474, and (b) Central Water Agency, Patission 147, Athens, 11251, Greece*

Abstract

This paper considers an integrated water management problem for a region lacking fresh and ground water resources, which comprises (a) the optimal placement of desalination, water reclamation and wastewater treatment plants, (b) the calculation of the optimal capacities of the above facilities, and (c) the calculation of the optimal conveyance system for desalinated, reclaimed water and wastewater. This problem is formulated as a mixed-integer linear programming (MILP) model, with an objective to minimise the annualised total cost including capital and operating costs. Finally, the proposed model is applied to a real case for the Greek island of Syros in the Aegean Sea.

Keywords: integrated water resources management, MILP, desalination, wastewater treatment, water reclamation

1. Introduction

In the last decade, optimisation techniques have widely been used in the field of integrated water resources management. Medellín-Azuara et al. [1] applied an economic-engineering optimisation model to explore and integrate water management alternatives such as water markets, reuse and seawater desalination in Ensenada, Mexico. Han et al. [2] presented a multi-objective linear programming model to allocate various water resources among multiusers and applied it to obtain reasonable allocation of water supply and demand in Dalian, China. Cunha et al. [3] presented an optimisation model for regional wastewater systems planning, together with a simulated annealing (SA) algorithm to optimise layout of sewer networks, location of treatment plants, etc.

With the increasing water consumption worldwide, study of using various water sources available to fulfill the water demand has become an important issue recently. In this work, we aim to develop an optimisation-based approach using mixed-integer linear programming (MILP) techniques for the integrated water resources management in a water deficient insular area, where fresh water importation is a particularly expensive and non-sustainable option [4]. The alternative water recourses, which can meet the demands for water, are seawater desalination and water reclamation from wastewater.

2. Problem Statement

In this work, we consider an insular area which is water deficient. The water demand is exclusively satisfied by desalinated seawater and reclaimed water from wastewater. Desalination yields potable quality water at a relatively high cost, while reclaimed water can be used for non-potable urban, industrial and agricultural applications at production cost significantly lower to that of desalinated water.

The area is divided into several sub-regions based on the population distribution and land terrain. The optimal allocation in each region of desalination, wastewater treatment and water reclamation plants are to be determined. Wastewater is collected from all sources and is conveyed to a wastewater treatment plant, where it is treated to meet the specific discharge limits. Then, part of the treated wastewater may undergo further treatment (at an extra cost) in order to meet the reclaimed water quality criteria, while the remaining is discharged into the sea. For simplicity, it is assumed that there is no water loss in the process of wastewater treatment and water reclamation. The desalinated water can be used as potable water; but it may also be augmented with reclaimed water for non-potable uses. The schematic graph of the water/wastewater flows in a region is given in Fig. 1. It is assumed that both qualities of water and wastewater are allowed to be freely distributed among most of the regions. Thus, the infrastructure needs for water distribution and storage, including the main pipeline network, pumping stations, and storage tanks, are also optimised.
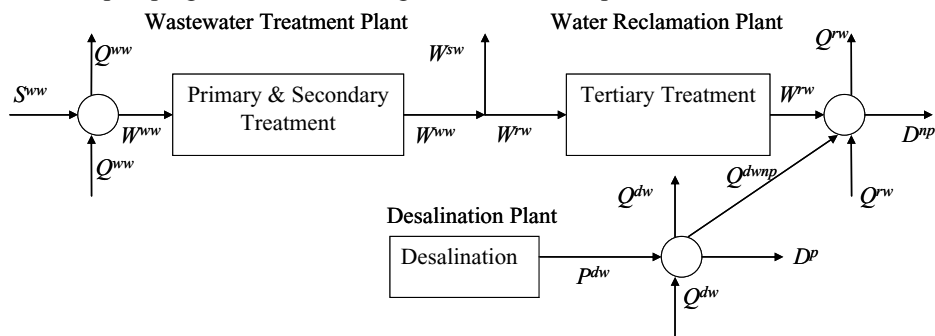
Fig. 1. Schematic graph of the flows of various types of water and wastewater

In this problem, given are the pairwise distances, pumping distances and elevations between the relative population centres of the regions, daily potable/non-potable water demand, wastewater production, capital costs of the relative plants with different sizes, production costs of desalinated and reclaimed water, treated wastewater, diameters and unit costs of pipelines (installed), unit costs, maximum flow rates and pumping elevations of pumps, unit storage cost and storage retention time, cost of electricity, and water/wastewater velocity to determine the locations and capacities of the plants, pipeline main network characteristics, daily production volumes of the plants, daily mains flows of desalinated water, reclaimed water and wastewater, and pumps (number and capacities) at each established link, so as to minimise the annualised total cost, including capital costs for plants, pipelines, pumping stations, and storage tanks, and operating/energy costs for water production, wastewater treatment and pumping.

3. Mathematical Formulation

The integrated water resources management problem is formulated as an MILP model, as described next:

3.1. Mass Flow Balance

From Fig. 1, we can see that before the wastewater treatment in region i, collected wastewater flows, Q_{jit}^{ww}, plus wastewater supply, S_{it}^{ww}, is equal to distributed wastewater flows, Q_{ijt}^{ww}, plus total wastewater for treatment, W_{it}^{ww}.

$$\sum_j Q_{jit}^{ww} + S_{it}^{ww} = \sum_i Q_{ijt}^{ww} + W_{it}^{ww}, \qquad \forall i, t \tag{1}$$

The treated wastewater production, W_{it}^{ww}, is equal to the summation of discharged wastewater amount, W_{it}^{sw}, and reclaimed water production, W_{it}^{rw}.

$$W_{it}^{ww} = W_{it}^{sw} + W_{it}^{rw}, \qquad \forall i, t \tag{2}$$

The reclaimed water production, W_{it}^{rw}, plus incoming reclaimed water flows, Q_{jit}^{rw}, and interactive desalinated water flow, Q_{it}^{dwnp}, is equal to outgoing reclaimed water, Q_{ijt}^{rw}, and non-potable water demand, D_{it}^{np}.

$$\sum_j Q_{jit}^{rw} + Q_{it}^{dwnp} + W_{it}^{rw} = \sum_j Q_{ijt}^{rw} + D_{it}^{np}, \qquad \forall i, t \tag{3}$$

In the desalination plant, the summation of desalinated water production, P_{it}^{dw}, and incoming desalinated water, Q_{jit}^{dw}, is equal to the summation of outgoing desalinated water, Q_{ijt}^{dw}, potable water demand, D_{it}^{p}, and desalinated water flow, Q_{it}^{dwnp}.

$$\sum_j Q_{jit}^{dw} + P_{it}^{dw} = \sum_i Q_{ijt}^{dw} + Q_{it}^{dwnp} + D_{it}^{p}, \qquad \forall i, t \tag{4}$$

3.2. Plant Capital Cost

The capital costs of the desalination plants (*DPCC*), wastewater treatment plants (*WTPCC*) and water reclamation plants (*WRPCC*), which are functions of plant capacities, are piecewise linearised in the model.

$$DPCC = \sum_i \sum_k CC_k^{dw} \cdot \lambda_{ik}^{dw}, \tag{5}$$

$$A_i^{dw} = \sum_k \widetilde{A}_k^{dw} \cdot \lambda_{ik}^{dw}, \text{ and } \sum_k \lambda_{ik}^{dw} = E_i^{dw}, \qquad \forall i \tag{6}$$

where CC_k^{dw} and $\widetilde{A}_k^{dw}$ are the cost and capacity at the break point k in the desalination plant capital cost function in region i, and A_i^{dw} are positive variables for the capacity and λ_{ik}^{dw} are SOS2 variables, while binary variable E_i^{dw} indicates whether a desalination plant is allocated in region i. Similar constraints and variables are also defined for the capital cost of wastewater treatment plants and water reclamation plants.

3.3. Pipeline Network Capital Cost

There are potentially three different pipeline networks, each used for desalinated water, wastewater or reclaimed water. The capital cost (*PipeCC*) is calculated by the unit cost of pipe type p multiplied by the length of pipes (L_{ij}). We also introduce binary variables Y_{ijp} to indicate if pipe type p is selected between regions i and j for water and wastewater transportation.

The flow rate of water/wastewater, $\widetilde{Q}_p$, is determined by the pipe diameter (d_p) and the flow velocity (v). The daily flow and flow rate of desalinated water are related by the operation time proportion variable, γ_{ijt}^{dw}. We introduce $YG_{ijpt}^{dw} \equiv \gamma_{ijt}^{dw} \cdot Y_{ijp}^{dw}$ to linearise the nonlinear term, as given below:

$$Q_{ijt}^{dw} = \sum_{p \in P} \widetilde{Q}_p^{dw} \cdot \gamma_{ijt}^{dw} \cdot Y_{ijp}^{dw} = \sum_{p \in P} \widetilde{Q}_p^{dw} \cdot YG_{ijpt}^{dw}, \qquad \forall i,j,t \tag{7}$$

Similar constraints for the daily flows and flow rates of wastewater and reclaimed water are also developed.

3.4. Pumping Station Capital Cost

The cost of a pumping station includes the cost of a pair of pumps (operating and standby) and a shell cost. By determining the number of each type of pumps selected, we can get the pumping station capital cost (*PStaCC*). For the selected pumps, the maximum allowable flow rate should be no less than the real flow rate, and the summation of the pump's maximum pumping elevations should be no less the elevation between regions plus the head loss, which is given by the Hazen-William Equation. Incorporating the pipe selection, the head loss from region *i* to *j* can be written as:

$$\Delta H_{ij} = b \cdot \alpha_{ij} \cdot \sum_p (\frac{\widetilde{Q}_p}{C})^{1.852} \cdot \frac{Y_{ijp}}{d_p^{4.87}}, \qquad \forall i,j \tag{8}$$

where b is conversion constant, α_{ij} is pumping distance between regions, and C is roughness coefficient.

3.5. Storage Capital Cost

The storage capital cost (*StorageCC*) is given by the unit storage cost, daily water demand (potable and non-potable), and water retention days.

3.6. Production Operating Cost

Similar to the plant capital costs in Section 3.2, the production operating costs are functions of the water/wastewater amounts. Piecewise linear approximations are used to calculate the above production operating costs of desalinated water, treated wastewater and reclaimed water. For desalination, $\widetilde{P}_k^{dw}$ and PEC_k^{dw} are the production amount and unit energy usage at break point k, EP is the cost of electricity and N_t is the duration of season t. P_{it}^{dw} are the daily production variables during season t and ξ_{ikt}^{dw} are SOS2 variables, while X_{it}^{dw} is a binary variable, which is equal to 1 if the desalination plant in region i has production during season t. The annual desalination production cost (*DC*) is given by:

$$DC = \sum_i \sum_k \sum_t N_t \cdot EP \cdot PEC_k^{dw} \cdot \widetilde{P}_k^{dw} \cdot \xi_{ikt}^{dw} \tag{9}$$

$$P_{it}^{dw} = \sum_k \widetilde{P}_k^{dw} \cdot \xi_{ikt}^{dw}, \text{ and } \sum_k \xi_{ikt}^{dw} = X_{it}^{dw}, \qquad \forall i,t \tag{10}$$

Similar constraints for annual production costs of treated wastewater (*WTC*) and reclaimed water (*WRC*) are also used here.

3.7. Pumping Cost

The pumping cost (*PumpingC*) is equal to the pumping energy multiplied by the cost of electricity. The daily required pumping energy is given by the daily water flow, elevation, head loss and pump efficiency. By using Eqs. (7) and (8), we have the following equations for the pumping energy for desalinated water:

$$PumpE_t^{dw} = \frac{1}{\beta^{dw}} \cdot \rho \cdot g \cdot \sum_i \sum_j \sum_p \widetilde{Q}_p^{dw} \cdot [H_{ij} + \frac{b \cdot L_{ij}}{d_p^{4.87}} \cdot (\frac{\widetilde{Q}_p^{dw}}{C})^{1.852}] \cdot YG_{ijpt}^{dw}, \qquad \forall t \tag{11}$$

where β^{dw} is the desalinated water pump efficiency, ρ is the water density, and g is the gravity. The required pumping energy for wastewater and reclaimed water is derived similarly.

3.8. Objective Function

The objective is to minimise the annualised total cost, including the operating costs and capital costs, which are annualised by the recovery capital factor (CRF):

$$OBJ = DC + WTC + WRC + PumpingC +$$

$$(DPCC + WTPCC + WRPCC + PipeCC + PStaCC + StorageCC) \cdot \frac{r \cdot (1+r)^n}{(1+r)^n - 1} \quad (12)$$

where r is the annual interest rate and n is the length of the project in years.

4. Case Study

Now, we apply the proposed model to Syros, an island in Aegean Sea, Greece. The existing infrastructure on the island is not considered in the problem. In the problem, the project is for 20 years, and an annual interest rate of 5% is used. The island is divided into 6 regions (R1-R6), and the population centre for every region is at sea level, apart from R1 which is at elevation of 250 m. The water demands and wastewater supply vary between summer time (high, 4 months) and the rest of a year (low, 8 months):

Table 1. Estimated daily water demands and wastewater supply in Syros (summer/winter) (m³/day)

	R1	R2	R3	R4	R5	R6
potable water demand	150/50	4000/2800	500/200	650/350	500/200	500/300
non-potable water demand	250/0	900/100	600/50	880/30	580/30	380/30
wastewater supply	150/50	3700/2600	200/100	300/150	300/150	450/250

The capital cost and operating production cost of the plants are piecewise linearised by 4 break points, where the capacity/production is 100, 1000, 2500 and 5000 m³/day, respectively. There are 8 types of available pipes (4 for desalinated and reclaimed water and 4 for wastewater), and 8 types of pumps (4 for desalinated and reclaimed water and 4 for wastewater). The water storage should satisfy two days' demand in summer time.

The proposed MILP model has been implemented in GAMS 22.8 using solver CPLEX 11.1 on a Intel Pentium 4 3.40 GHz, 1.00 GB RAM machine. The optimality gap was set to 5%. There are 1636 constraints, 1009 continuous variables, and 756 integer/binary variables in the model. After a computational time of 164 s, an objective of $2,954,339 is obtained. The breakdown of the optimal annualised total cost is given in Fig. 2, while the optimal allocations of plants in each region and pipeline links are presented in Fig. 3.
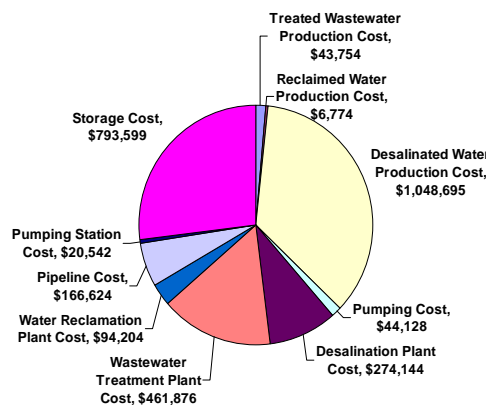
Fig. 2. Breakdown of the optimal objective   Fig. 3. Optimal plant allocations/pipeline links
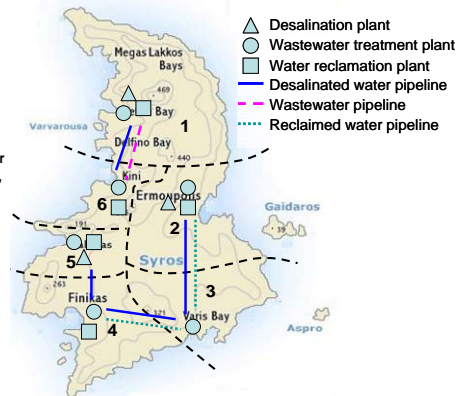
The optimal details of the optimal solution are shown in Table 2, including information for each established link (flow/direction, type of pipes, type and number of pumps). The optimal daily production of each plant is given in Fig. 4.

Table 2. Solution details for each established link

link	water type*	pipe diameter (in)	flow rate (m³/day)	direction	pump type (m³/day)	operating pump no.
1---6	dw	4	560.4	1→6	-	0
2---3	dw	6	1260.9	2→3	2400	1
3---4	dw	6	1260.9	3→4	2400	1
4---5	dw	4	560.4	4→5	1200	1
				5→4	1200	1
2---3	rw	6	1260.9	2→3	2400	1
3---4	rw	4	560.4	3→4	720	1
1---6	ww	4	700.5	1→6	-	0

* dw: desalinated water, rw: reclaimed water, ww: wastewater
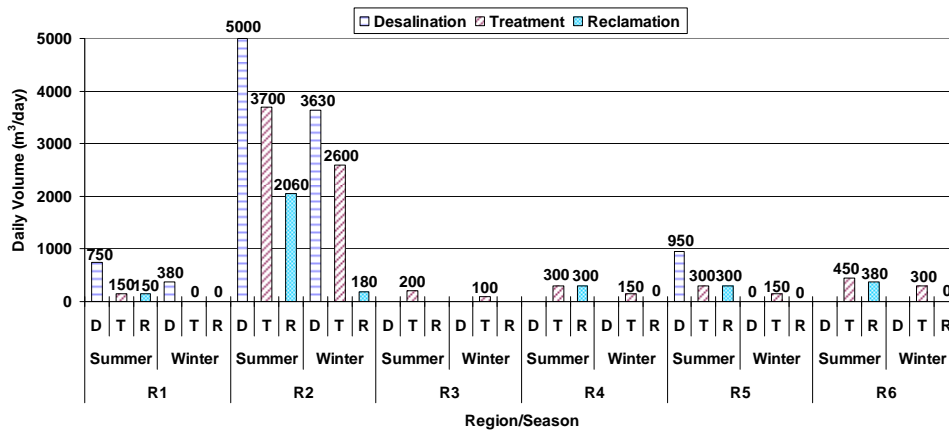
Fig. 4. Optimal daily production/treatment volumes of all plants in summer time and winter time

5. Concluding Remarks

The problem of integrated water and wastewater resources management of a water deficient island has been addressed. A mixed integer linear programming model has been proposed for the problem, by minimising the annualised total cost. The optimisation-based approach has successfully been applied to the Greek island of Syros.

6. Acknowledgements

References

[1] J. Medellín-Azuara, L.G. Mendoza-Espinosa, J.R. Lund, and R.J. Ramírez-Acosta, Water Sci. Technol., 55 (2007) 339.
[2] Y. Han, S. Xu, and X. Xu, Water Resour. Manage., 22 (2008) 911.
[3] M.C. Cunha, L. Pinheiro, J. Zeferino, A. Antunes, and P. Afonso, J. Water Resour. Plan. Manage.-ASCE, 135 (2009) 23.
[4] P. Gikas and G. Tchobanoglous, J. Environ. Manage., 90 (2009) 2601.

The Value of Battery Storage and Discharge Logic with Solar Microgeneration

Shisheng Huang, Bri-Mathias S. Hodge, Joseph F. Pekny, Gintaras V. Reklaitis

School of Chemical Engineering, Purdue University, 480 Stadium Mall Drive, West Lafayette, IN 47907, USA, [huang47, bhodge, pekny, reklaiti] @purdue.edu

Abstract

Microgeneration using solar photovoltaic systems is becoming increasingly popular in residential households as such systems allow for households use a renewable energy source while also reducing their reliance on the electricity grid to fulfill their electricity demand. In this study we explore the added value of a battery storage system with regards to a solar photovoltaic system during the summer months. A battery storage system is able to capture the excess electricity generated by a photovoltaic system and use it to displace some portion of the household electricity demand at a later period. California is used as a case study to determine the value of adding a battery storage system for a household with a solar photovoltaic array.

Keywords: Microgeneration, Electricity Demand Modeling, Agent-Based Modeling, Solar Power, Energy Storage

1. Introduction

As the emphasis on electricity production from renewable energy sources has increased, there has been an increasing proliferation of small-scale electricity generation devices. These generation units, such as small-scale solar cells and micro-wind turbines have begun to appear in residential applications where electricity from the grid is available, instead of solely being found in isolated locations. These small-scale generation devices help to meet a portion of household electricity demand and have the potential to reduce the reliance on electricity from the electrical grid. However, the issue of electricity supply intermittency arises when wind and solar energy sources are considered. Solar irradiance is highest around the noon timeframe and correspondingly that is when peak electricity generation from solar cells occurs. Unfortunately this creates a mismatch between supply and demand timing as the peak electricity demand for a residential household is generally in the evening when people typically finish their work activities and return home. This disparity between peak supply and peak demand can result in an inefficient usage of valuable energy resources, with excess generated electricity during the day being wasted for a lack of utilization [1]. Wind power poses a slightly different problem. Electricity generation from wind is unpredictable in both timing and output and thus one always encounters problems in being able to reliably match supply with demand.

One option to address these mismatches would be to allow for excess electricity produced by these small generators to be exported back to the grid when appropriate conditions are present [2, 3]. However, this scenario could potentially result in disincentives for the utilities since they would be assisting potential customers in reducing dependence on their own services [4, 5]. Another option, that is simpler from a regulatory perspective, is the addition of electricity storage to the microgeneration

systems so that excess electricity can be stored on location and used to reduce household demand at a time when the local demand outstrips supply. While numerous electricity storage mediums exist in practice, for example, pumped water storage, super-capacitors and flywheels, we will only consider the use of chemical storage in the form of batteries in this study. The value of these storage systems would hence come from the additional reduction in the amount of electricity consumed. Another added advantage would be to store electricity when it is cheap and to use it when electricity is expensive, though such a scenario would require a time of use (TOU) pricing schedule.

2. Background

The concept of microgeneration is not new, and with the recent emphasis on renewable electricity production, more people are starting to install residential solar panels and micro wind turbines to supplement the grid in fulfilling their electricity usage. Recent studies have suggested, however, that micro-wind turbines are only economical when the installation location is feasible and as such is not accessible to every household [2, 6-8]. Therefore, in this study, wind micro generation is excluded from the analysis and only micro solar systems at the household level will be examined.

Several models have been published in the literature on the performance of a microgeneration system coupled with an energy storage system. It has been shown that the addition of a lead-acid battery of sufficiently large storage capacity can reduce electricity exports back to the grid by over 90% in some cases and can store electricity at a round trip efficiency of 70-72%. However, the rigorous demands of on-site electricity storage may not be ideal for a lead-acid battery [9]. This is due to the constant small charge cycling found in microgeneration systems instead of deep charging and discharging. Using hydrogen as a storage medium could also potentially provide efficiency gains when compared to a conventional lead-acid battery setup [10, 11]. There exist a large number of energy storage options that have been described in the literature. Most of the storage options explored have the potential of fulfilling the energy storage need associated with renewable energy sources [12-14], though there exist disparate opinions as to the most suitable storage option. It has therefore been decided not to adopt any particular technology in this study, but instead to model a generic storage capability that can be revised to represent a particular storage technology selection if it is needed for future analyses. One aspect that is lacking in the above mentioned models is the ability for the model to be integrated into a larger system wide model in order to examine potential system wide effects of microgeneration adoption. An integrated system wide model for the electricity system was introduced by Huang et al to examine the effects of PHEV adoption on natural gas demand [15]. Such a model could easily be adapted to explore the effects of microgeneration on a system wide scale and is planned to be the subject of subsequent studies.

3. Modeling Approach

An agent based approach to household electricity demand modeling has been implemented for this study. The household model here can be decomposed into three parts. The first part is based on a discrete event framework for a household based around a set of appliances. Each appliance has at any given point in time a certain probability of being turned on [15]. This set of appliances and their usage patterns is used to simulate the electrical demands of a representative household. The second part of the model is an agent for the solar panel. In this study, it is assumed that the

household has installed a 2 kW rated solar panel. The electricity generated by the solar panel is dependent on the irradiation level experienced by the solar panel over the course of the day. Solar radiation data has been obtained from the National Solar Radiation Database [16] for four different California sites corresponding to the four most populous urban centers within the state. Hourly radiation mean and deviation data were obtained for each month. The monthly data are divided into summer months and winter months respectively. Summer months range from May until October while the winter months make up the balance of the year, an assumption consistent with California's meteorological environment. Every fifteen minutes each household is allocated a solar radiation value drawn from a normal distribution with the unique mean and standard deviation value associated with that particular hour for a selected month. The amount of electricity generated is used to supplement the household electricity demand, subject to an inverter efficiency loss. If a battery is present, the electricity generated in excess of household electricity demand is stored in the battery. .

The last part of the household model is thus the energy storage component. The storage medium in this study is assumed to be a generic battery capable of fulfilling the household storage needs. Examples of storage mediums that could potentially fulfill this need include lead-acid batteries, redox-flow batteries, nickel based batteries and hydrogen-based energy storage. In this study we assume that the battery efficiency is 80% with an inverter efficiency of 95%. The storage medium would only be charged when there is excess electricity produced by the solar panel and would not be charged otherwise with power from the grid. There are two plausible scenarios under which the battery may be called upon to discharge. In the first scenario the battery would discharge to meet any electricity usage whenever solar production cannot cover the household electricity demand. In the other scenario the battery would only discharge to cover electricity usage when the price of electricity is the highest, i.e during system wide peak electricity consumption periods. These two cases are based on the different electricity tariff regimes present in the California system. Under flat-rate tariff pricing the timing of electricity usage is unimportant, only the usage amount will have an effect on the customer's electricity bill. When TOU pricing is in effect the timing of electricity usage is more important, with higher rates being charged for each unit of electricity demand during periods of peak and near-peak demand. In the California system three periods are in place for the summer months: off-peak, shoulder peak and peak. During the winter months only shoulder peak and off-peak periods are considered.

4. Case Study

The combined electricity demand and solar microgeneration model described above has been used in order to gauge the economic savings of incorporating battery storage into a standard California household with a 2 kW capacity solar array. Eight different electricity demand profiles have been examined, incorporating the two different electricity pricing schemes available in California: flat-rate and TOU pricing. As mentioned in the preceding section two discharge operating scenarios for the batteries have also been examined. In one case, the battery is assumed to discharge and supplement household demand as and when it can while, in the second case, it is assumed that the battery would discharge only when shoulder and peak prices are experienced. The following scenarios have been examined, each under both pricing schemes: a representative household, the household with only the solar array and the

household with the solar array and three different battery sizes. Electricity demand has been simulated over a six month summer period to determine the electricity needs of a California household during the peak usage period that is the summer. The amount of electricity used by the household per summer day was predicted by the model to be approximately 16.05 kWh and the incremental savings of electricity usage are presented in Table 1 below.

It can be seen that with the addition of a battery of only 1 kWh capacity, the reduction in electricity consumption almost doubles from the case in which the household only owns a solar panel. However, the subsequent decrease in the magnitude of electricity consumption is not as great, suggesting that the savings generated by an electricity storage system does not increase linearly with storage capacity. The electricity load profiles under the different system configurations are shown below in Figure 1. It can be seen that the battery can capture the excess solar radiation and supplement demand whenever there is a shortfall in production. The battery's impact is felt even when there is solar radiation, indicating that the battery is actively storing and discharging throughout the day. This would indicate that the battery chosen must be able to tolerate high levels of charge/discharge cycles.

	Electricity Savings Per Day (Wh)	
	No Constraints	Discharge Constraint
Solar Panel only	2,615.32	2,615.32
1 kWh Battery	4,691.98	3,777.67
2 kWh Battery	5,652.83	4,624.03
4 kWh Battery	7,083.01	5,553.25

Table 1: Electricity savings from the installation of a solar panel and batteries of different storage capacities
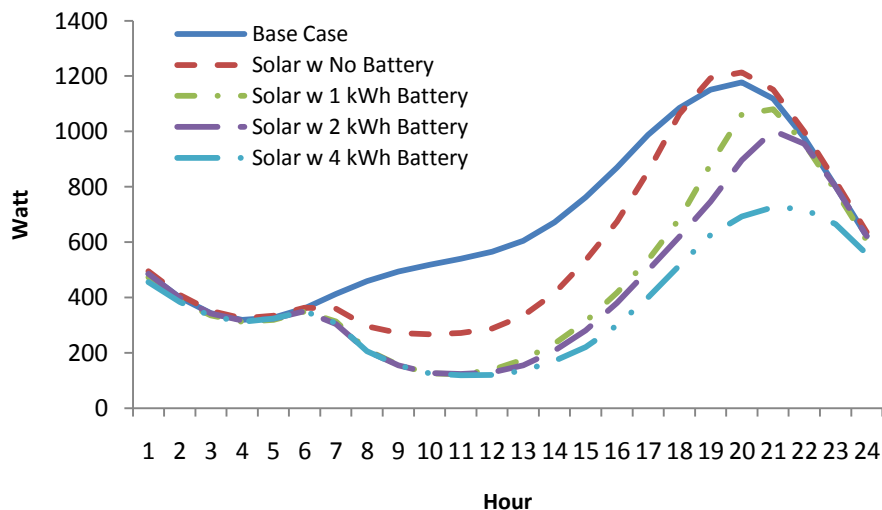
Figure 1: Electricity load profiles of households with the various combinations of solar panels and battery capacities

The Value of Battery Storage and Discharge Logic with Solar Microgeneration

Another interesting point to note is that when a discharge constraint is imposed on the battery in order to limit discharging to peak periods, the overall electricity savings are not as high as the corresponding case without discharge constraints. This counter-intuitive result is due to the fact that the battery could have been supplementing load during some periods where prices were still off-peak. The difference can be clearly seen in Figure 2. In the figure, the household has a 4kWh battery storage system coupled with a 2 kW solar panel. It can be seen that before 10 am, when there is already solar radiation, there would be a significant electricity import reduction if the battery were allowed to discharge. That is not the case, however, with a time of use restraint imposed on the battery, hence losing the potential electricity savings.
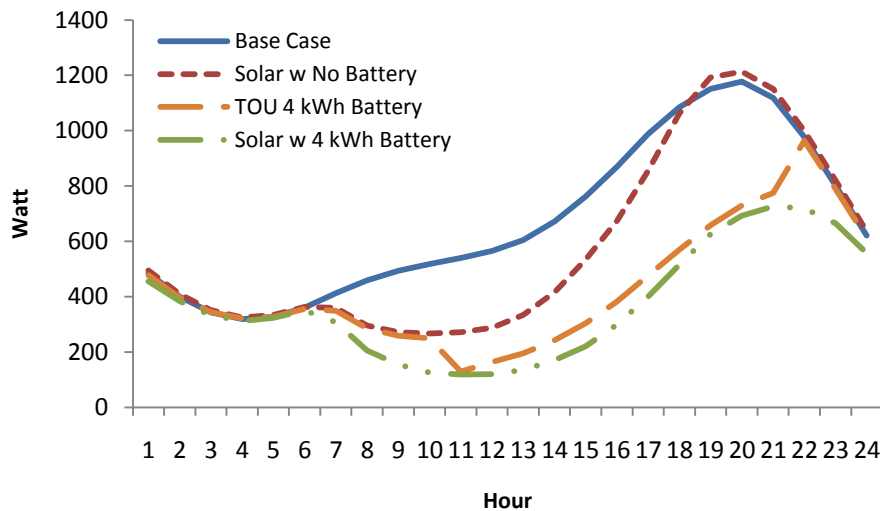
Figure 2: Difference in load consumption between a 4 kWh Battery storage with and without discharge constraints.

It can also be seen that the maximum electricity savings comes from a battery storage system operated under no constraint on discharging. Using a Californian electricity rate schedule [17], we can determine the cost savings per summer day of the household. Listed in Table 2 below are the daily savings afforded the household when a solar panel is installed together with the different battery storage systems. It can be seen that with just a 1kWh of battery capacity installed, the cost savings from electricity reduction is almost doubled when a time of use electricity pricing schedule is in effect. This shows that the choice of discharge operating heuristic for the battery will differ depending on the electricity pricing scheme in effect for the household. A household with the flat-rate electricity pricing scheme would benefit from the discharge heuristic that allows battery supplemented power at any time due to the decreased overall consumption. However, under the TOU pricing scheme the alternative discharge heuristic would be more economical for the household. Regardless of which discharge logic is ultimately used if a battery system costs less than the solar panel, which in most cases would be true, it would be economical for the household to install a battery electricity storage system of some sort to complement a solar panel.

	Cost Savings Per Day of Summer			
	Flat Rate		Time of Use	
No Battery, Just Solar Panel	$	0.34	$	0.43
1 kWh Battery	$	0.58	$	1.09
2 kWh Battery	$	0.70	$	1.44
4 kWh Battery	$	0.86	$	1.48

Table 2: Savings per day of summer under different pricing schedules

5. Conclusion

A household model of micro-solar generation has been used to determine the value of a battery storage system for a California household during the summer months. Results indicate that adding a constraint on battery operation to only discharge during periods of high electricity prices results in a less efficient usage of the battery. It can be seen that if a 1kWh battery costs less than a 2kW solar photovoltaic installation, it could be economical to install a storage system to complement the solar array since the resultant savings per day more than double.

References

[1] A.S. Bahaj, P.A.B. James, Renewable and Sustainable Energy Reviews, 11 (2007) 2121-2136.
[2] A.S. Bahaj, L. Myers, P.A.B. James, Energy and Buildings, 39 (2007) 154-165.
[3] J. Watson, R. Sauter, B. Bahaj, P. James, L. Myers, R. Wing, Energy Policy, 36 (2008) 3095-3106.
[4] J. Watson, Energy Policy, 32 (2004) 1981-1990.
[5] R. Bettle, C.H. Pout, E.R. Hitchin, Energy Policy, 34 (2006) 3434-3446.
[6] N. Mithraratne, Energy and Buildings, 41 (2009) 1013-1018.
[7] A.D. Peacock, D. Jenkins, M. Ahadzi, A. Berry, S. Turan, Energy and Buildings, 40 (2008) 1324-1333.
[8] Encraft, Encraft Warwick Wind Trials Project Final Report, in, Encraft, 2009.
[9] D.P. Jenkins, J. Fletcher, D. Kane, Energy Conversion and Management, 49 (2008) 2413-2424.
[10] S. Kélouwani, K. Agbossou, R. Chahine, Journal of Power Sources, 140 (2005) 392-399.
[11] J.D. Maclay, J. Brouwer, G.S. Samuelsen, Journal of Power Sources, 163 (2007) 916-925.
[12] J.K. Kaldellis, D. Zafirakis, Energy, 32 (2007) 2295-2305.
[13] M. Perrin, Y.M. Saint-Drenan, F. Mattera, P. Malbranche, Journal of Power Sources, 144 (2005) 402-410.
[14] C. Fabjan, J. Garche, B. Harrer, L. Jörissen, C. Kolbeck, F. Philippi, G. Tomazic, F. Wagner, Electrochimica Acta, 47 (2001) 825-831.
[15] S. Huang, B.-M. Hodge, J. Pekny, G.V. Reklaitis, The Impact of PHEV Adoption on Natural Gas Demand in Electricity Generation, in: H. Alfadala, G.V.R. Reklaitis, M.M. El-Halwagi (Eds.) 2nd Annual Gas Processing Symposium, Elsevier, Qatar, 2010.
[16] NREL, National Solar Radiation Data Base, in, 2009.
[17] Pacific Gas and Electric Company, Electric Rate Schedules, in: PG&E Tariff Book, Pacific Gas and Electric Company, 2009.

An Efficient and Fast General Optimization Model for a Sustainable Recovery Network of Industrial Polluted Wastes

J. Duque*[a], A. P. F. D. Barbosa-Póvoa[b], A. Q. Novais[a]

a Unidade de Modelação e Optimização de Sistemas Energéticos, LNEG,
 Estrada do Paço do Lumiar, 22, 1649-038 Lisboa, Portugal
b Centro de Estudos de Gestão, CEG-IST, Instituto Superior Técnico,
Universidade Técnica de Lisboa, Av. Rovisco Pais,   1049-001, Lisboa, Portugal

Abstract

The design and management of waste recovery networks involves the consideration of both economical and environmental issues, for which the authors developed a general network model that embeds the eco-indicator 99 methodology. Practical problems, which were tackled, involved the modeling of wide geographical networks and as a consequence, the superstructures of the resulting optimization models are highly dimensional, presenting in addition an inherent combinatorial nature. Hence, the resulting mathematical formulations, typically MILP problems, may undesirably take a few hours of CPU time to solve and yet only fall within a tolerance, albeit acceptable, of the best solution possible. These tend to be, however, hard cases corresponding to extreme operating conditions, such as imposing a discretionary green behavior while maximizing return, or conversely, imposing a high return while aiming for minimal environmental damages. This work explores alternative approaches to reduce CPU times by orders of magnitude and thus opening the way to handle effectively real-sized networks. These include a heuristic method that may be used to further simplify those approaches based on the ε-constraint method. Details of the implementation are described and results compared to solutions located along the Pareto front, obtained with the latter.

1. Introduction

The authors already reported work on the case study of a network for industrial wastes (e.g., J.Duque et al, 2009), where an adapted version of the maximal State Task Network (mSTN) (Barbosa-Póvoa and Macchietto, 1994) was employed for modeling the transportation and transformation processes. To account for the environmental issue the general network model embedded the eco-indicator 99 methodology (Pré Consultants, B.V.Amersfoort, Netherlands), which provides a damage indicator (EI99) that may be used as an environmental objective function along with profit. Due to the

---

- Corresponding author, Tel: +351 - 210 924 601. Fax: +351 - 21 7167016. E-mail: joaquim.duque@ineti.pt

nationwide scale of the networks, the superstructure of the resulting optimization model involves a high number of entities of various types, geographically dispersed. The transport network needed to interconnect them present an inherent combinatorial nature. Besides transports, a recovery network involves three other different types of entities: the producers (P) of recoverable materials, the entities that transform (T) them and their clients (C). Thus for a total of fifty four entities of the three types (18xP,18xT,18xC), if two recoverable materials and two alternative types of transport are considered for each possible route, the number of possible transport combinations is of the order of 10^6. Hence, the resulting MILP problem may take, if only in the hardest cases, up to six hours CPU time to achieve less than 5% of the best solution possible.

This work explores alternative approaches to reduce the CPU time by orders of magnitude. While considering an aggregation/ disaggregation approach, apart from the time variable, for which the model uses a cyclic operation over the time horizon, none of the remaining model variables were found to offer scope for model simplification. A decomposition method was also attempted, which took advantage of the network echelon basic structure and decoupled the problem into two parts, i.e. material transformation and demand supply, but it was found to lead repeatedly to local rather than to the global optimum. As a more promising alternative, a heuristic was devised, which was inspired by the use of the ε-constraint method and that takes to its advantage the conflicting tendencies imposed by the simultaneous environmental and economic optimization. The detailed analysis of this simplifying heuristic method is presented next and applied to an illustrative example.

2. Heuristic Rationale

The use of the ε-constraint method starts by defining upper and lower bounds for the Objective Functions (OFs). The application of the method involves a series of runs optimizing one OF while regularly incrementing the constraints on the other. Plotting the OFs optimal values against each other, the Pareto front can be drawn (see figure 1). If needed, more details can be added to that front by plotting additional intermediate points generated by further optimizations. Thus the challenge consists in finding a lower cost procedure alternative to the ε-constraint method, which guarantees the effective attainment of the preferred solution. The solution to this will strongly depend on the decision making goals but it can become apparent in some cases: in the case of a recovery network it is likely for management to expect the highest profit together with the least possible pollution.

With a view to the mathematical implementation and to make the OFs dimensionless, the respective bounds are used as normalization factors. The OFs bounds are those that have the highest absolute value. In our example these are X_P^{UB} and X_D^{LB}, respectively, i.e. the value that maximizes profit (X_P) with no constraints imposed on the EI99, and the (negative) value that minimizes the EI99 (X_D), with no constraints imposed on

An Efficient and Fast General Optimization Model for a Sustainable Recovery Network of Industrial Polluted Wastes

profit. Then, introducing a new positive variable $\lambda \mid (0 \le \lambda \le 1)$ we may write a new

optimization problem:
$$\begin{cases} \max \lambda \\ X_P \ge \lambda X_P^{UB} \\ X_D \le \lambda X_D^{LB} \\ 0 \le \lambda \le 1 \end{cases} \qquad (1)$$

Where the objective functions X_P and X_D are defined by the general model. This heuristic is deduced from convex space properties (see section.3). Figure 1 shows the new value V obtained from this new optimization problem against all the results obtained in our previous work (J.Duque et al, 2009) with the ε-constraint method.
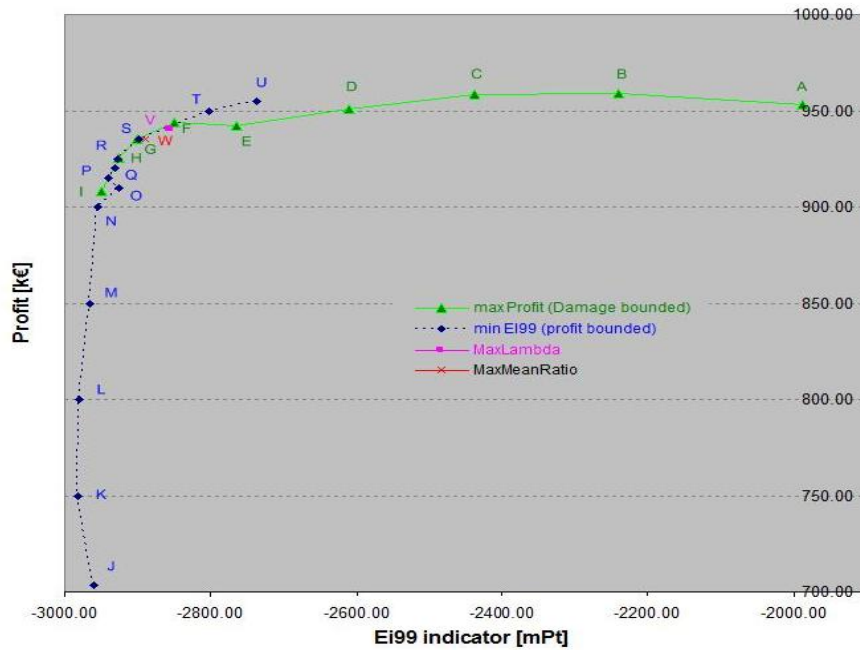
Figure1. – Plotting the new optimization values over the Pareto front

The W solution value is a limiting value obtained by forcing the equality on the constraints and solving for λ according to $\lambda = \left(X_P / X_P^{UB} + X_D / X_D^{LB} \right) \big/ 2$.

Table1. Comparison of solutions' statistics

ID	Number of Variables			Time to [seconds]			Value		
	NIter	NTotal	NInteger	Generate	Execute	Optim.	Best	Relative
V	5489496	61868	31208	975	5316	0.96	0.97	1.80E-02
W	40097	61868	31208	977	167	0.97	0.99	1.84E-02
G	11519276	61867	31208	9889	22000	935	959	2.45E-02
S	10001582	61867	31208	983	22000	-2900	2970	2.35E-02
A	7419	61867	31208	975	20	953	971	1.84E-02
J	4596	61867	31208	978	33	-2960	3010	1.66E-02

Note: The optimized values dimensions are: A, G - [k€], of J, S - [miliPoints] and V, W are adimensional.

Table 1 show the computational statistics for the new solutions V and W, together with solutions G, A, and S, J, that were obtained earlier, respectively, for profit maximization, with and without constraining EI99, and for EI99 minimization, with and without constraining profit. Table 2 presents the absolute values for EI99, profit and their relative values (solutions J and A, respectively).

Table 2 Comparison of absolute and relative values for profit and damage

ID	EI99		Profit	
	[miliPoints]	Rel.Deviation (Sol. J)	[k€]	Rel.Deviation (Sol. A)
V	-2868.11	0.97	942.27	0.99
W	-2889.71	0.98	935.37	0.98
G	-2900.00	0.98	935.38	0.98
S	-2898.25	0.98	935.00	0.98
A	-1989.11	0.67	953.12	1.00
J	-2960.10	1.00	703.44	0.74

The graphics show that both solutions V and W lie over the quasi-Pareto front (a termination gap $\approx 2\%$ was allowed) and the statistics presented on table 2 suggest that solution V slightly favors profit over damage, achieving 99% of the maximal profit and 97% of the minimal damage. As expected, when enforcing the constraint equalities the solutions are at 98%, both equidistant to their limit values. Furthermore, when compared to equivalent solutions (G and S), solution V performs about 4 times faster, while solution W is approximately 130 times faster for the optimized λ values shown in table 1, under the "Optim" heading.

3. Using general properties of convex spaces

In a convex space the solutions obey the following equation:

$$X = \lambda X^A + (1-\lambda)X^B \tag{2}$$

where the value of X depends on the parameter $\lambda \in [0,1]$ that defines a point in the line segment that connects the solution bounds (X^A - solution value at the first bound, $\lambda = 1$, and X^B - solution value at the second bound, $\lambda = 0$).
When applying the ε-constraint method to the example previously referred the upper limits $\left(X_P^{UB}, X_D^{UB}\right)$ were obtained by maximizing the profit OF, $\left(X_P\right)$, with no constraints on the damage OF, $\left(X_D\right)$, and the lower limits $\left(X_P^{LB}, X_D^{LB}\right)$ were obtained by minimizing the damage OF, $\left(X_D\right)$, with no constraints on the profit OF, $\left(X_P\right)$. Then:

$$\lambda_1, \lambda_2 = 1 \Rightarrow \begin{cases} \max(X_P) = X_P^{UB} \\ \min(X_D) = X_D^{LB} \end{cases} and \quad \lambda_1, \lambda_2 = 0 \Rightarrow \begin{cases} X_P^{LB} \\ X_D^{UB} \end{cases} \tag{3}$$

And thus, by replacing the actual bounds on equation 2:

$$\begin{cases} X_P = \lambda_1\left(X_P^{UB}\right) + \left(1-\lambda_1\right)\left(X_P^{LB}\right) \\ X_D = \lambda_2\left(X_D^{LB}\right) + \left(1-\lambda_2\right)\left(X_D^{UB}\right) \end{cases} \tag{4}$$

An Efficient and Fast General Optimization Model for a Sustainable Recovery Network of Industrial Polluted Wastes

The use of the damage lower bound when $\lambda = 1$ is due to the fact that the damage values are negative; the damages avoided by recovering the material surpass by far the network's damages. Equation (4) may be re-written as:

$$
\begin{cases}
\dfrac{X_P}{X_P^{UB}} = \lambda_1 + (1-\lambda_1)\left(\dfrac{X_P^{LB}}{X_P^{UB}}\right) \\[3mm]
\dfrac{X_D}{X_D^{LB}} = \lambda_2 + (1-\lambda_2)\left(\dfrac{X_D^{UB}}{X_D^{LB}}\right)
\end{cases}
\tag{5}
$$

Given the conditions

$$
0 \le (\lambda_1 \text{ e } \lambda_2) \le 1, \ 0 \le \left(X_P^{LB} \Big/ X_P^{UB}\right) \le 1 \ \text{ and } \ 0 \le \left(X_D^{UB} \Big/ X_D^{LB}\right) \le 1 \ , \text{ then:}
$$

$$
\begin{cases}
X_P \Big/ X_P^{UB} \ge \lambda_1 \\[2mm]
X_D \Big/ X_D^{LB} \ge \lambda_2
\end{cases}
\Rightarrow
\begin{cases}
X_P \ge \lambda_1 X_P^{UB} \\[2mm]
X_D \le \lambda_2 X_D^{LB}
\end{cases}
\tag{6}
$$

That by imposing the same λ variable to both equations leads to the maximization problem I already defined in equation (1):

$$
[\text{I}]
\begin{cases}
\max \lambda \\
X_P \ge \lambda X_P^{UB} \\
X_D \le \lambda X_D^{LB} \\
0 \le \lambda \le 1
\end{cases}
\quad
\overset{\text{By forcing equality}}{\underset{\text{and after rearrangement}}{\Longrightarrow}}
\quad
[\text{II}]
\begin{cases}
\max \lambda \\
\lambda = \left(X_P / X_P^{UB} + X_D / X_D^{LB}\right)\Big/ 2 \\
0 \le \lambda \le 1
\end{cases}
\tag{7}
$$

Problem II represents basically the mean of the normalized profit and damage functions. This case solves much faster when compared to the equivalent ε-constraint method solutions (see table 1). Problem II may therefore generate points equivalent to those obtained with the ε-constraint method, by conducting successive runs for values of a weight parameter C_λ spanning the range [0, 1].

$$
[\text{III}]
\begin{cases}
\max \lambda \\
\lambda = \left(C_\lambda \left(X_P / X_P^{UB}\right) + (1-C_\lambda)\left(X_D / X_D^{LB}\right)\right) \\
0 \le \lambda \le 1
\end{cases}
\tag{8}
$$

The optimized values from problem III are presented by small squares, over the quasi-Pareto front in figure 2 and show an excellent superposition. Furthermore the computational resources stay low all along the curve, attaining at most 200 seconds to execute and around 900 seconds to generate.
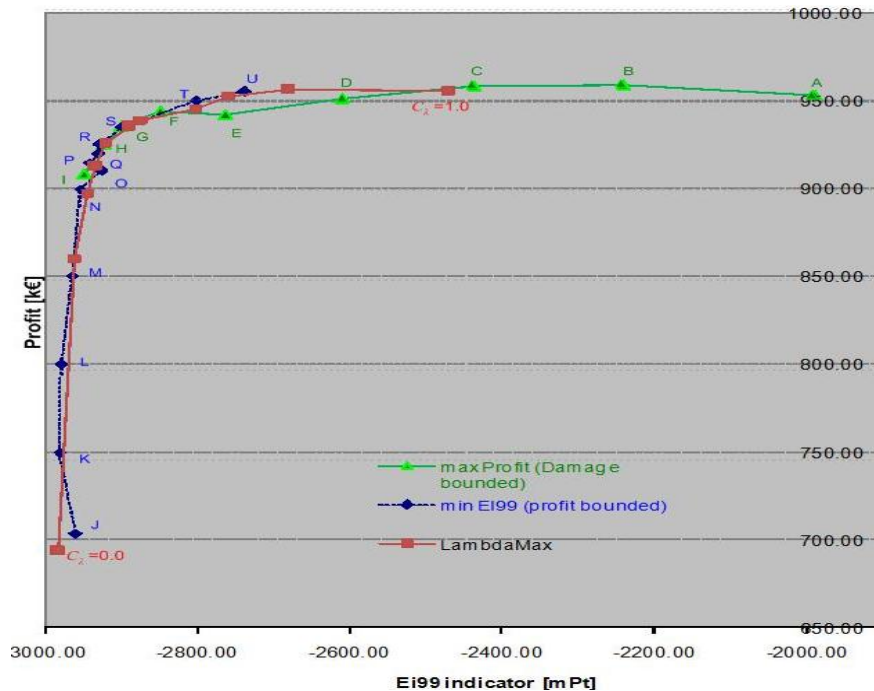
Figure2. –Optimization values of the new method overlapped on those by the ε-constraint method

4. **Conclusions**

A modified optimization procedure, based on a heuristic derived from the general concept of convexity is proposed to handle an optimization model, where two conflicting objectives are considered, i.e. profit and environment impact. This heuristic is applied to a generic model of a sustainable recovery network for industrial polluted wastes, previously solved by the ε-constraint method that proved effective, but often required numerous and lengthy calculations. In the present work the relative performance of these two methods is compared and the relative merits of the new method established, leading to the conclusion that it is equally effective, but a lot faster. However, while the ε-constraint method spans the search space, allowing Pareto type surfaces to be generated, the proposed heuristic leads to single points, admittedly at a fraction of the calculation time. This is not a limitation, though, since the heuristic procedure can be repeatedly employed over the validity range, under the steer of a weight-like parameter. On-going research explores the application of this heuristic-based procedure to other multiobjective problems, to assess its potential as both an optimization and an exploratory tool for the study of complex and highly dimensional systems.

References

Barbosa-Póvoa, APFD and Macchietto, S., 1994, Comput. Chem. Engng., 18, 11/12, 1013-1042.
Duque, J,, Barbosa-Póvoa, APFD and Novais, AQ, 2009, , Computer-Aided Chemical Engineering, 26, 1177 - 1182.
Pré Consultants, 2000, B.V.Amersfoort, Netherlands

Parametric Programming Technique for Global Optimization of Wastewater Treatment Systems

João Teles,[a,b] Pedro Castro,[a] Henrique Matos[b]

[a]*U. Modelação Optimização Sistemas Energéticos, LNEG, 1649-038 Lisboa, Portugal,*
 [b]*Dep. Engenharia Química e Biológica, IST, 1049-00, Lisboa, Portugal*

Abstract

This paper presents a parametric programming technique for the optimal design of industrial wastewater treatment networks (WTN) featuring multiple contaminants. Inspired in scientific notation and powers of ten, the proposed approach avoids the non-convex bilinear terms through a piecewise decomposition scheme that combines the generation of artificial flowrate variables with a multi-parameterization of the outlet concentration variables. The general non-linear problem (NLP) formulation is replaced by a mixed-integer linear programming (MILP) model that is able to generate near optimal solutions, fast. The performance of the new approach is compared to that of global optimization solver BARON through the solution a few test cases.

Keywords: Wastewater; Optimization; Mixed-Integer Linear Programming.

1. Introduction

Water is a resource that is used intensively for many different purposes in industry. Many of the processes are today subject to strict environmental regulations on discharge effluents due to increased water scarcity. Improved water management can effectively reduce freshwater demand and overall wastewater generation, and thus lower freshwater and effluent treatment costs.

Over the past decade, numerous research works have addressed this topic ranging from graphical pinch analysis techniques to mathematical optimization approaches. Graphical methods are easier to understand conceptually, while mathematical programming has a wider applicability scope. The mathematical programming approach relies commonly on the optimization of a superstructure for either integrated or separated problems. The optimal design of water-using networks (Teles et al., 2009), water treatment networks (Castro et al., 2009) or both integrated into one large system featuring regeneration and recycling (Gunaratnam et al., 2005; Karuppiah & Grossmann, 2006) can be formulated as (mixed-integer) non-linear programming problems (if logic constraints are used to prevent recycling). Such problems feature non-convex bilinear terms that make them very difficult to solve by gradient-based algorithms that are the basis of most commercial NLP solvers, which often cannot avoid getting trapped in suboptimal solutions. The complexity of non-convex NLP/MINLP problems is well documented in the literature and a number of different algorithms have been proposed for their solution such as: branch-and-bound, adaptive random-search, outer-approximation/equality relaxation, branch-and-reduce, generalized disjunctive programming, simulated annealing, MILP/LP heuristic search strategies, etc. While some approaches cannot ensure global optimality, others may require significant, if not prohibitively large, computational resources.

2. Problem statement

In this paper, we focus on the optimal design of a wastewater treatment system. Given a set of process wastewater streams W containing well-defined pollutants (set C) with known flowrates tf_w^{wwat} and concentrations $c_{w,c}^{wwat}$, the goal is to generate an effluent that meets discharge regulations c_c^{env} for all contaminants while minimizing the total flowrate going through the treatment units T. These are characterized by fixed removal ratios $rr_{i,c}$ and maximum inlet concentrations $c_{i,c}^{in\,max}$ ($i \in T$).

3. New parametric programming approach

To address this problem in a systematic way it is necessary to build a superstructure that embeds all possible flow configurations, similarly to any other optimization study in process synthesis: (i) each wastewater stream that enters the network can be sent to the treatment units or to the final discharge mixer (bypass); (ii) each unit is preceded by a mixer, which is fed by wastewater and reuse/recycle streams originating from the outlets of all treatment processes; (iii) each treatment unit is followed by a splitter that feeds the final discharge mixer, as well as other treatment processes; (iv) effluent streams from each unit and bypass streams are mixed in a final discharge mixer to ensure compliance with the environmental legislation.

To optimize these superstructures mathematical models are required and two alternative formulations have been proposed in the literature. They include non-convex bilinear terms in the mass balances of treatment units, involving either products of stream flowrates and concentrations in the mixers, or products of contaminant flowrates and split fractions in the splitters. Thus, if solved with local optimization solvers like CONOPT, suboptimal solutions are mostly likely to occur. To overcome this limitation, a new deterministic procedure is proposed. It is based on a piecewise decomposition scheme that approximates to a chosen accuracy level the non-convex terms in the original model. The resulting MILP generates a near optimal network that acts as an upper bound on the true global optimum. Furthermore, such output solution can easily be refined following initialization and solution of the general NLP with a local solver.

The required steps for converting the bilinear NLP into a MILP are described next.

3.1. Parameterization of the outlet concentration variables

Consider the well-known notation in the decimal numeral system that uses positions for each power of ten: units, tens, hundreds, thousands, etc., and ten different numerals, the digits 0, 1, 2, ..., 9 to represent any real number. It also requires a dot (decimal point) to represent decimal fractions. We follow this principle to approximate the outlet concentration for all contaminants in all treatment units. The dynamic construction process starts by fixing the number (n) of digits for the fractional part of the number (decimal region). The required number of digits ($k^{i,c}$) in the left-hand side of the decimal point (integral region) is then determined based on the maximum possible outlet concentration for contaminant c in unit i, which can be calculated from the given maximum inlet concentration and removal ratio. Thus, a total of $\xi_{i,c}=n+k^{i,c}$ digits are involved in the representation of the outlet concentration of contaminant c in unit i.

3.2. Disaggregation of the flowrate variables

The flowrate variables linked to the bilinear terms associated to the mass balances of unit i, give the flowrate from i to unit j, $F_{i,j}$, and the amount sent to the discharge mixer, F_i^{dis}. We now need to define, for each different contaminant c, multiple artificial

Parametric Programming Technique for Global Optimization of Wastewater Treatment Systems

variables resulting from the disaggregation of these base variables: $TF_{i,j,c}^{a,p}$ and $TD_{i,c}^{a,p}$, respectively. The two additional indices are critical, a identifies the numerical position $(a_{n+k,\dots}, a_{n+1}, a_n, a_{n-1},\dots a_1)$ and p the ten different possible digits $(0,\dots,9)$.

Fig. 1 illustrates the concept for variables $F_{i,j}$. The bilinear term $F_{i,j} \times C_{i,c}^{out}$, representing the contaminant mass in the stream linking the outlet splitter (circle) of unit i and the inlet mixer of unit j (diamond), is disaggregated into a sum of linear terms. Suppose that the optimal outlet concentration of contaminant c is equal to $1(\dots)0.8(\dots)3$. If 6 digits ($n=k=3$) are used, the number can be generated by: $1E(+3)+\dots+0E0+8E(-1)+\dots+3E(-3)$. Notice that we are selecting a single digit per position, with the chosen values being identified through non-zero values of the decision variables $Z_{i,c}^{a,p}$. Then, only the corresponding artificial flowrate variables (highlighted in Fig. 1) can assume positive values so that the accurate component mass flow is generated. Moreover, all artificial variables $TF_{i,j,c}^{a,p}$ will be forced to have the same value, $F_{i,j}$.

Overall, the new parametric programming approach is versatile since the number of significant digits can be increased for more accurate optimal solutions and decreased for lower computational effort.
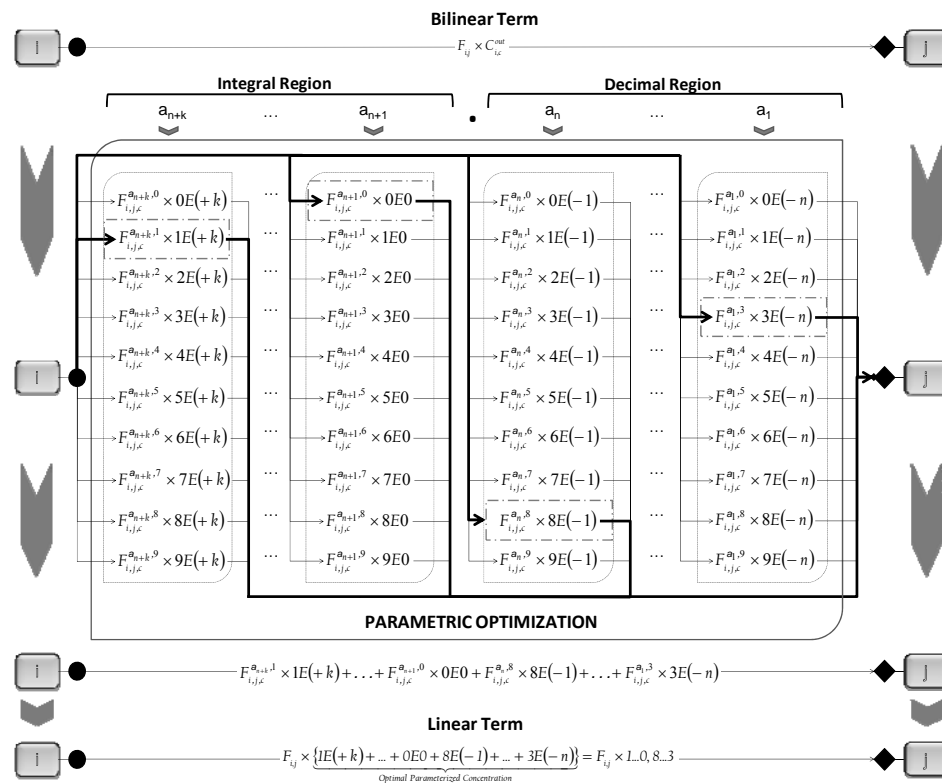
Figure 1. Illustration of parametric programming approach for the link between units i and j.

4. Mathematical formulation

The mixed integer linear programming formulation requires the additional positive continuous variables are required: F_i^{tot} gives the total flowrate entering/leaving treatment unit i; $F_{w,i}^{wwat}$ is the flowrate of wastewater w into treatment unit i; F_w^{byp} is the flowrate of wastewater w that bypasses the treatment system and goes directly to the final discharge mixer; $MT_{i,c}^{In}$ and $MT_{i,c}^{Out}$ are the inlet/outlet mass flows of contaminant c for treatment unit i.

The objective is to minimize the total flowrate, going through the treatment units, Eq. 1. Eq. 2 represents the flow balance over the splitters associated to the system's inlet wastewater streams. Eqs. 3-4 are the flowrate balances over the inlet mixer and outlet splitter linked to treatment unit i, while Eqs. 5-6 are the corresponding mass balances.

$$\min \sum_{i \in T} F_i^{tot} \tag{1}$$

$$tf_w^{wwat} = \sum_{i \in T} F_{w,i}^{wwat} + F_w^{byp} , \quad \forall w \in W \tag{2}$$

$$F_i^{tot} = \sum_{w \in W} F_{w,i}^{wwat} + \sum_{j \in T} F_{i,j}, \quad \forall i \in T \tag{3}$$

$$F_i^{tot} = \sum_{j \in T} F_{i,j} + F_i^{dis}, \quad \forall i \in T \tag{4}$$

$$MT_{i,c}^{In} = \sum_{w \in W} F_{w,i}^{wwat} \cdot c_{w,c}^{wwat} + \sum_{j \in T} \sum_{a=1}^{\xi_{i,c}} \sum_{p=0}^{9} TF_{j,i,c}^{a,p} \times \psi_{a,p} , \quad \forall i \in T, c \in C \tag{5}$$

$$MT_{i,c}^{Out} = \sum_{j \in T} \sum_{a=1}^{\xi_{i,c}} \sum_{p=0}^{9} TF_{i,j,c}^{a,p} \cdot \psi_{a,p} + \sum_{a=1}^{\xi_{i,c}} \sum_{p=0}^{9} TD_{i,c}^{a,p} \cdot \psi_{a,p} , \quad \forall i \in T, c \in C \tag{6}$$

Eqs. 7-8 ensure that all active artificial variables have the same value. Eq. 9 is the definition of the removal ratio of each contaminant within each unit. Eq. 10 ensures that the environmental discharge limits are not exceeded. Eq. 11 forces the artificial flowrate variables linked to non-selected parameters to be zero. On the other hand, the outlet flow cannot exceed a certain upper bound, where multiplicative factor φ is employed to allow for a recycled flow greater than the total amount of wastewater entering the treatment system. Eq. 12 ensures that a single digit is selected for a certain numerical position for contaminant c in unit i. Finally, Eqs. 13-14 guarantee that the inlet contaminants concentration does not exceed their maximum admissible values and that of outlet concentrations (LHS) are lower than the maximum outlet concentrations. Additional design constraints can be easily incorporated into the mathematical model.

$$F_i^{dis} = \sum_{p=0}^{9} TD_{i,c}^{a,p}, \quad \forall i \in T, c \in C, a \in \xi_{i,c} \tag{7}$$

$$F_{i,j} = \sum_{a=1}^{9} \sum_{p=0}^{9} TF_{i,j,c}^{a,p}, \quad \forall i,j \in T, c \in C, a \in \xi_{i,c} \tag{8}$$

$$MT_{i,c}^{Out} = MT_{i,c}^{In} \times \left(1 - rr_{i,c}\right), \quad \forall i \in T, c \in C \tag{9}$$

Parametric Programming Technique for Global Optimization of Wastewater Treatment Systems

$$\sum_{w \in W} F_w^{byp} \cdot c_{w,c}^{wwat} + \sum_{i \in T} \sum_{a \in \xi_{i,c}} \sum_{p=0}^{9} TD_{i,c}^{a,p} \times \psi_{a,p} \leq \left(\sum_{w \in W} F_w^{byp} + \sum_{i \in T} F_i^{dis} \right) \times c_c^{env}, \quad \forall c \in C \tag{10}$$

$$\sum_{j \in T} TF_{i,j,c}^{a,p} + TD_{i,c}^{a,p} \leq \varphi \cdot \min \left[\sum_{w \in W} tf_w^{wwat}, \left(\sum_{w \in W} tf_w^{wwat} \times c_{w,c}^{wwat} \right) / c_{i,c}^{in\,max} \right] \times Z_{i,c}^{a,p}, \tag{11}$$

$$\forall i \in T, c \in C, a \in \xi_{i,c}, p \in \{0,1,...,9\}$$

$$\sum_{p=0}^{9} Z_{i,c}^{a,p} = 1, \quad \forall i \in T, c \in C, a \in \xi_{i,c} \tag{12}$$

$$C_{i,c}^{in} \leq c_{i,c}^{in\,max}, \quad \forall i \in T, c \in C \tag{13}$$

$$\sum_{a=1}^{\xi_{i,c}} \sum_{p=0}^{9} \psi_{a,p} \times Z_{i,c}^{a,p} \leq c_{i,c}^{out\,max}, \quad \forall i \in T, c \in C \tag{14}$$

5. Computational results

The performance of the new technique is now illustrated through several samples. Their respective sizes and numerical results are given in Figs. 2-3. The hardware consisted of an Intel Core 2 Duo 2.4 GHz processor, with 2 GB of RAM memory, running Windows Vista. The underlying formulations were implemented and solved in GAMS 23.2, using CPLEX as the MILP solver, and CONOPT and BARON as NLP solvers, the latter being a global optimization solver.

The results in Figure4 show that solution quality increases with an increase in the number of decimal places used to represent the outlet concentrations. This is true both for CPLEX and CONOPT, which is initialized with the former solution. Thus, it is not surprising to find out that CONOPT normally improves the solution, by making it more accurate. The highest deviation in the range [1, 4] occurs for Ex5, with the gap between the outputs from CPLEX and CONOPT being equal to 2.79, 0.08, 0.02 and 0%, respectively. For two decimal places (n=2) the gap is always under 0.10%, which is a very small number for all practical purposes. Furthermore, with the exception of Ex5 for n=1 and Ex11 for n=1-3, the solution from CONOPT is the global optimal solution. And only in the former the difference is relevant, 231.881 vs. 229.701.

BARON could also find the global optimal solution for all examples. However, the optimality gaps for Ex9-12 and Ex14 are very significant (roughly >10%), particularly after one takes into account the computational effort (CPUs in Fig. 3). In contrast, our new CPLEX+CONOPT approach took less than one hour for all problems but Ex14 for n=4. Despite the fact that we cannot calculate an optimality gap since there is no relaxation providing a lower bound as in BARON, it is fair to assume that the optimality gap will be lower than the difference between the output values from CPLEX for n=3 and n=4, which are below 0.02%.

Finally, it should be noted that there is not a linear relation between the number of decimals and computational effort. Frequently the CPUs decreased for increasing problem sizes (directly related to n), which can be explained by: (i) increasing n may lead to lower integrality gaps, which contributes to a faster search; (ii) the branch & bound algorithm beneath CPLEX is heuristic.

| | Number of decimal digits for optimal approximation of variable concentrations (n) | | | | | | | | | | | | BARON | | |
| Aprox | 1 | | | 2 | | | 3 | | | 4 | | | | | |
Case Studies	CPLEX	CONOPT	Gap	CPLEX	CONOPT	Gap	CPLEX	CONOPT	Gap	CPLEX	CONOPT	Gap	Lower Bound	Sol.	Gap
Ex1	**179.798**	**179.798**	0.00%	**179.798**	**179.798**	0.00%	**179.798**	**179.798**	0.00%	**179.798**	**179.798**	0.00%	**179.798**	**179.798**	0.00%
Ex2	130.705	**130.703**	0.00%	130.705	**130.703**	0.00%	**130.703**	**130.703**	0.00%	**130.703**	**130.703**	0.00%	**130.703**	**130.703**	0.00%
Ex3	**99.495**	**99.495**	0.00%	**99.495**	**99.495**	0.00%	**99.495**	**99.495**	0.00%	**99.495**	**99.495**	0.00%	**99.495**	**99.495**	0.00%
Ex4	90.441	**89.836**	0.67%	89.929	**89.836**	0.10%	89.841	**89.836**	0.01%	89.837	**89.836**	0.00%	**89.836**	**89.836**	0.00%
Ex5	238.359	231.881	2.79%	229.849	**229.701**	0.06%	229.75	**229.701**	0.02%	229.708	**229.701**	0.00%	**229.701**	**229.701**	0.00%
Ex6	174.054	**173.478**	0.33%	173.529	**173.478**	0.03%	173.483	**173.478**	0.00%	173.479	**173.478**	0.00%	**173.478**	**173.478**	0.00%
Ex7	80.87	**80.779**	0.11%	80.783	**80.779**	0.00%	80.781	**80.779**	0.00%	80.78	**80.779**	0.00%	**80.779**	**80.779**	0.00%
Ex8	586.814	**586.68**	0.02%	586.702	**586.68**	0.00%	586.681	**586.68**	0.00%	586.681	**586.68**	0.00%	**586.68**	**586.68**	0.00%
Ex9	2127.168	**2127.115**	0.00%	2127.123	**2127.115**	0.00%	2127.116	**2127.115**	0.00%	2127.116	**2127.115**	0.00%	1534.628	**2127.115**	27.85%
Ex10	**1201.038**	**1201.038**	0.00%	**1201.038**	**1201.038**	0.00%	**1201.038**	**1201.038**	0.00%	**1201.038**	**1201.038**	0.00%	829.982	**1201.038**	30.89%
Ex11	1566.785	1564.958	0.12%	1564.992	1564.958	0.00%	1564.968	1564.958	0.00%	1564.959	**1564.957**	0.00%	1126.431	**1564.957**	28.02%
Ex12	513.875	**513.001**	0.17%	513.082	**513.001**	0.02%	513.054	**513.001**	0.01%	**513.001**	**513.001**	0.00%	404.629	**513.001**	21.13%
Ex13	2463.297	**2446.429**	0.69%	2447.114	**2446.429**	0.03%	2446.53	**2446.429**	0.00%	2446.43	**2446.429**	0.00%	**2446.429**	**2446.429**	0.00%
Ex14	1359.688	**1358.663**	0.08%	1358.792	**1358.663**	0.01%	1358.67	**1358.663**	0.00%	1358.668	**1358.663**	0.00%	1233.176	**1358.663**	9.24%

Figure 2. Solution quality as a function of the number of decimal digits (global solution in bold).

| | | | | SINGLE EQUATIONS | | | | SINGLE VARIABLES | | | | DISCRETE VARIABLES | | | | CPUs | | | | |
| | | | | Number of decimal digits for optimal approximation of variable concentrations (n) | | | | | | | | | | | | CPLEX + CONOPT | | | | |
Case Studies	#C	#WS	#TU	1	2	3	4	1	2	3	4	1	2	3	4	1	2	3	4	BARON
Ex1	1	3	1	37	50	63	76	72	102	132	162	20	30	40	50	0.14	0.07	0.07	0.07	0.05
Ex2	1	2	2	100	128	156	184	259	339	419	499	60	80	100	120	0.28	0.18	0.26	0.29	0.06
Ex3	1	3	2	87	115	143	171	222	302	382	462	50	70	90	110	0.18	0.18	0.19	0.18	0.08
Ex4	2	2	2	193	249	305	361	503	663	823	983	120	160	200	240	0.17	0.38	0.81	1.17	0.12
Ex5	3	3	3	514	649	784	919	1596	2046	2496	2946	310	400	490	580	38	111	130	245	10.95
Ex6	3	3	3	559	694	829	964	1746	2196	2646	3096	340	430	520	610	1068	274	347.97	421	40.26
Ex7	3	3	3	589	724	859	994	1846	2296	2746	3196	360	450	540	630	26	73	134.99	1335	303
Ex8	1	3	5	307	392	477	562	1184	1534	1884	2234	160	210	260	310	6	13	20	23	280
Ex9	1	5	7	486	619	752	885	2188	2818	3448	4078	230	300	370	440	6	11	11.056	27	8441[a]
Ex10	1	6	10	705	925	1145	1365	3676	4876	6076	7276	290	390	490	590	14	641	122	731	8295[a]
Ex11	1	6	15	1394	1799	2204	2609	8542	11092	13642	16192	480	630	780	930	3011	39	245	420	225780[a]
Ex12	2	3	5	583	753	923	1,093	2244	2944	3644	4344	310	410	510	610	2298	2927.5	722	2945	17262[a]
Ex13	3	4	3	469	604	739	874	1446	1896	2346	2796	280	370	460	550	153	112	435	1095	1254
Ex14	2	2	6	749	965	1181	1,397	3127	4087	5047	6007	380	500	620	740	131	1866.8	243	8028	78605[a]

[a]unable to prove global optimality (lower bound at time of interruption)

Figure 3. Computational statistics as a function of the number of decimal digits.

6. Conclusions

This work has presented a new strategy for the optimal design of wastewater treatment networks. It involves generating a set of artificial multi-parametric elements resulting from the decomposition of bilinear terms present in the general nonlinear mathematical formulation. The optimal elements are then chosen through the solution of a single MILP problem. The outcome is an upper bound on the global optimal solution that becomes increasingly tighter with an increase in the number of decimal digits used in the approximation. The same principle can be applied to other process synthesis problems. In particular, work is underway to extend the approach to the design of water-using networks and fully integrated water networks.

References

J. Teles, P. Castro, A. Novais, 2009, MILP-based initialization strategies for the optimal design of water-using networks. Chem. Eng. Sci. 64, 3736-3752.

P. Castro, J. Teles, A. Novais, 2009, Linear program-based algorithm for the optimal design of wastewater treatment systems. Clean Techn Environ Policy. 11, 83-93.

M. Gunaratnam, A. Alva-Argáez, A. Kokossis, J. Kim J & R. Smith, 2005, Automated design of total water systems. Ind. Eng. Chem. Res. 44, 588.

R. Karuppiah, I. Grossmann, 2006, Global optimization for the synthesis of integrated water systems in chemical processes. Comput. Chem. Eng. 30, 650.

Systematic approach for synthesis of intensified biodiesel production processes

Lene Fjerbaek Sotoft, Ben-Guang Rong, Knud V. Christensen, Birgir Norddahl

Inst. of Chemical Eng., Biotechnology and Environmental Tech., University of Southern Denmark, Niels Bohrs Alle 1, DK-5230 Odense M, Denmark

Abstract

Biodiesel production is a very hot topic within research, politics and investments worldwide. To intensify production and improve decision making and production evaluation, a systematic approach to describe biodiesel production processes is needed. The paper presents a systematic approach for process synthesis of biodiesel production and illustrates the importance of the approach by an example of biodiesel production from waste animal fats. The process synthesis method uses a step-by-step approach to construct, improve or evaluate biodiesel production processes.

Keywords: Biodiesel, process synthesis, conceptual design, process intensification

1. Introduction

Biodiesel has gained much focus in the recent years as a liquid renewable fuel for the transportation sector due to its many merits compared to the conventional diesel fuel. Biodiesel is the alkyl ester product of a transesterification reaction between a triglyceride and an alcohol (Fig.1).
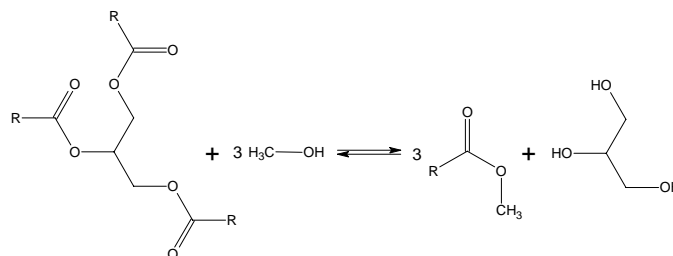
Figure 1: Formation of biodiesel and glycerol byproduct from triglyceride and alcohol.

However, there are many factors that determine the success of biodiesel as a final alternative transportation fuel. Among which, two major factors are substantial to determine the process economy for the biodiesel production processes. The first is that the process has to be able to use the cheapest raw materials as raw material costs are the major part of the total production cost. The second is that the process should be intensified with as few as possible processing steps and equipment requirements. This calls for a systematic approach for biodiesel production process synthesis.

2. Procedures for production of biodiesel today

Biodiesel production is today taking place with traditional homogeneous chemical catalysts such as NaOH, KOH and H_2SO_4. The catalysts are not reusable, but must be removed from the biodiesel before use. This leads to highly water consuming downstream processing for removal of the salts from the final product and subsequent large amounts of complicated waste water that needs treatment. It also requires high temperatures, large amounts of catalysts and additional chemicals for neutralization.

In order to reduce the complexity of the process, much research has been carried out investigating new catalysts, in particular heterogeneous catalysts without any water requirement in the downstream processing and thereby without waste water production. Many of these catalysts are promising, but methodological measures lack for the overall evaluation of their performance. They are mostly only evaluated based on the reactor performance, but not of price, complexity, special requirements i.e. solvents and what implications their use has to the overall process design and performance.

On the other hand, various raw materials are used in industrial scale for biodiesel production such as palm, rapeseed, soybean, castor and *Jatropha curcas* oil as well as various waste products such as grease, waste vegetable oils and animal fat, while numerous more or less exotic plant oils and fats worldwide also have been tested by researchers Fjerbaek et al. [4]. Algae oil attracts much attention also, and many industrial initiatives exist, but no commercially available algae biodiesel is yet seen. The raw materials vary highly in i.e. water content, free fatty acid content, fatty acid composition and properties derived from these. This has an impact on the product quality and process synthesis. Therefore, also the overall performance of the raw materials must be based on a systematic procedure for the synthesis of different process alternatives.

The aim of the paper is to formulate a systematic procedure for the synthesis of different process alternatives depending on choice of catalysts and raw materials as a basis for good reliable overall process evaluations. The key steps for determining the process structures are systematically being analyzed and integrated, including different raw materials, different catalysts, reactor design and solvents, reaction subsystem, separation unit operations as well as the interconnection between reaction and separation. The systematic procedure serves as a workflow to produce different process alternatives step-by-step. Different process alternatives are compared on the basis of conventional catalysts and enzyme-based catalysts.

This paper will briefly introduce the key elements of the systematic procedure illustrated by case studies, the comparisons of the different process alternatives from both process intensification viewpoint and process simulation evaluations will be presented in the full paper.

3. Synthesis of intensified biodiesel production process

Producing biodiesel is not an easy and constant way to make money. The market of biodiesel production is ever changing with respect to taxation, tolls, market prices and market demands.

Systematic tools are needed for researchers, managers, engineers and investors to propose and built new plants with the highest possible degree of success, if biodiesel is

to be produced and that viably. In order to improve process economy and process efficiency, a systematic procedure for an intensified biodiesel production process is required.

3.1. Process synthesis of biodiesel production

The major steps that determine the final outline of a biodiesel production process are put into a systematic procedure as shown on Figure 2.
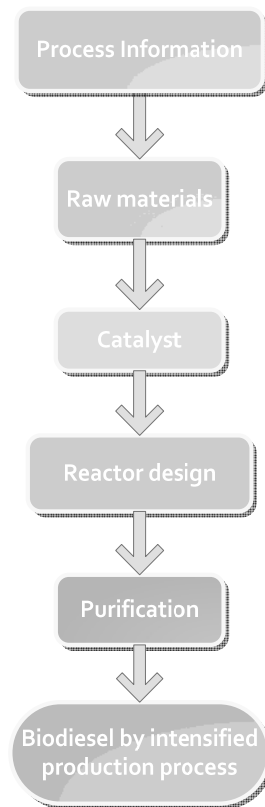
Figure 2: Procedure for process synthesis of biodiesel production

A plan for a biodiesel production includes a certain amount of process information. That is i.e. plant location, production size or usage of a specific raw material.

This information then forms a basis for firstly the choice of raw materials, both oil and alcohol and all the subsequent steps in the production process. This is best to be illustrated by an example, where a biodiesel production is planned to be based on waste animal fats. For a high quality vegetable oil, a one step process with a homogeneous caustic catalyst followed by phase separation, water washing and alcohol recovery is the obvious, traditional choice. This is very different from waste animal fats as raw material.

Waste animal fats in contrast have a high content of free fatty acids (FFA) with great seasonal variations in the content. The raw material also contains relative high amounts

of water and impurities (protein residues, phosphorlipids etc.). As Table 1 shows, the possible raw materials for biodiesel production vary a lot in composition, i.e. FFA.

Table 1: Content of free fatty acids in commonly used raw materials for biodiesel production.

Rapeseed oil	Unrefined soybean oil	Refined *Jatropha* oil	Used frying oil	Animal fat	Byproduct from oil production
[10]	[11]	[8]	[1]	[5]	[12]
0.3-2 wt%	0.3-0.7 wt%* < 0.05 wt%**	14 wt%	0.7-41.8 wt%	5-30 wt%	78 wt%

*unrefined, ** refined

The water and FFA are not working very well together with a caustic catalyst due to soap formation and severe emulsification, therefore a combination of catalysts must be used with a preesterification of the FFA with i.e. acid followed by caustic transesterification of the triglycerides in a two step process.
Reactors are then to be designed depending on reaction time, operational mode (batch/continuous, parallel or serial reactors) and cost.

Product specifications, composition after reactors and impurities determine the needed degree of purification of the biodiesel. Firstly, glycerol is removed by phase separation. This can be made complicated by the presence of emulsifying impurities. Afterwards, homogeneous catalyst residues are to be removed by water washing. Alternatively, glycerol, monoglycerides, soaps and catalyst residue can be removed in an ion-exchange column that can replace water washing.

The alcohol is normally added as the excess reactant, so excess alcohol is to be recovered by distillation. The biodiesel quality at this point is determined by the raw material. High quality vegetable oils require no further treatment to meet specifications, while low quality animal fat a.o. require distillation of the biodiesel. Here the product mixture is vacuum distilled, where the top product is the pure biodiesel, and the bottom product is a waste fraction. This is a necessary, but very costly purification step.
The glycerol can be sold as crude or refined glycerol. What is done depends on plant size. For small scale operation, crude glycerol is typically sold directly, while it can be worthwhile for larger scale operations to refine the glycerol by distillation.

All in all these requirements can be addressed as seen in the more detailed process synthesis procedure, see Figure 3.

The process synthesis procedure can describe a specific case with a degree of details depending on the purpose of the work. The procedure can be used to evaluate new catalysts, catalysts in new forms and new intensified reactor types with respect to influence on the overall production process.

Much work is put into enzymes as an environmental friendly alternative to traditional catalysts. This procedure is very useful for comparison of the two with respect how good an alternative the enzymes really are. Many catalysts, both caustic, acidic and enzymatic, are also made into heterogeneous catalysts to substitute the traditional homogeneous ones. The analysis of their performance and process impact (i.e. no water washing is needed) can also be carried out by this stepwise approach.
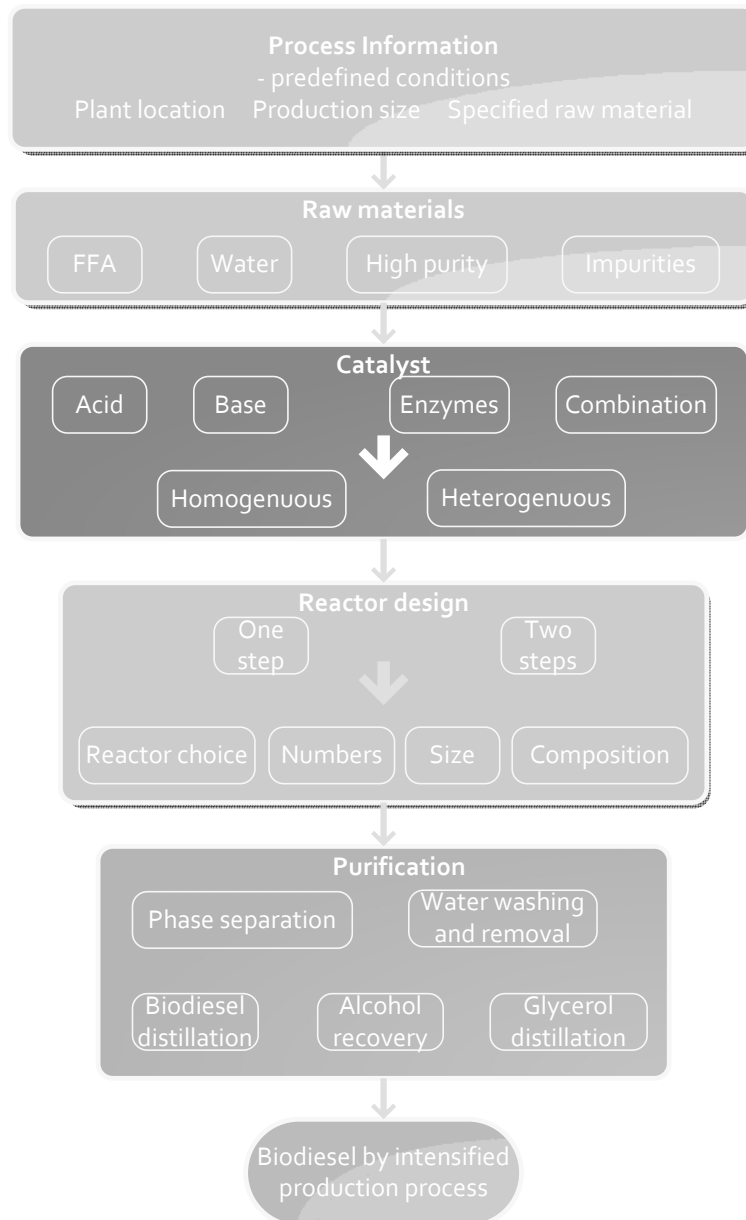
Figure 3: Detailed procedure for process synthesis of biodiesel production

Last, but not least new reactor types emerge and intensify the production process by combining i.e. reactive distillation [3,7], catalytic membrane reactors [2,9] and various novel reactor/catalyst configurations [6]. These need to be analyzed thoroughly to investigate their potential and also in this case, this systematic approach for process synthesis is highly applicable

4. Conclusions and future work

A systematic approach for biodiesel production process synthesis is presented, and the value of the approach is sketched on an overall basis.

At each step, different partial solutions can be generated which will produce different final process alternatives. The cases studied took different raw materials with various compositions as well as usage of different catalysts. The implications of the choices will then be outlined and clarified to a greater extent by the proposed procedure then otherwise possible with existing methods. The full paper will present systematic synthesis of the conceptual biodiesel process alternatives and their simulations.

5. Acknowledgements

References

[1] M. Canakci, 2007, The potential of restaurant waste lipids as biodiesel feedstocks, Bioresource Technol, 98, 183.

[2] P. Cao et. al., 2008, High-purity fatty acid methyl ester production from canola,soybean, palm, and yellow grease lipids by means of a membrane reactor, Biomass bioenerg, 32, 1028.

[3] AC. Dimian et. al., 2009, Innovative process for fatty acid esters by dual reactive distillation,Comput Chem Eng, 33, 743

[4] L. Fjerbaek et. al., 2009, A review of current state of biodiesel production using enzymatic transesterification, Biotechnol Bioeng, 102, 1298.

[5] JV. Gerpen, 2005, Biodiesel processing & production, Fuel Process Technol, 86, 1097.

[6] M. Hajar et. al., 2009, Solvent-free methanolysis of canola oil in a packed-bed reactor with use of Novozym 435 plus loofa, Enzyme Microb Tech, 45, 3, 188.

[7] AA. Kiss, 2009, Separative reactors for integrated production of bioethanol and biodiesel, Comput Chem Eng, doi:10.1016/j.compchemeng.2009.09.005

[8] A. Kumar Tiwari et.al., 2007, Biodiesel production from jatropha oil (*Jatropha curcas*) with high free fatty acids: An optimized process, Biomass Bioenerg, 31, 569.

[9] DE. Lopez et. al., 2007,Transesterification of triacetin with methanol on Nafion acid resins, J Catal, 245, 2, 381.

[10] M.Mittelbach, C.Remschmidt, 2006, Biodiesel - the comprehensive handbook.

[11] FT. Orthoefer,GR. List, 2007, Initial Quality of Frying Oil, Oil Mill Gazetteer, 113, 3.

[12] Y. Watanabe et. al., 2007, Conversion of acid oil by-produced in vegetable oil refining to biodiesel fuel by imm. *C. antarctica* lipase , J Mol Catal B-Enzym, 44, 99.

A systems platform for the optimal synthesis of biomass based manufacturing systems

Antonis C. Kokossis[a], Aidong Yang[b], Marinella Tsakalova[b], Ta-Chen Lin[b]

[a]*School of Chemical Engineering, National Technical University of Athens, Greece*
[b]*Faculty of Engineering and Physical Sciences, University of Surrey, UK*

Abstract

The gradual depletion of oil, uncertainties in energy supplies and a commanding requirement to reduce greenhouse gas (GHG) emissions have intensified worldwide interest in renewable forms of energy. By far, biomass is the most abundant, the fastest growing and the only carbon-based source of renewable energy. Its efficient use is the task of a chemical engineer and could lead to the development of new manufacturing systems or the retrofitting of existing ones. To rationalize its use, a large number of options are important to screen in relation to the type of feedstocks available, the appropriate production routes, and the product portfolios. The paper discusses a process synthesis method to address the systematic integration of flowsheets supporting decisions for investment and the selection of processing paths. The implementation takes the form of a systems platform with modelling and optimization capabilities at different levels. This methodology is part of a UK national programme devoted to valorise biorefinery capabilities in the North of England.

Keywords: biomass processing, modelling, optimization, synthesis

1. Introduction

The use of renewable biomass feedstocks is growing fast, induced by the gradual depletion of oil, uncertainties in energy supplies and a commanding requirement to reduce greenhouse gas (GHG) emissions. Sustainable use of biomass and the economical use of capital venture require the simultaneous assessment of factors in relation to the occasional regional strengths, the local availability of biomass, and the integration of existing manufacturing facilities (possible mix with conventional feedstock). From a systems perspective, one has to screen over several economical and environmental objectives optimizing the selection of feedstocks and products, as well as the selection of processing routes and technologies. A support system with capabilities to support high-level decisions would be highly useful to individual companies and public organizations.

Systems engineering methods have received strong encouragement and attention in connection to biomass-based manufacturing. Dimian (2007) discusses opportunities for computer-aided process engineering and renewable raw materials, particularly with regard to conceptual process design. Klatt and Marquardt (2009) highlight the processing of renewable feedstocks as one of the emerging application domains in process systems engineering. In addressing specific processes, flowsheeting and simulation studies dominate the literature, as in the work of Cardona and Sanchez (2006) and Piccolo and Bezzo (2009). In comparison to flowsheeting analysis, synthesis and process integration applications are rare but exist; see e.g. Gassner and Marechal

(2009). The potential for a systems approach to address biorefinery challenges is highlighted in recent literature (Kokossis and Yang, 2009).

As opposed to the study of individual processes, the Department of Environment, Food and Rural Area (DEFRA) in UK has supported the development of a systematic approach to screen biomass feedstocks and design options. The screening considers Compartments of Processing Blocks (CPB) that may include a single process or clusters of processes. Clusters may include the entire bioethanol or transesterification (biodiesel) process that, as a whole, convert raw material (including biomass and fossil based feedstock) into specific sets of products. The paper provides a brief summary of the methodology in Section 2. Section 3 outlines the systems platform and the following sections explain progress on the implementation of the methodology and illustrative case studies.

2. Methodology: purpose and objectives

The objective of the paper is to address the synthesis problem at a conceptual stage, namely by determining the optimal combination of processing blocks to handle different feedstocks. A formal description of the problem takes the form:
Given,

(i)	types, volumes, and distribution of biomass feedstocks,
(ii)	sets of applicable processing routes and technologies,
(iii)	sets of candidate products, and
(iv)	sets of economical and environmental constraints,

Optimally determine the

(i)	compartments of processing blocks required
(ii)	suitable feedstock (types, volumes, and locations) or the mix of feedstocks
(iii)	product portfolios and the production scales

The objective function accounts for economic and environmental targets (e.g. costs, profit, energy use, percentage consumption of non-renewable fuels, GHG emissions).
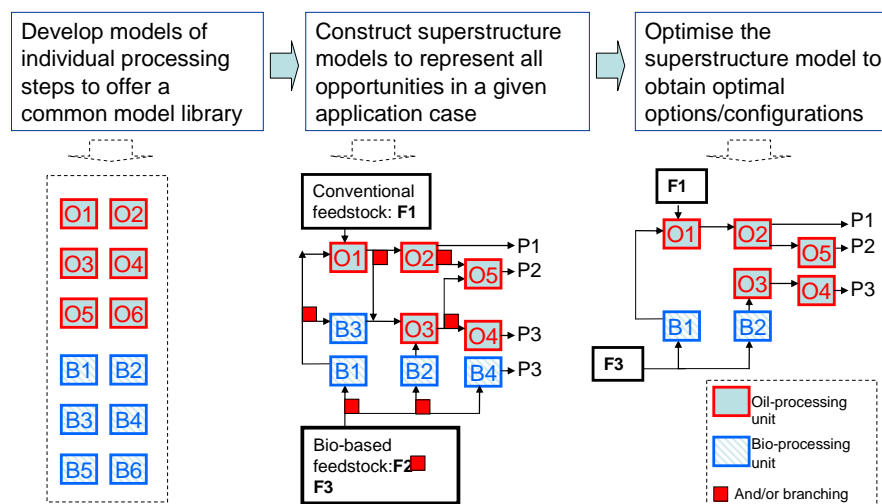
Figure 1. Methodology for the synthesis of biomass based manufacturing systems

A systems platform for the optimal synthesis of biomass based manufacturing systems

The methodology determines (a) subsets of processes to replace or modify (Compartments of Processing Blocks, CPB), and (b) libraries of input-output models each featuring process variables in relation to the performance of each block and its material and energy streams. Blocks include bio-based ('B' blocks) and conventional blocks ('O' blocks as it typically refers to an oil-based process) as shown in the illustrative superstructure of Figure 1. Blocks stand for synthesis components combined in superstructure schemes that account for all their compatible combinations. Superstructures are modelled as optimization problems whose solutions provide recommendations for integration schemes, feedstocks and processing steps to select.

3. Design of the systems platform

The paper summarizes key components of the systems platform further outlining two important concepts in relation to the level of analysis and the synthesis units.

3.1. Key components

The synthesis functions include four components highlighted in Figure 2:
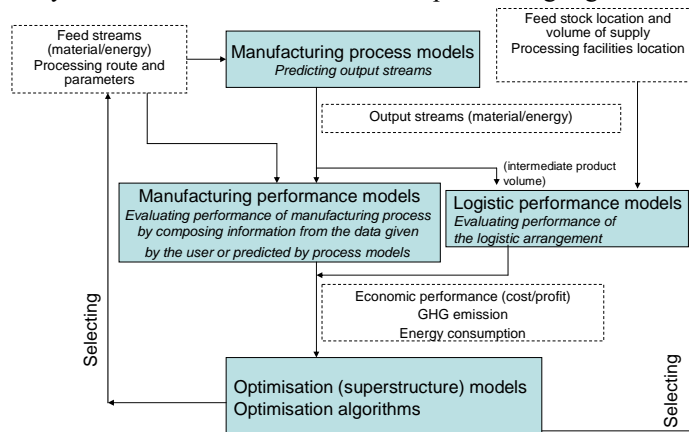
Figure 2. Key components of the systems platform.

- *Manufacturing process models* that predict the output of each CPB from input information (material and energy streams to the block)
- *Manufacturing performance models* that evaluate, given output from the models above, the economical performance, the energy consumption, and the greenhouse gas emissions (GHG)
- *Logistic performance models* addressing logistics and supply chain considerations
- An integrated scheme (*superstructure*) enabling a simultaneous and exhaustive assessment of design options and the mathematical modelling as an optimisation problem.

3.2. Level of analysis

The development of an optimal manufacturing system essentially determines the systems components to integrate. As shown in Figure 3, there exist several levels to

locate a particular component (Levels 2, 3, and 4) whereas, and as summarised in Table 1, the level of abstraction bears a significant impact on modelling and optimisation.
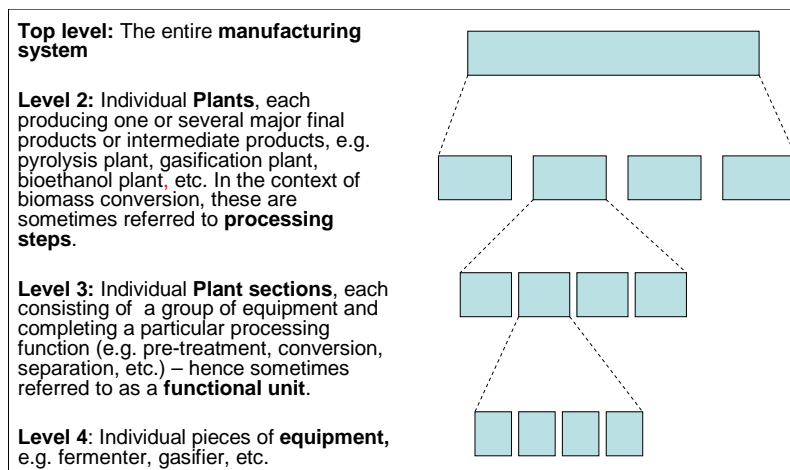
Figure 3. Different levels of analysis.

Table 1: Implication of level selection

Levels	Effort of modelling and optimization	Scope of assessment	Resolution of solution
Plants or Processing steps (Level 2)	Low	Large	Preliminary
Plant sections or Functional units (Level 3)	Medium	Medium	Medium
Equipment (Level 4)	High	Small	Fine

The paper focuses on modelling and analysis work addressing Level 2 and/or Level 3. Level 2 screens a wide range of processing options whereas Level 3 is intended to distinguish between processing technologies of different efficiencies and cost. Other application may determine other levels dictated by the intended scope of analysis and scrutiny. A mixed mode is also possible, suggesting a multi-resolution approach (e.g. part of the system based on Level 2 and the rest of analysis on Level 3). Considering the current uncertainties in technologies, uncertainties in markets and products, alongside the unknown potential for process integration, the modelling and the analysis work assumed at Level 4 can be hardly justified and is probably a task to address in the future.

3.3. Unit of synthesis
With multiple possibilities concerning the level of analysis, Compartments of Processing Blocks (CPB) account for the synthesis units to integrate in the superstructure. Depending on the analysis level, a CPB may represent a whole plant (e.g. pyrolysis, gasification), a plant section (or functional unit, e.g. pre-treatment, conversion, cleaning/purification), or a piece of equipment (e.g. fermentor, gasifier). A plant or a particular process to model is treated as a single CPB, or decomposed instead

into a number of interconnected CPBs (each representing the plant section or individual pieces of equipment). CPB boundaries can vary according to the intended use of the CPB. The different ways reflect on the particular scope to assess process integration benefits (i.e. emphasis on entire plant sections viz. a piece of equipment) and remains an input for the proposed approach.
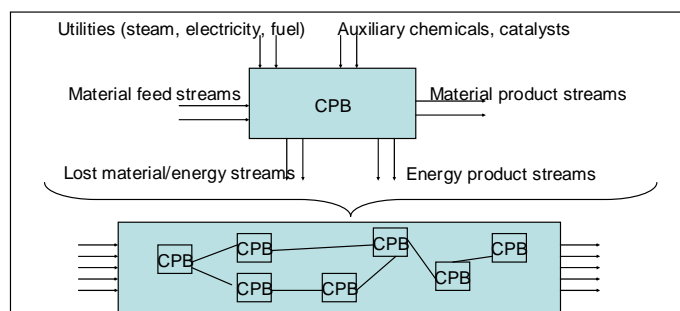
Figure 4. CPB superstructures

Figure 4 illustrates the modelling information to capture in the CPB as well as connections of CPBs towards superstructures. Reflecting on the occasional differences in boundaries, the modelling components may be pre-constructed (for individual CPBs) or take a form of configurable formulations (superstructures). In principle, the former practice applies to the *manufacturing process models* (Section 2) and the models estimating capital and fixed operating costs (part of the models used to assess the manufacturing performance). The latter practice applies to models estimating variable operating costs, sales, revenues, energy consumption or exporting, and GHG emissions. Such models are not constructed for each individual CPB as they measure performance of the entire system and are required to account for material and energy streams exchanged between the system considered and its environment.

4. Implementation and case studies

A GAMS-based implementation is being carried out for the systems platform. This involves the population of a CPB model library where each CPB model is currently formulated as a separate GAMS model; different CPB models can be selected and combined according to an application specific superstructure model. Although currently composed manually, it is envisaged that in the long run the generation of superstructure models may be assisted by a graphical user interface (GUI). The GUI may be also useful in the formulation of objective functions by means of interaction with the user. The ongoing development is particularly guided by case studies that address the manufacturing of two specific commodity chemicals, namely *ethylene* and *propylene glycol*. Figure 5 shows the processing routes that have been considered for producing ethylene from woody biomass. The case study is to identify the best combination of these known steps for maximizing the profit while subject to the constraints on biomass availability, expected productivity of main products, and levels of energy consumption and GHG emission.
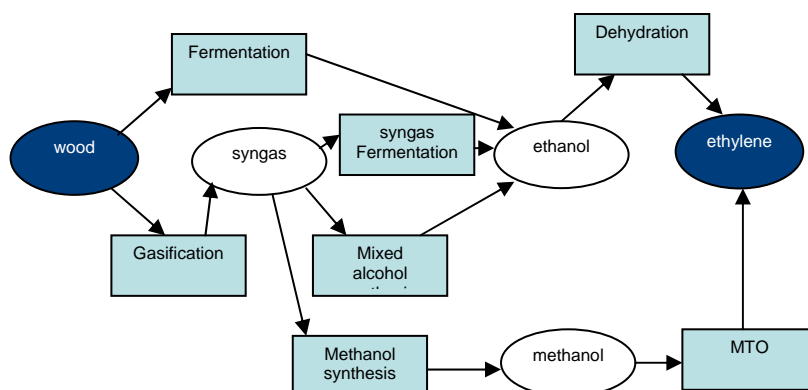
Figure 5. Routes considered for producing ethylene from woody biomass.

5. Concluding remarks

Introducing renewable biomass into the manufacturing of energy and chemical products poses great challenges as well as opportunities to process systems engineering. This requires extensions of established methods and tools to address new problems in this area. This paper reports the concepts developed in an ongoing project that assesses the optimal opportunities to produce chemicals from biomass, possibly in combination with more conventional feedstock. The methodology for synthesizing complex manufacturing chains or networks is presented, and the design of a systems platform to carry out this type of synthesis described. The experiences and outcomes of implementing and tesing this systems platform will be reported in the future.

Acknowledgement

References

Cardona, C.A., Sanchez O.J. (2006). Energy consumption analysis of integrated flowsheets for production of fuel ethanol from lignocellulosic biomass. Energy 31, 2447–2459.

Dimian, A. C. (2007). Renewable raw materials: chance and challenge for computer-aided process engineering. Computer Aided Chemical Engineering, 24, 309-318.

Gassner, M, Marechal, F. (2009). Methodology for the optimal thermo-economic, multi-objective design of thermochemical fuel production from biomass. Comput. Chem. Eng. 33, 769-781.

Klatt K.-U., Marquardt, W. (2009). Perspectives for process systems engineering—Personal views from academia and industry. Comput. Chem. Eng. 33, 536-550.

Kokossis A., Yang A. (2009). Future system challenges in the design of renewable bio-energy systems and the synthesis of sustainable biorefineries. FOCAPD 2009.

Piccolo, C., Bezzo, F. (2009). A techno-economic comparison between two technologies for bioethanol production from lignocellulose. Biomass and Bioenergy, 33, 478-491.

Heat-Integrated Process for Biodiesel by Reactive Absorption

Anton A. Kiss

AkzoNobel – Research, Development and Innovation, Velperweg 76, 6824 BM Arnhem, The Netherlands, tony.kiss@akzonobel.com

Abstract

This study proposes an innovative biodiesel technology based on heat-integrated reactive absorption. Rigorous simulations embedding experimental results were performed in AspenTech AspenONE engineering suite to design this novel process and evaluate the technical and economical feasibility. The main results are given for a plant producing 10 ktpy biodiesel from waste vegetable oil with high free fatty acids content, using solid acids as green catalysts. This innovative process eliminates all conventional catalyst related operations, and efficiently uses the raw materials and the reactor volume in an integrated setup that allows significant savings in capital and operating costs.

Keywords: Biodiesel, reactive absorption, reactive distillation, energy integration

1. Introduction

Due to obvious economic incentives, process integration and intensification are increasingly applied also in the biodiesel production. This study proposes an innovative heat integrated process based on reactive absorption (RA) thus further improving previous work on reactive-separation technology for biodiesel production. RA offers significant benefits over reactive distillation (RD), such as: reduced capital investment and operating costs due to the absence of the reboiler and condenser, higher conversion and selectivity as no products are recycled in form of reflux or boil-up vapors, as well as no occurrence of thermal degradation of the products due to a lower temperature profile in the column. Rigorous simulations embedding experimental results were performed in AspenTech Aspen Plus to design this novel reactive-absorption process and evaluate the technical and economical feasibility. The main results are given for a plant producing 10 ktpy biodiesel from waste vegetable oil with high free fatty acids content (up to 100%), using solid acids as green catalysts. By adding heat integration to this process, the heating / cooling requirements are significantly reduced by 85% and 90%, respectively.
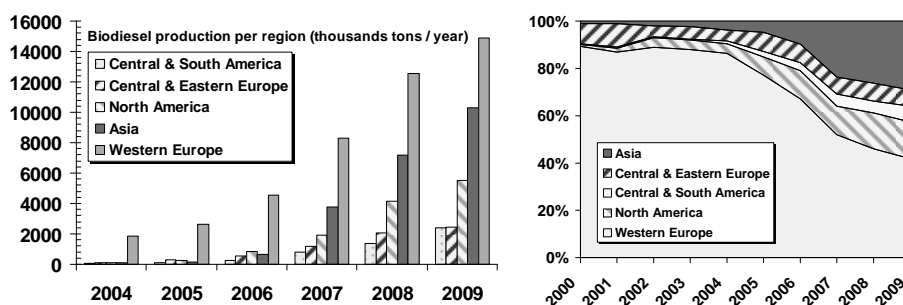
Figure 1. Biodiesel production in the world, and global market share per region.

2. Problem statement

Conventional biodiesel processes are all plagued among others by the use of an excess of alcohol to push the chemical equilibrium of (trans-)esterification towards fatty esters formation, and costly downstream processing steps associated to the use of homogeneous catalysts. Due to the increased costs of raw materials, the current trend is to use less expensive alternatives such as animal fat, waste cooking oil, or waste vegetable oil (Kulkarni and Dalai, 2006). Moreover, although the raw materials represent now the major part of the variable costs associated with the biodiesel production, the reduction of energy requirements remains still a very strong incentive.

The problem with waste oils is the very high content of free fatty acids (FFA) that lead to soap formation in a conventional base catalyzed process. Moreover, the homogeneous catalysts require neutralization, washing, separation, recovery, and waste disposal operations with severe economical and environmental penalties.

To solve these problems we propose a novel fatty acids esterification process based on reactive absorption using water-tolerant solid acids as catalysts and therefore eliminating the additional separation steps and the salt waste streams. In this work we selected sulfated zirconia as acid catalyst due to its thermal stability and high activity. Nevertheless, ion-exchange resins (Steinigeweg, 2003) are also suitable due to the moderate temperatures used in this particular process. Previously reported experimental results (Kiss, 2006) are embedded in the process simulations performed in this work.

3. Simulation results

The physical properties of the components present in this process were previously determined experimentally or estimated using state-of-the-art contribution methods such as UNIFAC – Dortmund modified. The ternary map shows that liquid phase split is possible, while the residue curve map (RCM) reveals no azeotropes (Figure 2).

Kinetic data for esterification of dodecanoic acid with methanol is available from previous work (Kiss et al., 2006; Dimian et al., 2009). Since the reverse hydrolysis reaction is negligible, a simple kinetic expression can be used for simulation purposes. The reaction rate is given by $r = k \cdot W_{cat} \cdot C_{Acid} \cdot C_{Alcohol}$, where C_{Acid} and $C_{Alcohol}$ are the mass concentration of reactants and $k = 250 \cdot \exp(-55000/RT)$. The reaction rate could be similar for different catalysts, by changing the weight amount of catalyst used (W_{cat}).
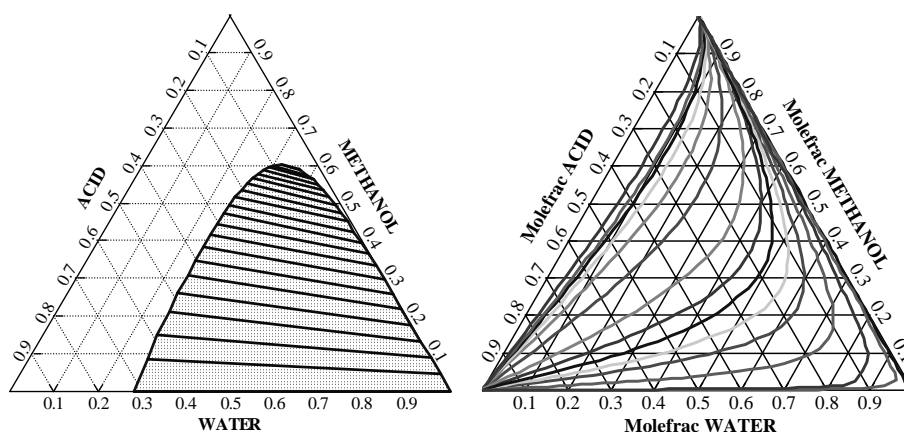
Figure 2. Ternary map (left) and Residue Curve Map (right) for the mixture methanol-acid-water.

Heat-Integrated Process for Biodiesel by Reactive Absorption
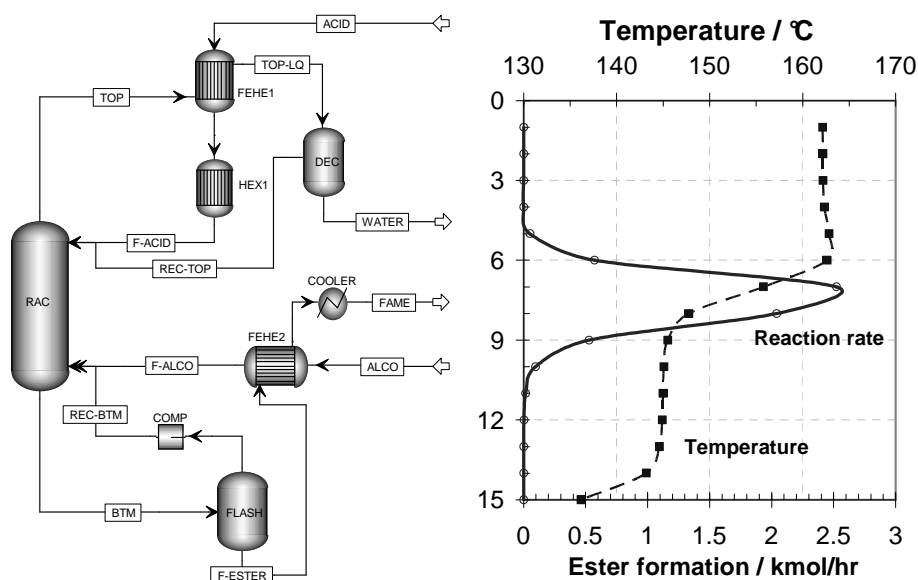
Figure 3. Simulated HI-RA flowsheet (left). Temperature and reaction rate profiles (right).

Figure 3 (left) presents the developed flowsheet of a biodiesel production process based on a reactive absorption column (RAC) as the central operating unit. The reactants feed streams are pre-heated using the product streams. The complete process was simulated in AspenTech Aspen Plus – using the rigorous RadFrac unit with RateSep model for the RAC – then, the sensitivity of the key operating parameters was also evaluated. Figure 3 (right) illustrates the temperature and reaction rate profiles along the column. The reactive separation column is operated in the temperature range of 135–160 °C, at ambient pressure, with a maximum reaction rate in the middle of the reactive zone – rate profile similar to a RD column. The production rate considered for the biodiesel plant designed in this work is 10 ktpy fatty acid methyl esters (FAME). The complete mass and energy balance is given in Table 1, while the main process design parameters – such as column size, catalyst loading, and feed condition – are listed in Table 2.

Table 1. Mass and energy balance of a 10 ktpy biodiesel process based on reactive-absorption.

	F-ACID	F-ALCO	BTM	REC-BTM	REC-TOP	TOP	WATER	FAME
Temperature C	160	65.4	136.2	146.2	51.8	162.1	51.8	30
Pressure bar	1.05	1.05	1.03	1.216	1	1	1	0.203
Vapor Frac	0	1	0	1	0	1	0	0
Mass Flow kg/hr	1166.755	188.306	1261.295	11.295	9.369	114.43	105.061	1250
Volume Flow cum/hr	1.492	157.565	1.417	8.949	0.011	213.042	0.109	1.259
Enthalpy Gcal/hr	-0.94	-0.279	-0.904	-0.013	-0.009	-0.337	-0.395	-0.957
Mass Flow kg/hr								
METHANOL	0	188.304	9.125	7.544	0.002	0.103	0.101	1.581
ACID	1166.739	0	trace	trace	9.218	9.233	0.016	trace
WATER	0	0	trace	trace	0.24	105.166	104.926	trace
ESTER-M	0	0	1252.183	3.764	0.846	0.846	trace	1248.419
Mass Frac								
METHANOL	0	1	0.007	0.667	172 ppm	894 ppm	965 ppm	0.001
ACID	1	0	trace	trace	0.894	0.08	148 ppm	trace
WATER	0	0	trace	10 ppb	0.023	0.912	0.999	trace
ESTER-M	0	0	0.993	0.333	0.082	0.007	513 ppb	0.999

Table 2. Design parameters for simulating the reactive absorption column.

Parameter	Value	Units
Total number of theoretical stages	15	–
Number of reactive stages	10 (from 3 to 12)	–
Column diameter	0.4	m
HETP	0.6	m
Valid phases	VLL	–
Volume liquid holdup per stage	18	L
Mass catalyst per stage	6.5	kg
Catalyst bulk density	1050	kg/m^3
Fatty acid conversion	>99.99	%
Fatty acid feed (liquid, at 160 °C)	1167	kg/hr
Methanol feed (vapor, at 65 °C)	188	kg/hr
Production of biodiesel (FAME)	1250	kg/hr
RA column productivity	19.2	kg FAME /kg cat / hr

High purity products are possible, the purity specifications exceeding 99.9 %wt for the final biodiesel product. The total amount of the (optional) recycle streams is not significant, representing only ~1.5% of the production rate. High conversion of the reactants is achieved, at a remarkable productivity of ~19 kg fatty esters/kg catalyst/h.
Figure 4 shows the molar composition profiles in both liquid and vapor phase. The concentration of the fatty acid and water increases from the bottom to the top of the column, while the concentration of fatty ester and methanol increases from the top to bottom. Consequently, in the top of the reactive separation column there are vapors of water and liquid fatty acids, while in the bottom there are vapors of the methanol feed and liquid fatty esters product (biodiesel). Note that the concentration of methanol in the liquid phase could be further increased by working at higher pressure.
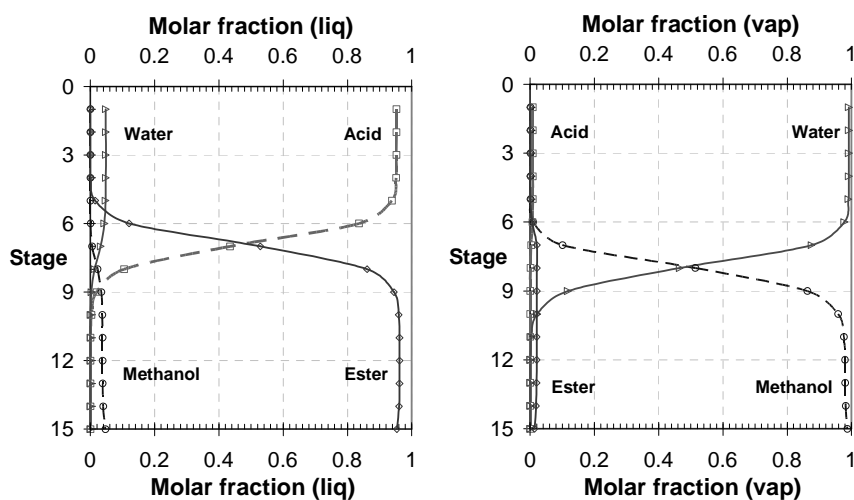
Figure 4. Liquid and vapor composition profiles along the reactive column.

Heat-Integrated Process for Biodiesel by Reactive Absorption
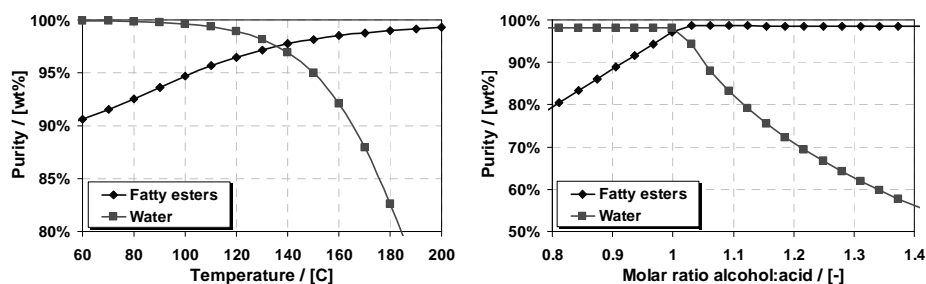
Figure 5. Purity of top (water) and bottom (FAME) products versus temperature of the fatty acid feed stream (left) or versus molar reactants ratio (right).

Afterwards, sensitivity analysis was used as a powerful tool to evaluate the optimal range of the operating parameters: temperature of feed streams, reactants ratio, recycle rates, decanting temperature, flashing pressure. The main results are shown in Figure 5. An optimal range exists for the temperature of the fatty acid feed stream (100-160 °C), in order to obtain concentrated fatty esters (min. 95%) in the bottom of the column. Remarkably, the optimal molar ratio of the reactants (alcohol:acid) is the stoichiometric value of one. In practice, using a very small excess of methanol (~0.5%) or an efficient control structure that can ensure the stoichiometric ratio of reactants, is sufficient for the complete conversion of the fatty acids (Bildea and Kiss, 2009).

Table 3 shows a head-to-head comparison of the novel heat-integrated RA process proposed in this study against the previously reported state-of-the-art RD process (Kiss, 2008, 2009), in terms of number of stages, operating parameters, productivity, as well as heating and cooling requirements – figures that ultimately translate into equipment size and cost. Remarkably, the energy usage is less than 22 kW/ton biodiesel, with significant reduction of the heating (–85%) and cooling (–90%) requirements.

Table 3. Comparison between the integrated reactive-absorption vs reactive-distillation processes.

Equipment / Parameter / Units	RD	HI-RD	RA	HI-RA
Reactive column – reboiler duty (heater), KW	136	136	n/a	**n/a**
HEX-1/FEHE heat duty (fatty acid heater), KW	95	0	108	**27**
HEX-2/FEHE heat duty (methanol heater), KW	8	0	65	**0**
Reactive column – condenser duty (cooler), KW	– 72	– 72	n/a	**n/a**
HEX-3/FEHE water cooler/decanter, KW	– 6	– 6	– 77	**0**
COOLER heat duty (biodiesel cooler), KW	– 141	– 38	– 78	**– 14**
FLASH heat duty (methanol recovery), KW	0	0	0	**0**
Compressor power (electricity), KW	0.6	0.6	0.6	**0.6**
Reactive column, number of reactive stages	10	10	10	**10**
Feed stage number, for acid / alcohol streams	3 / 10	3 / 10	1 / 15	**1 / 15**
Reactive column diameter, m	0.4	0.4	0.4	**0.4**
Reflux ratio (mass ratio R/D), kg/kg	0.10	0.10	n/a	**n/a**
Boil-up ratio (mass ratio V/B), kg/kg	0.12	0.12	n/a	**n/a**
Productivity, kg ester / kg catalyst / h	20.4	20.4	19.2	**19.2**

4. Conclusions

This study presented a heat-integrated biodiesel process based on FFA esterification in a reactive absorption column using solid acids as green catalysts. CAPE tools, such as AspenTech Aspen Plus, were successfully applied in the development and evaluation of this novel process. Sensitivity analysis was also used as a powerful tool to determine the optimal process parameters and explore the effect of the reactants ratio or the feed streams temperature on the purity of products. The most favorable results were obtained near the stoichiometric reactants ratio and relatively high temperature of the fatty acids feed stream. At optimal operation, the highest yield and purity can be achieved by using stoichiometric reactants ratio, with practically negligible amount of methanol lost in top (ppm level) and complete conversion of the fatty acids.

The innovative process proposed in this work significantly improves the biodiesel production and considerably reduces the number of downstream processing steps. Compared to conventional processes, the major benefits of this unique process are:

- Multifunctional plant suitable for a large range of light alcohols and fatty raw materials with very high FFA content, such as frying oils, animal tallow and wvo.
- Simple and robust process with no catalyst-related waste salt streams, no soap formation, and sulfur-free biodiesel as the solid catalysts do not leach into product.
- Elimination of all conventional catalyst-related operations such as: catalyst neutralization, separation and disposal of waste salts, waste water treatment.
- Efficient use of raw materials: stoichiometric reactants ratio, high conversion, no products recycled as reflux or boil-up vapors, and no thermal degradation.
- Effective use of the reactor volume leading to significantly high unit productivity, up to 6-10 higher than conventional biodiesel processes.
- Reduced CapEx and OpEx due to the integrated design with no reboiler or condenser. Compared to similar reactive distillation processes, about 20% savings on the total capital investment and 30% less operating costs are possible.
- Significant reduction of the heating (–85%) and cooling requirements (–90%).

5. Acknowledgements

References

C. S. Bildea, A. A. Kiss, 2009, Plantwide Control of a Biodiesel Process by Reactive Absorption.

A. C. Dimian, C. S. Bildea, F. Omota, A. A. Kiss, 2009, Innovative Process for Fatty Acid Esters by Dual Reactive Distillation, Computers & Chemical Engineering, 33, 743-750.

M. G. Kulkarni, A. K. Dalai, 2006, Waste cooking oil-an economical source for biodiesel: A review, Industrial & Engineering Chemistry Research, 45, 2901-2913.

A. A. Kiss, A. C. Dimian, G. Rothenberg, 2006, Solid acid catalysts for biodiesel production - towards sustainable energy, Advanced Synthesis & Catalysis, 348, 75-81.

A. A. Kiss, G. Rothenberg, A. C. Dimian, F. Omota, 2006, The heterogeneous advantage: biodiesel by catalytic reactive distillation, Topics in Catalysis, 40, 141-150.

A. A. Kiss, A. C. Dimian, G. Rothenberg, 2008, Biodiesel by Reactive Distillation Powered by Metal Oxides, Energy & Fuels, 22, 598-604.

A. A. Kiss, 2009, Novel Process for Biodiesel by Reactive Absorption, Separation & Purification Technology, 69, 280-287.

S. Steinigeweg, J. Gmehling, 2003, Esterification of a fatty acid by reactive distillation, Industrial & Engineering Chemistry Research, 42, 3612-3619.

Robust Thermodynamically-guided Algorithms for Synthesis of Energy Efficient Separation Networks

Gerardo J. Ruiz[1], Seon B. Kim[1], Daniel A. Beneke[2] and Andreas A. Linninger[1]

[1]*Department of Chemical Engineering, University of Illinois at Chicago, 851 S. Morgan St., Chicago, Illinois 60607, USA, linninge@uic.edu*
[2]*Department of Chemical Engineering,University of the Witwatersrand, 1 Jan Smuts Ave, Johannesburg, 2050, South Africa*

Abstract

Distillation has a particularly high potential for energy savings, as it accounts for around 40-70% of capital and operating costs in petrochemical and commodity industries. However, numerous existing techniques that have been proposed to facilitate the design often lack the robustness and reliability needed to rigorously solve the problem, largely due to simplifying assumptions. In this contribution, we present computational methods which exploit a novel thermodynamically motivated problem transformation, entitled temperature collocation, rather than using classical tray-by-tray models which fail in pinched regions. This novel methodology applies to ideal, non-ideal and azeotropic mixtures, and is independent of the number of components. Due to the generalised nature of the method, we have also addressed the synthesis of complex column configurations as well as heat integration of distillation trains. Furthermore, the robust synthesis algorithm is presented, a method that automatically synthesizes a distillation network for given product purity requirements. Entire separation flowsheets are generated with rigorous thermodynamic models without the need to introduce limiting simplifications as is the case with existing shortcut techniques. The computational approach guarantees to identify realizable columns with a finite number of trays and operating conditions, which may easily be validated with industrial flowsheet simulator software packages like *Aspen*Plus.

Keywords: Difference point equation, complex column, temperature collocation

1. Introduction

Distillation is one of most widely used and versatile separation techniques used in the petrochemical and commodity industry. It currently constitutes 40-70% of capital and operating costs of chemical manufacturing, and accounts for more than 60% of the total process energy for the manufacture of commodity chemicals [1]. Due to the low price of energy and less stringent environmental standards in previous years, mainly simple distillation column configurations were built. However, the recent rise in energy costs and environmental concerns have reshaped the design priorities for industrial separations with a new focus on energy conservation.

Numerous techniques, such as the Rectification Body Method and the Underwood Method, have been proposed to facilitate the design of distillation systems. However, these techniques are not properly equipped for solving industrial problems, as the simplifying assumptions lead idealizations which are inadequate for rigorous design. Recently, computer-aided separation synthesis methods for energy savings in distillation have been revisited critically, specifically that substantial energy savings can be

achieved by designing the entire network simultaneously [2-3]. Other options with high potential include complex column configurations which have the potential of achieving up to 70% energy savings over simple column networks [4]. The well studied Petlyuk complex column configuration is an excellent example to realize energy efficient separations. With full thermal coupling, this complex column configuration may achieve energy savings of up to 40% by reducing the numbers of reboilers and condensers.

In this work, we have developed a set of temperature collocation based methodologies to design distillation column networks, including simple and complex column configurations. The method utilizes a thermodynamic variable transformation which enables it to successfully solve columns design problems, even near pinch and saddle points, with acceptable accuracy. We also present some cases where some current design techniques like the Rectification Body Method fails as a sufficient condition to determine feasible designs. The key novelty of our procedure exploits a view of physically meaningful design variables to distinguish between feasible regions and unattainable composition ranges. The method is shown to be effective in constant relative volatility, ideal and even non-ideal equilibrium models. Furthermore, we present a case study for the separation of a quaternary mixture into high purity products using an energy efficient complex column network configuration.

2. Current Design Techniques

In distillation problems, it is frequently of interest to the designer to solve the inverse design problem, i.e. for a given set of product purities, determine operating and structural conditions that would render a feasible design. It is generally agreed that for a design to be feasible that the liquid composition profiles of adjacent columns sections intersect. The only way to ascertain whether the aforementioned condition is met is by rigorous simulation using the MESH equations. However, if the product purities and other column operating parameters are not set properly, it may cause the program to not converge and hence, it is impractical to use the MESH equations as a rigorous feasibility criterion.

Numerous papers have been published on shortcut methods for column design, and specifically the Rectification Body Method (RBM) has been particularly popular, having been applied to complex as well as reactive distillation column design. The method was developed by Bausa et al. [5] as a simplified means of determining minimum energy requirements for a proposed split involving non-ideal and azeotropic n-component mixtures. It requires solving the pinch point equation and connecting the resulting saddle and stable node pinch compositions with the product composition to form a rectification body. A design is then considered to be feasible if rectification bodies of adjacent column sections intersect. Although this method is useful for quickly estimating a distillation column's key operating conditions, it has several limitations, some of which are highlighted here:

Profiles intersect, but rectification bodies do not. One of the major assumptions of the method is that the rectification bodies are linear. As discussed by [6], the two major consequences of this assumption are that the estimation of minimum reflux will be inaccurate and that the method is unsuitable for use as a feasibility test. This shortcoming is illustrated in Figure 1 (a) for the Non-ideal mixture of Acetone/Benzene/Chloroform, which has a particularly curved distillation boundary. Notice in this case that the composition profiles intersect, rendering a feasible design, but the rectification bodies do not.

Robust Themodynamically-giuded Algorithms for Synthesis of Energy Efficient Networks

Rectification bodies intersect, but profiles do not. The simplified ideal cases shown in Figure 1 (b) and (c) shows that the rectification bodies intersect for a quaternary mixture, suggesting that the design is feasible, but liquid composition profiles show no intersection and thus an infeasible design.

Absence of stationary points. Another assumption of the method is that the Pinch Point Equation always has real solutions, and hence it is always possible to construct a rectification body. Although the design of simple columns will always have real solutions, it is entirely possible for the Pinch Point Equations to have complex roots when designing complex columns. Hence it is impossible to construct a rectification body to determine feasibility, as for example, a ternary system will only will only have 1 real root, as shown in Figure 1(d) for a constant relative volatility system.

In short, it may be concluded that the RBM is a useful tool for estimating column feasibility and key operating parameters, but it lacks the rigour needed to assess column feasibility for all cases. It has been shown that the RBM is neither a necessary nor sufficient condition for finding a feasible design, and hence an impractical method for rigorous distillation design.
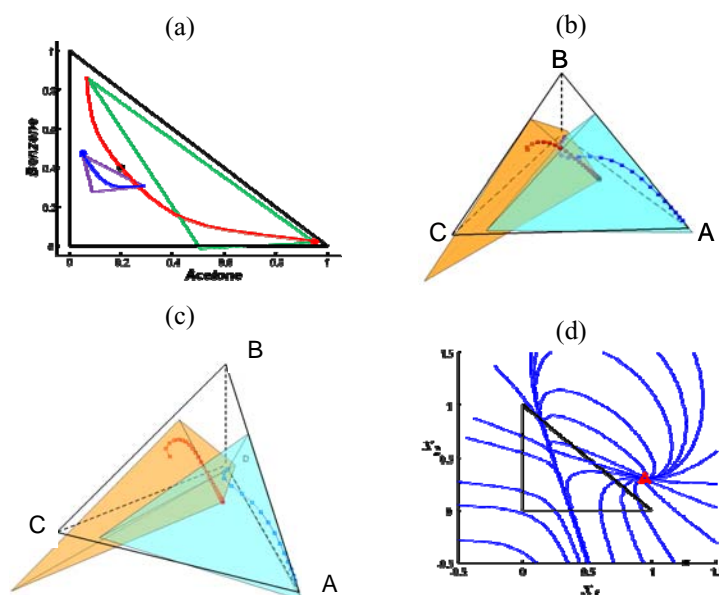
Figure 1 a-d: An illustration of several limitations of the Rectification Body Method.

3. Methodology

As evident from the preceding section, simplified design methods are not suited for rigorous column design. Instead, we propose searching for intersecting liquid composition profiles in adjacent Column Sections to guarantee feasibility. Our approach towards a general solution of the separation network synthesis problem has three major components: (i) systematic network generation, (ii) thermodynamic problem transformation, and (iii) advanced algorithm engineering to seek simultaneously in the structural and parametric design space. We intend to combine network generation methods such as those proposed by [7] with a robust global search in the bubble point

distance space to systematically identify optimum distillation schemes for given product purity targets. The solution seeks to determine both structural as well as parametric design decisions for multi-component feed and product streams columns with different vapor-liquid interactions. Furthermore, we propose the use of the inverse design procedure, whereby the designer sets product targets (purities or flowrates) from which all structural and operational variables are determined. Our methodologies are summarized in the following subsections:

3.1. Systematic Network Generation

Presently, our network generation method comprises two stages:

Step 1. Generic structure synthesis: In the first step, only structural information is used, thus yielding identical networks for any mixture of the same number of components. Structures may be generated using techniques described by [7].The superstructure could be enumerated exhaustively or incorporate equipment constraints when addressing retrofit problems with existing column inventory.

Step 2. Network Task Optimization: Here, rigorous optimization of all candidate solutions in their respective parametric design spaces is used to find the globally optimal structure. Network optimization task should take into account the specific feed composition and product targets as well as the specific thermodynamic properties of the chemical species to be separated.

3.2. Thermodynamic Problem Transformation

Thermodynamically transforming the column stage number, *n*, into the equilibrium bubble point temperature, *T*, offers massive size reductions in the column profile computations and is a key feature in a rigorous feasibility criterion based on minimum bubble point distance functions. The liquid profile equations in terms of temperature can be conveniently solved with finite element methods of orthogonal polynomials. Eq. (1) shows the composition profiles parameterized using temperature instead of stage height.

$$\frac{\partial x_j}{\partial T} = -\left[\left(1+\frac{1}{R_\Delta}\right)\left(x_j - y_j\right) + \frac{1}{R_\Delta}\left(X_{\Delta j} - x_j\right)\right] \frac{\sum\limits_{j=1}^{c}\frac{\partial K_j}{\partial T}x_j}{\sum\limits_{j=1}^{c}\left\{\left[\left(1+\frac{1}{R_\Delta}\right)\left(x_j - y_j\right) + \frac{1}{R_\Delta}\left(X_{\Delta j} - x_j\right)\right]K_j\right\}} \quad (1)$$

Where $X_\Delta = (VY^T - LX^T)/(V-L)$, $R_\Delta = L/(V-L)$, and K is an equilibrium constant which relates vapour and liquid compositions such that $y_i = K_i x_i$. X_Δ is known as the Difference Point which relates passing vapor and liquid streams in a Column Section and R_Δ is a generalised reflux ratio.

3.3. Advanced algorithm engineering

The advanced hybrid algorithms that we have adopted aim to combine advantages of a Newton-type technique with stochastic search techniques. Newton methods are particularly useful to satisfy constraints such as a large set of state variables, while open design variables such as structural decisions which change the problem topology and equations are better suited for stochastic methods. This method allows for solving the minimum reflux for n-components, even for highly non-ideal mixtures.

4. Separation of Quaternary Mixtures

In mixtures with four or more components, the design problem involves the search for one or more compositional degree of freedom of any product stream. This multidimensional global optimization problem is solved using a hybrid method that combines a stochastic genetic algorithm (i.e. product specification) with rigorous finite

Robust Themodynamically-giuded Algorithms for Synthesis of Energy Efficient Networks

element collocation of column profiles [8]. A complex network for the separation of an equal molar mixture of Pentane, Hexane, Heptane, and Octane into products with purity of at least 99% uses two simple columns and one complex column as shown in Figure 2(a). The composition profiles through the equivalent column equilibrium tray for all columns are showed in Figure 2 (a).

(a): x_i vs. N stage from *T collocation* | (b): x_i vs. N stage from *Aspen*Plus
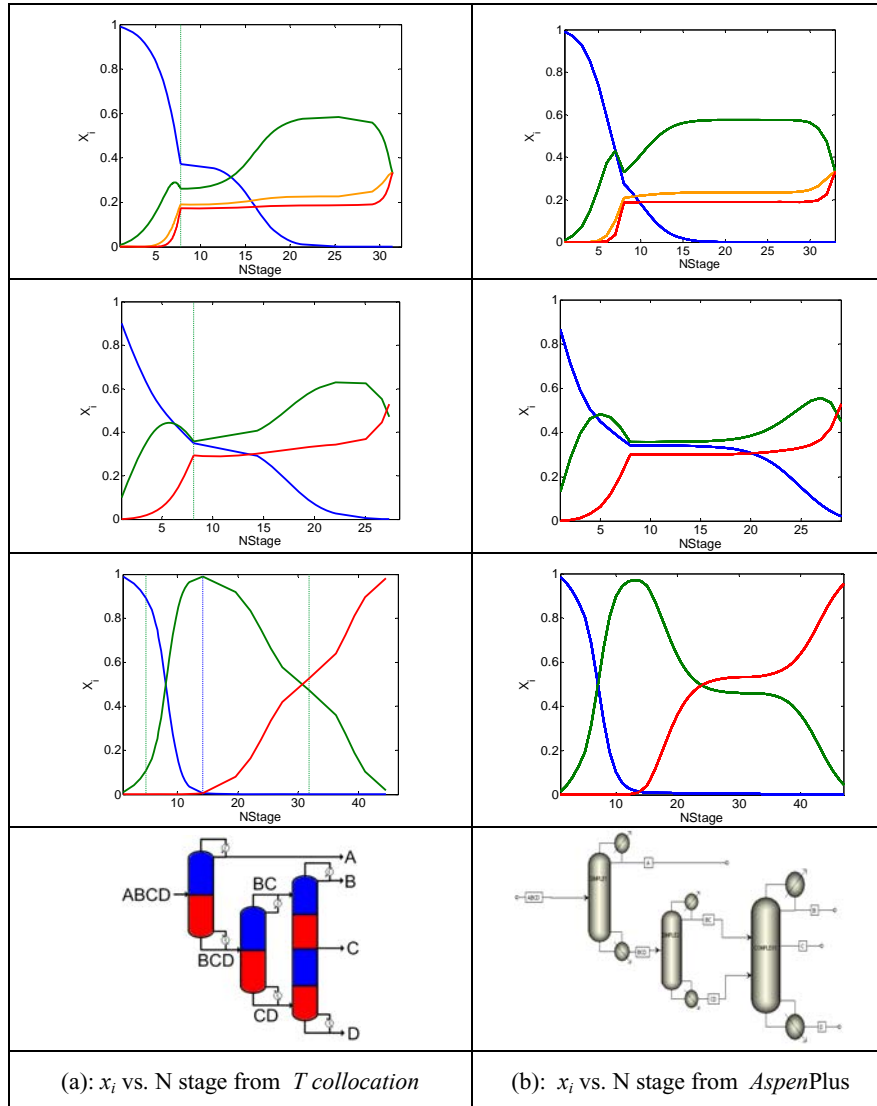
Figure 2. Initialization of *Aspen*Plus by the *Temperature Collocation* approach. Files (a) and (b) describe the profiles of the entire network in the stage number space for each column in the network found by *Temperature Collocation* and *Aspen*Plus respectively.

Once composition profiles have been calculated as functions of temperature, the column height can easily be recovered with the help of eq. (2). To complete the capital cost estimation, column height, feed stages, and product stages have been computed. Regarding the total cost of the complex network analyzed, we estimate that it saves up

to 70% of total vapor rate compared to the simple column configurations. This case study illustrates a feasible solution to the synthesis problem using a complex network, and potential energy savings realizable with complex columns. Also, shows the rigorous *Aspen*Plus simulation results initialized with our design solution. When using the predictions of our inverse design as initial guesses for *Aspen* performance calculation as described in Figure 2(f), the flowsheet simulator converged typically in a couple iterations of a few CPU seconds, thus confirming an essential agreement between the two computational results

$$\frac{dn}{dT} = -\frac{\sum_{i=1}^{c} \frac{\partial K_i}{\partial T} x_i}{\sum_{i=1}^{c} \left\{ \left[\left(1 + \frac{1}{R_\Delta} \right)(x_i - y_i) + \frac{1}{R_\Delta}(X_{\Delta i} - x_i) \right] K_i \right\}} \qquad (2)$$

5. Conclusions

In this article, we advocated the recent progress for synthesizing energy efficient complex distillation networks and have demonstrated some of the shortcomings of existing methods when rigorously assessing separation feasibility. An approach towards a generalized solution of the separation synthesis problem is presented, using the generalized difference point equation with temperature as an independent variable, instead of tray number. This method was employed to find composition profiles of complex column sections and the entire complex network. The temperature collocation and minimum bubble point distance algorithms were effective to find a feasible separation by intersecting profiles.

A novel hybrid synthesis procedure that generates structure synthesis networks with rigorous column profile computations for separating a mixture of given composition into desired product targets using an evolutionary algorithm and rigorous feasibility test was implemented, which showed the complete feasible design of complex networks to separate a quaternary mixture. Subsequently, these optimal feasible designs were validated by the initialization of *Aspen* flowsheet diagrams that converged in one or two iterations, showing the capacity and certainty of our hybrid synthesis methodology applied to the design of feasible complex network configurations.

6. Acknowledgements

References

1. DOE. *Technical Topic Description*. 2005; Available from: www.doe.gov.
2. Leboreiro, J. and J. Acevedo. Comp. & Chem. Eng., 2004. **28**(8): p. 1223-1236.
3. Linninger, A.. Comp. & Chem. Eng., 2009. **33**(12): p. 2018-2027.
4. Hilde K. Engelien, S.. AIChE J., 2005. **51**(6): p. 1714-1725.
5. Bausa, J., R. Watzdorf, and W. Marquardt. AIChE J., 1998. **44**(10): p. 2181-2198.
6. Thong, D. and M. Jobson. Chem. Eng. Sci., 2001. **56**(14): p. 4369-4391.
7. Agrawal, R.. AIChE J., 2003. **49**(2): p. 379-401.
8. Zhang, L. and A. Linninger. AIChE J., 2006. **52**(4): p. 1392-1409.

Process Synthesis under Uncertainty via Multi-parametric Programming

Luis F. Domínguez and Efstratios N. Pistikopoulos

Centre for Process Systems Engineering, Department of Chemical Engineering, Imperial College London, SW7 2AZ, UK, e.pistikopoulos@ic.ac.uk.

Abstract

In this work we propose a multi-parametric mixed-integer quadratic-approximation algorithm for the solution of convex multi-parametric mixed-integer nonlinear programming problems arising in process synthesis under uncertainty. The algorithm follows a decomposition procedure where a primal sub-problem is solved using multi-parametric nonlinear programming techniques and a master sub-problem is solved using a mixed-inter-nonlinear programming formulation. An example problem is presented to illustrate the proposed algorithm.

Keywords: Process synthesis, multi-parametric mixed-integer quadratic approximation, multi-parametric mixed-integer nonlinear programming.

1. Introduction

Mathematically, process synthesis problems (see Fig. 1) can be posed as a mixed-integer nonlinear programming (MINLP) problems [1, 2]:

$$z = \min_{x,y} f(x, y)$$
$$s.t. \quad g(x, y) \leq 0 \tag{1}$$
$$x \in X; y \in \{0,1\}^m$$

where f denotes a cost function to be maximized or minimized, g denotes the problem restrictions such as boundaries in quality specifications or safety limits, x is a vector of continuous decisions such as operating points in temperature or pressure and y is a vector of discrete decisions denoting the existence (1) or nonexistence (0) of process units.
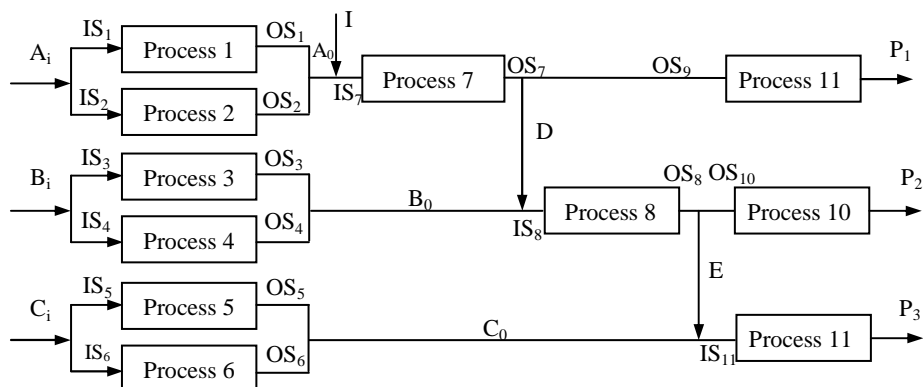
Figure 1. Process Synthesis problem [3]

The inevitable presence of fluctuations in some variables such as demand, prices or quality in raw materials, and the lack of exact knowledge of certain parameters such as kinetic data or physical properties transforms the synthesis problem, Eq. 1, into the following multi-parametric mixed-integer nonlinear programming problem (mp-MINLP)[4, 5]:

$$
\begin{aligned}
z(\theta) = \min_{x,y} \; & f(x) + d^T y \\
s.t. \quad & g(x) + Ey \leq b + F\theta \\
& \theta^{\min} \leq \theta \leq \theta^{\max} \\
& x \in X; y \in \{0,1\}^m; \theta \in \Theta
\end{aligned}
\tag{2}
$$

where θ is a vector of uncertain parameters, b and d are constant vectors and E and F are constant matrices.

Multi-parametric programming [6, 7, 8] is a natural tool for analyzing the effects of uncertain parameters in the solution of parametric problems. Using a multi-parametric programming framework, algorithms have been developed for addressing process synthesis problems under uncertainty. [4] proposed an mp-MINLP algorithm based on the outer-approximation/equality relaxation algorithm of [9] whereas [3] proposed three algorithms based on the reformulation of a master problem using (i) deterministic, (ii) outer-approximation (OA), and (iii) generalized benders decomposition (GBD) principles. While these algorithms proved to be successful they remain computationally intensive. In this work we propose a multi-parametric mixed-integer quadratic approximation (mp-MIQA) algorithm for the solution of mp-MINLP problems arising in process synthesis under uncertainty. The proposed algorithm is based on the recent advances in multi-parametric nonlinear programming (mp-NLP) [10]. In the next section the proposed algorithm is presented.

2. An algorithm for mp-MINLP

Consider the mp-MINLP problem Eq. 2. Assuming the optimization variables and the parameters are both convex and separable [11], a way of approaching this problem is by decomposing it into two sub-problems: (i) a primal and (ii) master sub-problems.

2.1. Primal sub-problem

Before the primal sub-problem is solved, an integer solution must be found. An initial integer solution can be obtained by considering the parameter vector, θ, as optimization variable and solving a MINLP formulation of Eq. 2, as follows:

$$
\begin{aligned}
z = \min_{x,y,\theta} \; & f(x) + d^T y \\
s.t. \quad & g(x) + Ey \leq b + F\theta \\
& \theta^{\min} \leq \theta \leq \theta^{\max} \\
& x \in X; y \in \{0,1\}^m; \theta \in \Theta
\end{aligned}
\tag{3}
$$

Fixing y at the integer solution point found in Eq. 3, $\bar{y} = y^*$, transforms the mp-MINLP problem into the following mp-NLP:

Process Synthesis under Uncertainty via Multi-parametric Programming

$$z(\theta) = \min_{x} f(x) + d^T \overline{y}$$
$$\text{s.t. } g(x) + E\overline{y} \leq b + F\theta \tag{4}$$
$$\theta^{\min} \leq \theta \leq \theta^{\max}$$
$$x \in X; \theta \in \Theta$$

A quadratic approximation of the objective function and an outer approximation of the constraints can then be constructed at the solution point found in Eq. 3 transforming Eq. 4 into the following multi-parametric quadratic programming (mp-QP) problem:

$$z_q(\theta) = \min_{x} f(x^*) + \nabla f(x^*)(x - x^*) + 1/2\nabla^2 f(x^*)(x - x^*)^2 + d^T \overline{y}$$
$$\text{s.t. } g(x^*) + \nabla g(x^*)(x - x^*) + E\overline{y} \leq b + F\theta \tag{5}$$
$$\theta^{\min} \leq \theta \leq \theta^{\max}$$
$$x \in X; \theta \in \Theta$$

where ∇f and $\nabla^2 f$ represents the gradient and second gradient of the objective function and ∇g represents the gradient of the constraints.

Eq. 5 can be solved efficiently, for instance, by using the mp-QP algorithm of [6], the solution of which provides a set of parametric optimizers $x(\theta)$ and the critical regions (*CR*) in the space of the parameters where they are valid.

On the other hand, due to the nonlinearity of the constraints some vertices of the critical regions obtained by the mp-QP algorithm may be infeasible. If that is the case, a feasible point must be found. The closest feasible point, $\theta \in CR^i$, can be obtained by solving the following NLP problem as shown in [3]:

$$\min_{x,\theta} \delta^2$$
$$\text{s.t. } g(x) + E\overline{y} \leq b + F\theta \tag{6}$$
$$\delta = \theta - \theta^{\inf}$$
$$x \in X; \theta \in CR^i$$

where $\theta^{\inf}$ corresponds to an infeasible vertex and δ is a vector of free variables.

Nevertheless, a single linearization of the nonlinear constraints may not be enough to avoid infeasibilities in the solution provided by the mp-QP algorithm. Therefore, it may be necessary to accumulate all linearization points, x^* until no infeasible vertices are found in the parametric solution of Eq. 5. At this point another mp-QP problem can be solved:

$$z_q(\theta) = \min_{x} f(x^*) + \nabla f(x^*)(x - x^*) + 1/2\nabla^2 f(x^*)(x - x^*)^2 + d^T \overline{y}$$
$$\text{s.t. } g(x^*) + \nabla g(x^*)(x - x^*) + E\overline{y} \leq b + F\theta \tag{7}$$
$$g(x_\omega^*) + \nabla g(x_\omega^*)(x - x_\omega^*) + E\overline{y} \leq b + F\theta, \quad x_\omega^* \in X_k^*, \; k = 1,...,K$$
$$x \in X; \theta \in \Theta$$

where X_k^* is the feasible sets of optimizers, x_ω^* for each unfeasible vertex θ_ω $\omega=1...,\Omega_k$ found at iteration k.

Although infeasibilities may have been eliminated at this point, the quadratic approximation in Eq. 7 may still produce large errors at the vertices of the critical regions of the mp-QP solution (see Figure 2a).

(a) Initial quadratic approximation   (b) Quadratic approx. after partitioning
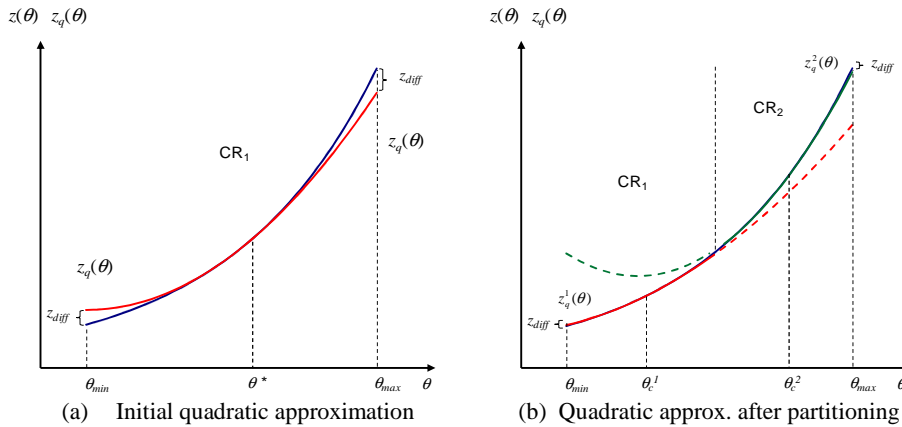
Figure 2. Progress of the quadratic approximation algorithm

If that is the case, another approximation point must be obtained. It has been shown [12, 13] that the best point to approximate an optimal value function is at the center of its critical region. Because of the convexity and polyhedral properties of the critical regions obtained by the mp-QP algorithm, it is possible to compute their center point by solving the following linear program (LP):

$$\theta_c, r = \arg\max_{\theta, r} r$$
$$s.t. \quad -F_i\theta + r\|F_i\| \le b_i, \quad \forall i = 1...I \tag{7}$$
$$\theta \in CR^j$$

At this point another check is made at the vertices of the critical regions produced by the new approximation. If the current quadratic approximation does not approximate the value function at the vertices of CR^i, $\forall i = 1,..,I$, with sufficient accuracy, i.e. $|z(\theta) - z_q(\theta)| \ge \varepsilon$, then CR^i can be split at its center and another mp-QP, Eq. 5, can be solved at this point. Figure 2b shows the progression of the quadratic approximation algorithm after partitioning an initial critical region.

Finally, parametric upper bounds, $\bar{z}$, CR^i $i=1,..,I$, can be obtained by substituting each parametric solution into the objective function of Eq. 2.

2.2. Master sub-problem

In order to obtain the integer solution for the next iteration, the following MINLP is solved for each of the critical regions obtained in the primal sub-problem:

$$\bar{z} = \min_{x,y,\theta} f(x) + d^T y$$
$$s.t. \quad g(x) + Ey \le b + F\theta \tag{8}$$
$$\left.\begin{array}{l} f(x) + d^T y - \bar{z}_q^{ik}(x(\theta),\theta) \le 0 \\ \sum_{j \in J^{ik}} y_j^{ik} - \sum_{j \in L^{ik}} y_j^{ik} \le |J^{ik}| - 1 \end{array}\right\} \quad k = 1,...,K^i$$
$$x \in X; \ y \in \{0,1\}; \ \theta \in CR^i$$

where $J^{ik} = (j/y_j^{ik} = 1)$ and $L^{ik} = (j/y_j^{ik} = 0)$, $|J^{ik}|$ is the cardinality of J^{ik} and K^i is the number of integer solutions that have been analyzed in CR^i. The inequalities $f(x,y,\theta) -$

$z_q^{ik}(\theta) \leq 0$ and $\Sigma_{j \in J^{ik}} y_j^{ik} - \Sigma_{j \in L^{ik}} y_j^{ik} \leq |J^{ik}| - 1$ correspond to parametric and integer cuts respectively. While the former constraint restricts the solution of Eq. 8 from taking on values which are higher than the current upper bound, $z_q^{ik}(\theta)$, the latter, restricts Eq. 8 from taking on integer values that have already been analyzed.

The parameter space, valid for the next integer solution, can be obtained following the procedure described in [14]. Consider the sets of inequalities describing the integer solutions that have been obtained, (CR^{IS}), the rest of the parametric space (CR^{Rest}) for which a new integer solution can be encountered is obtained by removing the intersection of the set of inequalities describing CR^{IS} and a given region (CR^{GR}) as follows:

$$CR^{Rest} = CR^{GR} - CR^{IS} \qquad (9)$$

Finally, by fixing the integer variables at the solution point found by the master problem in the rest of the regions, another quadratic approximation of the form of Eq. 5 can be performed. The algorithm terminates when there are no more integer solutions which are better than the already found in each region.

3. Process Synthesis Example

The following example problem [6] is concerned with obtaining the optimal configuration of a chemical plant for the processing of optimal amounts of raw materials A, B and C and an intermediate I in the face of uncertainty in the demand of three products: P_1, P_2 and P_3. An initial superstructure, which includes two alternatives of producing each product, is shown in Figure 1. The mathematical model of the process is as follows:

$$z(\theta) = \min_{x,y} \sum_{i=1}^{6} FC_i y_i + \sum_{i=7}^{11} FC_i + \sum_{i=1}^{11} OC_i IS_i^2 + 1.2A + 1.5B + 1.8C - 50P_1 - 60P_2 - 68P_3$$

$$s.t. \quad -IS_i + KC_i \exp(OS/PC_i) \leq KC_i, \quad i = \{1,..,6\}$$

$$\left.\begin{array}{l} -PC_i IS_i + OS_i = 0 \\ IS_i - MI_i y_i \leq 0 \end{array}\right\} i = \{1,...,11\}$$

$$\begin{array}{l|l} Ai = IS_1 + IS_2 & IS_7 = A_0 + I \\ A_0 = OS_1 + OS_2 & OS_7 = D + IS_9 \\ B_i = IS_3 + IS_4 & IS_8 = D + B_0 \\ B_0 = OS_4 + OS_4 & OS_8 = E + IS_{10} \\ C_1 = IS_5 + IS_6 & IS_{11} = E + C_0 \\ C_0 = OS_5 + OS_6 & \\ A_i \leq 1.5; & OS_9 = P_1 = \theta_1 \\ B_i \leq 1.6 & OS_{10} = P_2 = \theta_2 \\ C_i \leq 1.7 & OS_{11} = P_3 = \theta_3 \end{array}$$

$$\begin{array}{l|l} & MI = \{4, 3, 3.5, 5.5, 7.5, 6, 8, 6, 5, 5, 5\} \\ 0.80 \leq \theta_1 \leq 0.85 & PC = \{1.8, 2.0, 2.5, 2.2, 2.8, 3.0, 0.65, 0.9, 0.72, 0.68, 0.71\} \\ 0.75 \leq \theta_2 \leq 0.80 & KC = \{2.0, 2.1, 2.6, 2.1, 2.5, 3.0\} \\ 1.05 \leq \theta_3 \leq 1.10 & FC = \{1.9, 2.0, 3.0, 3.2, 3.7, 3.9, 1.1, 1.0, 5.0, 5.3, 6.2\} \\ & OC = \{1.1, 1.3, 1.6, 1.5, 2.1, 1.9, 1.7, 1.0, 3.0, 3.7, 4.1\} \end{array}$$

$$(8)$$

As shown above, the mathematical model involves 41 equations, 6 of which are nonlinear, 31 continuous variables, 6 binary variables, and 3 uncertain parameters.

The application of the mp-MIQA algorithm results in a single integer vector and the parametric solution shown in Table 1.

Table 1. Process synthesis example – multi-parametric solution

y	$x(\theta)$		CR
$(0,0,0,1,1,1,1,1,1,1,1)$	$IS_3 = 1.035\,P_2 - 0.106$	$OS_3 = 0.779\,P_2 - 0.011$	$-P1 \leq -0.8$
	$IS_4 = 1.120\,P_2 - 0.115$	$OS_4 = 0.855\,P_2 - 0.011$	$P1 \leq 0.85$
	$IS_5 = 0.860\,P_3 - 0.116$	$OS_5 = 0.722\,P_3 + 0.007$	$-P2 \leq -0.75$
	$IS_6 = 0.88\,P_3 - 0.119$	$OS_6 = 0.686\,P_3 - 0.007$	$P2 \leq 0.8$
	$IS_7 = 2.137\,P_1$	$OS_7 = 1.388\,P_1$	$-P3 \leq -1.05$
	$IS_8 = 1.634\,P_2$	$OS_8 = 1.470\,P_2$	$P3 \leq 1.1$
	$IS_9 = 1.389\,P_1$	$B_0 = 1.634\,P_2$	
	$IS_{10} = 1.470\,P_2$	$B_1 = 2.154\,P_2 - 0.222$	
	$IS_{11} = 1.408\,P_3$	$C_0 = 1.408\,P_3$	
		$C_1 = 1.739\,P_3 - 0.235$	
		$I = 2.137\,P_1$	

4. Concluding remarks

We have presented a multi-parametric mixed-integer quadratic-approximation algorithm for the solution of convex multi-parametric mixed-integer nonlinear programming problems arising in process synthesis problems under uncertainty. The proposed algorithm is based on a quadratic approximation of the objective function and outer approximations of the nonlinear constraints and was illustrated with an example. Our current work focuses on extending the above algorithm to the non-convex mp-MINLP case.

5. Acknowledgements

References

[1] L. T. BIEGLER and I. E. GROSSMANN, *Computers & Chemical Engineering* **28**, 1169 (2004).

[2] I. E. GROSSMANN and L. T. BIEGLER, *Computers & Chemical Engineering* **28**, 1193 (2004).

[3] V. DUA and E. N. PISTIKOPOULOS, *Ind. Eng. Chem. Res.* **38**, 3976 (1999).

[4] J. ACEVEDO and E. PISTIKOPOULOS, *Ind. Eng. Chem. Res.* **35**, 147 (1996).

[5] J. ACEVEDO, PhD thesis, Imperial College London, UK, 1996.

[6] V. DUA, N. A. BOZINIS, and E. N. PISTIKOPOULOS, *Computers and Chemical Engineering* **26**, 715 (2002).

[7] E. N. PISTIKOPOULOS, M. C. GEORGIADIS, and V. DUA, *Multi-Parametric Programming*, volume 1, WILEY-VCH, Weinheim, 2007.

[8] E. N. PISTIKOPOULOS, *AIChE Journal* **55**, 1918 (2009).

[9] G. R. KOCIS and I. E. GROSSMANN, *Industrial & Engineering Chemistry Research* **26**, 1869 (1987).

[10] L. F. DOMÍNGUEZ, D. A. NARCISO, and E. N. PISTIKOPOULOS, *Computers & Chemical Engineering* **doi:10.1016/j.compchemeng.2009.10.012** (2009).

[11] C. A. FLOUDAS, *Nonlinear and mixed integer optimization*, Oxford University Press, New York, 1995.

[12] T. A. JOHANSEN, *Proc. of IEEE Conf. Decision and Control, Las Vegas, NV*, pp. 12–5, 2002.

[13] L. F. DOMÍNGUEZ and E. PISTIKOPOULOS, Technical report, Imperial College London, 2009.

[14] V. DUA and E. N. PISTIKOPOULOS, *Annals of Operations Research* **99**, 123 (2000).

Surrogate-Based Process Synthesis

Carlos A. Henao, Christos T. Maravelias.

Department of Chemical and Biological Engineering University of Wisconsin,
1415 Engineering Drive, Madison – WI 53706, christos@engr.wisc.edu>

Abstract

Superstructure optimization-based process synthesis is generally regarded as theoretically powerful; however, it has not been widely used in practice since it typically results in large-scale non-convex Mixed-Integer Non-Linear Programs (MINLP) which are very hard to solve effectively. To address this limitation, we propose a framework leading to substantially simpler formulations through the replacement of complex first-principle unit models by compact and yet accurate surrogate models. We show how all the relevant variable relationships established by a unit model, can be expressed in terms of a subset of the original model variables. We discuss how this subset of variables can be identified, and we present a method to develop high quality surrogate models through artificial neural networks. Finally, we propose a tailored surrogate model reformulation to incorporate binary variables that allow activation/deactivation of particular units within the superstructure model. An example is presented to illustrate the application of the proposed framework.

Keywords: Process synthesis, Process optimization, Surrogate models.

1. Introduction

Methodologies for the synthesis of chemical processes mainly belong to one of two different groups: the more traditional sequential-conceptual methods, and the more recent and systematic superstructure optimization-based techniques. In theory, the latter approach is more powerful because, as opposed the sequential approach, it considers a large number of process alternatives and it involves the simultaneous determination of the optimum structure and operational conditions, taking into account all complex interactions between engineering decisions. However, this approach has its own challenges due to the mathematical complexity of the resulting optimization model, typically, a large-scale non-convex mixed-integer non-linear program (MINLP). Since most of this complexity comes from the use of detailed unit models, we present a framework where simpler superstructure MINLP formulations are obtained by using unit surrogates models.

Some authors (Jones, et al. 1998, Won and Ray 2005) have proposed the use surrogate models in engineering design to treat important aspects such as global optimality. This approach has been motivated by the existence of highly accurate although computationally expensive computer programs which capture the behavior of particular engineering systems. In the case of chemical process engineering, a commercial process simulator can be used to generate a set of simulation cases whose values are then fitted using a general purpose multivariable mapping.

Among the most popular techniques to generate such surrogates we find Kriging and Artificial Neural Networks (ANN) (Webster and Oliver 2007, Haykin 1999). Caballero and Grossmann (2008), and Davis and Ierapetritou (2007) have applied Kriging-based models to the solution of special classes of MINLP's in the synthesis-optimization of chemical process. The literature also presents several works applying ANNs to process

modeling and optimization (e.g. Fernandez 2006). However, surrogates in general, have not yet been used to reduce the complexity of superstructure models. Furthermore, no thorough and systematic treatment of aspects such as the selection of independent-dependent surrogate variables has been presented.

This paper covers the foundations and implementation aspects of a framework leading to realistic and simpler superstructure MINLPs. First, we present how the number of equations and variables can be greatly reduced by replacing detailed first-principle models of complex process units with surrogate models. To this end, we propose a novel unit variable analysis to determine the minimal set of variables which have to be included in a surrogate. Second, we discuss how surrogate models in the form of Artificial Neural Networks (ANN) can be generated using existing process simulators. Third, we present a novel reformulation to incorporate ANN surrogates in order to obtain a final superstructure MINLP model. Finally, we comment on the generation and use of surrogate models for entire plant subsystems instead of individual unit operations.

2. Surrogate model design

As already mentioned, the central idea of this work is to reduce the mathematical complexity of superstructure MINLPs by replacing original first-principle unit models with suitable surrogates. But, for this replacement to be helpful, such surrogates have to include as few of the original model variables as possible (making them compact) while being equivalent in terms of the constraints they enforce. Hence, the selection of the surrogate variables is a fundamental step in their design.

In essence, a unit model is a set of equations establishing relationships between unit variables which can be divided in two sets: the set of independent variables, $\mathbf{I_{UM}}$, and the set of dependent variables, $\mathbf{D_{UM}}$. In other words, the model implicitly imposes a mapping $Z^D = f(Z^I)$ where Z^I is a vector in the space defined by $\mathbf{I_{UM}}$, and Z^D is a vector in the space defined by $\mathbf{D_{UM}}$. Subsets of both independent and dependent variables form what we call here the set of connecting variables, $\mathbf{C_{UM}}$; that is, unit variables that appear not only in the unit model but also in other parts of the optimization problem (i.e. other constraints or the objective function). Finally, it is also important to recognize the existence of fixed unit variables, $\mathbf{F_{UM}}$; i.e., variables whose values are known and fixed from the problem statement or engineering specifications. For a surrogate model to enforce the same constraints the original unit model enforces, it is necessary and sufficient that it captures the relationship between the independent variables that are not fixed ($\mathbf{I_{UM}} \backslash \mathbf{F_{UM}}$) and the dependent connecting variables ($\mathbf{D_{UM}} \cap \mathbf{C_{UM}}$). The remaining dependent variables ($\mathbf{D_{UM}} \backslash \mathbf{C_{UM}}$) are internal and required to make the original unit model work, but irrelevant in the context of the superstructure optimization problem. In this way, the selection of dependent and independent variable for a surrogate mapping, henceforth denoted as $\mathbf{I_S}$ and $\mathbf{D_S}$, is given by:

$$\mathbf{I_S} = \mathbf{I_{UM}} \backslash \mathbf{F_{UM}} \tag{1}$$

$$\mathbf{D_S} = \mathbf{C_{UM}} \backslash \left(\mathbf{F_{UM}} \cup \mathbf{I_{UM}}\right) = \mathbf{C_{UM}} \backslash \mathbf{I_{UM}} \tag{2}$$

Here, the second equality in (2) considers $\mathbf{F_{UM}} \subset \mathbf{I_{UM}}$, since in principle no fixed variable can be dependent.

For a particular model, the sets $\mathbf{F_{UM}}$ and $\mathbf{C_{UM}}$ are uniquely determined by the process design requirements, the superstructure topology and the form of the objective function. However, set $\mathbf{I_{UM}}$ is not unique because many different subsets of variables can be fixed to close the degrees of freedom of a particular unit model. Henao and Maravelias

(2009) presented a systematic way to select the set $\mathbf{I_{UM}}$ that leads to a reduced total number of surrogate variables (i.e. $|\mathbf{I_S}|+|\mathbf{D_S}|$), while accounting for the unit model equation system structure.

Other than the presented tailored surrogate design, we can formulate general purpose surrogates. This is done by assuming set $\mathbf{F_{UM}}$ is empty (i.e. there are no fixed variables), set $\mathbf{I_{UM}}$ includes only inlet streams state variables plus the unit operational parameters, and set $\mathbf{C_{UM}}$ only includes outlet streams state variables plus the unit operational parameters related to capital and operational costs (e.g. unit capacities and energy consumption). With these assumptions we have formulated general surrogates for the most common process units, including reactors, separation units, mixers, splitters, heat exchangers, etc. Due to space limitations, in Table 1 we present the surrogate models for compressors, pumps, turbines and flash vessels. Note that these specific surrogates also incorporate the linear equations in the original unit models to calculate explicitly some of the variables in $\mathbf{D_S}$, while the remainder are related to $\mathbf{I_S}$ variables using a general purpose multivariable mapping.

Table 1. General purpose surrogate model for pumps, compressors and flash vessels.

3. Surrogate model generation

The variable analysis presented in the previous section is general and applicable to any kind of detailed process unit and surrogate model. In this work, we use Artificial Neural Networks (ANNs) due to their excellent fitting characteristics and low complexity. Particularly, the ANNs used here are denoted as multilayer perceptrons (MLP) (Haykin 1999). An MLP is composed by a series of connected "layers", each one composed itself by a set of elemental computational units called "neurons". Mathematically, every layer k in a network acts as a transformation between and input vector $\mathbf{u}^{k-1}$ and an output vector $\mathbf{u}^k$ according to:

$$\mathbf{u}^k = f\left(\mathbf{W}^k \cdot \mathbf{u}^{k-1} + \mathbf{b}^k\right) \quad , \quad k=1,...,K \tag{7}$$

The set of all relations (7) defines the mapping between the network input $\mathbf{u}^0$ and output $\mathbf{u}^K$. Here, parameters $\mathbf{b}^k$ and $\mathbf{W}^k$ of layer k (composed by n^k neurons) are known as the layer bias vector (n^k-dimensional) and weight matrix ($n^k_x n^{k-1}$ dimensional) respectively. The scalar function $f(\cdot)$, known as activation function, (generally $tanh(\cdot)$) acts on every

component of its argument vector to generate the output vector $\mathbf{u}^k$. The last layer ($k=K$) in an ANN is called the output layer, while the rest are called hidden layers.

Once $\mathbf{I_S}$ and $\mathbf{D_S}$ have been determined for a particular unit, the independent variable space is sampled using a variance reduction technique, such as Latin Hypercube. The sample points are used to specify different simulation cases, whose solution provide dependent variable values. The gathered information is then used to train a MLP.

For this work, a code in MATLAB interfacing with ASPEN Plus has been implemented to generate simulation information and train MLPs automatically. Figure 1 presents the details of such implementation.
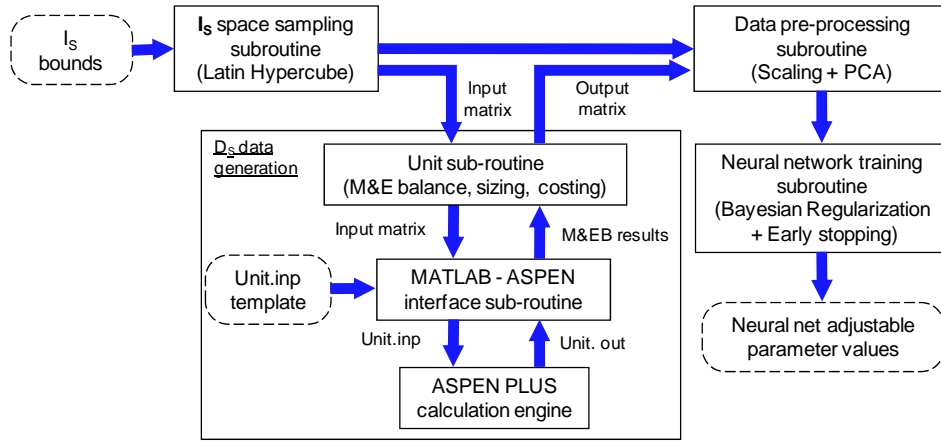
Figure 1. Automatic simulation data generation and network training.

4. Superstructure model and surrogate reformulation

Superstructure models are commonly formulated as Generalized Disjunctive Programs (GDM) (Lee and Grossmann 2000) and later reformulated as MINLPs, where binary variables are used to activate/deactivate particular unit models.

Taking advantage of the special structure of MLP surrogates, we propose a reformulation that avoids the introduction of additional non-linear terms keeping the activation functions as the only source of non-linearity. This contributes to the numerical tractability of the reformulated superstructure problems. Selecting *tanh(·)* as the activation function and taking advantage of its behavior (particularly the fact that *tanh(0)=0*) it is possible to reformulate the ANN surrogate presented in equation (7) as follows:

$$\mathbf{u}^k = f\left(\mathbf{v}^k\right) \tag{8}$$

$$\mathbf{v}^k = \mathbf{W}^k \cdot \mathbf{u}^{k-1} + \mathbf{b}^k \cdot y \tag{9}$$

$$\mathbf{L}^{k-1} \leq \mathbf{u}^{k-1} \leq \mathbf{U}^{k-1} \tag{10}$$

$$y \in \{0,1\} \quad , \quad k = 1,...,K \tag{11}$$

where $\mathbf{L}$ and $\mathbf{U}$ are lower and upper and bounds on the components of $\mathbf{u}^{k-1}$, and "y" is a binary selection variable. When $y=1$ the original surrogate relation is enforced; when $y=0$ the surrogate relation is deactivated, leading to $\mathbf{u}^{k-1}=0$, $\mathbf{u}^k=0$, $k=1,...,K$.

The simplicity of this reformulation is another reason why ANN surrogate models are used in this work. Furthermore, since this framework leads to the replacement of

multiple types of nonlinearities in the original unit models by only one type of non-linearity (i.e. the activation function) future work will include the development methods that deal specifically with this type of non-convexity in the final MINLP.

5. Surrogate modeling level

In the context of the process synthesis methodology presented in this paper, the use of surrogates at the unit operation level is convenient since unit models are the building blocks of any commercial process simulator, allowing a straightforward generation of the data required to build the surrogates. In addition, this strategy leads mostly to surrogates with continuous variables, leaving the use of discrete variables to manipulate process structure. However, surrogates can also be used to approximate the behavior of entire plant subsystems composed by several unit operations. This is particularly convenient when the structure of such subsystems is considered fixed (not subject to optimization), again leaving the use of discrete variables outside the surrogate level. Here, the obvious primary advantage is the reduction in the total number surrogate connecting variables (e.g. state variables of the streams connecting the units) and their associated surrogate expressions (in the case of the dependent variables).
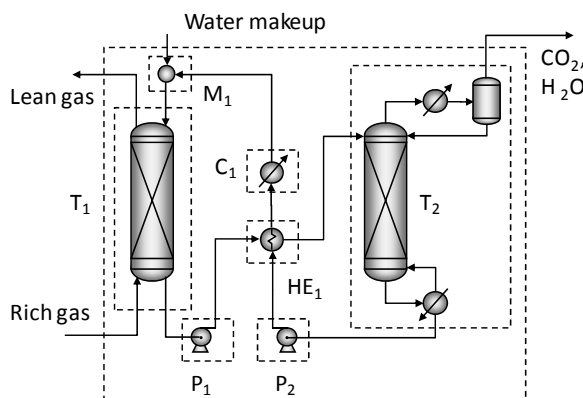
Figure 2. Absorption-based CO_2 capture system.

As an illustration, consider the CO_2 sequestration system in Figure 2. Assuming 5 components (O_2, N_2, CO, H_2O, MEA), the total number of variables when using surrogates for each piece of equipment approaches 99; however, considering a single surrogate for the whole system makes that number drop to 33. This reduction is the result of some streams becoming internal to the system, and the fact that unit coupling reduces the total number of degrees of freedom and hence the effective number of independent variables for the system.

6. Example

For the production on Maleic Anhydride from Benzene, consider the simple process superstructure presented in Figure 3, which consists of five units: a CSTR (R1), a PFR (R2), a flash tank (F), a splitter (S), and a mixer (M). Our goal is to find the reactor type (CSTR or PFR), reactor temperature and volume, flash temperature and recycle fraction, that maximize the annualized profit. The system is fed with 966 kmol/h of air and 34 kmol/h of benzene at 300 K, 1013 kPa, and we assume no pressure drop in all units.

Kinetics are as indicated by Westerlink (1988), and annualized profit calculations include revenue, operational and capital costs using standard algorithms.

The formulation involves the use of surrogate models for the mixer ($|\mathbf{I_S}|$=7, $|\mathbf{D_S}|$=7), CSTR ($|\mathbf{I_S}|$=9, $|\mathbf{D_S}|$=8), PFR, ($|\mathbf{I_S}|$=9, $|\mathbf{D_S}|$=8) and flash tank ($|\mathbf{I_S}|$=8,$|\mathbf{D_S}|$=15). The optimization was performed using GAMS 22.5 – DICOPT. The optimal solution involves a 40 m^3 PFR at 680K, a flash unit at 333 K, and recycle stream that is equal to 46% of the flash vapor stream.
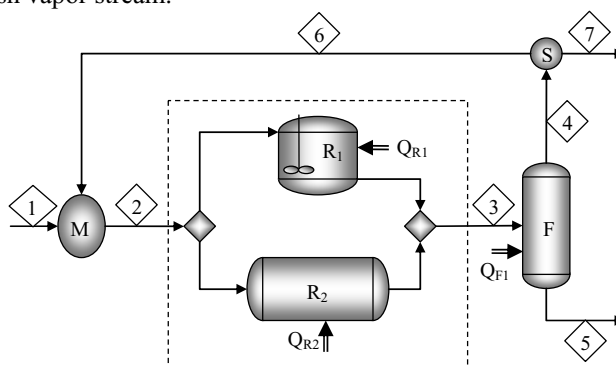
Figure 3. Maleic Anhydride process superstructure.

7. Conclusions

In this paper, we presented a framework for the formulation of computationally tractable superstructure optimization models. The key idea is the replacement of process unit models by ANNs. In achieving this, we presented a systematic method for the selection of surrogate independent and dependent variables; we discussed how surrogate models can be generated using process simulators; and illustrated the applicability of our framework using a small superstructure example.

References

Caballero, J.A., Grossmann, I. E. (2008). An Algorithm for the Use of Surrogate Models in Modular Flowsheet Optimization. AICHE J. 54, 2633.

Davis, E., Ierapetritou, M. (2007). A Kriging Method for the Solution of Nonlinear Programs with Black-Box Functions. AICHE J. 53,2001

Fernandes, FAN. (2006). Optimization of Fischer-Tropsch Synthesis Using Neural Networks. Chem. Eng. & Tech. 29,449.

Haykin, S. (1999). Artificial Neural Networks: A comprehensive foundation. Prentice Hall. Upper Saddle River, NJ.

Henao, C.A., Maravelias C.T. (2009). Process Superstructure Optimization Using Surrogate Models. Proceedings of the Seventh International Congerence on the Foundations of Computer-Aided Process Design. CRC Press.

Jones, D.R., Schonlau, M., Welch, W.J. (1998). Efficient Global Optimization of Expensive Black-Box Functions. J. Glob. Opt. 13,455.

Lee, S., Grossmann, I. E., (2000). New algorithms for nonlinear generalized disjunctive programming. Comp. Chem. Eng. 24,2125.

Webster, R., Oliver, M. A., (2007). Geo-statistics For Environmental Scientists. John Wiley& Sons Hall. Chichester, West Sussex UK.

Westerlink, E. J., Westerterp, K. R. (1988). Safe Design of Cooled Tubular Reactors for Exothermic Multiple Reactions: Multiple reactions networks. Chem. Eng. Sci. 43, 1051.

Won, KS., Ray, T. (2005). A Framework for Design Optimization Using Surrogates. Eng. Optim. 37,685.

Total Cost Targeting for Heat Exchanger Networks Including Pumping Costs

Medardo Serna-González, José María Ponce-Ortega, Oscar Burgara-Montero

Facultad de Ingeniería Química, Universidad Michoacana de San Nicolás de Hidalgo, Morelia, Mich. 58060, Mexico, E-mail: mserna@umich.mx; jmponce@umich.mx; obm_07@hotmail.com

Abstract

A total cost targeting method for the heat exchanger network (HEN) design problem is presented, which combines existing targeting methods for grass-roots design with a new method for targeting pumping costs (i.e., optimizing the pressure drops of streams). The overall targeting problem is formulated so as to minimize the total annual cost, which includes the capital costs of heat exchangers and pumping devices, as well as the utilities costs and the electricity costs to run the pumping equipment. This leads to a problem that can be solved as a non-linear optimization problem. The resulting solution of this mathematical model gives the optimal pressure drops and film heat-transfer coefficients of the streams. Therefore, neither the heat-transfer coefficients nor the allowable pressure drops are fixed, as is the case with previous targeting methods based on pinch technology. An example is presented to illustrate the targeting of heat exchanger networks using the proposed approach.

Keywords: heat exchanger networks, pumping cost, stream pressure drops, targeting

1. Introduction

Pinch technology is the most widely used technique for the grassroot design of HENs because it allows the design engineer to incorporate real plant situations easily for industrial scale problems. Pinch technology decomposes the HEN design problem into a sequential procedure. The first stage consists in generating economic trade-offs between capital and operating costs ahead of design to select the optimal value for the minimum allowed approach temperature, ΔT_{min} [1]. Then, in the second stage, the pinch design method [2] is applied to develop a network based on the optimal targets generated in the first stage. Targeting and synthesis stages constitute the conceptual design phase of pinch technology. The conceptual design is followed by the detailed design. In this last stage, the detailed heat-exchanger designs based on the stream pressure drops are performed. However, until recently, targeting and synthesis stages had been applied assuming constant film heat-transfer coefficients for all streams, which can lead to nonoptimal or infeasible networks due to inconsistencies with detailed designs of the heat exchangers at the last stage.

Polley and Panjeh Shahi [3] were the first to address the issue of considering allowable stream pressure drops in the conceptual design phase. They developed an area targeting algorithm that includes an expression that relates the stream pressure drop to the heat exchanger area and the stream heat transfer coefficient. Values of the allowable pressure drops of streams, fluid properties, volumetric flow rate and hydraulic diameters of the exchangers are specified to calculate the film heat-transfer coefficients of streams and network area. The calculated heat transfer coefficients are then used in the synthesis stage. This work improves the consistency between conceptual design and detailed

design phases for given allowable pressure drops of streams. Then, several procedures have been proposed to incorporate pressure drop effects into the grass-roots design of a HEN [4-10]. However, none of these works allowed targets to be set for optimum stream pressure drops, which are necessary to predict the capital and operating costs of the pumping equipment.

This paper presents a mathematical model that treats as optimization variables the stream pressure drops and film heat-transfer coefficients at the targeting stage of pinch technology. Equations based on the Kern method [11] that relate the stream pressure drops, the film heat transfer coefficients and the heat exchanger geometry for heat exchange matches are included in this formulation. The objective function minimizes the total annual cost (TAC), which includes the utility consumption cost, the capital investment for the heat exchangers and pumping devices, and the power cost.

2. Problem statement

The problem addressed in this paper can be stated as follows, given:

(i) a set of hot process streams and a set of cold process streams for heat exchange with their specified inlet and target temperatures, flow rates and physical properties (heat capacity, density, viscosity, thermal conductivity, and fouling factor);
(ii) a single heating utility and a single cooling utility to satisfy the energy requirements with their inlet and target temperatures as well as their physical properties;
(iii) operating and capital cost data involved in the network installation and operation.

determine the minimum total annual cost target for HEN design problems, including the target for the optimum power cost in addition to traditional HEN design targets for minimum utilities, area and number of units.

The following assumptions were used in this work:

- The film heat-transfer coefficient of a stream is match independent.
- Consideration is only given to heat exchange in a single phase.
- Only counter current and multipass shell-and-tube heat exchangers are considered.

3. Model formulation

The objective function is given in the following form:

$$min\ TAC = A_f\ CC + C_U + C_P \tag{1}$$

where A_f is the annualized factor for investment, CC is the capital cost for heat exchange units, C_U is the hot and cold utilities costs, and C_P is the power cost.

The annual utility cost of a network is given by:

$$C_U = H_Y\ (C_H\ Q_{Hmin} + C_C\ Q_{Cmin}) \tag{2}$$

where Q_{Hmin} (Q_{Cmin}) is the minimum hot (cold) utility target, C_H (C_C) is the unit cost of hot (cold) utility, and H_Y represents the hours of operation of network per year.

Assuming an equal distribution of area among all heat exchangers in the HEN, the network capital cost is given as:

$$C_{CAP} = N_{u,mer}\ [a + b(A_{min}/N_{u,mer})^c] \tag{3}$$

where $N_{u,mer}$ is the minimum number of units for a maximum energy recovery network, A_{min} is the network area target, and a, b, and c are cost law coefficients that depend on the type of heat exchanger, construction materials, and pressure rating.

In order to evaluate A_{min}, we develop a mathematical model that is based on the *spaghetti* design model provided by the composite enthalpy-temperature curves. For a

given value of ΔT_{min}, the first step to develop the *spaghetti* design of a process is to divide the balanced composite curves into $k = 1, \ldots, K$ enthalpy intervals defined wherever a change in slope occurs in either composite profile [1]. The *spaghetti* design assumes that each hot stream i is integrated with all cold streams j within each enthalpy interval k. This generates a vertical heat transfer structure featuring parallel stream splitting and isothermal mixing. Under the assumptions that each stream match of the *spaghetti* design represents one and only one heat exchanger and that the film heat-transfer coefficients of streams are match independent, the minimum area can be expressed as the sum of the stream contact areas:

$$A_{min} = \sum_i A_{ci} = \sum_j A_{cj} \tag{4}$$

$$A_{ci} = \sum_{j=1}^{J} \left(\frac{1}{h_i} + \frac{1}{h_j} \right) (UA_{i,j}) \qquad \text{for } i \in I \tag{5}$$

$$A_{cj} = \sum_{i=1}^{I} \left(\frac{1}{h_i} + \frac{1}{h_j} \right) (UA_{i,j}) \qquad \text{for } j \in J \tag{6}$$

$$UA_{i,j} = \sum_{k=1}^{K} UA_{i,j,k} \qquad \text{for } i \in I, j \in J \tag{7}$$

$$UA_{i,j,k} = \frac{q_{i,j,k}}{F_{Tk} \Delta T_{MLk}} \qquad \text{for } i \in I, j \in J, k \in K \tag{8}$$

$$q_{i,j,k} = \frac{CP_{i,k} CP_{j,k} \Delta T_{i,k}}{\sum_j CP_{j,k}} \qquad \text{for } i \in I, j \in J, k \in K \tag{9}$$

where A_{ci} (A_{cj}) is the contact area of the hot (cold) stream i (j) and, for each enthalpy interval k, $U_{i,j,k}$ is the *UA*-value for the match between streams i and j, $q_{i,j,k}$ is the amount of heat transferred from hot stream i to cold stream j, $CP_{i,k}$ ($CP_{j,k}$) is the heat capacity flow rate of hot (cold) stream i (j), $\Delta T_{i,k}$ is the temperature change of any hot stream i, F_{Tk} is the correction factor for the log mean temperature difference, and ΔT_{MLk} is the log mean temperature difference for any pair of streams.

If hot stream i is allocated in the shell side, the total stream pressure drop may be evaluated by the following equation [3, 9]:

$$\Delta P_i = K_i A_{ci} \left(\frac{1}{h_i} - R_{di} \right)^{-5.109} \qquad \text{for } i \in I \tag{10}$$

whereas if cold stream j is in the tube side,

$$\Delta P_j = K_j A_{cj} \left(\frac{D_{ti}}{D_t h_j} - R_{dj} \right)^{-3.5} \qquad \text{for } j \in J \tag{11}$$

where h_i (h_j) is the dirt film heat-transfer coefficient and R_{di} (R_{dj}) is the fouling factor for hot (cold) stream i (j). Constant K_i (K_j) depends on the stream physical properties and volumetric flow rate as well as the geometry of the heat exchanger [3, 9].

The total cost of power, which is necessary for pumping the fluid streams, is given by:

$$C_P = \sum_i C_{P,i} Q_i \Delta P_i + \sum_j C_{P,j} Q_j \Delta P_j \qquad (12)$$

where Q_i (Q_j) is the volumetric flow rate of the process stream i (j) and $C_{p,i}$ ($C_{p,j}$) is the unit cost of power used by the same stream.

For a given value of ΔT_{min}, eqs. (1)-(6) combined with eqs. (10)-(12) yield a non-linear programming (NLP) problem for minimizing the total annual cost of a HEN. The independent variables over which the minimization is performed are the minimum network area, the stream contact areas, the film heat-transfer coefficients and pressure drops of the streams The values of the fixed parameters $q_{i,j,k}$, $UA_{i,j,k}$, and $UA_{i,j}$ are calculated from equations (9), (8) and (7), respectively. Also, it should be noted that Q_{Hmin}, Q_{Cmin}, and $N_{u,mer}$ are determined before the NLP problem is solved.

By varying the value of ΔT_{min} used for utility targeting and recalculating the unit, area and power targets, the trade-off between capital and operating costs can be predicted to find the cost-optimal value of ΔT_{min} for HEN design.

4. A numerical example

The proposed procedure for predicting the minimum TAC of a HEN, including the target for optimum power cost, has been applied to solve a literature example recently considered by Mizutani et al. [5] and Ponce-Ortega et al. [7]. The stream data are shown in Table 1. A constant overall heat transfer coefficient for utility units is assumed as 444 W/m^2 K. Physical properties, which are assumed to be constant, are given in Table 2. The product $H_Y C_H$ was assumed as 60 ($/kW year), the product $H_Y C_H$ was taken as 6 ($/kW year), and the annualized factor for the capital cost as 0.2983/year. The values of *a, b,* and *c* were taken as 1000, 60 and 0.6, respectively. A total annual cost target is required at a $\Delta T_{min} = 10$ K

The solver CONOPT included in the GAMS optimization software [12] was used to solve the proposed NLP problem. With the value of ΔT_{min} (10 K) specified for utility targeting, the heating and cooling targets are 766.62 and 0 kW, respectively. The NLP model for this problem has 19 continuous variables. By solving the NLP model, the values of film heat-transfer coefficients and pressure drops of streams in Table 3 are obtained. Target values for area and minimum TAC for a network are given in Table 4.

Table 1. Stream data for Example

Stream	M (kg/s)	TIN (K)	TOUT (K)
H1	16.3	423	333
H2	65.2	363	333
H3	32.6	454	433
C1	20.4	293	398
C2	24.4	293	373
C3	65.2	283	288
HU		700	699
CU		300	320

Total cost targeting of heat exchanger networks including pumping costs

Table 2. Fluid physical properties for Example

Property	Value
μ [kg/m s]	2.4×10^{-4}
ρ [kg/m^3]	634
Cp [J/kg K]	2454
k [W/m K]	0.114

Table 3. Results for Example

Stream	h (kW/m^2 K)	ΔP (Pa)
H1	701	3701
H2	598.3	1578
H3	618	641
C1	707	4550
C2	682	4510
C3	469	214

Table 4. Comparison of results for Example

Concept	This work	Mizutani et al. [5]	Ponce-Ortega et al. [7]
TAC ($/year)	60,957	204,523	58,895
A_{min} (m^2)	868	442	513

Notice in Table 4 that the NLP targeting model predicts a minimum TAC of $60,957/year, which is only 3.5% greater than the optimum TAC of the detailed HEN obtained by Ponce-Ortega et al. [7] using genetic algorithms. However, the approach of Mizutani et al. [5] based on a rigorous MINLP formulation provides a detailed HEN with an optimum TAC of $204,523/year, thus overestimating the target for minimum TAC by 235.52%. Therefore, according to Ponce-Ortega et al. [7] it is evident that the solution provided by Mizutani et al. (2003), which was obtained with the use of MINLP techniques, reflects a case in which the algorithm got trapped into a local optimum point because of the nonconvexities of the problem. For these cases, the optimum heat-transfer coefficients and pressure drops of streams given by the NLP model could provide an excellent starting point for the optimum detailed design of HENs.

It should be noted that the proposed NLP model overestimates the network area target because it is based on the Kern method and assumes vertical heat transfer on the composite curves. To overcome these limitations, a better model for describing the behavior of shell-and-tube exchangers could be incorporated in the proposed targeting procedure.

5. Conclusions

This paper presents a mathematical model to set pressure drops or power cost targets in addition to traditional HEN design targets for the minimum utilities, area and number of

units. For a given value of ΔT_{min}, these targets are dependent on the economic data for the process but are independent on the network structure. Thus, this model allows the consideration of the trade-offs between all above HEN targets to obtain an optimum value of ΔT_{min} before an HEN is synthesized.

The example shows that the proposed targeting model can approach within a few percent the total annual cost target with the optimum TAC of a detailed HEN obtained by a rigorous methods like the one developed by Ponce-Ortega et al. [7]. Thus, the film heat-transfer coefficients and pressure drops of streams given by the solution of the NLP could provide a good starting point for rigorous methods for the synthesis of a detailed HEN.

Acknowledgements

References

[1] B. Linnhoff and S. Ahmad. Cost optimum heat exchanger networks: I. Minimum energy and capital using simple models for capital costs. Computers and Chemical Engineering, 14 (1990), 729.

[2] B. Linnhoff and E. Hindmarsh. The pinch design method for heat exchanger networks. Chemical Engineering Science, 38 (1983), 745.

[3] G. T. Polley and M.H.M. Panjeh Shahi. Interfacing heat exchanger network synthesis and detailed heat exchanger design. Chemical Engineering Research and Design, 69 (1991), 445.

[4] S. Fraustro-Hernández, V. Rico-Ramírez, A, Jiménez-Gutiérrez, S. Hernández-Castro. MINLP synthesis of heat exchanger networks considering pressure drop effects. Computers and Chemical Engineering, 27 (2003), 1143.

[5] F.T. Mizutani, F.L.P. Pessoa, E.M. Queiroz, S. Hauan and I.E. Grossmann. Mathematical Programming Model for Heat-Exchanger Network Synthesis Including Detailed Heat Exchanger Designs. 2. Network Synthesis. Industrial and Engineering Chemistry Research, 42 (2003), 4019.

[6] M.H. Panjeh Shahi and A. Khoshgard. Heat exchanger networks targeting and design with unequal heat transfer coefficient regarding allowable pressure drop of streams. Heat Transfer Engineering, 27 (2006), 36.

[7] J.M. Ponce-Ortega, M. Serna-Gonzalez and A. Jiménez-Gutiérrez. Heat exchanger network synthesis including detailed heat exchanger design using genetic algorithms. Industrial and Engineering Chemistry Research, 46 (2007), 8767.

[8] M.A.S.S. Ravagnani and J.A. Caballero. Optimal heat exchanger network synthesis with the detailed heat transfer equipment design. Computers and Chemical Engineering, 31 (2007), 1432.

[9] M. Serna-Gonzalez, J.M. Ponce-Ortega and A. Jiménez-Gutiérrez. Two-level optimization algorithm for heat exchanger networks including pressure drop considerations. Industrial and Engineering Chemistry Research, 43 (2004), 6766.

[10] X.X. Zhu and X.R. Nie. Pressure drop considerations for heat exchanger networks grassroots design. Computers and Chemical Engineering, 26 (2002), 1661.

[11] D.Q. Kern. Process Heat Transfer. McGraw Hill: New York, 1950.

[12] A. Brooke, D. Kendrick and A. Meeraus. GAMS User's Guide, The Scientific Press, USA, 2006.

Energy integration of industrial sites with heat exchange restrictions

Helen Becker, Luc Girardin, François Maréchal

Industrial Energy Systems Laboratory (LENI), Ecole Polytechnique Fédérale de Lausanne, CH-1015 Lausanne, Switzerland, helen.becker@epfl.ch

Abstract

Process integration methods aim at identifying options for heat recovery and optimal energy conversion in industrial processes. By applying pinch analysis methods, the first step is to calculate the maximum heat recovery between cold and hot streams. The second step consists in designing the corresponding heat recovery exchanger network, based on a fixed list of streams.

For the heat cascade, it is assumed that any heat exchange between cold and hot streams is possible, but due to industrial constraints, in many cases, this assumption cannot be accepted in practice and it is necessary to impose restricted matches. This introduces energy penalties, which could be reduced by using intermediate heat transfer systems. This paper introduces a targeting method including industrial constraints to ensure feasible solutions for the heat exchanger network. Intermediate heat transfer systems are integrated so that restricted heat exchanges become possible and heat recovery penalties, created by those constraints, can be reduced.

The problem is formulated as a MILP (mixed integer linear programming) problem, which considers not only restricted matches but also the optimal integration of the energy conversion system, like heat pumping and combined heat and power production.

The application of the method will be illustrated by an industrial example from the pulp and paper industry. The extension of the method to study the heat integration of semi batch processes will be discussed.

Keywords: pinch analysis method, utility integration, process design, process sub-systems, heat exchanger network

1.   Introduction

Pinch analysis is a promising tool to optimize the energy efficiency of industrial processes. To realize the maximum heat recovery and the optimal integration of utilities to supply heating and cooling requirement, a heat exchanger network has to be designed, considering process and utility streams. One major drawback is the assumption that any hot stream can exchange heat with any cold stream. In reality, heat exchanges become difficult or even impossible, due to constraints such as the distance between streams or quality and/or safety reasons, or due to system dynamics such as non-simultaneous operations.

The total site approach, presented by Klemeš et al. (1997), implicitly accounts for restricted matches. The cold and hot streams, resulting from sub-systems without considering self-sufficient pockets, are separated graphically. The sub-systems can only exchange heat via the steam system; the heat recovery is calculated, but there is no systematic approach to define the members of sub-systems and the integration of the energy conversion units is not considered. Especially, the self-sufficient pockets are suppressed, which penalizes the combined heat and power integration. Bagajewicz and Rodera (2001) propose

a single heat belt, which exchanges heat between process plants by an intermediate fluid. Only for special cases (3 process plants) this problem is solved by using a MILP formulation. Forbidden matches between certain pairs of process streams are considered by Papoulias and Grossmann (1983). They propose a mathematical formulation to minimize the energy penalty (without integrating utility streams) and the heat exchanger network design with restricted matches. Also Cerda and Westerberg (1983) studied heat exchanger networks with restricted matches and propose an algorithm which imposes constraints disallowing in part or in total the matching of stream pairs. Maréchal and Kalitventzeff (1999) propose a MILP strategy, which integrates forbidden heat exchange connections as constraints in the targeting phase, and allows the integration of heat transfer fluids.

For solving a site scale process integration problem, this paper presents an extension of this MILP strategy and introduces the approach of process integration by sub-systems, which makes the practical implementation easier and considers restricted matches between sub-systems.

2.   Method

The new methodology, proposed here, takes into account heat exchange restrictions at the targeting stage by dividing industrial plants into sub-systems. Heat can be exchanged inside a sub-system but no direct heat exchange with other sub-systems is allowed. The integration of a heat transfer system (e.g. hot water loop or steam network) gives the possibility of indirect heat exchange between sub-systems. The mass flows rates of the heat transfer fluids are optimized in order to minimize the cost of the energy penalty.

The problem is solved in three steps. In the first step, a MILP problem is formulated to define the composite curves with heat exchange restrictions, that represent the necessary enthalpy-temperature profiles for the heat transfer system. Then, a second MILP problem is solved to target the integration of heat transfer fluids together with energy conversion systems. The heat load distribution problem (HLD), proposed by Maréchal and Kalitventzeff (1989), is then adapted to incorporate the definition of sub-systems. The resolution of the HLD problem becomes much easier and corresponds to the first step of the heat exchanger network design. The major advantages of the presented method are:

- The process is divided into sub-systems (more realist than just heat restriction constraints for two streams); heat exchange inside sub-systems is favored.

- Contrary to the total site integration methodology, self-sufficient pockets are not suppressed. This allows the maximization of the combined heat and power production.

- The design of the heat exchanger network becomes easier and more flexible and implicitly includes topological constraints.

- Simultaneous optimization of the utility integration and the heat transfer system defines the complete list of streams and allows the HEN design.

- The combinatorial nature of the HEN design is reduced.

3.   Optimization algorithm

The objective is to minimize the operating costs (equation (1)). $\dot{E}^+_{fuel}$ is the energy delivered by the fuel (e.g. natural gas) and $\dot{E}_{el}$ is the electricity demand$^{(+)}$ or excess$^{(-)}$ of the

process. For the electricity cost, c_{el}^+ is the purchase cost and c_{el}^- is the selling price. c_{fuel} is the fuel price.

$$F_{obj} = min(c_{fuel}\dot{E}_{fuel}^+ + c_{el}^+\dot{E}_{el}^+ - c_{el}^-\dot{E}_{el}^-) \tag{1}$$

The normal heat cascade for each temperature interval k is given by equation (2), where $\dot{M}$ is the mass flow rate [kg/s] and q is the heat load per mass flow [MJ/kg]. The subscripts "c" and "h" indicate cold and hot streams respectively and k refers to the temperature interval k of the heat cascade. R_k is the cascaded heat from the temperature interval k to the lower temperature intervals.

$$\sum_{h_k=1}^{n_{h,k}} \dot{M}_{h,k}q_{h,k} - \sum_{c_k=1}^{n_{c,k}} \dot{M}_{c,k}q_{c,k} + \dot{R}_{k+1} - \dot{R}_k = 0 \qquad \forall k = 1...,n_k \tag{2}$$

$$R_k \geq 0 \qquad \forall k = 1...,n_k \tag{3}$$

When the industrial plant is divided into sub-systems following equations (4) - (7) are added to take account of heat exchange restrictions:

$$\sum_{h_{s,k}=1}^{n_{h,s,k}} \dot{M}_{h,s,k}q_{h,s,k} - \sum_{c_{s,k}=1}^{n_{c,s,k}} \dot{M}_{c,s,k}q_{c,s,k} + \dot{Q}_{hts,s,k}^- - \dot{Q}_{hts,s,k}^+ + \dot{R}_{s,k+1} - \dot{R}_{s,k} = 0$$

$$\forall k = 1...,n_k, \forall s = 1...,n_s \tag{4}$$

$$\dot{R}_{s,k} \geq 0 \qquad \forall k = 1...,n_k, \forall s = 1...,n_s \tag{5}$$

For each sub-system (s) the heat cascade is given by equation (4). When a sub-system has a deficit or a surplus of heat in the temperature interval k, the heat is supplied from the heat transfer system ($Q_{hts,s,k}^-$) or respectively removed by the heat transfer system ($Q_{hts,s,k}^+$). To ensure that heat is cascaded correctly, a second set of equations is necessary. Equation (6) express the heat balance of the hot streams and equation (7) express the heat balance of the cold streams in the heat transfer system ($_{hts}$). The flow rates of the heat transfer fluids have to be optimized in order to satisfy the remaining heat demand of all sub-systems.

$$\sum_{h=1}^{n_{h,hts,k}} \dot{M}_{h,hts,k}q_{h,hts,k} + \dot{R}_{hts,k+1} - \dot{R}_{hts,k} - \sum_{s=1}^{n_s} \dot{Q}_{hts,s,k}^- \geq 0 \qquad \forall k = 1...,n_k \tag{6}$$

$$-\sum_{c=1}^{n_{c,hts,k}} \dot{M}_{c,hts,k}q_{c,hts,k} + \dot{R}_{hts,k+1} - \dot{R}_{hts,k} + \sum_{s=1}^{n_s} \dot{Q}_{hts,s,k}^+ \geq 0 \qquad \forall k = 1...,n_k \tag{7}$$

4. Numerical example - drying process in paper industry

The humid pulp is first preheated and enters then in the dryer unit. Steam satisfies a big part of the heat demand. The hot air also introduces heat to the dryer but its main function is to evacuate the evaporated water from the pulp. Possible heat recuperation is modeled by a humid air stream which has to be cooled down to the final temperature of 30°C. The list of involved streams is given in table 1.

The pulping unit, drying unit and the boiler are considered as different sub-systems. Heat can not be exchanged directly between these sub-systems. This means, that the heat demand of sub-system 1 has to be satisfied by the heat transfer system, even if the excess heat of sub-system 2 is sufficient to satisfy the demand. The penalty of this constraint can be visualized in the integrated composite curve of the utility system (figure 2a), calculated with the constraint and without heat transfer fluids.

In order to reduce or eliminate this penalty, intermediate fluids can be used in the heat
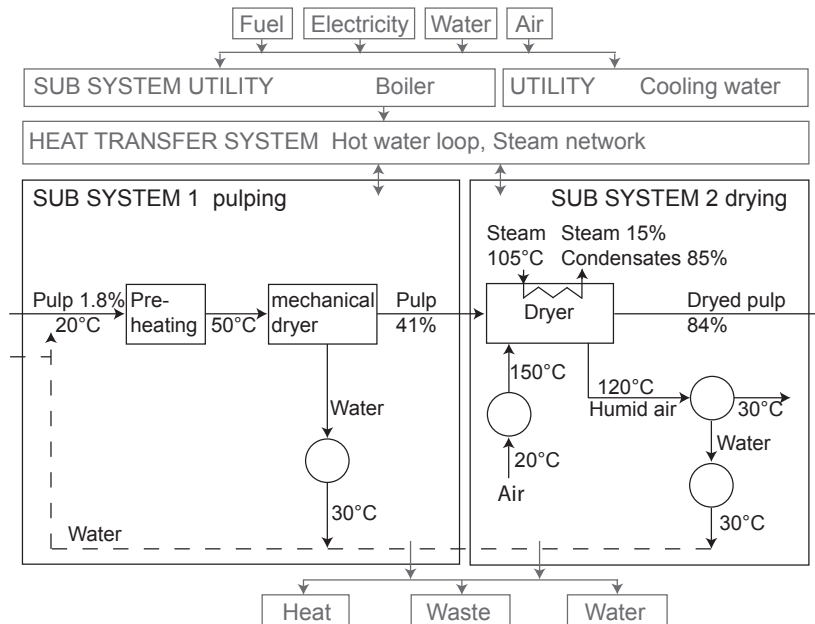
Figure 1: Representation of process

Table 1: Hot and cold streams of the process

Sub-system	Name	T_{in} [°C]	T_{out} [°C]	Heat Load [kW]	Remark
pulping	C1	20	50	11262	Preheating
	H2	50	30	-7297	water cooling
drying	C5	20	150	664	Air heating
	C6	95	105	6058	Steam demand
	H3	105	105	-892	Condensation of 15% steam
	H4	105	95	-112	cooling of condensates
	H9	120	30	-5319	Humid air cooling

transfer system. Figure 2b shows the integrated composite curve obtained by solving the problem with the integrated heat transfer system. Introducing a hot water loop, eliminates the penalty of the heat exchange restriction between the pulping and drying sub-system. Water is heated up from 35°C to 80°C with streams from the drying unit and heat is given back to the pulping unit by cooling down the water from 80°C to 35 °C. At higher temperatures a steam network is used as the intermediate heat transfer between the boiler (steam production at 80 bar) and the process demand (steam utilization at 7 bar and 2 bar).

The heat load distribution is then calculated for each zone delimited by a pinch point. Figure 3 shows the heat load distribution of zone 1 (12-103 °C), limited by the first pinch point at 103 °C, obtained with the integrated heat transfer system (hot water loop and steam network). The temperatures on the graph correspond to the corrected temperatures.

(a) Integrated composite curves of the utility system (b) Integrated composite curves of heat transfer system
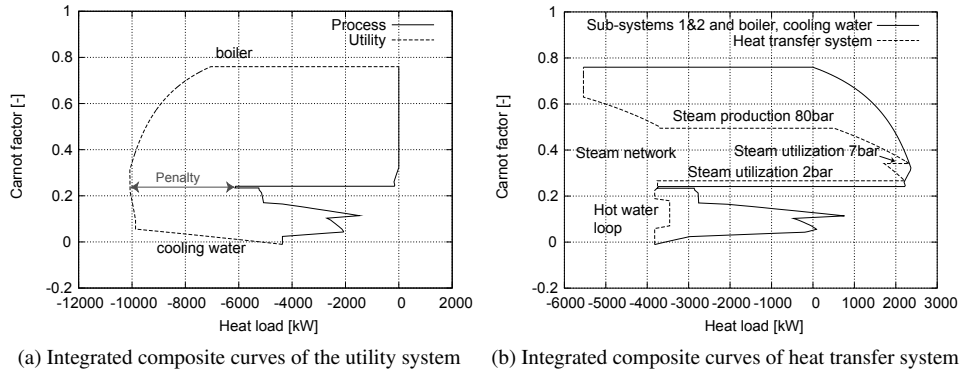
Figure 2: Composite curves with constraints

Table 2: Results

	Unit	No constraints	With constraints	Constraints and heat transfer system
Fuel consumption	[kW]	6014	9868	7760
Cooling water consumption	[kW]	1651	5505	1676
Electricity	[kW]			1684
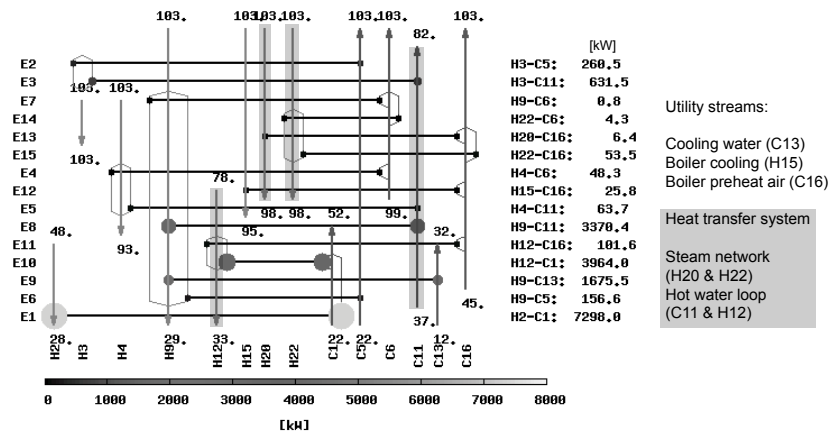
Figure 3: Heat load distribution with integrated heat transfer system

5. Discussion and perspectives

Although, the method presented in this paper is illustrated by a simple example with three sub-systems, the method aims at solving complex examples with multiple sub-systems (e.g. process units with different locations or other industrial site constraints). The sub-system concept is also considered for calculating the integration of utility systems, for example the produced heat in a boiler cannot exchange directly with process streams, but a steam network makes the heat exchange possible.

The method can also be used to solve batch problems (non simultaneously operations

in one period) considering that during one operation step the streams cannot exchange directly with another operation step. In this case, the heat exchange between two batch operations of the same period requires the use of a heat transfer system that will be optimized by the proposed method. The batch operations can exchange heat by storing temporarily this heat in vessels.

This approach combined with the MILP formulation for combined heat and power integration allows one to solve rigorously the problem of site scale integration. The proposed method can also be added to the approach proposed by Maréchal and Kalitventzeff (2003) to solve multi-period problems for site scale integration.

6.   Acknowledgements

References

Bagajewicz, M. and Rodera, H. 2001. On the use of heat belts for energy integration across many plants in the total site. *Canadian Journal of Chemical Engineering*, 79(4):633–642.

Cerda, J. and Westerberg, A. W. 1983. Synthesizing heat exchanger networks having restricted stream/stream matches using transportation problem formulations. *Chemical Engineering Science*, 38(10):1723–1740.

Klemeš, J., Dhole, V., Raissi, K., Perry, S., and Puigjaner, L. 1997. Targeting and design methodology for reduction of fuel, power and CO2 on total sites. *Applied Thermal Engineering*, 17(8-10):993–1003.

Maréchal, F. and Kalitventzeff, B. 1989. Synep1: a methodology for energy integration and optimal heat exchanger network synthesis. *Computers chemical Engineering*, 13(4/5):603–610.

Maréchal, F. and Kalitventzeff, B. 1999. Restricted mataches and minimum cost of energy requirements: tools and methodology for producing pratical solutions. *2-nd Conference on Process Integration and Optimisation for Energy Saving and Pollution Reduction - PRES'99*.

Maréchal, F. and Kalitventzeff, B. 2003. Targeting the integration of multi-period utility systems for site scale process integration. *Applied Thermal Engineering*, 23:1763–1784.

Papoulias, S. A. and Grossmann, I. E. 1983. A structural optimization approach in process synthesis - ii. *Computers and Chemical Engineering*, 7(6):707–721.

Synthesis, Design and Analysis of Downstream Separation in Bio-refinery Processes through a Group-Contribution Approach

Merlin Alvarado-Morales, Krist V. Gernaey, John M. Woodley, Rafiqul Gani

Department of Chemical and Biochemical Engineering, Technical University of Denmark, DK-2800 Kgs. Lyngby, Denmark, rag@kt.dtu.dk

Abstract

In this paper, a novel systematic approach to simultaneously model, design, and synthesize chemical and biochemical processes is presented. The core idea behind this approach is to apply the principles of the group-contribution approach for pure component property prediction to the synthesis and design of chemical process flowsheets. The method is highlighted through a bio-refinery case study involving the production of bioethanol (bioEtOH), succinic acid (SA) and diethyl succinate (DES), for which, energy efficient processing options have been identified.

Keywords: process synthesis, flowsheet modeling, systematic method, process-group

1. Introduction

In the group-contribution method for pure component/mixture property prediction, molecular identity is described by means of a set of groups bonded together to form a specified molecular structure. By analogy, for flowsheet property prediction, a process flowsheet can be described by means of a set of process-groups bonded together to represent a specified flowsheet structure. The process-groups represent either a single unit operation (such as a reactor, distillation, flash, etc.) or a set of unit operations (such as extractive distillation, pressure swing distillation, etc). The bonds among the process-groups represent the streams and/or recycles, in an analogous way to the bonds that link molecular groups. Consequently, each process-group provides a contribution to the properties of the flowsheet. The properties can be the performance in terms of energy consumption or operating cost or profit, etc. In this way, once the flowsheet is described by the process-groups, the property of interest can be calculated. Therefore, based on this premise we have applied the group-contribution approach to systematically model, synthesize/design and analyze downstream separation from chemical and biochemical processes. The core idea of the approach in this paper, is based on the process-group approach developed by d' Anterroches and Gani (2005) to solve synthesis/design problems related to chemical processes. In order to represent an extended set of unit operations, an extended set of corresponding process-groups has been developed. The application range of the new set of process-groups is highlighted by means of a bio-refinery case study involving the production of bio-ethanol (bioEtOH), succinic acid (SA) and diethyl succinate (DES).

2. Overview of the method

The method presented in this paper, the process-group contribution based approach, consists of the following seven steps: (1) synthesis problem definition, (2) synthesis

problem analysis, (3) process-group selection, (4) generation of candidates, (5) ranking/selection of candidates, (6) reverse simulation, and (7) final verification. Step 1 involves the structural definition of the process inputs (raw materials) and outputs (desired products) of the process flowsheet as well as the definition of the flowsheet property targets. In step 2, in order to gain usable knowledge for the subsequent steps, reaction and pure component/mixture property analysis are performed. By identification of the component identities in the desired product that are not available as reactants (raw materials), a database search is performed to find the chemical reactions yielding those components as products. The pure component/mixture property analysis is performed by means of the thermodynamic insights based method developed by Jaksland (1996). This method is based on the principle that for each process operation task, the properties of the species to be separated can be associated in order to provide information related to the feasibility of a process separation task for a given separation technique. In step 3, the process-groups are matched with the feasible operation tasks and the separation techniques identified in the previous step. The objective in step 4 is to combine the process-groups selected in step 3 according to the set of connectivity rules and specifications proposed by d' Anterroches and Gani (2005) to generate flowsheet structures. Each process-group has output specifications, which are guaranteed to be met if the connectivity rules of the process-groups are satisfied. In step 5, the generated flowsheet candidates are tested with respect to their target property values defined in step 1, using the corresponding flowsheet property model. Step 6 involves two tasks, the resolution of the mass balance for each process-group and the calculation of the flowsheet design parameters of the unit operations in the process flowsheet. Here, reverse simulation is used. In reverse simulation, knowledge of the state variables corresponding to the inputs and outputs of a unit operation, i.e., individual flowrates, pressures and temperatures, are used to back calculate the design parameters of the corresponding unit operation (e.g. number of stages, feed location, reflux ratio, residence time, volume). The reverse simulation for separation process-groups (such as distillation, extractive distillation, flash) is based on the driving force concept proposed by Bek−Pedersen and Gani (2004). For the reactor process-group it is based on the work of Horn (1964) on the attainable region concept. In step 7, all the necessary information to perform the final verification through rigorous simulation is available. The use of a commercial simulator allows further fine-tuning of the alternatives and the possibility to perform optimization of the design parameters for the most promising candidates.

3. Application of the method

3.1. Case study

The starting point of the case study is a slurry generated by a saccharification process where the conversion of cellulose to glucose is catalyzed by a cocktail of enzymes. The available glucose in the slurry to be converted into bioEtOH and SA is equal to 32110 *kg/h*. In order to calculate the amount of glucose that should be converted into SA, a mass balance is done based on a SA production capacity of 190.34 kg_{SA}/h and an annual load equal to 8406 *h/yr*. This SA production capacity corresponds to 10% of the current worldwide SA production according to the BREW Project (Patel, 2006). Assuming a SA yield on glucose equal to 0.775 (Song *et al*., 2007), the amount of glucose needed to produce this amount of SA is equal to 246 *kg/h*. From mass balance calculations, the amount of saccharified slurry containing this amount of glucose corresponds to 2811.6 *kg/h*. The remaining slurry is then sent to the bioEtOH production plant where bioEtOH is produced by fermentation.

Synthesis, Design and Analysis of Downstream Separation in Chemical and Bio−Processes through a Group−Contribution Approach

3.1.1. Synthesis problem definition

Based on the above analysis, the structural definition of the synthesis problem related to the downstream separation of bioEtOH from the effluent of the fermentor has been formulated as follows: one input process-group initialized with the mixture produced by the fermentor and one output process-group initialized with the desired product, bioEtOH. The flowsheet property target is the minimization of the energy consumption. The structural definition of the synthesis problem related to SA production is formulated as follows: one input process-group initialized with the mixture produced by the saccharification reactor, and one output process-group initialized with the desired product, SA. For the DES production process the structural definition is formulated as follows: one input process-group initialized with bioEtOH and SA streams resulting from their respective production processes, and one output process-group initialized with the DES. Figure 1 shows a schematic representation of the synthesis problem.
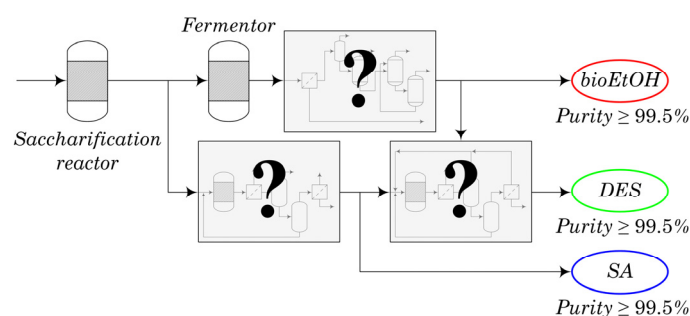
Figure 1. A schematic representation of the synthesis problem.

3.1.2. Synthesis problem analysis

The reactions in the SA production process taking place in the fermentor were reported by Song *et al*. (2007), and the reactions for the DES production process were taken from Kolah *et al*. (2008). From the pure component/mixture property analysis for the bioEtOH process, flash separation has been identified to separate CO_2 and O_2 from the other components. Flash/evaporation and distillation were identified as alternatives to perform the separation of water from the other components, except for the ethanol-water binary pair which forms an azeotrope. Further analysis of this binary azeotrope indicated that ethanol can be recovered by using liquid membrane, pervaporation, gas adsorption or extractive distillation. For the SA process, formic acid, pyruvic acid, CO_2, H_2, O_2, ammonia and microbial cells were found to be either products and/or reactants in the reactions. A pure component/mixture analysis was performed taking into account these components. Flash, distillation, and pervaporation have been identified to perform the separation of CO_2, H_2, O_2, and ammonia from the other components. Crystallization has been identified to separate SA from the rest of the components. However, due to the highly dilute nature of the mixture, liquid-liquid extraction has also been identified as suitable for the separation of water from the rest of the components in the mixture. As DES is produced from SA via the intermediate formation of monoethyl succinate (MES), this component was also taken into account for the pure component/mixture analysis. Flash/evaporation, liquid membrane, and pervaporation, have been identified to perform the separation of water and ethanol from the other components in the DES synthesis problem. Crystallization has been identified as a feasible separation technique to perform the separation of SA from DES and MES, while liquid adsorption, liquid

membrane, and pervaporation have been identified to perform the separation of DES and MES.

3.1.3. Process-group selection

The corresponding process-groups are selected from the database according to the feasible separation techniques identified in the previous step, and are matched with the components in the mixture.

3.1.4. Generation of candidates

For the downstream separation design of a bioEtOH production process 288 candidates have been generated through the combination of the process-groups selected in the previous step. However, out of the 288 design alternatives, only 4 candidates were of interest, since only they satisfied the connectivity rules as well as the structural definition of the synthesis problem. For the SA production process, 65 design candidates have been generated through the combination of the process-groups. Out of these 65 candidates, 16 were found to be feasible based on the structural constraints defined in the problem. For the DES production process, 54 feasible candidates have been generated.

3.1.5. Ranking/selection of candidates

The design candidate using the solvent-based azeotropic separation process-group in the downstream separation design for the bioEtOH process is considered to highlight the workflow in this step of the method. Note that the solvent identity is not known so far. Therefore, a CAMD (Computer Aided Molecular Design) problem formulation can be set up to find a matching solvent. The ProCAMD tool (Gani *et al.*, 1997) has been used to find potential candidates. With respect to ionic liquids (IL), the potential candidates were found through a search in the open literature (Seiler *et al.*, 2004; Jork *et al.*, 2004; Wang *et al.*, 2007). Table 1 shows the performance of the design candidates in terms of energy consumption.

Table 1. Design candidate results.

Candidate	Solvent fraction	Driving force	Predicted energy (*MkJ/h/kmole*)	Energy demand (*MkJ/h/kmole*) Seiler *et al.* (2004)
Glycerol	0.63	0.48	0.0322	
Ethylene glycol (EG)	0.52	0.48	0.0317	0.0335
Triethylene glycol	0.63	0.25	0.0618	
$[EMIM]^+[BF_4]^-$	0.375	0.37	0.0352	0.0333
$[BMIM]^+[Cl]^-$	0.45	0.42	0.0260	
$[EMIM]^+[EtSO_4]^-$	0.40	0.31	0.0386	
$[EMIM]^+[DMP]^-$	0.40	0.38	0.0318	

In the case of the SA production process, only one candidate using a liquid-liquid based separation process-group (representing liquid-liquid extraction) is selected for further analysis. For the bioEtOH process, the solvent identity is not known so far. Performing a database search, the following candidates have been identified as potential extractive agents: *n*-decyl acetate and *n*-butyl acetate. The performance of the potential candidates

Synthesis, Design and Analysis of Downstream Separation in Chemical and Bio−Processes through a Group−Contribution Approach

has been tested through *PT*-flash (multi-phase) rigorous calculation using ICAS (Gani *et al*., 1997). Based on this calculation, *n*-decyl acetate was selected as the best extractive agent since it is totally immiscible with water and it can promote a higher driving force than *n*-butyl acetate. Note that additional criteria such as cost of the solvent or toxicity of the solvent can be taken into account as well before making the final selection. Since the solvent needs to be recovered (for recycle), crystallization has been identified to separate the solvent from SA. In the case of the DES production process, the one candidate using the pervaporation process-group is selected for further analysis.

3.1.6. Reverse simulation

The feasible flowsheet shown in Figure 2 is considered for the reverse simulation step. In Figure 2 each process unit operation is represented by its corresponding process-group. Once the solvent identity is known, the solvent flowrate is calculated by mass balance.
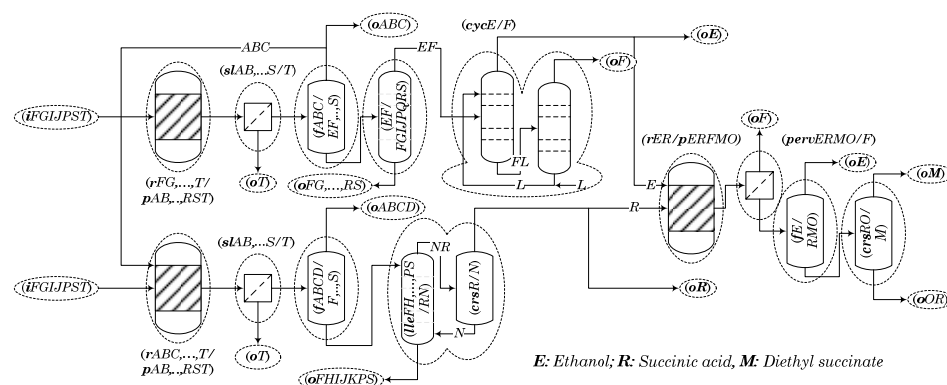
Figure 2. Process flowsheet for the selected design candidate.

A mass balance was made for each process-group for the process flowsheet in Figure 2. The downstream separation in the bioEtOH process using EG as an entrainer and the downstream separation in the SA process using *n*-decyl acetate as an extractive agent, were considered for reverse simulation. The reverse simulation of the flash and distillation process-groups was performed using the driving force based method (Bek-Pedersen and Gani, 2004) and the results are given in Table 2.

Table 2. Design parameters of the distillation columns.

	Distillation column	Extractive column	Recovery column
Number of stages	32	30	15
Feed stage	17	22	5
Reflux ratio	3.2	0.52	0.54
DF_{max}	0.35	0.48	0.59

In the case of the liquid-liquid based separation process-group, the number of stages can be determined by using the phase diagram, plotting *X* (*kg* solute/*kg* carrier) versus *Y* (*kg* solute/*kg* solvent) as illustrated Figure 3.
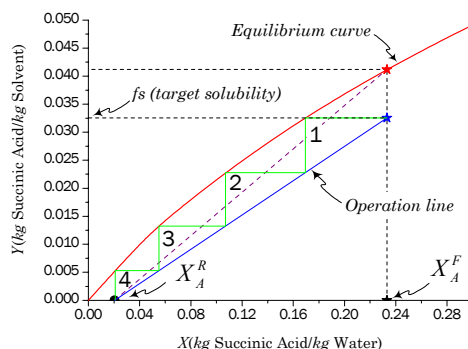
Figure 3. Graphical determination of the number of equilibrium stages for the liquid-liquid extraction process-group.

3.1.7. Final verification

The candidates with the best performance in terms of energy consumption were verified through rigorous simulation using an appropriate simulator. The optimal integrated sequence uses extractive distillation with IL for BioEtOH recovery, liquid-liquid extraction and crystallization for SA recovery, and pervaporation for DES recovery. The total predicted energy consumption is equal to 0.329 *MkJ/h/kmole*.

4. Conclusions

A novel systematic approach based on the group-contribution concept has been presented. One important feature of the method is its versatility, since it can be extended by adding new process-groups representing all types of process unit operations. Thus, it is possible to simultaneously model, design, and synthesize novel products and processes as is demonstrated in this paper. On the other hand, the ability to predict a flowsheet property − energy consumption − without the need for rigorous simulation offers a lot of advantages as it opens the possibility to screen a lot of process options very quickly and with high accuracy, a feature which has been demonstrated for downstream separation from the bioEtOH and SA processes. Finally, the results also show that the method provides a fast, efficient, and systematic process design approach by first solving the mass balance based on the process-group specifications, followed by calculation of the design parameters of the unit operations through reverse simulation.

References

E. Bek-Pedersen, R. Gani, Chem. Eng. Process., 43 (2004) 251.

L. d'Anterroches, R Gani, Fluid Phase Equilib., 228-229 (2005) 141.

R. Gani, G. Hytoft, C. Jaksland, A.K. Jensen, Comput. Chem. Eng., 21 (1997) 1135.

F. Horn, (eds.), In Proc. 3rd Eur. Symp. on Chemical Reaction Engineering, 1964.

C. Jaksland, PhD Thesis, Department of Chemical Engineering, DTU, Denmark, 1996.

C. Jork, M. Seiler, Y.A. Beste, W. Arlt, J. Chem. Eng. Data, 49 (2004) 852.

A.K. Kolah, N.S. Asthana, D.T. Vu, C.T. Lira, D.J. Miller, Ind. Eng. Chem. Res., 47 (2008) 5313.

M. Seiler, C. Jork, A. Kavarnou, W. Arlt, R. Hirsch, AIChE Journal, 50 (2004) 2439.

H. Song, Y.S. Huh, S.Y. Lee, H.W. Hong, Y.K. Hong, J. Biotechnol., 132 (2007) 445.

M. Patel, The BREW Project, Final report, Utrecht University, 2006.

JF. Wang, CX. Li, ZH. Wang, ZJ. Li, YB. Jiang, Fluid Phase Equilib., 255 (2007) 186.

Optimal Production Scheduling and Lot-sizing In Yoghurt Production Lines

Georgios M. Kopanos,[a] Luis Puigjaner,[a] Michael C. Georgiadis[b]

[a]*Universitat Politècnica de Catalunya-ETSEIB, Diagonal 647, Barcelona 08028, Spain*
[b]*University of Western Macedonia, Department of Engineering Informatics & Telecommunications, Karamanli & Lygeris, Kozani 50100, Greece, mgeorg@otenet.gr*

Abstract

The lot-sizing and production scheduling problem in a real-life multi-product yoghurt production line is addressed. A new discrete/continuous time representation mixed-integer linear programming model, based on the definition of families of products, is presented. It is mainly optimized the packaging stage while timing and capacity constraints are imposed with respect to the pasteurization/homogenization and fermentation stage. Packaging units operate in parallel and may share common resources. Sequence-dependent times and costs are also explicitly taken into account. An industrial case study is presented wherein production bottlenecks are revealed and several retrofit design options are proposed to enhance the production capacity and flexibility of the plant.

Keywords: Production scheduling, lot-sizing, dairy industry, MILP model

1. Introduction

Yoghurt production could be considered as a particular case of a batch or a semi-continuous production process. The PSE research community has studied these types of production processes during the last 20 years. One of the main features of batch processes is that a large number of products are produced from a few initial product recipes. The same holds for yoghurt production. Final yoghurt products may differ in at least one of the following features: (i) fermentation recipe type origin, (ii) total cup weight, (iii) number of cups per piece, (iv) labeling depending on their customer destination, (v) flavors, and (vi) packaging cup type (material, shape, etc.). Packing rates may significantly vary from one product to another.

A plethora of contributions addressing production scheduling problems can be found in the Operational Research and PSE communities' literature. However, the use of optimization-based techniques for scheduling dairy plants is still in its infancy. This can be mainly attributed to the complex production recipes, the large number of products to be produced under tight operating and quality constraints and the existence of mixed batch and semi-continuous production modes. An excellent review covering the short-term batch and continuous process scheduling can be found in Méndez et al. (2006).

Few attempts regarding production planning in yoghurt production lines can be found in the literature (Entrup et al., (2005), Marinelli et al., (2007)). In most works, changeover times/costs or fermentation stage restrictions were ignored. To the best of our knowledge, this work is the first that address the main processing features in yoghurt production lines in tandem.

2. Yoghurt Production Processes Description

The two main yoghurt product types are set and stirred yoghurt. Both types are subsequently subjected to cooling and packaging. Additionally, fruit and nuts may be added to stirred yoghurt where applicable. The principal difference between these two yoghurt types is that set yoghurt first passes from the packaging lines and afterwards is fermented in the final retail container. Figure 1 illustrates the main processing steps for producing stirred yoghurt.
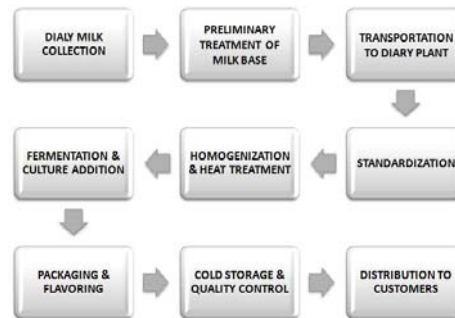
Figure 1: Yoghurt production process.

3. Problem Statement

3.1. Yoghurt production line description

The production line produces set, stirred or flavored yoghurt. Note that flavored yoghurt is stirred yoghurt with additional fruit (or other type) flavor. Thus, flavored yoghurt production should pass through fruit-mixer equipments in order to perform the addition and the mixing of fruit substances. The yoghurt production line consists of: (i) a set of cooling tanks (set yoghurt), (ii) a set of fermentation tanks (stirred and flavored yoghurt), (iii) 4 packaging units, and (iv) 2 fruit-mixer equipment units (see Fig. 2).

The short-term scheduling time horizon for yoghurt production is one week. Regular production is performed from Monday to Friday. Overtime is permitted on Sunday and/or on Saturday. Period production time is equal to 24 hours. Daily scheduled plant cleaning operations, *shutdown$_{jn}$*, last 2 hours, while before the start of the fermentation stage (including pasteurization, homogenization, etc.), which reflect the total plant setup time, *su$_{jn}$*, equal to 3 hours. Product demand data are packaging stage production targets and they are provided from the Logistics department.

3.2. Retrofitting alternatives

Some alternative retrofitting options revealed by having a closer look on the current yoghurt production line:

(i) Fruit-mixers are common resources that limit the total plant production capacity. For instance, packaging unit *J3* and packaging unit *J4* cannot package flavored yoghurt simultaneously (see Fig. 2); the same statement holds for packaging unit *J1* and packaging unit *J2*. Therefore, a relatively low cost fruit-mixers investment seems an alternative to increase the yoghurt production line capacity.

(ii) Each product p and/or product family f can be produced only to one packaging unit. Thus, the production process seems to be lacking of flexibility. Discussions with the plant manager revealed that it is possible to install a low-cost manifolds' investment in packaging unit *J4* in order to package more product families. With the current operating polocyte, packaging unit *J4* can only process product family *F23*. Note that packaging unit *J4* could process 5 product families (*F6*, *F8*, *F9*, *F10*, and *F23*) instead of just 1 product family (*F23*); if this investment takes place.

The current plant configuration will be referred as NFM, the fruit-mixer retrofit design option as FM, and the joint fruit-mixer and manifolds investment as FM&M.
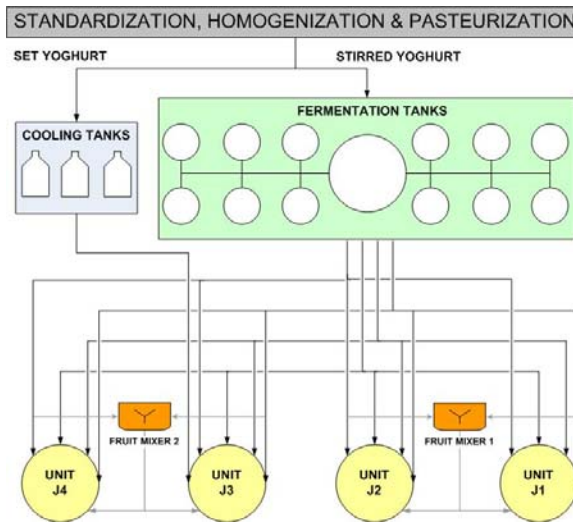
Figure 2: Yoghurt production line layout.

4. Conceptual Model Design

Production scheduling in dairy plants typically deals with a large number of products. Fortunately, many products illustrate similar processing characteristics. Products that share the same processing characteristics could be treated as a product family group. Products grouping significantly reduces the size of the underlying mathematical model and, thus, the necessary computational effort; without sacrificing any feasibility constraint. In the proposed approach products belong to the same product family *if and only if*: (i) they have the same fermentation recipe origin, (ii) there is no sequence-dependent setup time among them, and (iii) they share the same processing rate.

When changing the production between two products that are not based on the same recipe, it is always necessary to perform changeover cleaning and/or sterilizing operations. In dairy plants, a "natural" sequence of products often exists (e.g. from the lower taste to the stronger or from the brighter color to the darker) thus the sequence of products within a product family is known a priory. Therefore, when changing the production between two products of the same product family, the cleaning and sterilizing can be neglected. Hence, in dairy plants not only the sequence of products belonging to the same product family may be fixed but also the sequence of product families in each packaging line. In that case, different product families are enumerated according to their relative position in the production day. Note that a product families' demand is the aggregated demands of the products that belong to it.

5. Mathematical Formulation

The proposed model is a crossbreed between a continuous-time and a discrete-time model. Concretely, the production horizon is divided to production periods (days) whose material balances are modeled with a discrete time representation, while within each production day a continuous-time representation is adopted. It follows a brief description of our model. For lack of space, are not presented all equations here.

5.1. Lot-sizing and timing constraints

Product families' packaging times T_{fjn} lower and upper limits are gievn by Eq. (1). Packaging rates ρ_{fj} are fixed. Lower and upper bounds for the completion time C_{fjn} are given by Eqs. (2) and (3), respectively. Fermentation times t_f^{ferm} are included in Eq. (2).

$$t_{fjn}^{min} Y_{fjn} \leq T_{fjn} = \frac{Q_{fjn}}{\rho_{fj}} \leq t_{fjn}^{max} Y_{fjn} \quad \forall f, j \in (J^{pack} \cap FJ_f), n \tag{1}$$

$$C_{fjn} \geq (setup_{jn} + t_f^{ferm})Y_{fjn} + T_{fjn} + \sum_{f' \neq f, f' \in JF_j} sd_{f'fj} X_{f'fjn} \quad \forall f, j \in (J^{pack} \cap FJ_f), n \tag{2}$$

$$C_{fjn} \leq (hor_{jn} - shutdown_{jn})Y_{fjn} \quad \forall f, j \in (J^{pack} \cap FJ_f)n \tag{3}$$

5.2. Timing and sequencing constraints

Eq. (4) forces the starting time of a product family f' that follows another product family f on a packaging line j (i.e. $X_{ff'jn} = 1$) is greater than the completion time of product family f plus the necessary changeover time $sd_{ff'j}$ between them. M is a big number.

$$C_{fjn} + sd_{ff'j} \leq C_{f'jn} - T_{f'jn} + M(1 - X_{ff'jn}) \quad \forall f, f' \neq f, j \in (J^{pack} \cap FJ_f \cap FJ_{f'}), n \tag{4}$$

5.3. Allocation and sequencing constraints

Eqs. (5) and (6) state that if a product family f is allocated to packaging unit j at period n, at most one product family f is processed after and/or before it, respectively.

$$\sum_{f' \neq f, f' \in JF_j} X_{ff'jn} \leq Y_{fjn} \quad \forall f, j \in (J^{pack} \cap FJ_f), n \tag{5}$$

$$\sum_{f' \neq f, f' \in JF_j} X_{f'fjn} \leq Y_{fjn} \quad \forall f, j \in (J^{pack} \cap FJ_f), n \tag{6}$$

Eq. (7) denotes that the packaging unit j is used in period n, (i.e. $YJ_{jn} = 1$) if at least one product family f is assigned at period n. Eq. (8) states that the total number of sequencing binary variables plus the unit utilization binary variable should be equal to the total number of allocation binary variables in a packaging unit j at period n.

$$YJ_{jn} \leq Y_{fjn} \quad \forall f, j \in (J^{pack} \cap FJ_f), n \tag{7}$$

$$\sum_{f \in JF_j} \sum_{f' \neq f, f' \in JF_j} X_{ff'jn} + YJ_{jn} = \sum_{f \in JF_j} Y_{fjn} \quad \forall j \in J^{pack}, n \tag{8}$$

5.4. Fermentation stage constraints

Fermentation and pasteurization stage restrictions must be included to the mathematical model in order to obtain feasible production schedules. Eq. (9) states that the cumulative packaged quantity of product families f that come from the same fermentation recipe r should be greater than the minimum produced fermentation recipe amount in the pasteurization/fermentation stage and lower than the maximum production capacity. Moreover, a fermentation recipe r is produced at period n, (i.e. $YR_{rn} = 1$), if at least one product family $f \epsilon RF_r$ is packaged in a packaging unit j at the same period n.

$$QR_r^{min} YR_{rn} \leq \sum_{f \in RF_r} \sum_{j \in (J^{pack} \cap FJ_f)} Q_{fjn} \leq QR_r^{max} YR_{rn} \quad \forall r, n \tag{9}$$

Optimal Production Scheduling and Lot-sizing In Yoghurt Production Lines

$$YR_{rn} \geq \sum_{j \in (J^{pack} \cap FJ_f)} Y_{fjn} \quad \forall r, f \in RF_r, n \tag{10}$$

5.5. Common resources constraints

Appropriate common resource constraints regarding fruit-mixer equipments are also included in order to cope with the current production line configuration (NFM).

5.6. Mass balance constraints

In fresh food industry, backordering is not allowed; unsatisfied demand is lost. The total product family quantity produced at period n should not exceed the cumulative demand for the same product family f for all periods equal to or greater than actual period n, as Eq. (11) states. Customers' demand satisfaction is forced by Eq. (12). The inventory St_{fn} of product family f is estimated through the traditional mass balance of Eq. (13).

$$\sum_{j \in (J^{pack} \cap FJ_f)} \sum_n Q_{fjn} \leq \sum_{n' \geq n}^N dem_{fn'} \quad \forall f, n \tag{11}$$

$$\sum_{j \in (J^{pack} \cap FJ_f)} \sum_n Q_{fjn} \geq \sum_n dem_{fn} \quad \forall f \tag{12}$$

$$St_{fn} \geq St_{fn-1} + \sum_{j \in (J^{pack} \cap FJ_f)} Q_{fjn} - dem_{fn} \quad \forall f, n \tag{13}$$

5.7. Objective function

The optimization goal to be minimized is the total cost that includes several cost-related factors such as: (i) inventory costs, (ii) operating costs, (iii) fermentation recipes preparing costs, (iv) unit utilization costs, and (v) product families' changeover costs.

6. Industrial Case Study

An industrial case study of 93 final products that have been grouped into 23 product families is addressed here. All models have been resolved using CPLEX 11.0 solver via the GAMS 22.8 interface. The optimal solution has been reached in all cases at a very low computational time. The bigger-size NFM model gave the optimal solution in 22 CPU seconds while FM and FM&M cases were solved in half a second.
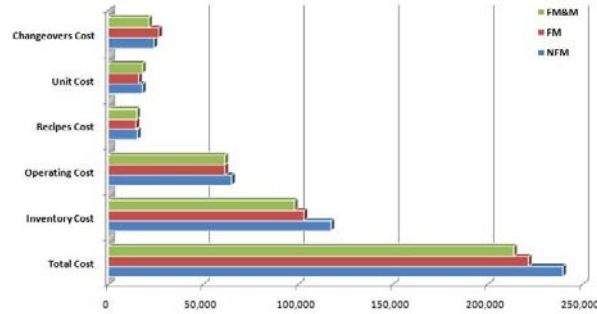
Figure 3: Cost comparative analysis (€).

FM model's solution shows a *7.6%* improvement over that of the NFM. The FM&M total cost is *10.8%* lower than that of the NFM model. A visual representation is illustrated in Fig. 3. Note that FM and FM&M configurations lead to lower total inventory cost compared to that of the NFM (*12.2%* and *16.4%*, respectively). Dairy plant opens on Sunday (*n0*) in the FM case, in order to achieve full demand satisfaction,

resulting into higher operating costs. The other alternatives are capable of satisfying the demand profile without overtimes. Fig. 4 presents the Gantt charts for all cases.
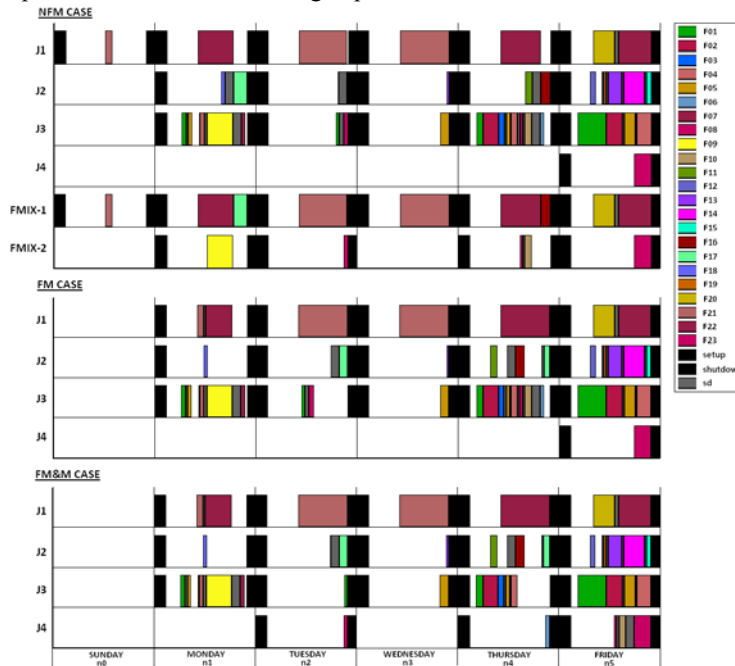
Figure 4: Production schedules for all plant configurations.

7. Final Considerations

This model aims at being the core element of a computer-aided advanced scheduling and planning system in order to facilitate decision-making in dairy plant industrial environments. A salient feature of the dairy industry is that the customers usually confirm (i.e. change) their order quantities just prior to dispatch, thus the production should be changed on the fly. Yoghurt is a perishable product and strategies of building up inventories are inappropriate since they compromise its quality, its selling price, its freshness, and its goodness. Therefore, a reactive production planning approach to address these problems in this kind of industries is a challenging research direction.

8. Acknowledgements

References

[1] C.A. Méndez, J. Cerdá, I.E. Grossmann, I. Harjunkoski and M. Fahl, Comput. Chem. Eng. 30 (2006) 913.
[2] M.L. Entrup, H.-O. Günther, P. Van Beek, M. Grunow and T. Seiler, International Journal of Production Research 43 (2005) 5071.
[3] F. Marinelli, M.E. Nenni, A. Sforza, Annals of Operations Research 150 (2007) 177.

Refining Scheduling of Crude Oil Unloading, Storing, and Processing Considering Production Level Cost

G. Robertson[a], A. Palazoglu[b],J.A. Romagnoli[a]

[a] *Louisiana State University, Baton Rouge, LA 70803 U.S.A., jose@lsu.edu*
[b] *University of California, Davis, CA, 95616 U.S.A., anpalazoglu@ucdavis.edu*

Abstract

This paper presents an integrated optimization approach for crude operations scheduling and production for refineries. The production process is composed of a pre-fractionator, crude, and vacuum distillation columns. It is modeled as an NLP. The scheduling problem is composed of unloading operations and simultaneous blending and charging of CDUs. It is modeled as a MILP. The nonlinear simulation model for the production process is used to derive individual crude costs for the two crudes considered (Dubai and Masila). This is performed using multiple linear regressions of the individual crude oil flow rates around the crude oil percentage range allowed by the production facility. These individual crude costs are then used to derive a linear cost function that is optimized in the MILP scheduling model, along with logistics costs. Results show that this integrated approach can lead to a 0.53 M$ decrease in production and logistics costs in a 15 day time horizon.

Keywords: multilevel simulation, refinery scheduling, process optimization

1. Introduction

Process industry supply chains are striving to improve efficiency (Shah, 2005). Resources can be used more efficiently by ensuring local objective functions along the supply chain are not undermining overall goals. Combining objectives is computationally burdensome so heuristics are used to narrow the solution space before integration can fine-tune the overall objective function. Along the crude oil supply chain, there exists an unloading/loading scheduling problem followed by a refinery operating problem.

Modern refining has become an extremely competitive business because of the deteriorating quality of crude oil coupled with tighter product specifications and more stringent environmental regulations. Therefore, refineries today receive many different types of crude from a variety of places. Refineries frequently change a unit's operating conditions to reduce operating expenses including environmental impact. A general description of the oil production system considered in our analysis, which consists of vessels, docks, storage tanks, and a separation train, is shown in Figure 1.
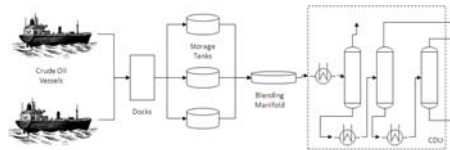
Figure 1: Short Term Schedule

Along the oil supply chain, vessels or tankers carrying crude arrive at the docking station according to a schedule whose duration is known as the time horizon. It is then decided in what manner the crude oil is unloaded from the vessels to storage tanks during the time horizon. The crude oil in the storage tanks is transferred to a blending manifold where it enters the separation train in a desired composition ratio range determined at the production level. We identify two local objectives along the crude oil supply chain. The crude scheduling problem (*CSP)* receives the shipping vessel's schedule, including arrival times, type, amounts, the *CDU* demand, and blend range. The scheduler of the crude loading and unloading decides which tanks to store the incoming crude oil and which tanks should feed the refinery distillation units (*CDU/VDU*). The scheduler does so in a fashion which minimizes logistical costs such as inventory holding, sea-waiting, and setup cost, while feeding the separation train with the proper blend of crude oil. The production unit planner manipulates unit operating conditions in order to optimize the energy integration of the fractionating section of the refinery as well as minimize environmental cost such as the burning of high sulfur fuels in furnaces.

The traditional approach to the crude scheduling problem *(CSP)* for a refinery is a discrete time optimization formulation where the scheduling horizon is split into time intervals of equal size and binary variables are used to indicate if an action starts or terminates during this time interval (Saharidis et al., 2009). Various mathematical models have been developed to solve the CSP. The objective functions of these solutions include cost incurred for waiting sea vessels, unloading cost, inventory cost, etc. Yet, the blend of the crude affects the refining cost even when in the operational range. The effect of combining these objectives has not yet been published.

2. Problem Statement

This section contains the specifics of a typical scenario which was analyzed in this work.

2.1 The Process: Primary Units of a Crude Oil Refinery

The separation train considered in this study consists of pre-fractionator, atmospheric, and vacuum distillation columns each with a preheat train ending with a furnace to elevate the crude feed temperature. Masila crude is blended with lighter Dubai crude for refining. This crude oil is separated by vapor pressures into fluids with differing properties including naphtha, kerosene, diesel, gas oil, etc. The crude is first heated to approximately 245°C before entering the 30 stage pre-fractionator distillation column which takes off light gas and light naphtha in order to reduce the vapor load in the distillation column. The pre-topped crude is then further heated before entering the *CDU* where heavy naphtha, two kerosene grades, and diesel side draws are taken off. The bottoms are put under a vacuum, heated further and separated in the *VDU* into another diesel stream, light vacuum gas oil, heavy vacuum gas oil, sour diesel, and vacuum residuum. Product side draws lead into steam strippers and steam is blown up though each main column. In order to recover as much heat as possible from the

distillation units, pump around streams exist and along with the product stream, compose the preheat trains. Some separated products are finished while others must be further treated, but each can be assigned a value.

On the production level, personnel must maximize the amount of valuable product considering the environmental and utility cost of the separation train by manipulating steam to the main column, the steam to the side strippers, pump around flow rates, side cuts, as well as operating conditions such as temperature. The complex heat integration schemes and the interactive nature of the process due to the presence of pump around and side-strippers make it difficult to operate at the optimal conditions. Each product streams must meet certain specifications such as density and compositions, but side cuts of the distillation columns have room for slight manipulation while still meeting these specifications. Product draws are then stripped with steam in side strippers. The heat integration strategy recovers as much heat as possible from the distillation process by recovering heat from final products as well as pump around streams. *The Decision Variables* are the stripping steam mass flow rates, product flow rates, pump around flows, reflux rate, and atmospheric & vacuum furnace outlet temperatures (25 in our model). The constraints are the quality parameters, temperature of the product flows, furnace duty, and the bounds on the Decision Variables. A more detailed problem description can be found in Yela and Romagnoli (2008).

2.2 Short Term Refinery Scheduling problem

In our case study, one dock is used to feed one refinery CDU with six tanks as intermediate storage. Tanks 1,3, and 6 initially contain 90,000 m^3 Dubai, 40,000 m^3 Masila, and 99,000 m^3 Masila respectively.The Masila crude is blended with lighter Dubai crude for refining. The time horizon is 15 days with discrete time intervals of 1 hour. A 90,000 m^3 shipment of Dubai is scheduled to arrive in the first time interval and will continuously unload for 36 hours. Similarly, a ship is scheduled to arrive on the 51st time interval and unload an addition 90,000 m^3 of Masila. The demand for the CDU is 765 m^3/hr throughout the duration of the time horizon, and the flow rate from the tanks to blender CDU has a limitation of 600 m^3. The setup of a pipeline is a lengthy procedure including filling the pipeline and sampling its contents, so the setup cost plus penalty are accessed at $5,000. The CDU component concentration ranges between 32.5 vol%-37.5vol% Dubai.

The following assumptions can be made for our problem: the amount of crude oil remaining in the pipeline is neglected; due to their small value in comparison with the scheduling horizon change over times are neglected; it is also assumed that there is perfect mixing in the blending manifold tanks and any additional mixing time is neglected in this model.

The decision variables that are determined at this level are: flow rates from vessel to storage tank for each vessel; flow rates from storage tank to CDUs for each storage tank; crude oil inventory levels in storage tanks for each time interval; series of crude oil blends to be charged in each CDUs under optimum costs; periods where connections are established (or setups) or broken.

3. Refinery Simulation/optimization and MILP scheduling model

This section presents models and objective functions used in the solution approach to solving the presented problem.

3.1 Production Level Model/Optimization

Objective Function: The goal in the production planning level is maximizing revenue. The objective function used includes the cost associated with the feed, products, utilities, and environmental effects. Due to the rising concerns on global warming and with implementation of emissions trading programs ("cap and trade"), the triple bottom line objective function was used (Yela&Romagnoli, 2008).

$$TripleBottomLine = ProfitFuction - EnvironmentalCosts + SustainableCredits \qquad (1)$$

$$ProfitFucntion = RevenueofProducts - UtilityCosts \qquad (2)$$

*In Eq. (1) *Environmental Costs* is the cost required to comply with environmental regulations including permits, monitoring emissions, fines, etc; *Sustainable credits* represents the credits given to the processes that consume CO_2. In this study, sulphur dioxide (SO_2), carbon dioxide (CO_2), and nitrogen oxides (NOx) are chosen as Environmental Load.

The modelling equations include thermodynamic relationships, mass balances, and energy balances obtained using Hysys software. The NLP optimization model is solved with Frontline Systems' Premium Solver Platform, and a bridge code is programmed in Visual Basic Application allowing the user to import and export selected variables between the HYSYS model and Excel worksheet.

3.2 Scheduling Solution Approach

The objective function consists of logistical costs such as loading/unloading and refining costs of the blend entering the separation train. In our scenario sea-waiting times and inventory costs are negligible making setup the only the logistical cost.

Model equations are constructed by combining material balances for the vessel, storage tanks, and operation rules for arrival and departure of vessels as well as for crude oil charging. GAMS optimization software with CPLEX solver was used for the bilinear mixed integer linear loading /unloading scheduling problem. The complete model can be found in Saharidis et al. (2009) using flexible recipe/ blending in manifold model.

3.3 Integration of the two models

To account for the operational cost of crude oil blend entering the CDU, a linear relationship between individual crude flow rates and total refining cost was embedded into the MILP's objective function. For several crude flow rates Eq. (1) is maximized and the associated total costs are tabulated. The refining cost from the separation optimization is calculated for the desired range of operation set for the production level and a cost equation as a function of the individual crude feed flow rates, Y_i, is created:

$$Cost(Y_1, Y_2, ..., Y_N) = \prod_{n=1}^{N} a_n Y_n \qquad (3)$$

Refining Scheduling of Crude Oil Unloading, Storing, and Processing Considering Production Level Cost

The coefficients, a_i, of Eq. 3 are determined by multiple linear regressions around the range allowable for the production facility. This refining cost function is then embedded into the scheduling and planning MILP objective function.

The optimization process first takes the variables of the triple bottom line objective function from the Hysys library for the production layer model. The VBA bridge code then embeds them into an Excel spreadsheet where Frontline chooses the next set of variables to insert into the Hysys model. In each of the iterations, the total cost of the refinery operations is embedded into the Excel spread sheet. After the decision variables have reached a maximum, the cost of the profit optimized refinery are tabulated over the production level range.

4. Results and Discussion

Table 1 illustrates an optimal production unit operation conditions for particular feed flow rates.

Table 1: Summary of decision variables for production level

Decision Variables	Optimal value	Constraints (Min)	Constraints (Max)
LN rate, m3/hr	12.05	10.00	14.00
HN rate, m3/hr	27.51	26.00	30.00
Kerosene-1 rate, m3/hr	98.57	95.00	99.00
Kerosene-2 rate, m3/hr	46.12	44.00	49.00
Diesel rate, m3/hr	106.03	102.00	107.00
ADU feed Temp., degC	372.00	372.00	373.00
VDU feed Temp. degC	398.00	398.00	400.00

The refining cost for the maximized profit (Eq. 1) is tabulated for each of the twelve Dubai-Masila flow rate sets, each whose sum is 756 m^3/hr and whose ratio is within the production allowable range. Table 2 shows data points created using the production optimization/simulation to determine the relationship between the crude blends, refining cost, revenue, and predicted values from that fitting function.

Table 2 describes some of the key process (decision) variables and the product flows at a specific feed flow rate of Masila and Dhubai.

*Note difficult conditions to change such as temperature are tightly constrained.

Table 2: Data points for correlation

Dubai Flow (m^3/hr)	Masila Flow (m^3/hr)	vol%	Refining cost	Predicted cost	Residuals
286.6	478.6	0.371	33981.81	33928.5	53.31315
283.8	481.5	0.370	34240.47	33954	286.4732
278	487.3	0.363	34096.08	33997.06	99.02241
275.2	490.1	0.359	33621.26	34017.85	-396.586
269.4	495.9	0.352	33654.13	34060.91	-406.776
260.8	504.6	0.340737	34288.87	34129.47	159.4029
258	507.4	0.337	33794.59	34150.26	-355.665

The residuals from the total cost function were randomly distributed within the blend range. The linear Regression provides the following equations used for predicting cost on the scheduling level:

$$Total\ Cost = 39.7\$/m^3 * Y_{Dubai} + 47.1\$/m^3 * Y_{Masila} \tag{4}$$

Table 3 demonstrates the result of the scheduler adding the operational cost to the MILP's objective function. A tabulation of the number of setups comprises the first row's entries when the scheduler minimizes the logistical costs (setup cost in our scenario), while still abiding by the production level blend specifications.

Table 3: Effects of Operational Cost

Objective	Setup	Total cost (M$)
Min(setup cost)	4	4.52
Min(total cost)	4	3.99
Benefit	0	.53

The returned total cost was then calculated. It was found even when the setup was the same, i.e. the logistical costs were similar, the operational cost could be minimized further. In other words, there exists a solution space for the general objective function. Therefore adding the cost function picks a particular solution with the lowest refining cost. The second row indicates the results from minimizing the sum of refining and setup costs which gave the minimal total cost. Although there are many solutions to one problem, this particular solution can lead to a total cost of 4.52M$. Therefore, there is a potential reduction in cost of $530,000 by combining the different levels.

5. Conclusions

An optimization approach was presented, integrating the optimal schedule for the short-term refinery loading, unloading, and production problems. Production optimization and production level costs were considered in the tactical scheduling problem of a refinery. The strategy proposed accounts for these costs in a simple yet effective manner. The results show a significant decrease in the total operational cost for the refinery.

References

[1] N. Shah, Process industry supply chains: Advances and Challenges, Computers and Chemical Engineering 29. Elsevier., 1225–1235 (2005)

[2] G. Saharidis., M.Minoux., Y. Dallery, Scheduling of Loading and Unloading of Crude Oil in a Refinery using Event Based time Formulation, Computers and Chemical Engineering, 2009

[3] S. Yela, O. Galán, J.A. Romagnoli (2008), Framework for Operability Assessment of Production Facilities: An Application to a Primary Unit of a Crude Oil Refinery, *AIChE Spring Meeting*, Apr 6-10, New Orleans, LA

An Application of a Cocitation-Analysis Method to Find Further Research Possibilities on the Area of Scheduling Problems

T. Kocsis[a-b], S. Negny[a], P. Floquet[a], X. Meyer[a], E. Rév[b]

[a]Université de Toulouse LGC UMR 5503 – INPT ENSIACET, 4 allée Emile Monso BP 74 2333, 31432 Toulouse Cedex 4, FRANCE
[b]Budapest University of Technics and Economics, Department of Chemical and Environmental Process Engineering, Budafoki út 8, 1111 Budapest, HUNGARY, kocsis.tibor@gmail.com

Abstract

In this article we will give firstly a classification scheme of scheduling problems and their solving methods. The main aspects under examination are the following: machine and secondary resources, constraints, objective functions, uncertainty, mathematical models and adapted solution methods.

In a second part, based on this scheme, we will examine a corpus of 60 main articles (1015 citation links were recorded in total) in scheduling literature from 1977 to 2009. The main purpose is to discover the underlying themes within the literature and to examine how they have evolved. To identify documents likely to be closely related, we are going to use the cocitation-based method of Greene et al. (2008).

Our aim is to build a base of articles in order to extract the much developed research themes and find the less examined ones as well, and then try to discuss the reasons of the poorly investigation of some areas.

Keywords: Scheduling, Classification, Cocitation Analysis, Review

1. Introduction

Scheduling is a critical issue in process operations for improving production performance. In the last twenty years, there have been significant research efforts regarding this area, and several excellent reviews have been published. Esquirol – Lopez (1999) showed the areas of scheduling with a detailed review on them, Mendez (2006) and Pinedo (2008) delivered articles classifying scheduling problems and their solution methods.

A brief description of the problem is the following (Taillard (1989)): n jobs have to be performed on m unrelated machines; usually every job consists of m non-preempting operations. Every operation of a job uses a different machine during a given time, and usually (but not always) may wait before being processed.

The aim of this paper is to deliver a classification of these problems, and to find out which areas are well examined, and which ones less. To find the interesting areas we are going to use the method described by Greene et al. (2008), who used a special technique based on cocitation-analysis to give a classification and show the connecting areas of their research area, Case-Based Reasoning (CBR) problems.

2. Classification of scheduling problems and their solution methods

We classify the problems under the following main aspects: machines and secondary resources, constraints, objectives, uncertainty, mathematical models and adapted solution methods.

The first aspect is the number and connection scheme of **machines**. We can speak about one machine, parallel machines and more machines. A job is a set of operations visiting different machines and sequenced in a linear manner along a chain. By the way followed by jobs there are 3 main categories: FlowShop Scheduling Problems (FSSP), JobShop Scheduling Problems (JSSP) and OpenShop Scheduling Problems (OSSP). Certainly, flow shop problem is the most examined one. We can find details, and also methods in Taillard (1989) to generate, examine and solve problems.

Sometimes tasks require multiple **resources**, such as a machine operator. These resources can be renewable (disjunctive or cumulative), or consumable. An example is employee timetabling.

We can distinguish **constraints** of functionality, techniques, societals, economics, life-cycle. Among them most frequently we meet functionality constraints, defining special cases, which are in some way different from the typical one, i.e. they do not use, partially or at all, the assumptions applied in general. For example: preemption, stocking (capacity or waiting time), process-time dependence on size, transit time, overlap, due-dates, cleaning/maintenance, availability of resources, etc.

Objective functions can be time-based, resource-based, cost-based, income-based, environmentals and even multi-objectives.

Many research studies focus on optimizing the makespan under static conditions and do not take into consideration **dynamic disturbances** such as machine breakdown or new job arrivals. Regarding to this aspect we can distinguish static and dynamic shop scheduling problems. There are many attempts to describe stochastic problems also, i.e. when we work with attended values and variances instead of fixed starting times and durations.

The **mathematical model** of a problem contains an objective function (to be minimized or maximized), and several constraints. Regarding to linearity there are linear and nonlinear models, regarding to variables we can speak about integer, continuous, or mixed integer problems. By difficulty it can be polynomial (i.e. in function of the number of variables there are an algorithm that can find the solution in a polynomial time) or NP (i.e. the time needed is depending on the size of the problem exponentially). Mainly in the process scheduling models we can distinguish also the time representation (discrete-continuous) models, material balance (network flow or lots), and event-representation (time intervals, time events, time slots) models.

The most important **solution methods** are the following ones: exact methods, artificial intelligence methods, heuristics, metaheuristics and decomposition strategies.

Exact methods use the technique of mathematical programming, especially the mixed-integer linear programming methods (most scheduling problem can be formulated as an MILP problem). *AI methods* are constraint-satisfaction programming (CSP), case-based reasoning applications, neural network, multiagent-systems, expert systems, logic flow, etc. For more details see Gabot (2009). *Heuristic methods* can be constructive, which from zero construct a new solution, and ameliorative creating a solution as a skeleton, and step by step try to ameliorate it. We mention here the well-known NEH heuristic (from the names of the authors: Nawaz, Encore and Ham).

. Most heuristics are due-date or process-time based ones. *Metaheuristic methods* in general need some existing solutions to ameliorate. So, they can used to complete heuristic methods. They are inspired by some *analogy* for example from physics

An application of a cocitation-analysis method to find further research possibilities on the area of Scheduling problems

(simulated annealing, simulated diffusion), biology (genetic algorithms, taboo search), or ethology (ant colony optimization). A good review in this area is the book of Xhafa and Abraham (2008). Finally, there are *decomposition strategies* regarding to time, machines and tasks or resources. A famous example of decomposition by time is the so-called Rolling Horizon technique.

3. Cocitation analysis

To identify documents likely to be closely related, two techniques have been devised: bibliographic coupling and cocitation analysis (Smith (1981)). Two documents are bibliographically coupled if their reference lists share one or more of the same cited papers, and two documents are cocited when they are jointly cited in one or more subsequently published documents.

In their work Greene et al (2008) set out to examine the themes that have evolved in their domain research as revealed by the implicit and explicit relationships between the conference papers, after 15 years of conferences. They have found that a clustering based on co-citation of papers appears to produce the most meaningful organization. Their idea inspired us to apply a similar method to the area of scheduling.

The concept of cocitation analysis is illustrated in Fig. 1. Let for example P1 and P2 be two related articles. The fact that P3 and P4 are cited both by P1 and P2 indicates a strong relationship between these papers. In this example cocitation analysis suggests a weaker relationship between P3 and P5 and P4 and P5 based on cocitation in P2.
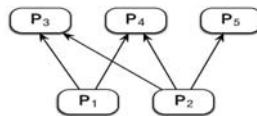
Fig. 1. The concept of cocitation

Rather than using raw co-citation indices as a basis for measuring the similarity between papers, they use a so-called CoCit-Score which has been shown to be a particularly effective choice for clustering co-citation data. The similarity between a pair of papers (Pi, Pj) is given by normalizing their co-citation frequency with respect to the minimum and mean of the pair's respective citation counts:

$$S_{ij} = \frac{C_{ij}^{2}}{\min(C_{ii}, C_{jj}) \times mean(C_{ii}, C_{jj})} \tag{1}$$

To illustrate this equation, suppose that Paper A is cited 26 times, Paper B 18 times, and they are cited together 7 times. In this case $S_{ij} = \dfrac{7^{2}}{\min(18,26) \times mean(18,26)} = 0.124$

Each entry is now in the range [0,1], where a larger value indicates a stronger association between the papers. When we know these association values, we can arbitrarily choose a value (for example 0.05) which we treat as an association „strong enough". Using these values we classify the articles into groups, and regarding these groups we can distinguish the larger research areas existing in the literature. Regarding to the timeline it is possible by this way to show the intensity of research, i.e. the importance of an article in the research area. It can also be seen whether an area is under research, because in this case we see the dynamism on the diagram, or it is well researched, when we see a large intensity during a couple of year, and fewer and fewer articles later.

If we represent papers by bubbles on the diagram, the size of a bubble can respect to the total number of citation of an article, and by that we can deduct to the relevance of the adequat paper.

In the next chapter, we will show an example of application of this method, in our case to the scheduling literature.

4. Application of the method to scheduling literature

To examine the connections among articles in the literature of scheduling problems, we used Elsevier Science Direct©, to find works in this area. Arbitrarily we have chosen 60 articles, and followed the 1015 cocitation links in the search motor of Science Direct. Several articles have not yet been cited because they are new ones. The results for each article have been saved by Reference Manager©. The databases have been exported to Microsoft Excel, and sorted in order to find out cocitations, and to calculate the CoCitScore values. The sum of these values for an article we call global centrality. There is a local degree centrality too, based on co-citation counts from the CoCit Score values. Firstly, for each cluster we assign the papers if their previous membership weight for that cluster exceeds a given threshold (we used 0.1 as Greene et al (2008) proposed). This yields to scores in the range [0,1], where a higher score indicates that a paper is more influential in the area of research.

4.1. Analysis

First we will make our remarks on the articles of scheduling literature in general, and later we will show the results for the two special areas. On Fig. 2. gray bubble means a production scheduling article, and white a process scheduling one. We can see centrality values of the examined articles on the timeline (the size of a bubble shows the total citation number of the corresponding article).

Scheduling literature is far from being a finished, well-researched area yet, there is significant activity in last years also. After some very important basic article from the early nineties, the area is in dynamic evolution, and this trend seems to be continued.

Now, if we regard the cocitations, among the articles we can find two significant groups of papers. The first one belongs to process scheduling, and the second one to production scheduling. For each area there is actually significant research activity, and these two groups are being examined in a parallel way.
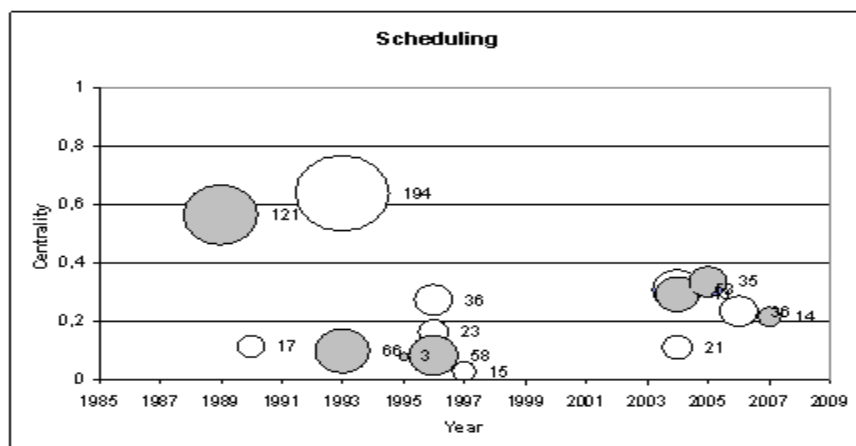
Fig.2. Some relevant paper in scheduling literature

An application of a cocitation-analysis method to find further research possibilities on the area of Scheduling problems

4.2. Comments on results and referring to classification

In this part we will make the connection between our classification scheme, and the founded two domains in the literature. We are going to try to discuss why certain types are typical to both of these domains, to one and not the other, or to neither of them.

4.2.1. Machines

Regarding to the number of machines both area focuses mostly to more machine problems. Of course there are several theoretical studies for the special cases of one or parallel machines, but in real life most often we meet with more machine problems.

In the **production scheduling** (PDS) area, the permutation flowshop problems are clearly dominating the others. Non-permutation flow shop and job shop problems are also well treated, but there are only a few articles on open shop problems. Closed circle is rarely appears in the network. Poorly are examined yet the hybrid shops (flexibility).

On the other hand, in the process engineering, **process scheduling** (PSS) area multiproduct and multipurpose batch plant scheduling problems are the most examined ones; literature calls them flow shop and job shop problems also (but these expressions in this context means not exactly the same as generally). Several excellent reviews can be found, for example Mendez et al. (2006). Flexibility is widespread, machines are usually independents.

4.2.2. Secondary resources

In PDS, secondary resources are rarely treated; an exception is integrated scheduling and employee timetabling problems. However, in PSS we can find more examples, especially energy-requirements of machines, raw materials or other resources.

4.2.3. Constraints

Regarding functionality constraints an unexamined aspect is the limited wait time, for both areas. We can find several examples for unlimited and zero wait problems, but only few for the limited ones. Nevertheless, stocking and waiting-time constraints are more often in PSS, because of the problem structure, and the evident need of material treating requirements. There is no differentiation made between finite intermediate storage and limited waiting time, maybe for practical reasons.

Process times are usually invariable for PDS but often depend on size or resource linearly (or sometimes even exponentially) in PSS. Exponential dependence is much less often, because of the need of an MINLP model in this case, which would be much more difficult than an MILP one.

Overlapping and connections between jobs are also rare especially for PDS area (only few exception, maybe because it is a really hard problem), maintenance or cleaning are rarely taken into account, and we've found no example for the time-dependent maintenance. Process dependent maintenance occurs almost only in PSS. Overlapping or availability information can be included in the superstructure representation of a PSS much easier, on the other hand transit time is easier to deal with in PDS.

Another unexamined area is the one of non-functional constraints, practically no example (in our base) for technical, societal or lifecycle constraints in neither area. However, there are some exceptions on the taking into account environmental considerations for PSS.

4.2.4. Objective functions

In PDS problems the objective function most often is the makespan, or some other function of time-based type in PDS area, and there is practically no example for resource, revenue, environmental or cost-based criteria. In PSS cost or revenue-based objective functions are typical, most frequently the total profit.

4.2.5. Uncertainty

Dynamic disturbances such as machine breakdown or new job arrival are taken into account sometimes in PDS problems, but not widespread in PSS. However it is not too difficult to deal with, for example by modifying the superstructure and use a Rolling Horizon technique.

Stochastic approaches on the other hand we found for some PSS problems.

4.2.6. Mathematical models and adapted solution methods

For both areas mathematical models are mixed integer linear ones. As we mentioned before, in PSS the exponential dependence causes an MINLP model. In PDS precedence-based ones dominate, most frequently articles treat N~M problems. On the other hand, in PSS problems event representation is an important characteristic of the model, time-interval based models are the oldest ones, global and unit-specific time point based models are newer. First ones are sometimes (if we have not enough intervals) not enough accurate, but later ones usually need more computational time.

Material balances are important in PSS, but not so in PDS, as here we have to deal usually only with the continuity of a job, and not with quantities. In PDS we also don't need time-representation equations. Despite of that, because of the objective function, PDS problems are usually hard to deal with exact methods. Most often we find heuristic and metaheuristic solution methods for them.

On the other hand, for the PSS problems we can meet more often exact methods. Heuristics are appearing only for larger problems. A specific method family is the decomposition strategy, which is missing in production scheduling. The most famous one is the so-called Rolling Horizon.

For both great area can be said, that Artificial Intelligency methods are not widespread yet. Except for the CSP the application of AI methods, especially CBR or neuron network techniques deserve further research without doubt.

5. Conclusion

We discussed the possibilities and classification of scheduling problems and their solution methods, and regarded several articles. To find poorly examined areas we used the cocitation-based method of Greene et al. (2008). By constructing a base of articles we have shown the much and the less examined areas.

We could see that still exist problem types to be focused on. The application of AI methods and the effects of technical, societal and environmental constraints are to be researched. Some special possibilities have been proposed to further examination. We see continuous research which is far from end, so deserves further work.

References

Esquirol – Lopez: L'ordonnancement. Paris, Ed ECONOMICA, 1999

Bernard Grabot: Ordonnancement d'ateliers manufacturiers. http://www.techniques-ingenieur.fr/book/ag3015/ordonnancement-d-ateliers-manufacturiers.html (2009)

Greene, Freyne, Smyth, Cunningham: An analysis of research themes in the CBR conference literature In: K-D Althoff et al: ECCBR 2008, LNAI 5239 pp 18-43, 2008 Berlin, Springer

Mendez Carlos A. ; Cerda Jaime ; Grossmann Ignacio E. ; Harjunkoski Iiro ; Fahl Marco: State-of-the-art review of optimization methods for short-term scheduling of batch processes Computers & Chemical Engineering 30 (2006) 913-946

Pinedo: Scheduling: Theory, algorithms and systems. Prentice Hall, 2008

Smith: Cocitation analysis. In: W.G. Potter: Bibliometrics 83-106 Illinois, 1981

Taillard: Benchmarks for basic scheduling problems ORWP89/21 Dec 1989

Xhafa – Abraham: Metaheuristics for scheduling in industrial and manufacturing applications. Springer, 2008

Rolling-Horizon Algorithm for Scheduling under Time-Dependent Utility Pricing and Availability

Pedro M. Castro,[a] Iiro Harjunkoski,[b] Ignacio E. Grossmann[c]

[a]Energy Systems Modeling and Optimization Unit, LNEG, 1649-038 Lisboa, Portugal
[b]ABB Corporate Research Center, Wallstadter Str. 59, 68526 Ladenburg, Germany
[c]Dept. Chemical Engineering, Carnegie Mellon University, Pittsburgh 15213, USA

Abstract

This work addresses the scheduling of continuous single-stage multiproduct plants with energy intensive processing tasks and time-dependent electricity cost and power supply. A new rolling horizon algorithm is proposed that consists of a planning model to predict the production levels and a continuous-time model for detailed scheduling. The results from a set of test problems from the literature show that the algorithm can generate global optimal solutions much more rapidly than standalone discrete or continuous-time formulations in problems involving unlimited power availability.

Keywords: Continuous-time, Resource-Task Network, Planning, Electricity

1. Introduction

Modern enterprises are complex global networks consisting of multiple business units and functions. In order to remain competitive in the global marketplace, companies need to optimize the various functions that comprise the supply chain. One of the focuses of Enterprise-Wide Optimization is the operation of manufacturing facilities, where essential operational items include planning and scheduling (Grossmann, 2005). A major challenge is the discovery of what type of models to use to render the effective solution of real-life problems.

This paper builds on our recent work (Castro et al. 2009), which has proposed a new continuous-time short-term scheduling formulation for continuous plants subject to energy constraints related to electricity pricing and availability. Despite the major breakthrough of effectively handling discrete events resulting from time-dependent utility cost/availability profiles and multiple intermediate due dates, the number of event points needed to represent a solution rapidly increases with the number of energy levels and demand points. As a consequence, the mixed-integer linear programming (MILP) formulation can only solve very small problems to optimality, when considering a one-week horizon with end-of-the-day demands and frequent energy changes.

Discrete events are handled more naturally with a discrete-time formulation but there are two important issues when addressing problems featuring continuous tasks. First, slightly suboptimal solutions may result since the task duration is approximated by a multiple of the prespecified interval length. Second, multiple instances will typically need to be executed to meet the daily demands, leading to high solution degeneracy. Thus, only by chance will one get solutions with a minimum number of changeovers that can actually be implemented in practice.

Clearly, discrete and continuous-time formulations have complementary strengths and the ideal approach should combine the advantages of both (Maravelias, 2005; Westerlund et al., 2007). In this paper, we propose a new rolling-horizon algorithm that can tackle the full problem in a sequence of iterations. The scheduling horizon is

divided into detailed and aggregate time blocks. The aggregate model comprises a non-uniform discrete-time model to predict the production levels without determining the timing of events. Detailed scheduling is achieved with the continuous-time model for a time window between consecutive demand points, while simultaneously considering the remaining horizon with the aggregate model.

2. Problem Definition

Consider the last processing stage of a continuous multiproduct plant. The intermediate material is transformed into one of a few possible products (*P*) in equipment units (*M*) through the use of electricity. These are then sent to storage units (*S*) where they wait for dispatch, see Fig. 1. Units are characterized by power requirements $pw_{p,m}$ [MW] and processing rates $\rho_{p,m}$ [ton/h]. Products may have multiple demands, occurring at any hour of the day, $d_{p,hr,dy}$ [ton]. The maximum capacity of storage units, which are shared, is cap_s [ton], while $im_{s,p}$ [ton] are the initial amounts in storage.

The objective will be to minimize total energy cost, for a given energy contract between the plant and electricity provider that specifies time-dependent electricity cost, $ec_{hr,dy}$ [€/kWh] and maximum power levels, $pwx_{hr,dy}$ [MW].
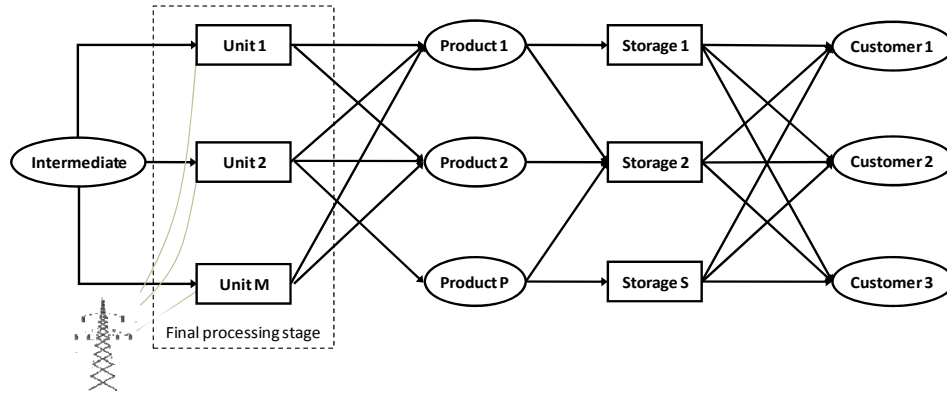
Figure 1. Generic representation of industrial case study
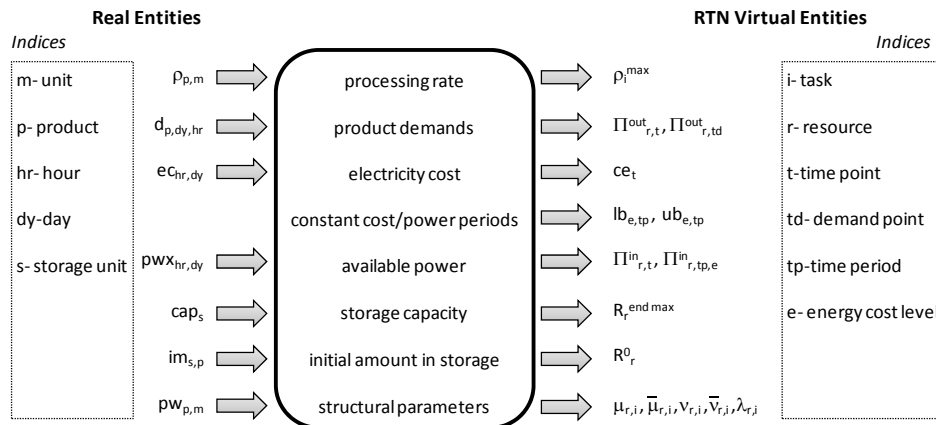
Figure 2. Process entities are converted into RTN model entities by a special purpose algorithm

Rolling-Horizon Algorithm for Scheduling under Time-Dependent Utility Pricing and Availability

3. From Process to Resource-Task Network Entities

The industrial process under study can be described as a Resource-Task Network (Castro et al., 2009). We can automatically generate the sets of tasks and resources as a function of problem data. Tasks can either be continuous (I^c), hybrid batch-continuous (I^s) or instantaneous (I^t). Resources include equipment units (R^{EQ}), divided into storage and processing units (R^{TC}), electricity (R^{UT}) and raw-material (R^{RM}). In addition, we distinguish the product location: immediately after processing (R^{LM}); in storage (R^{LS}); inside the clients' transportation vessels (R^{FP}). Resources continuously produced are given by $R^{CT}=R^{RM}\cup R^{LM}\cup R^{LS}$, whereas the final product (R^{FP}) is produced/consumed instantaneously. The RTN representation is then brought into the mathematical model by the structural parameters, see Fig. 2.

4. Mathematical Formulation

We propose a new combined aggregate/continuous-time model as the center of the rolling-horizon algorithm for scheduling. It uses a time grid that is partly continuous and partly discrete ($T= T^c\cup T^{ag}$). The first time points will be continuous, with the last being the element of $T^{c,\ last}$, a boundary time point also part of the set of discrete time points. The discrete part of the grid is non-uniform, meaning that the duration Δ_t is not the same for all time slots t. The number of slots and their exact location in the grid is calculated from the product demand and electricity cost and availability data.

Three sets of variables are used to characterize tasks. Binary variables $N_{i,t}$ identify the execution of task i during slot t; continuous variables $\xi_{i,t}$ give the amount processed; the amount continuously sent to storage by hybrid tasks is given by $\xi_{i,t}^*$. We keep track of resource availability over time through excess resource variables $R_{r,t}$ and $R_{r,t}^{end}$. Slack variables $S_{r,t}$, representing the extra power to be purchased [MW], allow the violation of utility balances but are penalized in the objective function through parameter cs.

$$\min pcost + \sum_{r\in R^{UT}}\sum_{e\in E}\sum_{t\in T^c\wedge t\notin T^{c,last}}\sum_{i\in I^\xi}c_e\cdot(-\mu_{r,i})\cdot\frac{\xi_{i,t}}{\rho_i^{\max}}$$
$$+\sum_{r\in R^{UT}}\sum_{t\in T^{ag}\wedge t\notin T^{last}}\sum_{i\in I_t^c}ce_t\cdot(-\mu_{r,i})\cdot\frac{\xi_{i,t}}{\rho_i^{\max}}+cs\sum_{t\in T}S_{r,t} \tag{1}$$

$$R_{r,t} = R_r^0\Big|_{t=1} + R_{r,t-1}^{end}\Big|_{r\in R^{CT}} + R_{r,t-1}\Big|_{r\notin(R^{CT}\cup R^{UT})} - (\Pi_{r,t}^{out}+\Pi_{r,td^*}^{out}\Big|_{t\in T^{c\,last}})\Big|_{r\in R^{FP}} +$$

$$(\Pi_{r,t}^{in}+\sum_{e\in E}\sum_{tp\in TP_e}\Pi_{r,tp,e}^{in}Y_{t,tp,e}\Big|_{t\in T^c\setminus T^{c,last}}+S_{r,t})\Bigg|_{r\in R^{UT}}+$$

$$\sum_{i\in I}(\mu_{r,i}N_{i,t}\Big|_{i\in I_t^c\vee(i\notin I^c\wedge t\neq|T|)}+v_{r,i}\xi_{i,t}\Big|_{i\in I_t^c\vee i\notin I^c}+\bar{\mu}_{r,i}N_{i,t-1}\Big|_{i\in I_{t-1}^c\vee i\notin I^c})+\sum_{i\in I^t}\bar{v}_{r,i}\xi_{i,t} \tag{2}$$

$$\forall t\in T, r\in R,[r\notin(R^{TC}\cup R^{UT})\vee t\in T^c\setminus T^{c,last}]$$

$$R_{r,t}^{end} = R_{r,t}+\sum_{i\in I_t^c}\lambda_{r,i}\xi_{i,t}+\sum_{i\in I^s}(\bar{v}_{r,i}\xi_{i,t}+\lambda_{r,i}\xi_{i,t}^*)\quad \forall r\in R^{CT}, t\in T, t\neq|T| \tag{3}$$

$$Ts_t \geq T_t\Big|_{t\notin T^{first}}+tfx_{td^*-1}\Big|_{t\in T^{first}}\quad \forall t\in T^c\setminus T^{c,last} \tag{4}$$

$$tfx_{td^*}\Big|_{t+1\in T^{c,last}}+T_{t+1}\Big|_{t+1\notin T^{c,last}}-Ts_t \geq \sum_{i\in I_t^c}\frac{\bar{\mu}_{r,i}\xi_{i,t}}{\rho_i^{\max}}\quad \forall r\in R^{TC}, t\in T^c\setminus T^{c,last} \tag{5}$$

$$Ts_t \geq \sum_{e \in E} \sum_{tp \in TP_e} lb_{e,tp} Y_{t,tp,e} \quad \forall t \in T^c \setminus T^{c,last} \tag{6}$$

$$Ts_t + \sum_{i \in I_t^c} \frac{\overline{\mu}_{r,i} \xi_{i,t}}{\rho_i^{max}} \leq \sum_{tp \in TP_e} ub_{e,tp} Y_{t,tp,e} + tfx_{td*} \cdot (1 - \sum_{i \in I_e^c} \overline{\mu}_{r,i} N_{i,t}) \forall r \in R^{TC}, e, t \in T^c \setminus T^{c,last} \tag{7}$$

$$\sum_{e \in E} \sum_{tp \in TP_e} Y_{t,tp,e} = 1 \quad \forall t \in T^c \setminus T^{c,last} \tag{8}$$

$$\Delta_t \geq \sum_{i \in I_t^c} \frac{\overline{\mu}_{r,i} \xi_{i,t}}{\rho_i^{max}} \quad \forall r \in R^{TC}, t \in T^{ag}, t \neq |T| \tag{9}$$

$$\Delta_t \cdot (\Pi_{r,t}^{in} + S_{r,t}) \geq \sum_{i \in I_t^c} \frac{-\mu_{r,i} \xi_{i,t}}{\rho_i^{max}} \quad \forall r \in R^{UT}, t \in T^{ag}, t \neq |T| \tag{10}$$

$$\xi_{i,t} + \xi_{i,t}^* \big|_{i \in I^s} \leq H \cdot \sum_{m \in M} \max_{p \in P} \rho_{p,m} \cdot N_{i,t} \quad \forall t \in T, i \in I_t^c \vee (i \in I^s \wedge t \neq |T|) \tag{11}$$

In Eq. 1, *pcost* is a parameter that accounts for the partial electricity cost [k€] in previously scheduled periods, the second term gives the electricity cost from processing tasks executed in the continuous part of the grid (sum over all electricity levels *e* of the product of electricity cost c_e [€/kWh], power consumption [MW] and task duration [h]) and the third, the cost from the discrete part of the grid.

The complexity of the model is reflected by the excess resource balances (Eqs. 2-3). They are very similar to those of the pure continuous-time model (Castro et al., 2009). Parameter $\Pi_{r,td*}^{out}$ holds the product demands at the end of the period *td* currently being scheduled in detail (superscript *). In the grid, such interaction occurs at $T^{c,last}$. There may be multiple energy pricing levels within a demand period and even different time periods *tp* within the same energy level (TP_e). To identify during which time period tasks executed at slot *t* are processed at, binary variables $Y_{t,tp,e}$ are used.

Eqs. 4-8 are for the continuous part of the model. Timing variables T_t and Ts_t provide the absolute time of event point *t* and the earliest starting time amongst all processing tasks executed during slot *t*. The starting time must be greater than the event point's absolute time, which for the first point is given by the location of the previous demand point (Eq. 4). Likewise, tasks must end before *t*+1 (Eq. 5). Furthermore, tasks executed during slot *t* must lie within a single time period of an energy level (Eq. 8). Hence, the starting time must also be greater than the time period's lower bound ($lb_{e,tp}$) and there must be enough time for the tasks to end before its upper bound, $ub_{e,tp}$ (Eqs. 6-7).

Note that the domain of Eq. 2 does not include processing units or utilities for the discrete part of the grid. The purpose is to allow multiple products to be processed on a given time slot. Resource balances for units are replaced by timing constraints that ensure that the total processing time in slot *t* does not exceed its duration (Eq. 9). For the utility, the instantaneous balances on power are replaced by energy balances (Eq. 10). Finally, the last set of constraints activate the binary variables whenever there is material being handled by the task (Eq. 11).

5. Rolling-Horizon Algorithm

The discrete part of the grid is handled by an aggregate (planning) model that predicts the production levels without being concerned with the timing of events. It is completely accurate for cases with unlimited power availability, generating a lower bound on cost otherwise. More importantly, the output solution can be analyzed to

Rolling-Horizon Algorithm for Scheduling under Time-Dependent Utility Pricing and Availability

predict how many event points are required by the continuous-time formulation to generate the detailed schedule. Consider an aggregate time slot *t* resulting from merging 3 time periods of length 2, 8 and 3 h for some energy level *e* (the lighter boxes in Fig. 3, which correspond to a medium cost level, see Castro et al. 2009 for further details). If the maximum processing time over all units *M* is 11 h, we know that the production will spread at least 2 slots (8+3 h) in the continuous time formulation. Step 0 of the rolling-horizon algorithm will thus involve solving the full problem with the aggregate model.

Fig. 3 illustrates the main components of the algorithm for a simple example. In step 0, $T^c=\varnothing$ and $T^{ag}=\{1,\dots,13\}$, while in step1, we use the same 13 events: $T^c=\{1,\dots,4\}$ and $T^{ag}=\{4,\dots,13\}$. Notice in the latter that the exact location of the energy levels is being considered in the first demand period but not all are active. Provided that a processing task occupies a single level (for low production rates), a slot can span across multiple levels (Castro et al., 2009). If the objective function value (OBJ) remains constant between iterations, we can fix the schedule and proceed to the next demand period.

In step 2, the aggregate model predicts 3 events for demand period 2: $T^c=\{4,\dots,6\}$ and $T^{ag}=\{6,\dots,12\}$. Now OBJ is greater than the lower bound (LB) so we need to resolve the problem following a single increase in $|T|$. If OBJ remains unchanged, the available power is insufficient to meet the production levels predicted by the aggregate model in the cheaper cost periods (see step 3). The lower bound can thus be reset, the added event point removed, and the algorithm can proceed to the next demand period until eventually tackling all of them and generating the full schedule (final step).
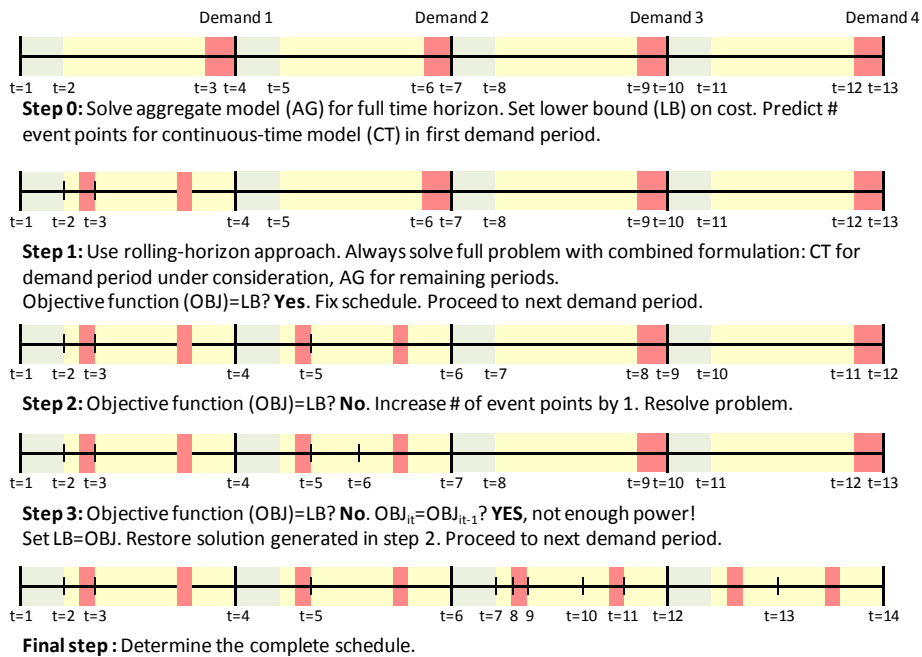
Figure 3. Illustration of rolling-horizon algorithm (boxes represent different energy cost levels)

6. Computational Results

The performance of the rolling-horizon (RH) algorithm and standalone aggregate model (AG) were evaluated through the solution of 7 example problems (Castro et al., 2009).

We have used GAMS 22.8 (CPLEX 11.1) to solve the resulting MILP problems. The termination criteria were: relative tolerance=10^{-6}; maximum computational time=3,600 CPUs per iteration. The hardware consisted on a laptop with an Intel Core2 Duo T9300 processor (2.5 GHz) with 4 GB of RAM running Windows Vista Enterprise. In order to prioritize schedules that meet the maximum power constraints, we set cs=10 k€/MW.

From the results in Table 1, it is clear that the aggregate model (AG) is very powerful. On the one hand, it can solve all problems in a few seconds, which represents orders of magnitude reduction in computational time when compared to a traditional discrete-time (DT) formulation employing 1-hour intervals (Castro et al., 2009). On the other hand, it gives very good total cost predictions, which are accurate for unrestricted power (U). However, the results for EX5a and EX8 are below those obtained by the rigorous DT model (29657 vs. 31798 and 14375 vs. 104622). In such problems, the RH algorithm generates solutions that violate the power constraints due to scheduling in detail a single demand period per iteration, rather than the full 1-week horizon. This is the reason for the large optimality gap. Besides EX8, RH could find solutions in less than half an hour, with the schedules for EX9-10 being slightly better than those obtained by DT in 2-hours time. Notice in the last column that we were able to go up to 30 event points, whereas with the full-space continuous-time approach the typical limit is roughly 10.

7. Conclusions

This paper has proposed a new rolling-horizon algorithm for the scheduling of single stage continuous multiproduct plants subject to energy constraints related to time dependent utility pricing and availability, and multiple due dates. The elements for decomposition are the demand periods, with the one period being tackled with a continuous-time scheduling model while simultaneously considering the remaining horizon with a non-uniform discrete-time planning model. The algorithm was shown to effectively solve real-life problems for the case of unrestricted power availability to global optimality. In the other cases, it may lead to suboptimal solutions due to the unnecessary violation of the power constraints, a consequence of the planning model underestimating the total electricity cost.

Table 1. Computational Statistics

Approach Case	(P,M,S)	Power	AG |T|	AG CPUs	AG Total cost [€]	RH	RH |T|	RH CPUs
EX2	(2,1,1)	R	17	0.12	21575		11	1.11
EX5	(3,2,2)	U	19	0.33	26758		13	2.26
EX5a	(3,2,2)	R	20	0.24	*29657*	41124	17	7.06
EX7	(3,3,4)	U	18	0.7	68282		12	3.12
EX8	(3,3,5)	R	19	2.05	*104375*	151257	31	17330
EX9	(4,3,4)	U	19	0.71	87817		25	917
EX10	(5,3,4)	U	19	3.57	86550		23	1508

References

P. Castro, I. Harjunkoski, I. Grossmann, 2009. Ind. Eng. Chem. Res. 48, 6701.

I. Grossmann, 2005. AIChE J., 51, 1846.

C. Maravelias, 2005. Ind. Eng. Chem. Res. 44, 9129.

J. Westerlund, M. Hastbacka, S. Forssell, T. Westerlund, 2007. Ind. Eng. Chem. Res. 46, 2781.

Operation scheduling of batch autothermal thermophilic aerobic digestion processes

Elisabet Capón-García[a], Jaime Rojas[b], Toshko Zhelev[b], Moisès Graells[a]

[a]*Chemical Engineering Department,Universitat Politècnica de Catalunya (UPC), Barcelona 08032, Spain.*
[b]*Chemical Engineering Department, University of Limerick (UL), Limerick, Ireland.*

Abstract

The autothermal thermophilic aerobic digestion (ATAD) is an exothermic batchwise operated sludge treatment process, in which sludge undergoes heating and pasteurization as temperature attains values about 50-60 ºC. Based on detailed process models, different approaches have been suggested (Rojas and Zhelev, 2009) to minimize the specific energy requirement of existing ATAD designs by altering the operating conditions, but less attention has been paid to process optimization by means of operation planning and timing. This work aims at minimizing the specific energy requirement of existing ATAD processes considering simultaneously operational conditions, planning and timing. Hence, a scheduling formulation including process model variables and a flexible recipe is proposed. Energy savings are reported for a case study of a simplified treatment plant. The proposed approach results in energy consumption reduction by optimizing batch sizes and operating conditions.

Keywords: batch processing, wastewater treatment, ATAD, process integration

1. Introduction

Autothermal thermophilic aerobic digestion (ATAD) is an activated sludge process used in wastewater treatment with two objectives: the stabilization and pasteurization of the sludge. Metabolic activity produces both biodegradation as well as heat release for maintaining necessary thermophilic temperatures. The ATAD treatment is a robust process that allows handling variable quality sludge and obtaining biologically stable products. The ATAD was first introduced at the end of the 1960s and the technology has been quite investigated and improved since then. Comprehensive review papers on ATAD origin, design and operation can be found in USEPA (1990), LaPara and Alleman (1999), and Layden et al. (2007a; 2007b).

ATAD is an energy intensive process because of the high oxygen uptake rates of thermophilic microorganisms. However, there are conflicting reports in current literature regarding the energy efficiency and cost effectiveness of ATAD systems (Layden et al., 2007b). Certainly, operating conditions play a key role for each single batch run. Hence, the optimization of the operating conditions for consecutive batches may lead to a significant improvement in energy management.

Physical, chemical and biological transformations occurring during the operation of the ATAD reactor are described by means of dynamic models, which are suitable tools to optimize process operation. Input process parameters are usually uncertain, and they must be estimated or assumed to be known in advance. Some of them can only be completely specified immediately before process implementation. Batch processes are

operated consecutively along time; therefore, considering a single or cyclic batch operation optimization may disregard possible tradeoffs derived from the interaction among different batches.

Scheduling aims at optimizing operational resources allocation along time. Ideally, scheduling models should include detailed process models, so that processing times and operating conditions would be fully described and considered at this level. However, the high complexity of dynamic models usually prevents them from being included in process scheduling; and consequently, they are usually simplified or even fixed to standard conditions. Therefore, decision variables such as processing times are fixed or simplified to a linear function of the processed quantity (Mishra et al., 2005). As a result, degrees of freedom are removed from the system and suboptimal solutions may be eventually implemented.

This work aims at minimizing the energy requirements of the ATAD process by identifying optimal scheduling strategies. Given the fact that including dynamic process models in the scheduling level leads to large optimization problems, an alternative modeling framework is proposed. Specifically, the differential equations that describe process operation are used to build a multivariable linear regression model around an ideal optimal operation strategy, which results from considering cyclic operation. As a result, a flexible recipe of the ATAD process is obtained, and decisions regarding operational conditions, planning and scheduling can be simultaneously made.

2. Problem statement

The work deals with the optimization of the operating conditions of consecutive batches in a single ATAD reactor, in order to minimize the specific energy consumption, measured as energy per unit of treated volume, by considering process operational issues in the planning and scheduling problem.

From the scheduling point of view, the main goal consists of treating all the sludge quantity expected to arrive at the treatment plant in a week, which is divided in six time periods (t), using the minimum necessary energy. The reactor operation is divided into three operations (o): load (o1), pasteurization and stabilization (o2) and discharge (o3). The main energy requirement stems from aeration in the second operation, which is described by a dynamic model. Pumping energy for loading and discharge is neglected.

A maximum of one batch can be performed at every time period. In addition, the quantity of incoming sludge, TrDem [m^3], and its temperature, T_{feed} [ºC], are considered initial parameters of every period, whereas the specific amount to be treated at each time period is a variable to be optimized by the planning problem ($V_{in}(t)$ and $V_{out}(t)$).

3. Solution approach

The ATAD process can be described by a simplified model (Burke et al., 2009) containing five differential equations that represent the behaviour of the main process variables along time, namely biomass concentration, X_B [g/l], substrate concentration, X [g/l], oxygen concentration, S [mg/l], reactor temperature, T [ºC] and volatile solids concentration, X_{vs} [g/l]. This model is used to derive a multivariable linear regression model to be included in the scheduling problem. In Burke et al. (2009), the model parameters and feed sludge properties are summarized.

First, the optimal operating conditions for cyclic scheduling are obtained, namely the aeration rate, A(t) [1/d], the batch reaction time [d] and reactor variables profiles, considering a fixed exchanged volume fraction of 0.1 (V_{in}/V). Next, several simulations of the dynamic model are run in order to describe the behaviour of the process around the former operational conditions. Consequently, the reactor final conditions and batch reactor time needed to achieve sludge pasteurisation and stabilisation are derived from a range of reactor initial conditions and aeration rates.

A total of 22133 initial conditions have been explored to build the multivariable linear regression model, considering the limiting initial and final values of the reactor variables that appear in Table 1. Final values reactor variables and batch time are adjusted to initial reactor conditions and aeration rate as shown in Table 2. The minimum residual variances for the regressions are over 0.80 in all cases. Hence, the data variability explained by the model is quite significant.

Table 1. Limiting initial and final conditions and steady state values of problem variables.

	A [1/d]	X_B [g/l]	X [g/l]	S [mg/l]	$T_{reactor}$ [ºC]	X_{vs} [g/l]	time [d]
Min. ini.	100	0.2	1.4	2	55	12.1	-
Max. ini.	150	0.5	2.1	8	60	12.6	-
Min. end	-	0.3	0	0	59	11	-
Max. end	-	0.5	1.5	1	60	12.3	-
Steady state	-	0.2794	1.5231	7.87	59.06	12.23	0.86
Feed	150	1	15	1	13	20	-

Table 2. Adjusted parameters of the multivariable linear regression process model.

	A [1/d]	X_{Bini} [g/l]	X_{ini} [g/l]	T_{ini}[ºC]	ADJ
time [d]	-0.00095	-	-	-0.06685	4.262
X_{Bout}[g/l]	0.00264	0.28791	-	-	0.02043
X_{out} [g/l]	-	-	0.98709	0.1553	-9.6053
T_{out} [ºC]	0.02028	-	-	0.42066	33.3134

In addition, the planning and scheduling problem must be defined. The objective function (see Eq. (1)) consists of minimizing the total specific energy consumption per cubic meter of treated sludge. The energy consumption of each time period, and its value depends on the decisions made in the scheduling problem, namely aeration rate, loaded and discharged volumes along time.

$$\min z = \min \frac{\sum_t A(t) \cdot time(o2,t)}{\sum_t TrDem(t)} \qquad (1)$$

Operating times for o1 and o3 depend on the loaded and discharged quantities, V_{in} and V_{out}, and the pumping rates, Q_1 and Q_2, respectively (see Eq. (2) and (3)). In addition, the whole processing time, which includes load, batch reaction time and discharge, cannot exceed the total time of each time period, H= 1d.

$$time(o1) \cdot Q_1 = V_{in}(t) \tag{2}$$

$$time(o3) \cdot Q_2 = V_{out}(t) \tag{3}$$

The initial reactor temperature (T_{ini}) at the start of a batch is obtained from the energy balance of the reactor (see Eq. (4)). Identical expressions are applied to X_B and X using the material balances. As for reactor mass balance, the remaining volume of treated sludge at the start of time period t, is equal to the remaining volume at the start of the previous period, plus the quantity that is charged in that period, minus the quantity that was discharged in the previous one (see Eq. (5)).

$$T_{ini}(t) \cdot V(t) = T_{out}(t-1) \cdot \left(V(t-1) - V_{out}(t-1)\right) + T_{feed}(t) \cdot V_{in}(t) \quad \forall t > 1 \tag{4}$$

$$V(t) = V(t-1) + V_{in}(t) - V_{out}(t-1) \quad \forall t > 1 \tag{5}$$

The quantity of sludge treated at each time period (V_{in}) is a problem variable. Therefore, the mass balance for a storage vessel previous to the reactor is posed through Eq. (6). The stored quantity of sludge at the start of period t, plus the new sludge income, is equal to the sludge that is pumped to the reactor at that time period plus the quantity of sludge in the storage tank remaining for the start of the next period.

$$ST(t) + TrDem(t) = V_{in}(t) + ST(t+1) \quad \forall t < T \tag{6}$$

Finally, the whole quantity of sludge that arrives at the plant along the time horizon must have been processed along the whole time horizon (see Eq. (7) and (8)).

$$\sum_t V_{out}(t) = \sum_t TrDem(t) \tag{7}$$

$$\sum_t V_{in}(t) = \sum_t TrDem(t) \tag{8}$$

4. Results

In order to check to improvement regarding the cyclic operation, a whole week is scheduled. The average temperature and incoming sludge volume correspond to the values considered in the cyclic operation (13ºC and $0.1 \cdot V_{reactor}$, respectively). Tables 3 and 4 contain the initial and limiting values of process variables and parameters.

Table 3. Sludge temperature and quantity to be treated along time periods.

	1	2	3	4	5	6
Te_{ini} [ºC]	15	10	9	14	15	15
TRDem [m^3]	9	10	9	12	9	11

Table 4. Minimum, maximum and initial values for some scheduling variables.

	ST [m^3]	V_{in} [m^3]	V_{out} [m^3]	V [m^3]
Minimum.	0	5	5	85
Maximum	100	15	15	125
Initial	50	-	-	100

Four different case studies have been solved for the scheduling of the whole week:
- The implementation of the cyclic policy, using batch times, loaded and discharged volumes, and optimizing aeration rates (CS1).
- The implementation of the optimal scheduling and operating conditions solving the dynamic model simultaneously with the scheduling problem (CS2).
- The use of the approximated process model in the scheduling problem (CS3).
- The use of the approximated process model in the scheduling problem with an additional condition: the volume loaded in each time period must be equal to the quantity discharged in the previous one (CS4).

The dynamic process model is implemented, solved and optimized in Matlab (CS1 and CS2). The scheduling problem with the regression model consists of a NLP formulation which is implemented in GAMS 22.9 and solved using BARON 8.1 (CS3 and CS4).

The overall minimum specific energy consumption for each case study is shown in Table 5. The cyclic operating policy (CS1) turns to be the worst choice, since the loaded and discharged quantities and batch times are fixed and disregard the possible tradeoffs that may arise from considering the interaction among batches and uncertain parameters, such as feed temperature and exchanged volume fraction, which are fixed for the cyclic scheduling problem. On the other hand, the optimal operating scheduling policy (CS2) has been solved for this small case study and sheds light to the maximum energy savings that may be obtained; which represent over 60% regarding the cyclic policy. As for the proposed approach, which considers the multivariable regression model to solve the scheduling problem (CS3), savings about 50% regarding the cyclic policy are reported. Even though the optimal operating policy has not been identified, a significant improvement is obtained, and process variables are not identical for all batches (Figure 1). When considering additional restrictions on scheduling (CS4), the energy savings can be even better than those from the cyclic scheduling.

Table 5. Objective function value for the case studies: (1) cyclic policy, (2) optimal operating policy, (3) optimal operating policy with regression model, (4) optimal operating policy with regression model and restrictions on loading and discharged volumes.

	1	2	3	4
OF [kW·h/m^3]	8.6216	3.3556	4.2346	4.4640
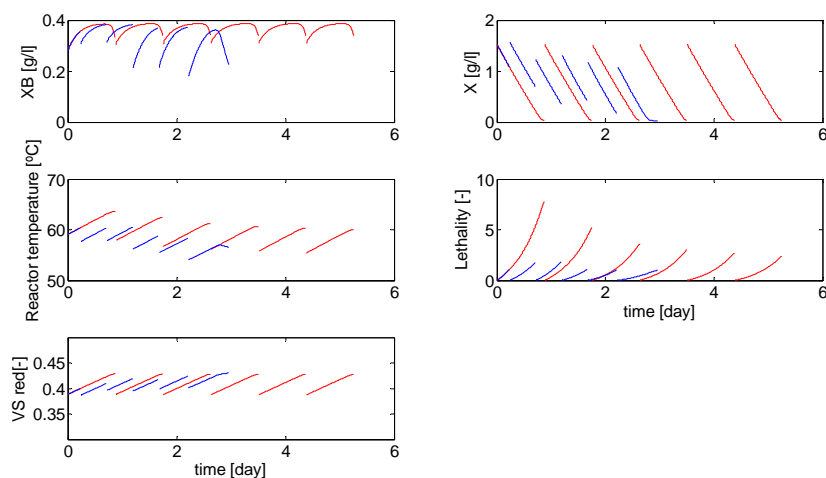
Figure 1. Process variable profiles along time in the ATAD reactor for CS1(red) and CS3(blue).

5. Conclusions

The consideration of process operation information at the scheduling level may lead to significant overall performance enhancement. Savings about one half in energy consumption have been achieved in the addressed ATAD case study, by considering temperature disturbances and sludge income along the whole time horizon, instead of treating the sludge to be processed under the assumption of a nominal recipe (fixed processing times and operating conditions). The savings obtained stem from the adjusted optimal processing times and processing conditions, given a flexible model and processing policy. The proposed framework can be extended to deal with other plant configurations and the obtained results support further work towards including in the scheduling model more detailed process models and resource integration opportunities.

6. Acknowledgements

References

M. Burke, M. Chapwanya, K. Doherty, I. Hewitt, A. Korobeinikov, S. McCarthy, M. Meere, V. Tuoi, M. O'Brien, J. Rojas, H. Winstanley, T. Zhelev. Submitted to Mathematics-in-Industry Case Studies Journal, (2009).

J. Gomez, M. de Gracia, E. Ayesa, J.L. Garcia-Heras. Water Research, 41 (2007) 959.

T. M. LaPara, J.E. Alleman, Water Research, 33 (1990) 895.

N.M. Layden, H.G. Kelly, D. S. Mavinic, R. Moles, J. Barlet. Journal of Environmental Engineering and Science, 6 (2007) 665.

N.M. Layden, H.G. Kelly, D. S. Mavinic, R. Moles, J. Barlet. Journal of Environmental Engineering and Science, 6 (2007) 679.

B.V. Mishra, E. Mayer, J. Raisch, A. Kienle. Industrial and Engineering Chemistry Research, 44 (2005) 4022.

J. Rojas, T. Zhelev. Computer Aided Chemical Engineering, 26 (2009) 1269.

U. S. Environmental Protection Agency. Environmental regulations and technology: Autothermal thermophilic aerobic digestion of municipal wastewater sludge. Technical report, (1990).

Integrated Campaign Planning and Resource Allocation in Batch Plants

Naresh Susarla, I. A. Karimi

Department of Chemical & Biomolecular Engineering, National University of Singapore, 4 Engineering Drive 4, Singapore 117576

Abstract

In this work, we develop a simple MILP model for simultaneous campaign planning and resource allocation in multi-stage batch plants. We capture several real life scenarios including maintenance planning, product outsourcing, and NPIs and study the effect of various process decisions on the solution of the integrated and resource constrained planning problem. Given the products, their projected demands, and available resources for a given time horizon, our model determines campaign lengths, product schedules on different production lines, and resource allocation profiles. Also, we consider sequence-dependent changeover times between two campaigns. To demonstrate the performance of our mathematical formulation, we consider a case study from a typical multistage specialty chemical batch plant. We validate our approach considering a series of dynamic business scenarios.

Keywords: Campaign scheduling, multiproduct batch plants, resource allocation, MILP

1. Introduction

Batch-wise manufacturing is very popular for specialty products (pharmaceuticals, cosmetics, polymers, biochemicals, food products, etc.) which are of high added value, low volume or require close control of process conditions. Operational planning seeks inputs and is reviewed by several departments such as process, maintenance, laboratory, suppliers, sales, and higher management. This is mainly because operational planning is often constrained by the availability of resources (manpower, utilities, laboratory, parts, etc.). Thus, planning is a collaborative activity of several departments. For this reason, a simple planning tool or strategy is required that can quickly cater to the needs of all the stakeholders.

In general, the problem of operational planning in multiproduct batch plants has been addressed by several researchers. Recently, Corsano et al. (2009) developed a MINLP model for the design and planning of multiproduct batch plants. Stefansson et al. (2006) presented a 3-level hierarchical framework for the planning and scheduling in pharmaceutical plants. Sundaramoorthy and Karimi (2004) studied the effect of new product introductions in the medium-term planning in the context of a pharmaceutical production facility. Sundaramoorthy et al. (2006) developed a simple LP model as a decision support tool for medium term integrated planning decisions to the managers in specialty chemical industry. Suryadi and Papageorgiou (2004) considered a production planning problem and incorporated maintenance planning and crew allocation constraints. Clearly, campaign planning problem has been well studied in batch plants, few work study the effect of integrating resource allocation

In this work, we use a basic model of Sundaramoorthy and Karimi (2004) and modify it to develop a multiperiod MILP planning model. Our model is less complicated, in terms

of solution strategy and model structure, and can address the needs of higher management readily. Our model captures several real life scenarios such as the effect of resource (manpower, utilities, laboratory, and waste-treatment capacity) availability in process planning, routine maintenance, new product introductions (NPIs), outsourcing of intermediate products, and sequence-dependent cleaning times. Furthermore, to capture the dynamic changes in the plant, we propose a reactive scheduling strategy for our model. Finally, to demonstrate the performance of our approach, we consider a case study from a typical multistage specialty chemical batch plant. Also, we evaluate our model considering various business scenarios.

2. Problem statement

A multiproduct specialty chemical manufacturing facility (F) produces several products using J processing units/lines ($j = 1, 2, …, J$). Operations in F involves I ($i = 1, 2, …, I$) tasks, which includes both processing (I^p) and maintenance (I^m) tasks. A recipe diagram of the manufacturing process gives the information on processing tasks, material states ($s = 1, 2, …, S$), and mass ratios (σ_{sij}) (Susarla et al., 2009). The planning problem in F can be described as follows. Given the (1) production recipes, (2) fixed batch sizes and processing and cycle times, (3) planning horizon, (4) demands and their due-dates, (5) cost and revenue details, (6) sequence-dependent cleaning times and costs, (7) resource availability, costs, & effects on process performance, (8) preventive-maintenance timings, (9) potential new products and their demands, we determine (1) allocation of tasks to production units/lines, (2) resource allocations, (3) campaigns, schedules, and number of batches, (4) material inventory profiles, (5) outsourcing strategy, assuming (1) deterministic scenario, (2) stable intermediate materials, (3) instantaneous procurement of raw materials (zero inventory cost), (4) all demands due at due dates, (5) one campaign per interval. We consider the maximization of gross profit (revenue through sales – cost of goods sold) as the optimization objective.

3. MILP formulation

We model the planning horizon **H**, on each unit j (1, …, **J**), in **NT** (1, …, **NT**) discrete intervals of length h_t ($h_1, h_2, …, h_{NT}$) each. An interval t is then referred to the time between two product delivery dates [DD_{t-1} - DD_t] and is of length h_t. Furthermore, to model the interval h_t we use a separate local time axis on every unit j and define **KT**$_j$ ($k = 1, 2, …, $ **KT**$_j$) slots, using a multi-grid continuous time approach (Susarla et al., 2009). Let T_{jkt}^s and T_{jkt}^e [$k = 1, 2, …, $ **KT**$_j$; $T_{j1t}^s \geq 0$; $T_{jKTt}^e \leq h_t$; $T_{jkt}^e \geq T_{jkt}^s$; $T_{j(k+1)t}^s \geq T_{jkt}^e$] denote the start and end time of the slot k on unit j for the interval t. Thus, the slot length is [T_{jkt}^e - T_{jkt}^s]. We use T_{ijkt}^c [$T_{jkt}^s = \sum_{i \in \mathbf{I}_j} T_{ijkt}^c$] to denote the start of a campaign of task i in slot k of unit j for interval t.

3.1. Campaign allocation

Each campaign involves several batches, every slot must have a campaign, and each campaign can be allocated to only one slot. To allocate each campaign in interval t to a slot, model transition of campaigns and extension of a campaign to the next interval we define the one binary (ys_{ijkt}) and two 0-1 variable

$$ys_{ijkt} = \begin{cases} 1 & \text{if a campaign of task } i \text{ is allocated to slot } k \text{ on unit } j \text{ in interval } t \\ 0 & \text{Otherwise} \end{cases}$$

$$i \in \mathbf{I}_j, \ 1 \le j \le J, \ 1 \le k \le \mathrm{KT}_j, \ 1 \le t \le \mathrm{NT}$$

$$x_{ii'jkt} = \begin{cases} 1 & \text{if campaign of task } i \text{ in slot } k \text{ precedes campaign of } i' \\ 0 & \text{Otherwise} \end{cases}$$

$$i, i' \in \mathbf{I}_j, \ 1 \le j \le J, \ 1 \le k < \mathrm{KT}_j, \ 1 \le t \le \mathrm{NT}$$

$$y_{ijt} = \begin{cases} 1 & \text{if a campaign of task } i \text{ in unit } j \text{ stretches from interval } t \text{ to } t+1 \\ 0 & \text{Otherwise} \end{cases}$$

$$i \in \mathbf{I}_j, \ 1 \le j \le J, \ 1 \le t \le \mathrm{NT}$$

Now, a slot k on unit j cannot perform more than 1 campaign. Also, we do not allow multiple campaigns of a task in the same processing unit, within a time interval.

$$\sum_{i \in \mathbf{I}_j} ys_{ijkt} \le 1 \qquad\qquad 1 \le j \le J, \ 1 \le k \le KT_j, \ 1 \le t \le NT \qquad (1a)$$

$$\sum_{k \in \mathbf{KT}_j} ys_{ijkt} \le 1 \qquad\qquad i \in \mathbf{I}_j, \ 1 \le j \le J, \ 1 \le t \le NT \qquad (1b)$$

$$\sum_{\substack{i' \in \mathbf{I}_j \\ i \ne i'}} x_{ii'jkt} \le ys_{ijkt} \qquad\qquad i \in \mathbf{I}_j, \ 1 \le j \le J, \ 1 \le k < KT_j, \ 1 \le t \le NT \qquad (2a)$$

$$\sum_{\substack{i' \in \mathbf{I}_j \\ i \ne i'}} x_{i'ijkt} \le ys_{ij(k+1)t} \qquad\qquad i \in \mathbf{I}_j, \ 1 \le j \le J, \ 1 \le k < KT_j, \ 1 \le t \le NT \qquad (2b)$$

$$ys_{ijkt} + ys_{i'j(k+1)t} - 1 \le x_{ii'jkt} \qquad i, i' \in \mathbf{I}_j, \ 1 \le j \le J, \ 1 \le k < KT_j, \ 1 \le t \le NT \qquad (2c)$$

A campaign can stretch over to the next interval only if it is the last campaign for the current interval in unit j. Also, if the campaign is stretched over to the next interval from the current one, it will be the first campaign in the next interval.

$$y_{ijt} \le ys_{ijkt} \qquad\qquad i \in \mathbf{I}_j, \ 1 \le j \le J, \ k = KT_j, \ 1 \le t \le NT \qquad (3a)$$

$$y_{ijt} \le ys_{ij1(t+1)} \qquad\qquad i \in \mathbf{I}_j, \ 1 \le j \le J, \ 1 \le t < NT \qquad (3b)$$

3.2. Campaign and Slot lengths

A campaign of a processing tasks i in unit j for interval t consists of nb_{ijkt} number of batches of constant processing time (pt_{ij}) and cycle time (ct_{ij}) (or maintenance time, mt_{jt}), the sequence-dependent changeover/set-up time ($\tau_{ii'}$), and constant batch size (bs_{ij}). We incorporate following constraints for the timings of campaigns and for ensuring minimum campaign lengths (MCL_{ij}).

$$T^e_{jkt} - T^s_{jkt} \ge pt_{ij} ys_{ijkt} + n_{ijkt} ct_{ij} + \sum_{i'} \tau_{ii'} x_{ii'jkt}$$

$$i, i' \in \mathbf{I}_j, \ 1 \le j \le J, \ 1 \le k < KT_j, \ 1 \le t \le NT \qquad (4a)$$

$$T^e_{jkt} - T^s_{jkt} \ge \sum_i MCL_{ij} ys_{ijkt} - \sum_i H(y_{ijt} + y_{ij(t-1)})$$

$$i \in I_j, \ k \le KT_j \qquad (4b)$$

$$T^e_{jkt} - T^s_{jkt} + T^e_{j1(t+1)} - T^s_{j0(t+1)} \ge \sum_i MCL_{ij} ys_{ijkt} - \sum_i h_t(1 - y_{ijt})$$

$$i \in I_j, \ k = KT_j \qquad (4c)$$

3.3. Operation times

We demand that the start time for a campaign of task i to be zero whenever slot k is not allocated to task i. So,

$$T_{ijkt}^c \leq h_t\, ys_{ijkt} \qquad\qquad 1 \leq j \leq J,\ 1 \leq k \leq KT_j,\ 1 \leq t \leq NT \qquad\qquad (5)$$

A campaign of task i cannot start unless all of the tasks i', which precede i in the product recipe, have produced sufficient amounts of material states that are required by i. Now, if both i and i' occur in the same interval, we demand the following.

$$\sum_{k \in KT_j} (T_{ijkt}^c + pt_{ij}\, ys_{ijkt}) \leq \sum_{k \in KT_j} T_{i'j'kt}^c + h_t\Big(1 - \sum_{k \in KT_j} ys_{i'j'kt}\Big)$$

$$i \in \mathbf{I}_j,\ i' \in \mathbf{I}_{j'},\ 1 \leq j, j' \leq J,\ \sigma_{si} > 0,\ \sigma_{si'} < 0 \qquad\qquad (6a)$$

$$\sum_{k \in KT_j} (T_{ijkt}^c + pt_{ij}\, ys_{ijkt} + nb_{ijkt}\, ct_{ij}) \leq$$

$$\sum_{k \in KT_j} (T_{i'j'kt}^c + pt_{i'j'}\, ys_{i'j'kt} + nb_{i'j'kt}\, ct_{i'j'}) + h_t\Big(1 - \sum_{k \in KT_j} ys_{i'j'kt}\Big)$$

$$i \in \mathbf{I}_j,\ i' \in \mathbf{I}_{j'},\ 1 \leq j, j' \leq J,\ \sigma_{si} > 0,\ \sigma_{si'} < 0 \qquad\qquad (6b)$$

3.4. Inventories

Let I_{st} $(I_{st} \leq I_s^{max})$ denotes the inventory of the material state s, I_s^{os} is the amount of material state s outsourced, I_s^{sup} is the amount of material state s supplied to the customers, and I_{st}^v is the safety stock violation at the end of a interval t. We write a balance on the inventory of material state s in the storage/supplied to customers/carry over to the next interval, similar to those of Sundaramoorthy and Karimi (2004, p.8293).

3.5. Resources

All chemical plants in general and specialty chemical plants in particular require several other utilities and resources for their general operations. These resources broadly include human, utilities, waste-treatment capacity, catalysts, and laboratory. For this reason, a typical scenario in the plants is that the initial plan is reviewed by various other departments (maintenance, process, and laboratory). We capture this variability of productivity depending on the availability of resources

$$nb_{ijt} \leq a(mp_t)\ ;\ nb_{ijt} \leq b(u_t) \qquad\qquad\qquad\qquad (7)$$

where, mp_t is available number of human resource for the period t, u_t is available quantity of each of the utilities u, and a, b are the conversion constants which are specific to each plant and can be calculated based on the experience or plant logs.

Let, uc_{uit} monitors consumption of utility u. Given the specific consumption rate (μ_{uit}) and the total available amount (U_{ut}) of utility u for the interval t, we write

$$uc_{uit} = \sum_j \sum_k \mu_{uit}\,(pt_{ij}\, ys_{ijkt} + nb_{ijkt}\, ct_{ij}) \qquad\qquad (8a)$$

$$\sum_i uc_{uit} \leq U_{ut} \qquad\qquad\qquad\qquad (8b)$$

3.6. New Product Introductions (NPIs)

Following Sundaramoorthy and Karimi (2004), we define the following 0-1 continuous variable

$$yv_{ijkt} = \begin{cases} 1 & \text{if unit } j \text{ begins campaign of task } i \text{ in the slot } k \text{ of interval } t \text{ for the first time} \\ 0 & \text{Otherwise} \end{cases}$$

Now, since the validation is one-time, and happens in the beginning of the first campaign of a task i of the new product in unit j.

$$\sum_t \sum_k yv_{ijkt} \leq 1 \qquad\qquad i \in \boldsymbol{I}_j,\ 1 \leq j \leq J \qquad\qquad (9a)$$

$$yv_{ijkt} \geq ys_{ijkt} - \sum_{t'<t} \sum_{k'} yv_{ijk't'} \qquad i \in \boldsymbol{I}_j,\ 1 \leq j \leq J,\ 1 \leq k < KT_j,\ 1 \leq t \leq NT \qquad (9b)$$

To include the validation time (vt_{ij}) into our timing constraints, we modify 4a, 4b, and 6a accordingly.

3.7. Strategies to include maintenance and reactive scheduling
The time intervals and the duration for routine maintenance are known a priori. So, we fix binary variables in our model to perform the maintenance.
Our model can easily handle uncertainties and revisions of plan. For this, we redefine the intervals from the current time and update model status by fixing the current values of variables as the initial condition for the revised model.

3.8. Planning Objective: costs and profits
The most preferred objective in planning process is the maximization of gross profit (revenue through sales – cost of production).
Cost of production includes processing cost (pc_{sj}), sequence-dependent changeover/cleaning cost ($cc_{ii'}$), maintenance cost (mtc_j), inventory holding cost (hc_{st}), material procurement cost (mc_{st}), waste treatment/disposal cost (wc_s), utility usage cost (ut_{ut}), loss for delaying order delivery (dc_s), and a penalty (spc_{st}) for the violation of safety stock limit. Let v_s denote the revenue per unit sale of material state s. Thus,

$$\max \text{NGP} = \sum_t \sum_t v_s I_{st}^{sup} - \text{cost} \qquad\qquad (10)$$

This completes our model (SK$_{planning}$, eqs. 1-10 and few other constraints) for operational planning.

4. Model Evaluations and Results

We present a case study from a multiproduct specialty chemical plant, to demonstrate the performance of our model. Our case study involves 13 tasks (9 process tasks, 3 maintenance tasks), 13 material states ($m1 - m13$), 3 units ($j1 - j3$), 25 operators, and 1 potential new product (tasks 7, 8, 9). We consider a planning horizon of 6 months. Table 1 consolidates the model and solution statistics for all the scenarios. For our evaluation, we used CPLEX 11/GAMS 22.8 on a LENOVO computer with AMD Athlon™ 64X2 Dual Core Processor 6000+ 3 GHz CPU, 3.25 GB RAM, running Windows XP Professional. Also, we present results by solving this case study for 4 different scenarios.

4.1. Scenario 1: 25 Operators
This is the base scenario and involves scheduling of product campaigns, maintenance, sequence-dependent changeover times, and 25 available operators. This scenario does not include the introduction of the new product.

4.2. Scenario 2: 19 Operators
We solve scenario 1 again with a limiting human resource. In this scenario, numbers of operators available are only 19. As expected, here the gross profit is less than the scenario 1 (5408.04 Kg vs. 11349.5 Kg). Similarly, we can solve our model for other limiting resources such as laboratory, waste treatment, parts, and utilities.

4.2.1. Scenario 3: Outsourcing

For this scenario, we allow outsourcing for one of the intermediates (material - $m6$). We consider the base scenario with 25 operators. The gross profit for this scenario is 11642.5 Kg.

4.3. Scenario 4: Outsourcing + NPIs

In this scenario, we allow both outsourcing of intermediate (material - $m6$) and the introduction of the new product (material - $m13$).

Table 1 Model and Solution Statistics

Statistics	Scenario 1 25 Operators	Scenario 2 20 Operators	Scenario 3 Outsourcing	Scenario 4 Outsourcing+NPIs
binary variables	274	274	274	385
continuous variables	1327	1327	1327	2287
constraints	2825	2825	2825	4529
non-zeros	9781	9781	9781	18621
MILP objective (Kg)	11349.5	5408.04	11642.5	14071.84
Relative gap (%)	0.08	0	0.06	0.24
CPUs	5000	296	5000	5000

% Relative gap [(best estimate - best integer) / best integer] - represents the upper bound for the distance between the best integer and optimal solution

5. Conclusions and Future Work

We successfully modify the model of Sundaramoorthy and Karimi (2004) to develop a simpler MILP model for campaign planning. We also demonstrate the usefulness of our model by evaluating 4 scenarios for a multiproduct specialty chemical plant and a planning horizon of 6 months. Our model successfully captures process variability with limited resources, sequence-dependent changeover times, and several real-life resources, features, and scenarios. We are currently working on further improving the solution efficiency of our model and interact further with a local company to improve the utility and acceptability of our model by the industry. Also, we are developing a planning tool that can readily generate several scenarios and adapt to the dynamic requirements of various stakeholders of a company.

References

Corsano, G., Aguirre, P. A., Montagna, J. M., 2009. Multiperiod design and planning of multiproduct batch plants with mixed-product campaigns. AIChE Journal 55, 9, 2356-2369.

Stefansson, H., Shah, N., Jensson, P., 2006. Multiscale planning and scheduling in the secondary pharmaceutical industry. AIChE Journal 52, 12, 4133-4149.

Sundaramoorthy, A., Karimi, I. A., 2004. Planning in pharmaceutical supply chains with outsourcing and new product introductions. Industrial and Engineering Chemistry Research 43, 8293-8306.

Sundaramoorthy, A., Xianming, S., Karimi, I. A., Srinivasan, R., Presented in PSE-2006, July 09-13. An integrated model for planning in global chemical supply chains.

Suryadi, H., Papageorgiou, L. G., 2004. Optimal maintenance planning and crew allocation for multipurpose batch plants. International Journal of Production Research 42, 2, 355-377.

Susarla, N., Li, J., Karimi, I. A., 2009. A novel approach to scheduling multipurpose batch plants using unit-slots. *In press* AIChE Journal

Order acceptance for revenue management and capacity allocation in make-to-order batch plants

Ernesto Martínez, Facundo Arredondo

INGAR(CONICET-UTN), Avellaneda 3657, Santa Fe 3000, Argentina,
ecmarti@santafe-conicet.gob.ar

Abstract

Challenging issues for optimal capacity allocation in make-to-order (MTO) batch plants are fixed manufacturing capacity and a highly diversified product portfolio compounded with pronounced fluctuations in demand and profitability. In this work, revenue management is carried out by maximizing profits under uncertainty in MTO production systems using an intelligent decision rule to dynamically control the inflow of orders. A novel approach for learning and update of an order acceptance policy using data is proposed. Comparisons made with threshold heuristics for capacity control in a multiproduct batch plant highlight superior performance of the dynamic order admission policy resulting from selective order acceptance using revenue management.

Keywords: batch plant management, capacity control, make-to-order production, order acceptance, revenue management.

1. Motivation and scope

Order acceptance is a key success factor for revenue management in make-to-order (MTO) manufacturing systems where a fixed manufacturing capacity and a great variety of offered products are challenged by pronounced fluctuations in demand and profitability. Also, in MTO batch plants swarming demand caused by a seasonal factor or a new product launching cause stringent capacity situations which requires selective order acceptance. This constitutes a classical environment for employing techniques of revenue management (RM) in manufacturing, capacity control in particular (Barut and Sridharan, 2005; Defregger and Kuhn, 2007). RM provides a systematic approach to quantitatively assess the opportunity costs of accepting or rejecting an arriving order. In this work, the objective of capacity control is focussed on maximizing the overall profit obtained from the (fixed) installed capacity by adequately selecting the best orders. To cope with different sources of uncertainty involved in order acceptance (arrival times, order types, profit and resources required) a reinforcement learning approach is integrated with RM to develop and update an order acceptance policy using data (Arredondo & Martínez, 2010).

2. Order acceptance for revenue management

Let's consider the order acceptance problem of a multiproduct batch plant which uses a MTO order fulfillment strategy and faces excess demand for a highly diverfied product portfolio. One of the common characteristics of MTO companies is the rigidity of available capacity. Although manufacturing capacity can be considered as flexible in the long term, increasing the capacity substantially in the short term is not possible. Spot market sales are considered such that for every decision epoch (e.g. two days or one week) there is a defined planning horizon T. Time can be divided into a sequence of

non-overlapping decision epochs *t, t+1,.., t+T* over a rolling planning horizon where processing capacity is fixed. Since production activities are strictly linked to customer orders, unused capacity is lost if it is not allocated. Each order is characterized by distinctive attributes such size, product mix, due date, divisibility and price. These attributes typically vary over a continuum which prevents *a priori* classification of any arriving order as belonging to a given type. Moreover, in MTO manufacturing it is generally not possible to obtain accurate forecast information about the timing and attributes of expected future orders over the planning horizon (Wu & Chiang, 2009).

The success of MTO manufacturing is heavily dependent on the selectivity of an order acceptance policy that seeks to maximize –under different sources of uncertainty– the average revenue obtained per unit cost of requested capacity. In Fig. 1, the hierarchy of functions involved in an order management system is shown. The decision an MTO plant manager has to make for an incoming order is whether to *accept, negotiate* or *reject* it depending on the revenue contributing value of the order and the possibility of inserting the order in the shop-floor schedule. As a guideline for selectivity, it is proposed here that when demand exceeds production capacity the logic for order acceptance should be based on the following simple rule: *only accept orders that help increasing the short-term average revenue obtained per unit cost of installed capacity*. This average will be referred to as ρ hereafter. A negotiation process of an order is only worth pursuing if, by accepting it, the average ρ is going to be increased in the near future. Otherwise, the order should be rejected. The rejection of an order may have important repercussions from the perspective of customer relationship management (CRM), though. Thus, there must be a cost for rejecting orders that may increase ρ, but that cannot be inserted into the current schedule due to previously accepted orders.

To formulate the order acceptance problem more formally, let's assume there exist a rolling planning horizon which is divided into T time intervals of the same length which will be referred to as decision epochs. At each epoch *t*, there exists a list of pending orders ℓ^t and the decision variable is the threshold ζ^t for acceptance or rejection of each candidate order. This threshold is defined as the minimum revenue per unit cost of capacity required by any order in ℓ^t to be eligible for shop-floor scheduling. To this aim, the list of orders ℓ^t is assumed ranked based on the revenue contributing *value* of each order. In what follows, let $r^t(\zeta^t, \ell^t)$ be the revenue obtained from all orders accepted during time window *t* and $Q^t(\zeta^t, \ell^t)$ is the cumulative revenue that can be obtained over the planning horizon if the threshold ζ^t is chosen for the time window *t* and the optimal thresholds are chosen thereafter. Finally, let $R^t(\zeta^t)$ be the cumulative revenue generated over the planning horizon if the optimal sequence of order acceptance thresholds is used from time epoch *t* onwards.

By definition, $\zeta^t(\bullet)$, $Q^t(\bullet)$, and $R^t(\bullet)$ are all random functions whose values depend on attributes of arriving orders and the actual realization of the demand epoch-wise. For revenue management under uncertainty, at each decision epoch *t*, the following stochastic dynamic program must be solved

$$E[R(\ell^t)] \equiv \max_{\zeta^t} E[Q^t(\zeta^t, \ell^t)] = \max_{\zeta^t} E[r^t(\zeta^t, \ell^t) + R^{t+1}(\ell^{t+1})] \tag{1}$$

for the optimal sequence of thresholds $\zeta_*^t, \zeta_*^{t+1}, ..., \zeta_*^{t+T}$, where the expectation is taken with respect to a stochastic demand process. In practical terms, solving (1)

Order acce ptance for reve nue ma nagement a nd ca pacity a llocation in MTO ba tch plants
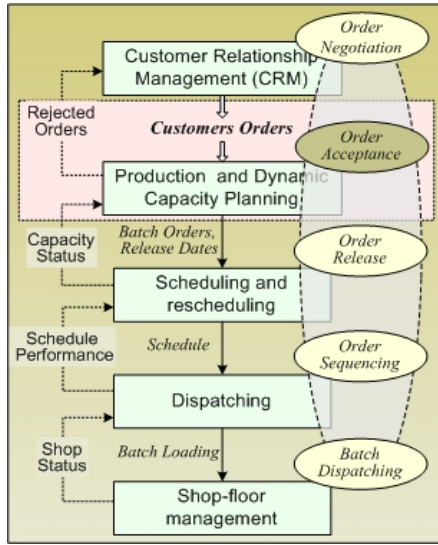
Fig. 1. Hierarchy of functions in order acceptance.   Fig. 2. Decision logic using $\mathcal{R}$-values.
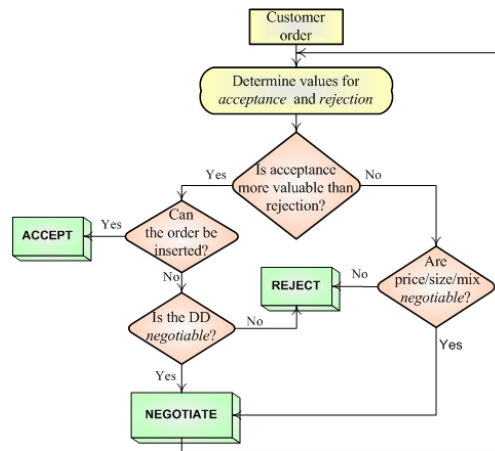
requires a prediction of which orders are expected for future decision epochs in the planning horizon. Due to a highly diversified product portfolio alongside with pronounced demand fluctuations this forecast is subjected to significant estimation errors (Wu & Chiang, 2009). Even for a simple model of the demand, $Q^t(\zeta^t, \ell^t)$ cannot be expressed in closed form and needs to be evaluated via simulation.

3. Policy learning and adaptation

To solve (1) using a reinforcement learning approach (Sutton & Barto, 1998), instead of determining the optimal revenue threshold for each decision epoch *t*, let's consider each order o_i in the ranked list ℓ^t for an aceptance or rejection decision based on the *gain* values $\mathcal{R}(o_i, acc)$ and $\mathcal{R}(o_i, rej)$, where each $\mathcal{R}$-value is simply the net effect on the average ρ following the acceptance or rejection of the order o_i, and acting optimally thereafter (Arredondo & Martínez, 2009). Should these $\mathcal{R}$-values be known *a pri ori*, order o_i is elegible for shop-floor scheduling if and only if $\mathcal{R}(o_i, acc) > \mathcal{R}(o_i, rej)$, otherwise order o_i must be rejected. The decision-making logic in Fig. 2 summarizes the rationale of resorting to gain values for order acceptance. If $\mathcal{R}(o_i, acc) > \mathcal{R}(o_i, rej)$, insertion of the order ''o_i'' into the current schedule is then attempted. If order insertion for shop-floor processing is feasible, the order is then accepted; otherwise, the due date may be, if possible, negotiated with the client to avoid rejecting it. If the gain values for the order o_i, are such that $\mathcal{R}(o_i, acc) < \mathcal{R}(o_i, rej)$, negotiation of order attributes is mandatory in order to increase $\mathcal{R}(o_i, acc)$ so as to exceed its rejection value $\mathcal{R}(o_i, rej)$.

To learn $\mathcal{R}$-values from simulated or actual data, the ARLOA algorithm proposed by Arredondo & Martínez (2010) is used here. For each order in ℓ^t, the greedy policy chooses the action *a* (accept/reject) with the highest $\mathcal{R}$-value when exploiting current knowledge. $\mathcal{R}$-values are updated following each action taken using the learning rule:

$$\mathcal{R}_{new}(o_i, a) \leftarrow \mathcal{R}_{old}(o_i, a) + \alpha \left[r - \rho + \max_b \mathcal{R}_{old}(o_j, b) - \mathcal{R}_{old}(o_i, a) \right] \tag{2}$$

where r is the reward resulting of accepting or rejecting o_i ℓ^t, whereas $0< \alpha < 1$ is the learning rate. Using revenues obtained from recently *accepted* orders, the average reward ρ is increasingly updated using the cumulative revenue obtained per installed capacity over T previous decision epochs

$$\rho = \sum_{t-T}^{t-1} revenue(orders) \big/ (T \times max\_cap) \tag{3}$$

where *max_cap* is the installed capacity available at every t. As no revenue is get from rejected orders, $\rho \equiv 0$ is enforced in (2) to update $\mathscr{R}(o_i, rej)$ whenever an order is rejected. For any order o_i which has been accepted, the reward r is exactly the same revenue obtained per unit cost of processing capacity requested (p_i). However, for rejected orders the reward r must be defined so as to account for opportunity costs (if any), based on current $\mathscr{R}$-values and the logic in Fig. 2

$$r(o_i) = -p_i \text{, if } \mathscr{R}(o_i, acc) \geq \mathscr{R}(o_i, rej) \text{ ; } r(o_i) = 0 \text{, otherwise.} \tag{4}$$

As can be seen a penalty (negative reward) is given to the decision of rejecting an order o_i which, according to its estimated $\mathscr{R}$-values, should be accepted.

Often orders have attributes that vary over a continuum. However, there always exist clusters of similar orders that favor inductive learning of $\mathscr{R}$-values. To generalize $\mathscr{R}$-values based on order similarity locally weighted regression (LWR) is used. LWR is a variation of standard regression techniques, in which training points close to the query point have more influence over fitted regression surface than those which are further away. LWR attempts to fit $\mathscr{R}$-values only in a small region around the location of the query point (attributes for an arriving order o_i). Data points are weighted according to a function of their distance to the query point. This function is typically a kernel function (k), where the 'width' parameter is known as the bandwidth (h):

$$k(d, h) = e^{-(d/h)^2} \tag{5}$$

where d is the Euclidean distance between the query point and each point in the training dataset. To effectively learn an order acceptance policy in the face of uncertainty the dilemma of exploitation vs. exploration must be addressed. This means that to discover better orders to profit from it is necessary to accept orders whose rejection seems, at first glance, the best thing to do. The trade-off between exploitation and exploration can be achieved in different ways. The easiest alternative is the so-called *ε-greedy*. With probability $(1-\varepsilon)$ the order with highest value $\mathscr{R}(o_i, acc)$ is selected whereas with probability ε, a sub-optimal order in ℓ^t is chosen. Also, orders in ℓ^t are classified as belonging either to the acceptance set $\mathscr{R}(o_i, acc) \geq \mathscr{R}(o_i, rej)$ or to the rejection set $\mathscr{R}(o_i, acc) < \mathscr{R}(o_i, rej)$. For exploitation, only orders in the acceptance set must be chosen for shop-floor insertion. For exploration some orders from the rejection set are also considered as candidates for acceptance.

4. Case study and results

To illustrate the proposed approach, order acceptance in a single-stage process plant adapted from Musier and Evans (1989) is used. Plant equipment items are semi-continuous extruders which process orders for four different products. Each order consists of a mix of products, has a due date and generates a revenue following acceptance. There is a maximum regular capacity of hours per working day and it is

assumed that all accepted orders must be completed without violating their due dates. The demand stochastic process is simulated such that the product mix of an arriving order is randomly generated from up to five well-differentiated order clusters or types. Orders of all types arrive following a Poisson process with mean λ=10 orders/day. Each order has a dominant product and the remaining product mix is evenly distributed among the other three products. Order types #1, #2, #3 and #4 have as dominant products P1, P2, P3 and P4, respectively. For each order type, the percentage of the dominant product is randomly obtained from a normal probability distribution with mean μ= 60% and standard deviation σ=5%, whereas for the other products in the order are evenly distributed. For order type #5, the mix of products is roughly balanced. Based on their p_i's preferences by order types (assumed unknown) are as follows: $4 \prec 3 \prec 2 \prec 5 \prec 1$. For all order types, due dates are generated using a (discrete) uniform distribution between 5 and 10 working days. Demand data are shown in Table 1.

Table 1. Demand data for the Base Case

Product	P1	P2	P3	P4	
Processing rate (kg/day)	30	25	25	30	
Revenue ($/kg)	5	2	1	0.5	
Order type	**#1**	**#2**	**#3**	**#4**	**#5**
Size (kg)	15	15	15	15	15
Composition (% Dom. Prod.)	$N\sim(60, 5)$	$N\sim(60, 5)$	$N\sim(60, 5)$	$N\sim(60, 5)$	$N\sim(22, 3)$
Due date (days)	$U\sim(5, 10$)	$U\sim(5, 10)$	$U\sim(5, 10)$	$U\sim(5, 10)$	$U\sim(5, 10)$
Arrival rates (% of λ)	10	20	30	30	10

Table 2. Demand changes (prices and arrival rates) for Case 1.

Product	P1	P2	P3	P4	
Revenue (before change)	5	2	1	0.5	
Revenue (After change)	4	1	0.5	0.5	
Order type	**#1**	**#2**	**#3**	**#4**	**#5**
Arrival rates (% of λ)					
Before change	15	20	25	25	15
After change	5	20	35	35	5

In Fig. 3, the learning curve for this illustrative example is shown for the order admission policy (ARLOA) based on the weekly revenue obtained from accepted orders. Learning curves were obtained by averaging five independent simulation runs. A comparison is made between learning $\mathcal{R}$-values, an optimal threshold heuristic $r \geq 55$ (ζ=55), the Q-learning rule and the first-come-first-served (FCFS) acceptance policy. The sub-optimality of both FCFS and Q-learning rules is quite noticeable. To assess the selectivity of the order acceptance policy, an independent simulation run was made without allowing further updates of the $\mathcal{R}$-values. Results obtained for the acceptance rate of each order type are shown in Fig. 4. As can be seen, the admission policy is highly selective since nearly 90% of type #1 orders are accepted. To assess the

adaptiveness of the admission policy, a significant reduction in the arrival rates and revenues of order types #1 and #5 are made as it is shown in Table 2. In Fig. 5, the learning curve of the adaptive policy (ARLOA) is compared to different thresholds.

5. Conclusions

A novel approach to capacity control under uncertainty in make-to-order batch plants using revenue management and reinforcement learning was presented. An adaptive policy for order acceptance based on the revenue contributing value of an order with regards to the average revenue obtained per installed capacity has been proposed.

References

F. Arredondo & E. Martínez, 2010, Learning and adaptation of a policy for dynamic order acceptance in make-to-order manufacturing, Computers and Industrial Engineering, 58, 70-83.

M. Barut & V. Sridharan, 2005, Revenue management in order-driven productions. Decision Sciences, 36, 2, 287–316.

F. Defregger & H. Kuhn, 2007, Revenue management for a make-to-order company with limited inventory capacity, OR Spectrum, 29, 137–156.

R. Musier & L. Evans, 1989, An approximate method for the production scheduling of industrial batch processes with parallel units, Computers and Chemical Engineering, 13, 229-238.

R. Sutton & A. Barto, 1998, Reinforcement learning: An introduction. London, UK, MIT Press.

A. Wu & D. Chiang, 2009, The impact of estimation error on the dynamic order admission policy in B2B MTO environments, Expert Systems with Applications, 36, 11782–11791.
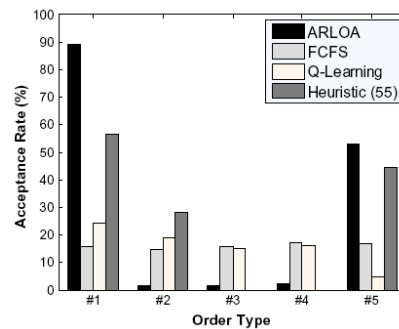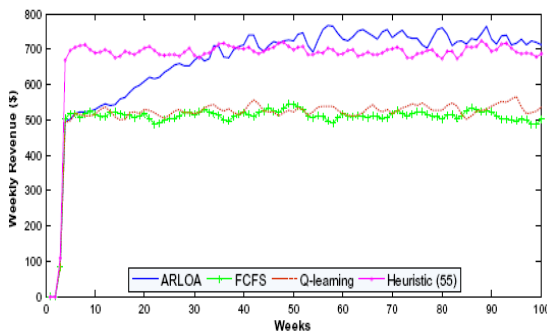
Fig. 3. Learning curves for the Base Case.

Fig. 4. Policy selectivity comparisons.
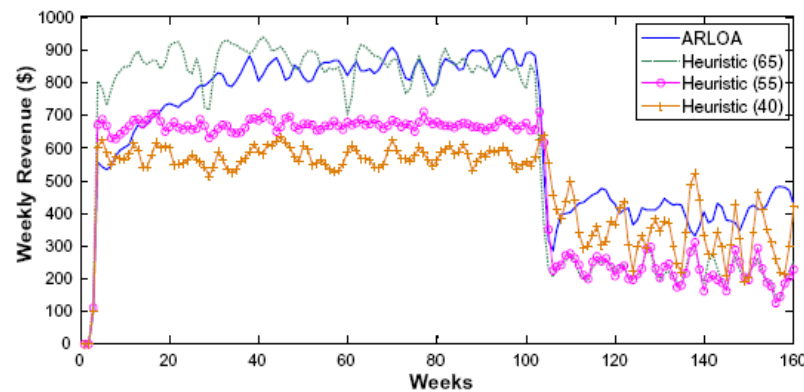
Fig. 5. ARLOA vs. different thresholds ($\zeta=40$, $\zeta=55$ & $\zeta=65$) for Case 2.

Scheduling and control decision-making under an integrated information environment

E. Capón-García[a], E. Muñoz[a], M. Moreno-Benito[a], A. Espuña[a], L. Puigjaner[a]

[a]*Chemical Engineering Department, Univesitat Politècnica de Catalunya, ETSEIB, Av. Diagonal 647, E08028, Barcelona, Spain. Email: luis.puigjaner@upc.edu*

Abstract

The complexity of decision-making in process industries and the need of highly competitive organizations require new supporting tools to coordinate and optimize the information flow among decision levels. This work presents a framework for integrating the scheduling and control decision levels by means of an ontology, which allows and coordinates the information exchange among the different modeling paradigms/conventions currently used for the enterprise-wide optimization (EWO). A multiproduct batch plant rescheduling case study is presented that illustrates the proposed working procedure.

Keywords: integrated process control and scheduling, multiproduct batch plants, ontology, decision-levels integration

1. Introduction

In a global but continuously changing production scenario, the importance of enterprise-wide optimization (EWO) and the need for development of new supporting tools to coordinate decision making at the different time and scale levels involved in EWO is highlighted by recent work in this research area [1, 2].

The first requirement to achieve such coordination is to define standardized information structures and more sophisticated information tools to exploit them, in order to improve availability and communication of data between different decision levels and the models behind the corresponding decision support tools. Under this condition, the different objectives of the different decision levels may converge to achieve an optimal action from the overall system point of view.

The aim of this work is to provide a framework to improve the decision making in a production process at the manufacturing (planning/scheduling) and basic control levels, consisting of an efficient information system that integrates online and historical information between these two typically independent decision levels. These decision levels have different time and space scales; however, the adopted decisions in one level influence the performance of the other one. Hence, an efficient communication framework is provided.

2. Proposed framework

On the one hand, scheduling decisions are taken at two time points: an initial off-line production scheduling and on-line rescheduling that updates the decisions taken at this level according to actual plant operation, and higher decision levels. On the other hand, control system level performs a supervision and regulation of process variables along

time, while implementing upper level decisions such as the schedule ones, in order to achieve the production goals. The operation conditions and set points in each process operation and stage are usually fixed by the control recipe, which is determined by the optimization action of the scheduling system when a master recipe is loaded.

However, the master recipe usually includes fixed processing times for the processing stages, so the scheduling optimization cannot take any action over process variables. In this framework, we propose the use of approximated dynamic models to be included in the scheduling stage. In any case, approximated models have an obvious deviation from real operation, and initially computed process times might not remain valid during the complete decision making process, and also online rescheduling should be considered along the time horizon. The use of ontology as integrator and supporter of a standardized information structure allows establishing a relationship among the models behind the different decision levels and makes the proposed approach feasible from an implementation point of view, ensuring semantic interoperability between different modelling paradigms. In addition, it allows them to express their objectives comprehensibly, resulting in a straightforward integration of both levels (Figure 1).
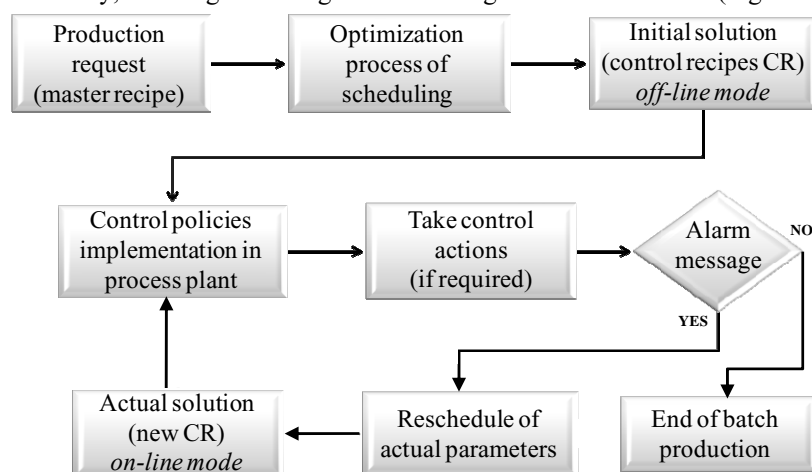
Figure 1: Diagram flow of the proposed solution procedure.

2.1. Ontology

The proposed ontology should support different activities by streamlining information gathering, data integration, model development and decision making [3,4]. Site, master and control recipes, representing a chemical flexible process, have been developed and distributed inside the ontology. These recipes contain a variety of information about available raw materials, processing requirements, the manufacturing of a single batch of a specific product, etc. Once this information has been created, the schedule and lot sizing information is made available by the corresponding optimization algorithms.

2.2. Scheduling formulation

In this work, a mathematical program based on the general precedence notion has been adopted to represent the scheduling problem. However, the proposed framework is suitable regardless of the chosen mathematical formulation.

The scheduling optimisation problem consists of maximising the profit of the plant along the time horizon, considering the income and the energy costs, and penalizing unaccomplished demand (Eq. (1)). In any case, the objective function depends on the

decision maker criteria and could be easily included as a part of the ontology information system.

$$z_{profit} = \sum_i pbatchprice(i) \cdot W(i) - \sum_{i,p} W(i) \cdot py(i,p) \cdot C\_energy(p) \cdot Temp(p) -$$
$$- penalty \cdot \sum_{i,p} pbatchprice(i) \cdot (1 - W(i)) \tag{1}$$

The dynamic stage model is represented by an expression that relates time with the decision process variables that have influence over the scheduling decisions. For instance, in the presented case study the operation time of product p, *operT(p)* is assumed to be related to the process temperature, *Temp(p)*, by means of an exponential correlation:

$$opertT(p) = A0(p) \cdot \left(Temp(p)^{B0(p)}\right) \quad \forall p \tag{2}$$

In order to implement online rescheduling strategies, the same formulation is used, but the batches that have already been produced are disregarded in the new problem, and the production time availability of the different production units is controlled.

2.3. Process control actions

On the one hand, control actions aim to implement in the process the control recipes, which are determined by a coordinating control action accordingly to the master recipes given by the scheduler. The specific values of the process variables of each unit as well as transition logics between each phase inside an operation or procedure are usually defined in the control recipe. However, by introducing dynamics in the scheduling level, some operational set points are defined then, instead of being fixed by control recipe.

The second function in process control level is the performance of the basic control, such as regulator control loops and supervisor system. This is implemented in the same process, with classical PID controllers appropriately situated and tuned, as well as alarm generators if delay in batch execution exceeds a specific value.

2.4. Integration

The integration is obtained by the use a set of rules which are defined within the ontology (e.g. ProductionPlanListEntrySlot ∃ ProductionPlanListEntry, this rule ensures that a production plant list entry must exist for the production plan list entry slot). Therefore, the former set of rules ensures that the attributes in the master recipe class and control recipe class, described within the ontology, are relating the specific parameters that the latter class needs from the former. When the model needs to be filled with new information, a Remote Procedure Call (RPC) module on the demanding agent will invoke a request to the RPC module of the server, which interacts with the ontology and the collaboration application programming interfaces (APIs) to provide the requested data, which are stored in a remote data base.

3. Results

The case study consists of a multiproduct batch plant that processes two products, through three stages. The first stage is an isothermal reaction process where a total of 95% conversion must be achieved, and whose processing time is a function of the temperature. Product batch sizes, optimal production times and product demands are given in Table 1. A single unit is available for each stage and unlimited intermediate

storage policy is adopted. Product changeover time and cost are disregarded. The time horizon is considered 6 hours.

Table 1. Product batch sizes, demand and production rates.

Processing time[h]		A	B
Stage	Equipment		
1	R1	Dynamic	Dynamic
2	P1	0.5	0.8
3	C1	0.3	0.4
Batch size [ton/batch]		5	6
Demand [ton]		20	24
Price [m.u./batch]		30	50

In unit R1, the reaction process stage for each batch is performed. Its dynamics are fully characterized in Matlab for both products (A, B). The reaction system consists of a continuous stirred tank reactor, and its corresponding reaction equations and rate laws are defined in Eq. (4) and (5). A temperature control loop with a PI controller (K_P = 0.964 and K_I = 0.030 s^{-1}) is used to lead reactor temperature to the desired value through the mean temperature in the jacket content.

$$C1 + C2 \xrightarrow{k_A} A \; ; \quad r_A = k_A \cdot C_{C1} \cdot C_{C2} \qquad (4)$$

$$C1 + C3 \xrightarrow{k_B} B \; ; \quad r_B = k_B \cdot C_{C1} \cdot C_{C3} \qquad (5)$$

Table 2. Kinetic parameters in reaction process.

$k_{A0} = 2.1 \cdot 10^5$ kg A·s^{-1}·m^{-3}	$E_A = 59.029 \cdot 10^3$ J·mol^{-1}
$k_{B0} = 1.8 \cdot 10^8$ kg B·s^{-1}·m^{-3}	$E_B = 74.826 \cdot 10^3$ J·mol^{-1}

The procedure logic in each batch, and actuation variables values, are defined using notation from ISA standard [5] in its control recipe, as shown in Figure 2.

The control recipe is defined by the coordinating control, based on the master recipe, which is defined by the scheduler. In this case, the process variable introduced in the scheduling level is temperature (see Eq. (2)). The adjusted parameters are in Table 3.

Table 3. Adjusted constants and values for Eq. (1) and (2).

	A	B
A0 [·10^7 s]	2.2325	6.4355
B0	-2.9081	-2.5180
C_energy [m.u./ºC]	4	4

Process plant operations and control actions are simulated using Matlab-based models, and the scheduling model is implemented in GAMS and solved using BARON 8.1.
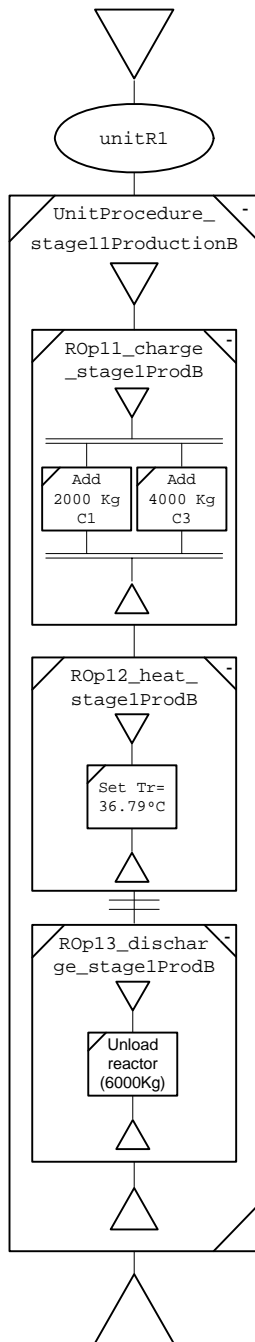
The Gantt chart of the solution of the previous case study is shown in Figure 3a. The objective function and corresponding temperatures for the batches are given in Table 4 ("initial case" column).

Table 4. Results for the case study.

Variable	Initial case	After rescheduling
OF [m.u.]	650.62	$-1.58 \cdot 10^4$
Temp(1) [ºC]	35.05	35
Temp(2) [ºC]	36.79	35
time_op(1) [h]	0.8	0.8
time_op(2) [h]	0.5	0.58

Next, the simulation of the process is carried out, including an unexpected event: the first batch resulted in a contaminated batch, and the reaction rate of product B changes as it would happen if k_B decreased to $k_B/2$. Hence, the reaction operation time of the first batch increases to 1.12 h. For this reason, the control system sends an alarm signal and rescheduling is started. The first action can only be taken from the third batch on. The ontology is responsible for transferring the information and process constraints between the lower control level, such as foreseen time of unit availability and processed orders, the scheduling level, which will eventually reschedule the orders, and even other decision levels, such as planning, if it was necessary. The resulting scheduling is shown in Figure 3b. Profit is negative because demand cannot be fully accomplished, which is highly penalized (400 m.u./batch). Since a change in the reaction temperature (among the safety limits) would not be enough to increase the fulfilled demand, the operation temperature is the minimum to maintain costs to the minimum, accomplishing the maximum possible demand. New actions are taken in the reactor control set points, as shown in Figure 3d.

In Figure 3, the difference between predicted reaction times (R1 in Figures 3a and 3b), and the actual implementation time (Figures 3c and 3d) can be clearly observed. The first batch was contaminated, but a difference between the expected reaction time of the adjusted model and real reaction time for the other batches is detected as well.

Figure 2. Procedure function chart for the reaction process stage in the control recipe during the production of 1 batch of B.
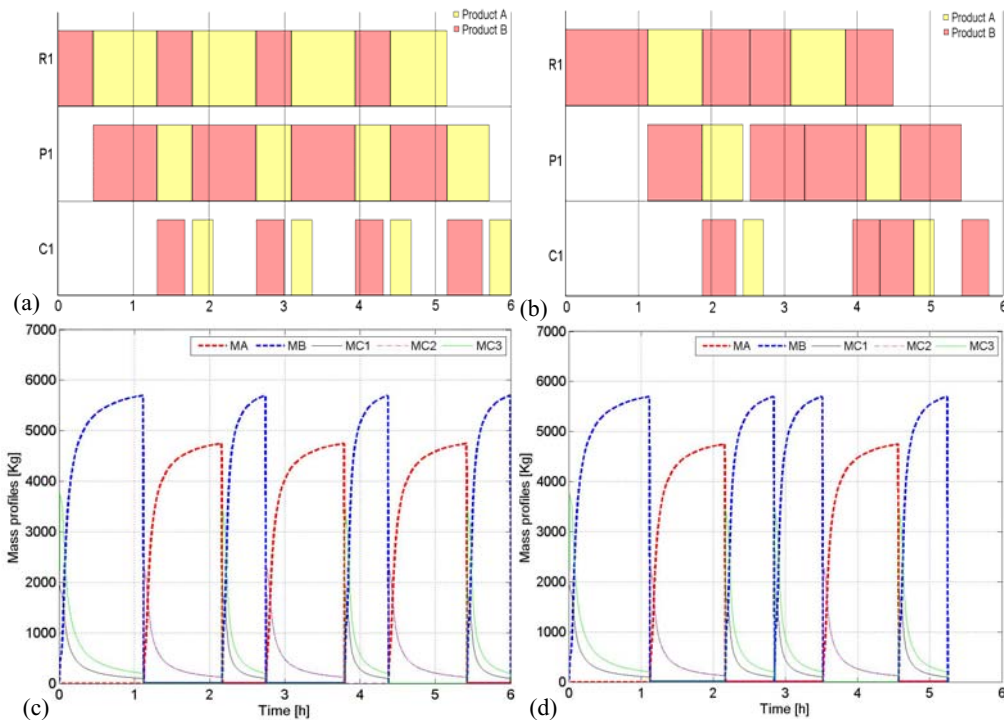
Figure 3. Gantt chart for initial optimal solution (a) and after the first rescheduling action is ordered (b), and mass profiles in reactor R1 for initial solution (c) and after rescheduling (d).

4. Conclusions

The proposed framework allows the integration of the scheduling and the control decision levels by means of an ontology which provides adequate tools to communicate both levels, which may lead to improved corporate decisions. In addition, process recipes including process decision variables have been successfully introduced in the scheduling level. Current work aims at extending the proposed framework to include other decision levels, as well as further solution strategies, overall objective functions and control actions. Further work should also focus on improved approximations of dynamic models at the different decision making levels.

5. Acknowledgements

References

[1] I. Grossmann. AIChe Journal, 51 (2005) 1846.
[2] R.Rengaswamy, (1995). Ph.D. thesis, Purdue University.
[3] V. Venkatasubramanian, C. Zhao, G. Joglekar, A. Jain, L. Hailemariam, P. Suresh, P. Akkisetty, K. Morris, G.V. Reklaitis. Comp. and Chem. Engng., 30 (2006) 1482.
[4] T.R. Gruber. An Int. J. of Knowledge Acquisition for Knowledge-Based Syst., 5 (1993) 199.
[5] Int. Soc. for Measurements and Control. ANSI/ISA-88. Batch Control Part 2: Data Structures and Guidelines for Languages. ISA Society (2001).

Short-Term Scheduling of Multi-Product Batch Plants with Sequence-Dependent Changeovers Using Timed Automata Models

Subanatarajan Subbiah[a] and Sebastian Engell[a]

[a] *Process Dynamics and Operations Group, Technische Universität Dortmund, 44227 Dortmund, Germany. {s.subbiah@bci.tu-dortmund.de, s.engell@bci.tu-dortmund.de}*

Abstract

In this contribution we propose an approach to model and solve scheduling problems with sequence-dependent changeover procedures using timed automata (TA). The processing units and the recipes are modeled as sets of interacting TA in a modular fashion. The setup and changeover procedures are modeled explicitly as operations in the recipe. The problem modeled as TA is then solved to derive schedules by performing a cost-optimal reachability analysis starting from an initial location where no recipe operation is started to a target location where all required demands are met. The proposed approach is applied to two case studies and the results are presented. The comparison of the experimental results from the TA-based approach with established MILP techniques shows that the proposed approach is efficient in the sense that optimal schedules are found within a limited computation time that are comparable to or better than those obtained with the MILP formulations and the solvers considered.

Keywords: Batch scheduling, sequence-dependent changeovers, timed automata, reachability analysis.

1. Introduction

Multi-product and multi-purpose batch plants are widely used in the production of high valued fine or special chemicals and in the food industries where a variety of products have to be produced. In such plants one of the challenges that plant managers face is to schedule the recipe operations such that the resources are optimally utilized. The processing units which process different materials require setup and changeover times to switch the production from one product to another. In the presence of significant sequence-dependent changeovers, the utilization times of the processing units are strongly influenced by the sequence in which the products are produced. To tackle the combinatorial nature of the problem it is necessary to have efficient scheduling models and solution techniques. Most of the solution approaches proposed in the last years solve such problems by modelling them as mathematical programming formulations and applying commercial solvers to solve them. Continuous time formulations based on multiple-time grid, and slot based approaches were proposed in [2] and [3], respectively. However, the time and effort needed to formulate the problem as a MI(N)LP or as a CP model are obstacles to the application of these approaches in industry [13].

An alternative approach to model scheduling problems is to use the framework of timed automata and to solve the optimization problem using reachability analysis [9]. A particular appeal of this approach is the modular and partly graphical modeling technique which enables inexperienced users to build models. Another advantage is the availability of powerful search algorithms that can be modified and extended for special purposes. In this contribution, we discuss an application of the TA framework to model and to solve batch scheduling problems with sequence-dependent changeovers.

2. Background of Timed Automata

Timed Automata (TA) are finite state automata extended by the notion of clocks to model discrete event systems with timed behaviour [7]. A short and informal definition of timed automata is given here, for complete definition of the syntax and semantics please refer to [8]. A timed automaton is defined by a tuple $A = (\ L,\ C,\ \Theta,\ inv,\ l_0,\ F\ )$ in which, L represents the finite set of discrete locations, with an initial location l_0 and a set of final locations F. C represents the finite set of clocks assigned to the TA. The set of transitions between the locations is represented by $\Theta \subset L \times \gamma \times Act \times U(C) \times L$ where, γ is a set of guards specified as conjunctions of constraints of the form $c_i \otimes n$ or $c_i - c_j \otimes n$, where $c_i, c_j \in C$, and $\otimes \in \{\leq, ==, \geq, <, >, \neq\}$ and $n \in \mathbf{N}$. Act represents the set of actions (e.g. invoking a new event or changing the value of a variable) while a transition is fired. $U(C)$ represents the set of clocks that are reset to zero after an enabled transition fires. inv represents a set of invariants that assign conditions for staying in locations and the invariant conditions must evaluate to true for the corresponding location to be active. A transition between a source location and target location can occur only when the guard conditions are satisfied and the invariant conditions of the target location evaluate to true. The automaton is forced to leave the location when the invariant evaluates to false. An extension of TA with the notion of costs is known as weighted or priced TA. Apart from the features given above, a priced TA has an additional function that assigns cost rates to locations and costs to transitions.

3. Modeling Changeover Procedures Using TA

In our previous contribution [12] the setup and changeover procedures were modeled *implicitly* in the resource automaton by introducing individual locations for each configuration and the changeovers were modeled as transitions with guards between these locations. In contrast to our previous work, the changeovers are modeled *explicitly* in the recipes as operations in this paper and this is explained using an illustrative example. Consider a simple example process where two products A and B have to be produced using resources M_1 and M_2. The recipe for product A consists of operations op_1 and op_2 which are executed in resource M_1 and M_2 with processing duration dur_1 and dur_2, respectively. Recipe for product A is represented as $(op1, M_1, dur_1) \rightarrow (op_2, M_2, dur_2)$. Similarly, the recipe for product B is represented as $(op_3, M_2, dur_3) \rightarrow (op_4, M_1, dur_4)$. The resources M_1 and M_2 have to undergo a setup and changeover procedure depending on the sequence in which the recipes are processed in the resources.

3.1 Recipe Automata

For each product a separate recipe automaton with a set of clocks is created. The recipe automaton for product A is shown in Fig. 1. The recipe operations modeled are shown in the top block of the figure and the setup and changeover operations of the resources M_1 and M_2 modeled are shown in the bottom part of the figure. Each recipe operation is represented by two locations namely *wait* - where the operation is waiting to be executed in the corresponding resource and *execute* - where the operation occupies the respective resource for the corresponding process duration. The *wait* and the *execute* location of the operation op_1 are labeled as *wait op_1* and *exec op_1*, respectively. An additional location *finish A* is defined to indicate the termination of the recipe A. Starting the execution of an operation by occupying a resource is represented by a transition labeled α and finishing an operation by releasing a resource is represented by a transition labeled φ. The clock c_1 is introduced to model the timing behaviour of the recipe. The invariants in the *execute* location of the corresponding operation force the automaton to leave the location once the operation durations have expired. The guard conditions on the transitions labelled with φ ensures that the task is executed for the corresponding duration only. The changeover table for M_1 and M_2 is shown in Fig. 1.

The bottom left part of the automaton depicted as *Changeover for M_1* models the setup and the changeover operations of the resource M_1 with respect to recipe *A* and the bottom right part depicted as *Changeover for M_2* models the same for resource M_2. The state that a resource is waiting to perform a setup or a changeover is defined by a common *wait* location and the state that a resource is executing a setup or a changeover operation is defined by an *execute* location for each operation. The common *wait* location of resource M_1 is labeled as *wait set_{1A}*. Executing the setup operation, the changeover operation from *A* to *A* and the changeover operation from *B* to *A* are labeled as *exec Set_{1A}, exec CO_{1AA}* and *exec CO_{1BA}*, respectively. The location *sync loc_1* represents that a resource has finished performing a setup or a changeover operation and is prepared to execute the recipe operation op_1. Since the changeover from *A* to *A* takes null duration the *exec CO_{1AA}* location is defined as an *urgent* location meaning that the location is not active for more than 0 time unit. Starting and finishing the execution of the setup operation is modeled by the transition labelled σ_{1A} and δ_{1A}, respectively. Similarly, starting the execution of a changeover operation by occupying a resource is represented by a transition labeled $\alpha\chi$ and finishing it by releasing a resource is represented by a transition labeled $\varphi\chi$. The transition labelled γ_1 ensures that op_1 can be executed only after the resource M_1 has undergone a setup or a changeover procedure.

A shared variable is assigned to each resource and their values represent the current configuration of the resource. In the illustrative example, the shared variable S_1 of resource M_1 has values 0, 1 and 2 that represent the configurations *Global*, *A* and *B*, respectively. The guards on the shared variables ensures that only the respective setup or changeover operations are executed and they are updated with the new configuration after executing the corresponding operation. The clocks c_2 and c_3 models the timing behaviour of the changeover operations for resource M_1 and M_2, respectively and all the clocks in the recipe automaton are reset to 0 in every transition.

3.2 Resource Automata

After the recipes are modeled as individual automata for each processing unit a resource automaton is created. The resource automaton for M_1 is shown in Fig. 2. Each resource automaton consists of a single *idle* location representing the resource is unoccupied and a *busy* location for each of the recipe operation, setup and changeover operation that it could perform. The *busy* location represents that the resource is executing the respective operation. An *urgent* location labelled *sync* is present in the resource automaton for each recipe operation and acts as a synchronization point between the considered resource automaton and the corresponding recipe automaton. Allocation of the resource to perform an operation is represented by a transition from the *idle* to the *busy* location of the corresponding operation and release of the resource after executing the operation is represented by a transition from the *busy* location back to *idle*. The communication between the resource automaton M_1 and the recipe automaton *A* is realized by the synchronization labels of the transitions. This ensures that the resource does not perform more than one operation at a time.

Figure 1: Recipe automaton for product *A*

CO matrix for M_1		
From \ To	A	B
Global	Set_{1A}	Set_{1B}
A	0	CO_{1AB}
B	CO_{1BA}	0

CO matrix for M_2		
From \ To	A	B
Global	Set_{2A}	Set_{2B}
A	0	CO_{2AB}
B	CO_{2BA}	0

Changeover for M_1 Changeover for M_2

Once the processing units and the recipes are modelled individually as sets of interacting TA they are composed by *parallel composition* which is performed on-the-fly to reduce space complexity. Composing the sets of automata results in a single global automaton that presents the complete model of the scheduling problem. A cost-optimal reachability analysis is performed on the composed global automaton starting from the initial location, that

Figure 2: Resource automaton for unit M_1

represents all recipes are waiting to be executed and all resources are idle, and evaluates the successor states created by a successor relation. This enumerative process is continued until the specified final target location that represents all the required demands are met is reached with minimal cost. The composed automaton can be explored emplying various search and reduction techniques that reduce the computational effort. For a detailed description of the reduction techniques see [10].

4. Numerical Study

The TA-based scheduling tool TAOpt developed at the Process Dynamics and Operations Group [10, 11] was used to perform the reachability analysis. The proposed TA-based approach is applied to 2 different case-studies. From the plant description with information on the resources, the recipes in the form of RTN and a table of production orders, TAOpt creates sets of recipe automata and resource automata in a modular fashion and the reachability analysis is performed on-the-fly. The search algorithm used to explore the reachability tree is a combination of depth-first search and best-first search. The reduction techniques sleep-set method and non-laziness scheme that does not prune the optimal traces were used to reduce the search-space. The computational equipment for the experiments was a Xeon machine with 3GHz and 2GB of memory with a Linux OS. For all tests the computation time was restricted to 3600 CPU seconds and the node limit was restricted to 5 millions.

4.1 Plastic Compounding Case-Study

This example was originally introduced by [1] and deals with a single stage plastic compounding facility with 4 extruders having different capacities, processing rates and unit-dependent setup times. Within a time horizon of 30 days a total of 12 single-batch orders have to be scheduled. Each of the orders has a specific due date which has to be satisfied and an early production of the order leads to inventory costs. The goal is to minimize the overall summed up earliness. To produce the orders, discrete renewable resources (man power) are required during the process. Based on the number of workers required to run the extruders, 5 different cases (Case A – Case E) of the problem instances are considered. In cases- A, B and C of the problem instances it is assumed that a single worker is required to run each extruder. In case - A unlimited manpower is available, in cases- B and C at most 3 workers and 2 workers are present to run the extruders, respectively. In cases- D and E, unit-dependent manpower requirements are considered. The total number of workers is limited to 5 in case - D and to 4 in case - E. The discrete resources are modeled using additional shared variables in TAOpt. The problem data for all the instances considered was taken from [6].

Short-Term Scheduling of Multi-Product Batch plants with Sequence-Dependent Changeovers Using Timed Automata Models
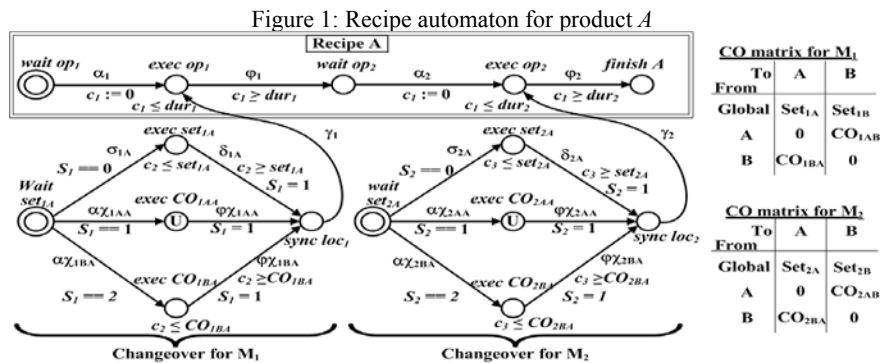
Table 1: Model statistics and computational results - (T_{CPU}) CPU time in seconds to prove optimality. ($T_{optimal}$) CPU time in seconds to reach the optimal solution, *sub-optimal solution – The computation time limit of 3600 CPU seconds was reached in these cases.

Model	Case	Binary Var.	Cont. Var.	Constraints	T_{CPU}	Opt. Sol.
[4] Janak, et. al. (2004)	A	150	513	1389	0.07	1.026
	B	458	2137	10382	6.53	1.895
	C	444	2137	10382	236.87	7.334
[5] Mendez & Cerda (2002)	A	82	24	241	0.03	1.026
	B	127	24	622	2.93	1.895
	C	115	24	490	13.28	7.334
	D	151	24	886	3600*	6.234*
	E	139	24	754	3600*	11.12*
[6] Marchetti & Cerda (2009)	A	82	24	214	0.03	1.026
	B	223	24	622	0.66	1.895
	C	223	24	622	7.39	7.334
	D	223	72	897	6.63	5.276
	E	223	72	897	8.59	11.12
		Created Nodes	Visited Nodes	$T_{optimal}$	T_{CPU}	Opt. Sol.
TAOpt	A	715	474	0.04	0.08	1.026
	B	778	509	0.06	0.09	1.895
	C	46711	27403	1.08	1.16	7.334
	D	2025	1315	0.12	0.14	5.276
	E	8465	4859	0.23	0.33	11.12

4.2 Experimental Results

The results obtained by different MILP models and by TAOpt are shown in Table 1. Comparing the CPU times to find the optimal solutions by TAOpt with the MILP model presented in [4] shows that TAOpt is faster by a substantial factor than the MILP model for all instances. In general for all instances TAOpt could found the optimal solution much quicker than any of the MILP models considered. The results in [5], [6] were computed using a Pentium 4 1.8 GHz with ILOG OPL studio 3.6. Certainly the CPU times cannot be compared directly, however it should be noted that for cases - D and E TAOpt found the optimal solution within fraction of a second whereas in [5] the optimal solution was not reached.

4.3 Ice-cream Manufacturing Case-Study

The second example considered is a medium size ice-cream manufacturing plant described in [13]. The case study is a two-stage process where 8 different products are produced (SKU A – SKU H). In the first stage, the raw materials stored in the warehouses are transported to the mixing department where they are processed in the processing line according to the product recipes and stored in the six intermediate storage units (S_1 - S_6). After a minimum waiting period has elapsed, the products are transferred to the packing lines (P_1 - P_2) where they are packed and delivered to the customers. The main features of the case study are: (a) the processing unit and the packing lines are subject to changeover procedures which are sequence-dependent (b) the intermediate storage units are limited and have to be shared and (c) the intermediate products are unstable and have a minimum waiting time and a maximum shelf-life period, introducing a timing constraint between the processing and the packing task. A routing constraint exists such that only vessels S_1, S_2 are coupled to packing line P_1 and the other four vessels (S_3 - S_6) are coupled to the packing line P_2. The main challenge in the case study is to decide on the allocation to and on the sequence of the operations on the resources such that a total demand of 91 single-batches can be produced within a week excluding a 48 hour weekend (i.e. within 7200 minutes).

Table 2: Results for the test on makespan minimization - (T_{cpu}) CPU time in seconds, (CN) total number of nodes created, (VN) total number of nodes explored, (C_{max}) makespan in min.

Set A					Set B				
Orders	T_{CPU}	CN	VN	C_{max}	Orders	T_{CPU}	CN	VN	C_{max}
20	0.23	551	195	1999.50	60	0.30	1077	358	4996.62
30	0.39	844	262	2602.07	70	1.02	1363	434	5647.52
40	0.43	971	304	4063.77	80	1.47	1609	490	6299.46
50	0.46	1041	354	4939.94	**91**	**13.13**	**124462**	**48203**	**7139.44**

4.4 Experimental results

Two sets of problem instances with different numbers of orders were solved using TAOpt (see table 2). **Set A** consists of instances with smaller numbers of orders. Here the recipes were modeled as explained earlier. **Set B** consists of instances with larger numbers of orders. When solving the instances in **set B,** a plausible heuristic was implemented that reduces the model. In the reduced formulation the recipe automata are modeled such that the transitions that represent the starting of changeover operations with longer changeover times in comparison to other changeover operations are excluded and only changeover operations with short durations are allowed to start. For the original instance with 91 batches the recipes are modeled with the reduced formulation and scheduled on the packing lines in a campaign fashion in the following sequence D-C-B-A in P_1 and H-G-F-E in P_2 since the changeover times in the packing lines are minimal in this sequence [13]. The sequencing on all other units and the complete timing was computed by TAOpt. While it cannot be ruled out that the heuristics employed prunes the optimal solution it is remarkable that a feasible solution which schedules 91 batches within a makespan of approximately 7140 minutes was computed by TAOpt within less than 15 CPU seconds.

5. Conclusion and Future Work

This paper describes an application of the TA-based approach to model and to solve batch scheduling problems which are subject to sequence-dependent changeovers and limited discrete resources. The numerical studies show that the presented TA-based approach is efficient in the sense that optimal solutions are found within limited computation time that are similar to or shorter than those required by established MILP techniques. The results on applying the approach to the case study proposed in [13] show that heuristics can be implemented easily and that the approach is applicable to practical large-scale problems. The graphical formalism and the modular definition of the elements lead to transparent and user-friendly modeling and the applicability of the technique for daily use inside factory environments with frequent changes to the model seem possible. Future work will investigate other objective functions and the embedding of rule-based techniques into the reachability analysis of TA.

References

1. J. M. Pinto and I.E. Grossmann, 1997. *Computers and Chem. Engg.,* 21, 801 – 818.
2. P. M. Castro, et al. , 2006, *Ind. Eng. Chem. Res.*, 6210 - 6226.
3. M. Erdirik-Dogan and I. E. Grossmann, 2008, *Ind. Eng. Chem. Res.*, 1159 - 1183.
4. S. L. Janak, X. Lin and C. A. Floudas, 2004. *Ind. Eng. Chem. Res.*, 43, 2516 - 2533.
5. C. A. Mendez and J. Cerda, 2002. *Computer Aided Chem. Engg.*, 10, 721- 726.
6. P. A. Marchetti and J. Cerda, 2009. *Computers and Chem. Engg.,* 33, 871 – 886.
7. R. Alur and D. L. Dill, 1994, *Theoretical Computer Science*, vol. 126, 183-235.
8. K. Larsen et al. , 2001, *Proc. 13th Int. Conf. on Computer Aided Verification,* 493 - 505.
9. Y. Abdeddaim and O. Maler, 2006*, Theoretical Computer Science,* 354, 272 -300.
10. S. Panek, S. Engell, and O. Stursberg, 2006, *Control Engg. Practice*, 1183 - 1197.
11. S. Subbiah, et al. , 2009, *Computers and Chem. Engg.,* 33, 1661 – 1676.
12. S. Subbiah and S.Engell, 2009, *Proc. 10th Int. Symposium on PSE, Salvador.*
13. Peter M. M. Bongers et al. , 2006*, Proc. 16th ESCAPE and 9th PSE,*1917-1922.

A Decomposition Approach for the Operational Scheduling of a Multiproduct Pipeline

Suelen Neves Boschetto,[a] Susana Relvas,[b] Ana Paula F.D. Barbosa-Póvoa,[b] Flávio Neves Jr[a]

[a]CPGEI, Federal University of Technology, Parana, Av. Sete de Setembro, 3165, 80230-901, Curitiba, Brazil, suelen@cpgei.ct.utfpr.edu.br, neves@utfpr.edu.br
[b]CEG-IST, Instituto Superior Técnico, UTL, Lisboa, Av. Rovisco Pais, 1049-001, Lisboa, Portugal, apovoa@ist.utl.pt, susanaicr@ist.utl.pt

Abstract

The present paper proposes an optimization structure to aid the operational decision-making of scheduling activities for a real-world multiproduct pipeline. The problem involves a distribution system composed by a unidirectional pipeline that is used to pump oil derivatives between a refinery and a distribution centre. The main goal is to obtain a pipeline schedule that accounts for end-of-pipe operations in a reasonable amount of time. A decomposition technique is developed where two MILP models, with different levels of detail at the pipeline scheduling problem, are considered (master and sub-problem). The MILP sequencing model (master) determines the volume of each batch and the sequence to be pumped. These results are then used in the MILP detailed model (sub-problem) in order to obtain the complete desired scheduling for the multiproduct pipeline. The decomposition approach is tested in typical operational scenarios involving more than 40 batches for a 30-days scheduling horizon.

Keywords: Scheduling, Multiproduct Pipeline, MILP, decomposition technique

1. Introduction

The supply chain in the petroleum industry comprises many intermediate steps starting from the exploration phase at the wellhead, going through trading and transportation, refinery and finally the distribution and delivery of the final products at the retail level. As stated by Grossmann (2004), process industry companies to remain competitive and economically viable should optimize their supply chain where costs and inventories are reduced and where the operation is performed efficiently while continuously looking to improving customer service quality. Under this perspective supply chain problems, at different planning levels, have been studied extensively in recent years. In particular, in the petroleum supply chain, oil products distribution scheduling problems have been addressed by some authors such as Cafaro and Cerdá (2009), Rejowski and Pinto (2008), Magatão et al. (2004) and Relvas et al. (2006, 2009). These authors addressed multiproduct pipelines using different approaches and dealing with different challenging aspects. Although,, it was noticed that some of the problems addresses still need further developments. This is the case of models' performance so as to guarantee good solutions in reasonable amounts of time so as to be used in real world scenarios. Medium-term pipeline scheduling problems (e.g monthly time horizons) are a good example where more efficient solution techniques are required.

In this paper we address such problem by proposing a decomposition method, based on the specific characteristics of the system under study, where the solution is obtained from two sequential MILP models. The developed models are applied to real-case studies of a Portuguese distribution company.

This paper is organized as follows: section 2 presents the generic problem under study, specifying data, desired results and assumptions. Section 3 describes the decision support tool architecture, where it is revised previous work by the authors and the decomposition method is briefly characterized. This method is then described in detail in section 4. Section 5 provides the case study and results. Conclusions and directions for further work are given in section 6.

2. Problem Statement

The problem considered in this work is based on the work by Relvas et al. (2006, 2009). The system considers a unidirectional pipeline used for pumping various types of oil products between a refinery and a distribution centre (Fig. 1). The distribution centre consists of a tank farm, where each tank has a fixed product service. Storage capacity by product is tackled in an aggregated strategy. The local market is supplied by the distribution centre through the in-site loading of tank-trucks. The problem constraints include supplying a daily requirement of client demands, management of the inventory level between capacity limits of storage, as well as operational restrictions. Operational restrictions include forbidden sequence of products within the pipeline as well as the existence of quality control tasks with associated settling periods that constrain the product availability for clients. The problem can be generically defined as:

Given:

- a. The number of products to be transported;
- b. The matrix of possible/forbidden sequences between pairs of products;
- c. The maximum storage capacity by product;
- d. The pipeline capacity;
- e. The pumping rate limits;
- f. The minimum settling period for each product;
- g. The time horizon extent and the total number of days to be considered;
- h. The maximum number of allowable batches to be pumped;
- i. The initial inventory of each product; and
- j. The daily demands for each product.

It is desired to obtain an optimal pipeline schedule and the associated inventory management at the distribution centre that guarantees client demands.

3. Decomposition Approach

One of the main conclusions from the literature review is that the sequence of products to be pumped to the pipeline has great impact on the performance and solution of the highly combinatorial MILP problem under study. Therefore, if the definition of the sequencing variables is made a priori, the remainder of the combinatorial problem should be easier to solve considering the reduction on the number of variables to be determined. One important issue to refer is that the sequencing variables have direct impact on the continuous time scale of the model. In this work, this idea is explored and a decomposition method is proposed, in order to obtain medium-term problem solutions. The decomposition is based on two sequential steps, where at each level a MILP model is solved. In a first instance, the master problem is addressed at a macro detail level, where the objective is to obtain a scenario-oriented sequence and the

volume of products to pump. The results from this level are then fed, as initial data, to the second MILP model, sub-problem, which explores the problem in a more detailed way judging whether the sequence and volume of products used returns a feasible solution. In this way, the first model defines the product sequence to transport, the batch volumes and the adequate number of batches to pump. The second one determines timing issues, beginning and ending times of each lot pumping and storage levels while considering all operational and costumer restrictions. Both models are MILP based. The simplified computational framework is presented by Fig. 2.
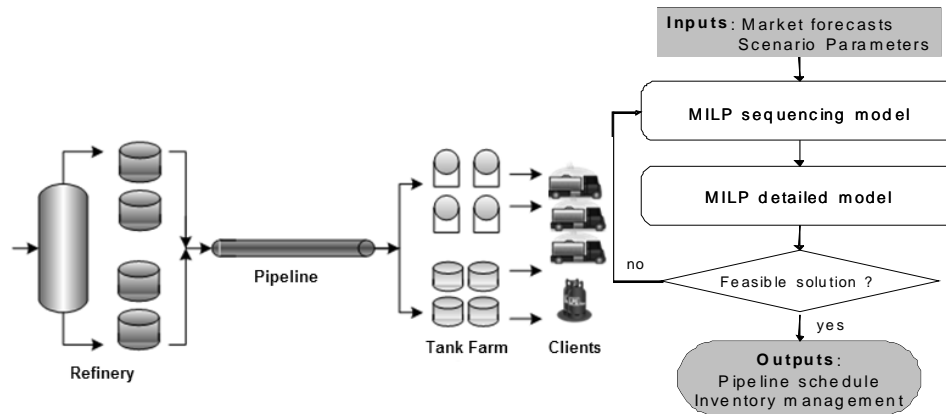

Fig 1. Problem schematic overview   Fig 2. Decomposition scheme

4. MILP Models

4.1. MILP Master model

The MILP master model formulation is based on a continuous volume and discrete time representation, where time is represented on a daily basis, but events are allowed to occur on any time instant of the time horizon. Timely based model information includes the receiving time of each batch at the tank farm (continuous) and inventory balance equations (on a daily basis).

This model defines the batch size of each product, the sequence of products to be pumped and the number of batches to pump. The volume of each batch is chosen from a continuous interval limited by lower and upper batch volume limits. The inventory is considered in order to obtain the better sequence to be pumped, respecting the storage capacity. This model has the advantage of not limiting batch sizes to pre-fixed values, which makes the solution more flexible, contrasting to the previous assumption made by the authors in Relvas et al. (2006). A third outcome of this model is the establishment of an adequate number of batches to pump, eliminating the iterative procedure of spanning a set on the number of batches and analyzing the value of the objective function. The drawback is the necessity of not limiting the scenario to a reduced number of batches, using a high number of batches margin. Taking as example a scenario with 6 products, 30 days and 60 batches, the model would have 10800 binary variables generated from the binary $ls_{i,p,k}$, that states that batch i is of product p will be received at the tank farm at day k. The binary variable index domain is reduced by developing cuts based on information such as maximum number of batches to pump and receiving time windows

per batch. These results are then used in the MILP detailed model in order to obtain the complete desired scheduling for the multiproduct pipeline.

The MILP sequencing model objective function (equation (1)) is set equal to the one of the MILP detailed model, which considers three normalized and dimensionless terms. These terms take into account the optimization of operational indicators. The first one minimizes the difference between total quantity transported and total demand (*dif*).

When the quantity transported is close to the total demand ($\sum_{p\in P}\sum_{k\in K} Dem_{p,k}$) it is obtained a solution with a minimized pumping flowrate, which is not a model variable due to linearity issues, and indirectly considering energy consumption targets. This is accomplished by including the second term that maximizes (by adding a negative sign) the amount of time used to pump products, maximizing pipeline usage. $T_{|I|}^{rec}$ is the receiving time of the last batch *i* and $h^{\max}$ is the time horizon extent. The third term maximizes (by adding a negative sign) the final inventory level of the product with the lowest level (using a minmax strategy through the variable *minid*) and balancing inventory levels. The weights presented result from previous studies regarding desired objectives for the model solution (Relvas, 2008).

$$\min \quad 3\times \frac{dif}{\sum_{p\in P}\sum_{k\in K} Dem_{p,k}} - 2\times \frac{T_{|I|}^{rec}}{h^{\max}} - 0.5\times minid \tag{1}$$

4.2. MILP Sub-problem Model

The MILP sub-problem model used is the one proposed by Relvas et al. (2006) and revised in Relvas et al. (2009). The MILP sequence model, section 4.1, determines the sequence and the volume size of each batch. These results are used as input data for the MILP detailed model. All operational constraints are here defined and the objective function is maintained equal to equation (1). The results obtained include the optimized pipeline schedule and the tank farm inventory management policy.

5. Results

The case study studies the CLC – *Companhia Logística de Combustíveis*, a Portuguese company that transports and distributes six different oil products (P1-P6). The pipeline to be scheduled has 18000 volumetric units (vu). The case under study contemplates a scenario of one month and the MILP sequencing model is applied consecutively, providing the data to feed as the detailed model (Table 1).

Table 1. Common data used in all scenarios

| | $lots_p^{\min}$ (vu) | $lots_p^{\max}$ (vu) | $ID_p^{\max}$ (vu) | stl_p (days) | Possible sequences | | | | | |
					P1	P2	P3	P4	P5	P6
P1	17300	21800	81500	1	–	✓	✓	✓	–	–
P2	8000	16000	32000	2	✓	–	–	–	–	–
P3	8000	16000	24000	1	✓	–	–	✓	✓	–
P4	3800	11800	27800	1	✓	–	✓	–	✓	–
P5	800	3200	10320	1	–	–	✓	✓	–	✓
P6	3100	6200	13120	1	–	–	–	–	✓	–

The minimum and maximum batch sizes values ($lots_p^{\min}$ and $lots_p^{\max}$) are given for each product. The operating flowrates can span from 450 to 650 vu/h. Due to product quality

issues, some sequences are not allowed inside of the pipeline. Table 1 represents the possible matches to combine two consecutive products: allowed ($\checkmark$) or not ($-$), maximum inventory capacity ($ID_p^{\max}$) and the settling period (stl_p).

5.1. Decomposition Approach Results

The model was run for 3 different scenarios based on CLC's typical operations. The scenarios used different time horizons: 1 week (S1), 2 weeks (S2) and 1 month (S3). The considered time limit for computation is 15 min, using GAMS 22.8, CPLEX 11.1 on an Intel Pentium Core 2 Duo P9400, 4 GB Ram. The results are presented in table 2. For each scenario, a single iteration was enough to obtain a solution.

Table 2. Decomposition Approach Results

	Scenario 1 (S1)		Scenario 2 (S2)		Scenario 3 (S3)	
	Master Problem	Sub-Problem	Master Problem	Sub-Problem	Master Problem	Sub-Problem
Time (s)	1.59	0.09	27.46	0.58	903.797	5.108
Objective function	-2.3041	-2.0746	-2.3052	-2.0726	-2.3052	-2.1325
Relative gap (%)	0	0	0	0	0	0
Iterations	23124	125	364731	1084	4020401	7375
Nodes	254	0	568	17	3217	140
Continuous Variables	947	2193	979	6878	11223	23804
Integer variables	180	433	756	1825	2844	6658
Constraints	2743	4042	9343	16380	24998	57852
Maximum Batches	20	-	40	-	60	-
Batches Pumped	11	11	22	22	43	43

From the analysis of table 2 it is possible to check how the master problem model complexity increases with time horizon. For a 1 month scenario (scenario 3), the optimal solution was found within the 15 min time limit (using in the end some time to polish the solution). For scenarios with this size, it is essential to effectively generate an index domain for the variables using the three indexes of the model. For instance, in S3 the total number of integer variables would be 10800 without index domain reduction. Another issue translated in table 2 is that an effective number of batches is returned, which increases similarly to the time horizon increase. Finally, it is easily checked how the objective function deteriorates from the master to the sub-problem, as expected, since there are some variables fixed *a priori*. The sub-problem presents solutions within 7-10% difference to the master problem.

5.2. Decomposition Approach Validation

In this section we compared the results obtained for scenario 3 (S3) with the previous decomposition strategy proposed by Relvas *et al* (2009) and with the real operation occurred in CLC during this period. For this month, the Relvas *et al* (2009) approach, which used a sequencing heuristic integrated with the same sub-problem model, proposed four different sequences and all of them were used in the sub-problem model. A total time of 248.7 minutes was spent in computation, having solutions for 2 of the 4 pre-determined sequences. For the solutions found, it were used 8.7 min, whereas the remaining two sequences had no solution after 2 h of computation, each. Additionally, the sequences with solutions did not have the optimality proven within the computational time limit. Still regarding timing issues, in such approach the user spent some time computing sequences and analyzing the best options, which was not considered in the computational time reported above. Another disadvantage is that once having results for several sequences, the user need to choose which is the best sequence

to apply for the operation. In this case, the selected solution had an objective function value of -1.8667 pumping a total of 39 batches. Table 3 summarizes the operational results for the developed approaches and the real operation.

Table 3. Decomposition Approach Comparison

	This Approach	S. Relvas et al. (2009)	Real Operation
# Batches	43	39	33
Max. Batches	58	41	-
Min. Batches	25	26	-
Medium Flowrate (vu/h)	516.0	543.1	511.0
ΔInventory (vu)	0	+384	-4146
Pipeline Usage (%)	98.18	93.33	98.00
Final Inventory Level (%)	51.51	51.53	
Minimum final inventory (%)	38.96 (P1)	25.24 (P5)	35.45 (P5)

In table 3, it can be seen that the medium flowrate was improved using the present approach, by replacing the exact amount consumed by clients and making use of a higher fraction of the time horizon. When comparing the final inventory levels, the present approach manages to increase the level of the product with least final inventory level. CLC's schedulers did not take into account reposition issues in their planning.

6. Conclusions and future developments

The present work proposes an optimization structure to aid the decision-making of scheduling activities for a real-world multiproduct pipeline. The decomposition approach was tested in real world scenarios to analyze performance and improvement issues, namely giving promising solutions for scenarios with more than 40 batches for a 30-day time horizon. The advantage of the present framework is the usage of a single run to compute the number of batches, sequence of products and continuous batch volumes. The disadvantages constitute focus for future research, namely the limitation of the domain index of the decision variable so as to improve the computational effort and the development of valid cuts whenever the solution of the master problem becomes infeasible at the sub-problem level. Cuts can be added at the sequence level (integer cut) or at the batch volume level (domain reduction for a continuous variable). A battery of tests should be developed to further validate the proposed approach.

7. Acknowledgements

References

D. Cafaro; J. Cerdá, *Ind. Eng. Chem. Res.*, 48, (2009) 6675.

I. Grossmann, *Computers and Chemical Engineering*, 29, (2004) 29.

L. Magatão; L.V.R. Arruda; F. Neves-Jr, *Computers and Chemical Engineering*, 28, (2004) 171.

R. Rejowski; J.M. Pinto, *Computers and Chemical Engineering*, 32, (2008) 1042.

S. Relvas, *Optimal Pipeline Scheduling and Inventory Management of a Multiproduct Oil Distribution Centre*, PhD Thesis, Instituto Superior Técnico, Technical University of Lisbon (2008).

S. Relvas; H.A. Matos; A.P.F.D. Barbosa-Póvoa; J. Fialho, *Computers and Chemical Engineering*, 33, (2009) 712.

S. Relvas; H.A. Matos; A.P.F.D. Barbosa-Póvoa; J. Fialho; A.S. Pinheiro, *Ind. Eng. Chem. Res.*, 45, (2006) 7841.

The retrofit of a closed-loop distribution network: the case of lead batteries

Ana Serra Fernandes[a], Maria Isabel Gomes-Salema[*b], Ana Paula Barbosa-Povoa[a],

[a]*Centre for Management Studies, CEG-IST, UTL, Av. Rovisco Pais, 1049-101 Lisboa,*

Portugal, apovoa@ist.utl.pt

[b]*Centro de Matematica e Aplicacoes, FCT, Universidade Nova de Lisboa, Monte de*

Caparica, 2825-114 Caparica, Portugal, mirg@fct.unl.pt

Abstract

Recent advances in global competition with the exhaustion of natural resources, and the increased society awareness towards environment created a new way of thinking when managing supply chains. Companies are now facing the need of seriously considering within their supply chain the presence of their end-of-life products. The concept of closed-loop supply chains has emerged where optimized structures are required. In this paper the design and planning of a real closed-loop supply chain is studied considering the production of lead batteries its distribution to the final clients and its collection at the end-of-life period. The company with a wide distribution network, at Portugal level, and a fleet of owned vehicles reached effectiveness in the delivery service to the final costumer. It is now vital to reach efficiency at costs level. To help reaching this goal the present work looks into the retrofit of the existing structure so as to achieve the optimized design of the closed-loop supply chain. A Mixed Integer Linear Programming (MILP) model is developed which simultaneously designs the forward and reverse chains. Various scenarios are built from which decisions regarding the retrofit of the existing network (namely, elimination, addition and replacement of warehouses) are obtained with a significant reduction of costs. The contemplated costs are: the cost of opening warehouses, the cost of the raw materials acquire to suppliers and of used products acquire to customers, as well as the cost of the different transportation resources. Besides strategic design, plans of supply, production, storage and transportation are also given by the model. The results obtained are compared with the existing network and important conclusions are drawn.

Keywords: Closed-loop Supply chain, Retrofit, Network design, Network Planning.

1. Introduction

Today companies, to remain competitive, have to provide a good service with very short delivery times and, simultaneously, at the lowest possible cost. To respond to the challenge of cost reduction and service enhancement, companies need to take a close look into the design and planning of their supply chain. Such logistics systems are however quite complex, especially when the recovery of end-of-life products is also at stake. So, it is difficult to take good decisions without the use of efficient tools to help

---
** Corresponding author

the decision support process.

The traditional supply chains, which start at raw materials and end at the final customers (forward flows), have been extensively discussed in the literature and reviews on the models studied can be found in Beamon (1998), Min and Zhou (2002), Goetschalckx et al. (2002), Shah (2005), Klose and Drexl (2005). It should be noted that these models are focused on strategic structural aspects of the supply chain, being the work developed by Sabri and Beamon (2000) an exception, as it allows simultaneously the strategic and operational (incorporates the production and the distribution) planning using a multi-objective analysis for of the supply chain.

The reverse supply chains start at the collection of the products from the costumers and end when products are adequately recovered or eliminated (reverse flows). These networks also need strategic planning concerning the location of the facilities, such as collection and/or recycling centres. A large number of published models for design or retrofit of reverse networks is reviewed in Barbosa-Povoa et al. (2007). Most of the reviewed models were developed to solve specific problems and not of general application.

When the supply chain integrated the forward and reverse networks, the *closed-loop supply chain (*CLSC) appears. The first generic model for CLSC was proposed by Fleischmann et al. (2001). In this work, the authors compared the simultaneous design of forward and reverse networks, with the design of the reverse network from an existing and operating forward network. Other published models on CLSC only contemplate strategic decisions regarding facility location, such as the case of Lu and Bostel (2007) where a model for the design of a network of producers, customers, intermediate centres and remanufacturing sites is proposed. However, very few models integrate within a single formulation the design of forward and reverse supply chain and, planning of activities such as production, acquisition, distribution, among others. The generic model presented by Salema et al. (2007) and later on generalised by the same authors (Salema et al., 2010) contributed to overcome such lack. In these models, within a single formulation, decisions concerning facilities site location and production, storage and distribution planning are taken into account simultaneously, within a predefined time horizon. The integration of strategic and tactical decisions is achieved through a multiperiod formulation, where two interconnected time scales are modelled. In this paper, this generic model is adapted and applied to the real case of a Portuguese company that produced and distributes lead batteries.

2. Case-study

As referred above the case-study in analysis is of a Portuguese company, A. A. Silva, that produces and distributes lead batteries. This company wants to redesign and plan its current closed-loop supply chain, in order to minimize the total supply chain cost. The company has one factory operating in Oeiras that is responsible for supplying the entire distribution network. This network has in total 12 warehouses located at: Porto, Beja, Coimbra, Santarém, Tondela, Lisboa, Almada, Setúbal, Sines, Loulé, Viseu and Mirandela. All warehouses have maximum capacities of storage and a monthly fixed cost. These facilities work simultaneously as warehouses for the distribution to the customers and as a direct sales point to the public.

Besides fulfilling the customers demand (forward flows), the company also collects the end-of-life batteries from the customers and takes them up to the factory (reverse flows), through the warehouses. Even though the demand of their 2300 clients has to be totally satisfied, the same is not imposed for collected. Nowadays, the company collects

15% of the total batteries sold. These reverse flows represent raw materials to produce new batteries. In terms of return freights, they are completely dependent on the existence of direct delivery freights. This is to mean, EOL batteries are collected only if new batteries were delivered.

The transports used for the forward and reverse flows between the factory and the warehouses are subcontracted. Concerning the forward and reverse flows between the warehouses and the customers, the company uses its owned fleet. Each type of vehicle has its own capacity limit.

3. Model and Data Characterization

A graph representation is used in this work to characterize the CLSC structure that goes from the factory to the customers and back to factories. This generic representation was presented by Salema et al. (2010). Nodes represent any supply chain entity, such as factories, warehouses, customers, distribution centres, while arcs between two nodes define an existing flow.

The mathematical model integrates the forward chain, which links the factory to the customers through the warehouses, and the reverse chain, which deals with the return of the products from the customers up to the factory, again through the warehouses. Thus, the supply chain comprises three types of entities: factory, warehouses and customers. Each one of these entities is defined by its geographical location, Due to computational difficulties customers have to be clustered. Thus, the 2300 customers (in year 2008) are grouped into 237 customer-clusters with respect to the municipality they belong to (from now on, customer-cluster will be referred as customers). The same locations are assumed as possible candidates to locate warehouses.

The model considers a distribution network formed by a single factory. As mentioned above, this factory is already operating and its location is not to be redefined. At the planning level, the macro-time unit is assumed to be equal to one year, and is divided into smaller units representative of one month each (micro-time unit).

Customers are characterized by a known demand, for each month (micro-time unit), that needs to be totally satisfied within that month. The amount of products that can be collected at each customer, per year (macro unit), is modelled as fraction of the total product supplied. At the factory, returned products are used, together with new components, to produce new batteries. These returned products are disassembled and the resulting components represent the metals that are needed to produce of a new battery. Four collected batteries are needed to fulfil the necessary amount for the production of a new one.

Inventories are modelled in all facilities and are limited to maximum level. The initial stock level at warehouses is assumed as zero. At the factory, the stock level is also limited to minimum value, equal to 10% of storage capacity.

Two types of vehicles are considered according to the connection in question: one for the flows between the factory and the warehouses and another one for the flows between warehouses and costumers. Each type of vehicle is characterized by a cost and a transport capacity, by trip. No more collect trips exist than the delivery ones. When owned vehicles are used, the return trip always exists, since the vehicles must return to the warehouses where they belong.

The different costs considered in the model are known or estimated. These are: warehouses fixed costs, supplying costs, transportation costs and the amount paid to customers for collecting end-of-life products. The transportation costs are calculated by different approaches, if primary or secondary distribution is considered. The model

objective is to design the distribution network structure that minimizes the total supply chain cost.

4. Results

The generic model formulation presented by Salema et al. (2010) is applied to the case-study previously presented. Both the network structure and planning are analyzed.

4.1. Network Structure

When comparing the results for the optimal network with the existent structure it can be seen that the former one is composed by one more warehouse than the latter one (14 and 13 warehouses, respectively). Six out of the thirteen current warehouses are kept unchanged (Viseu, Coimbra, Lisboa, Oeiras, Setúbal and Sines), two are closed (Almada and Tondela), three new locations are added (Braga, Guarda and Bombarral) and five of the current warehouses are replaced by new locations which are relatively close to the old one: Porto is replaced by Vila Nova de Gaia, Mirandela by Vila Real, Santarém by Torres Novas, Beja by Évora, and Loulé by São Brás de Alportel.

The obtained optimal network presents a cost of €1638,8*10^3, which translates into a reduction of €47700 on total costs when compared to the cost of the current network.

4.2. Network Planning

In terms of the planning a large number of results are obtained. Due to the lack of space only an illustrative case (Barreiro customer) will be shown. This customer is supplied by two warehouses (Setúbal and Oeiras), and sends its returns to both warehouses. When analyzing with more detail the flow between the factory and Oeiras and Setúbal warehouses (figure 1) it can be seen that the flow to Setúbal occurs in every micro time unit, which is not the case for the Oeiras warehouse. The supplying and return plans of Barreiro customer are also presented (figures 2 and 3), followed by the reverse flows from Setúbal and Oeiras warehouses to the factory (figure 4). Here it can be seen that the Barreiro customers is essentially supplied by the Setúbal warehouse (figure 2), while the amount collected at Barreiro goes mainly to Oeiras warehouse (figure 3). It should be noted that the collection only occurs when the customer is supplied by the warehouses, since the return freights can only occur when associated with freights of demand satisfaction.
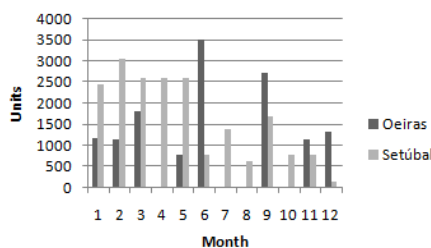

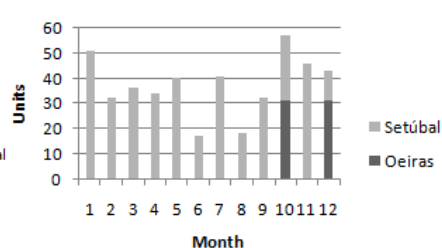
Fig.1: Flows from factory to Setúbal and Oeiras.   Fig. 2: Barreiro customer supplying plan.

The outbound flows from the Oeiras and Setúbal warehouses to the factory are shown in figure 4. The forward and reverse flows between the factory and the Setúbal and Oeiras warehouses (figures 1 and 4, respectively) are much larger when compared to the flows that occur between these warehouses and Barreiro (figure 2 and 3, respectively). This is explained by the fact that the warehouses of Oeiras and Setúbal are responsible for the supplying/collection of other 11 and 8 municipalities, respectively, besides Barreiro.

The retrofit of a closed-loop distribution network: the case of lead batteries
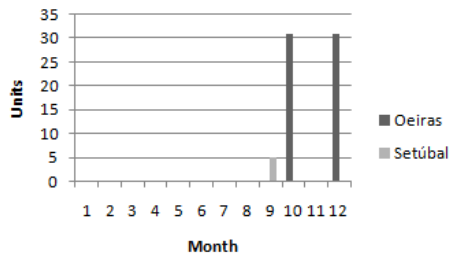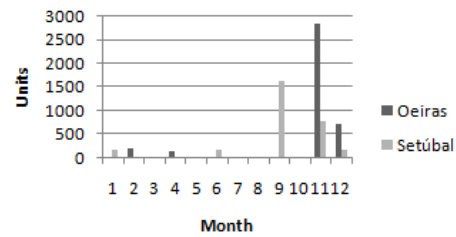

Fig. 3: Barreiro customer collection plan.

Fig. 4: Oeiras and Setúbal warehouses collection plans.

Figure 5 shows the number of freights for the referred flows that go through the Setúbal warehouse. These include the freights from the factory to Setúbal warehouse ($z1$) and from this warehouse to Barreiro ($z2$), from Barreiro back to Setúbal ($z3$) and from this warehouse to the factory ($z4$). The return freights ($z3$ and $z4$) do not occur every month, although the supplying freights ($z1$ and $z2$) do.
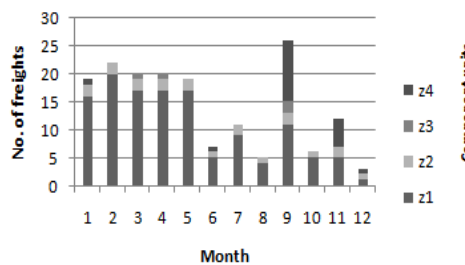

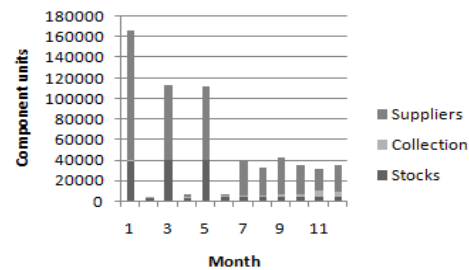Fig. 5: Number of freights between the factory and Barreiro customer, through Setúbal warehouse.

Fig. 6: Supplying and stock plans for the factory.

Finally, supplying and storage plans of the factory are shown in figure 6. These results differentiate the amounts that come from the collection of the end-of-life products at the customers from the ones acquired to the suppliers. It also shows the levels of stock created at the factory, in each month which varies according to the minimum and the maximum limits allowed. The recycled raw materials assume a greater importance in the second half of the year (see 'collection'), which is when return freights, from the warehouses to the factory, bring sufficient amounts that make the subcontracting of transport cost-effective.

4.3. Other Studies

Additional analyses to the ones presented above were also conducted through the use of the developed mathematical model. An important one was the study on the maximum amount that could be paid for the EOL batteries using the current network. Taken as basis the current value paid for each collected battery (0,70 €) and estimating a possible increase on the number of batteries collected (if an higher price would be paid) it was concluded that the optimal amount that could be supported by the actual network is of 1,05 € so as to guarantee a minimum value for the total network cost.

4.4. Computational results

The resulting MILP model was solved by GAMS/CPLEX (built 22.8), in a Intel (R) Core ™ 2 Duo CPU, 3.00 GB, 2.26 GHz. The model is characterized by 954579 variables (631394 binary) and 1432609 constraints, and took about 20835 CPU seconds to reach a 1% gap solution.

5. Conclusions

In this work, it was developed a mathematical formulation to model the real problem for the design and planning of a Portuguese supply chain with forward and reverse flows (*CLSC*). Within a single formulation, strategic decisions concerning warehouses number and location as well as tactical decisions concerning the planning level (supplying, production, storage and distribution planning) are taken simultaneously.

Through the developed model, decisions on the distribution network retrofit were obtained which show the possibility of reducing significantly the current network costs.

The carried out study illustrates the great advantage of using optimization models as they allow different test conditions to study and comparative results analysis. It also shows how a mathematical model can be a useful tool to solve complex logistic problems which corroborates the generic model applicability to study different supply chains and its adequacy to real world problems.

As future work improvements of the present model are mentioned since computational difficulties led to the need of modelling, not only time but also space, in aggregated ways. When doing so, the model will allow more detailed planning throughout time and greater accuracy concerning the locations chosen for the warehouses.

References

Barbosa-Povoa, A.P.F.D., M. I. Salema, M.I.G, and A. Q. Novais, (2007) Design and Planning of Closed-Loop Supply Chains. In Supply Chain Optimization in Supply Chain Optimization, Editores L. Papageorgiou e M.C. Georgiadis, Wiley-VCH, 7, 187-218.

Beamon, B., (1998), "Supply Chain Design and Analysis: Models and Methods", International Journal of Production Economics, 55 (3), 281-294.

Fleischmann, M., P.Beullens, J.Bloemhof-Ruwaard e L.Van Wassenhove,(2001),"The Impact of product recovery on logistics network design", Production and Operations Management, 10(2), 156-173.

Goetschalckx, M., C. J. Vidal, and K. Dogan, (2002). Modeling and design of global logistics systems: A review of integrated strategic and tactical models and design algorithms. European Journal of Operational Research 143:1-18.

Klose, A. e A. Drexl, (2005), "Facility Location Models for Distribution System Design", European Journal of Operational Research, 162 (1), 4-29.

Lu, Z. e N. Bostel, (2007), "A facility location model for logistics systems including reverse flows: the case of remanufacturing activities", Computers and Operations Research, 34 (2), 299-323.

Min, H. e G. Zhou, (2002), "Supply chain modeling: past, present and future", Computers and Industrial Engineering, 43, 231-249.

Sabri, E. e B. M. Beamon, (2000), "A Multi-Objective Approach to Simultaneous Strategic and Operational Planning in Supply Chain Design", Omega, 28, 581-598.

Salema, M.I.G., A. P. F. D. Barbosa-Povoa, e A. Q. Novais (2007), An Optimization Model for the Design for a Capacitated Multi-product Reverse Logistics Network with Uncertainty, European Journal of Operations Research (EJOR), 179,1063-1077

Salema, M. I. G., Barbosa-Povoa, A. P., and Novais, A. Q. (2010). "Simultaneous design and planning of supply chains with reverse flows: A generic modelling framework." European Journal of Operational Research, 203(2), 336-349.

Shah, N., (2005), "Process Industry Supply Chains: Advances and Challenges", Computers and Chemical Engineering, 29 (6), 1225-1235.

Decentralized vs. centralized management of abnormal situations in a multi-plant enterprise using an agent-based approach

Behzad Behdani[a,1], Zofia Lukszo[a], Arief Adhitya[b], Rajagopalan Srinivasan[b,c]

[a] *Delft University of Technology, Technology, Policy and Management, the Netherlands*
[b] *Institute of Chemical and Engineering Sciences, Singapore*
[c] *National University of Singapore, Dept of Chemical and Biomolecular Eng, Singapore*
[1]*Corresponding author: B.Behdani@tudelft.nl*

Abstract

A multi-plant enterprise is a complex network of different geographically distributed production plants producing, handling, and distributing specific products. Comparing with a single plant, a multi-plant enterprise is much more complex considering the high level of inter- and intra-organizational interactions and consequently it is subject to many disruptions. While facing an abnormal situation, many different possible solutions can be defined to mitigate the negative impacts of the disruption and recover the system to its normal operation. Basically, these solutions can be at the plant level or at the enterprise level. To study which solutions are more effective in coping with a disruption, appropriate quantitative models are necessary to evaluate the effect of local behaviors and policies on the overall system performance. In this paper, we will demonstrate how an agent-based model can be used to evaluate the dynamic behavior of a multi-plant enterprise particularly during abnormal situations. Using this model, different mechanisms for coping with abnormality – centralized and decentralized – are also modeled and compared.

Keywords: Agent based modeling, multi-plant enterprise, abnormal situation management, centralized/decentralized decision making

1. Introduction

Modern day enterprises operate in a global scale with production plants spread out around the world. Further, each of these production plants operates their own global supply chain and sources and supplies globally. Each production plant itself has several internal functional departments that are responsible for various internal activities (such as planning, scheduling, inventory management). As a result, an enterprise is a multi-level and multi-actor system that needs coordination at different levels. Each plant should coordinate all its internal activities at the plant-level; at the supply-chain level, the interactions between a plant and its suppliers and customers are managed; the relations between the different plants are optimized at a higher level, the enterprise-level. However, because of high level of dependencies and interactions in different levels of an enterprise, this system is highly complex and also vulnerable to different failures. Consequently, disruptions in one part of the system are rarely local; they may cascade through the system and affect the performance of other parts as well as overall performance of the whole enterprise. In addition, the emphasis on increasing efficiency during recent years, in some cases, has resulted in more vulnerable industrial networks, mainly because of very little excess capacity in the system, and, therefore, any

disruption can have a rapid impact on the system performance. Hence, disruption in enterprises is often not avoidable and an effective abnormal situation management system to cope with them is called for.

Managing abnormalities in a multi-plant enterprise can be very challenging, because a multi-plant enterprise has a distributed structure in which the local production plants have their own local goals that are sometimes in conflict with others and even with enterprise goals. This conflict may become tougher when faced with an abnormal situation that may result in the lack of some resources at the plant- or enterprise-level. Meanwhile, the complex system structure makes forecasting cascading disruption as well as evaluating possible solutions extremely difficult. To handle these challenges, we need appropriate modeling and simulation tools for managing disrupting events that take into account the distributed nature of the organizations and respect the actors' autonomy. Most of the previous modeling works in disruption management are mainly in the OR domain, as presented [1]. The general idea is that the plan (e.g., production plan or schedule) produced by an OR-based model has to be adjusted to take the disruption effects (e.g., machine breakdown) into account. With a similar idea, Xia *et al.* analyze the real-time disruption management in a production-inventory system using a mathematical programming model in which different penalty cost terms for deviations of the new production and inventory plan from the original one is considered [2]. The disruption recovery solutions with short time windows spanning one or two production cycles are also discussed. Tomlin presents a mathematical programming problem for a dual-sourcing model in which orders may be placed with either a cheap but unreliable supplier or an expensive but reliable supplier [3]. He evaluates the firm's optimal strategy under various realizations of the problem parameters.

Besides OR-based models, other modeling approaches are used for modeling the disruption management in supply chains. Wilson studies the effect of a transportation disruption on supply chain performance using system dynamics simulation by comparing a traditional supply chain and a vendor managed inventory system (VMI) [4]. Schmitt and Singh present a hybrid model in which risk profiles for the locations and connections in the supply chain are developed using Monte Carlo simulation and the flow of material and network interactions are modeled using discrete-event simulation [5]. This model is used to get a better view of the impact of disruptions on the system and to study various strategies for coping with the risk in the system.

Most of these analytical models do not explicitly take into account the social aspects of the system such as the negotiation process between actors during the disruption. An alternative approach that overcomes this is offered by agent-based models which adopt an actor-centric perspective instead of the activity-centric one [6]. A good example of application of agent-based modeling for disruption management in supply chains is presented by Mishra *et al*. [7, 8]. Based on the agent-based framework for modeling a supply chain developed by Julka *et al.*[9, 10], they present an agent-based decision support system to manage disruptions in a refinery supply chain. Faced with a disruption, agents collaborate to identify a holistic rectification strategy using heuristic rules.

In this paper we present an agent-based model for a multi-plant enterprise as a decision support tool to evaluate and compare different alternative abnormal situation policies. Section 2 describes the problem of abnormal-situation management in multi-plant enterprises. In Section 3, an agent-based model for a chemical enterprise is discussed, followed by experiments on different disruption management policies. Finally, Section 5 gives some concluding remarks and directions for future works.

2. Problem statement: abnormal-situation management in a multi-plant enterprise

Many internal and external factors may result in abnormal situations in a multi-plant enterprise. Considering the multi-plant enterprise as a system, the internal factors can be changes in internal system parameters (e.g., changes in processing time because of a problem in production plant, delay in raw material shipping from suppliers, etc.) or changes in the availability of resources (e.g., plant/machine failures, resign of key personnel, etc.) [1]. Some of the most important external factors are the sudden changes in external parameters (e.g., sudden increase in the market price for raw materials), some new restrictions (e.g., new government laws and standards) or unpredictable external events (e.g., union strikes, power outages, bad weather and its effect on transportation, etc.). However, regardless of the cause of disruption, an abnormal situation is characterized by its effects on the system performance and the predefined plans [1, 11]. Accordingly, if there is a disruption that does not push the system out of acceptable range or where the deviation from original plan is not so large that the plan has to be changed substantially, it is not an abnormal situation [12]. Consequently, "abnormal-situation management" involves defining and executing some policies to mitigate the negative impacts of the disruption and return the system to an acceptable range of operation. This can be provided through redundancy, improving the flexibility of system or improving collaboration among system components and coordinating different policies of different actors [11, 13]. For instance, if the single supplier of a production plant faces a temporary problem in its production facilities and cannot supply the plant with the raw material as was planned before, the plant would have an abnormal situation in its supply chain. To manage this abnormality, it may go to a secondary supplier to provide its raw material shortage in the short time (probably, even with a higher price). However, if resourcing is not an immediately available option, it may start negotiation with its customers to change the contract details (e.g., postpone the delivery time for orders) or may negotiate with other production plants to share resources or exchange orders. Taking into account distributed nature of multi-plant enterprise, most of the solutions to abnormal situations can be classified in two main types:

− *Plant-level (decentralized) disruption management:* faced with a disruption in its internal facilities or its supply chain, the productions plant itself solves the problem (e.g., through negotiation with other actors).
− *Enterprise-level (centralized) disruption management:* some types of coordination at the enterprise-level can be used to manage the abnormal situations in the plant-level.

Both these approaches can be studied using agent-based models as discussed in the next section.

3. Agent-based model for a global lube additive manufacturing enterprise

In this section, a global chemical manufacturing enterprise is introduced and its agent-based model is presented. The performance of the enterprise under an abnormal situation and also policies to cope with this disruption is studied through experiments.

The lube additive manufacturing enterprise studied here comprises three main actors: customers, suppliers, and the enterprise itself [14, 15]. The enterprise has multiple plants at different locations. Each of these plants has its own functional departments, each with a specific role (Figure 1). All production plants can produce various types of products from different raw materials. The goal is to fulfill a set of customer orders by assigning them to different plants and coordinating the behavior of the different departments in each plant.

The actors in this supply chain can be viewed in three levels:
- *Global level:* The actors in this level are "customers", "enterprise", and "suppliers".
- *Enterprise level:* The manufacturing enterprise consists of the global sales department and a number of plants.
- *Plant level:* Each plant has six different functional departments – scheduling department, operations department, storage department, packaging department, procurement department and logistics department.

The behavior of each of these actors is described in detail in Figure 1.

In the plant level, each production plant has some local customers that send their orders directly to that plant. There are also some customers in the enterprise level sending their orders to global sales department (GSD) and GSD assigns their orders to the plants based on the "First possible time" of fulfilling. Some of the customers in the enterprise-level are "Important customers" whose orders have special priority, especially during disruptions. In the situation that one production plant receives two orders from its local customer and GSD and they have similar conditions in scheduling, the order that is received from an "Important customer" has priority in the fulfillment; otherwise, their position in the current schedule of plant will be determined randomly.
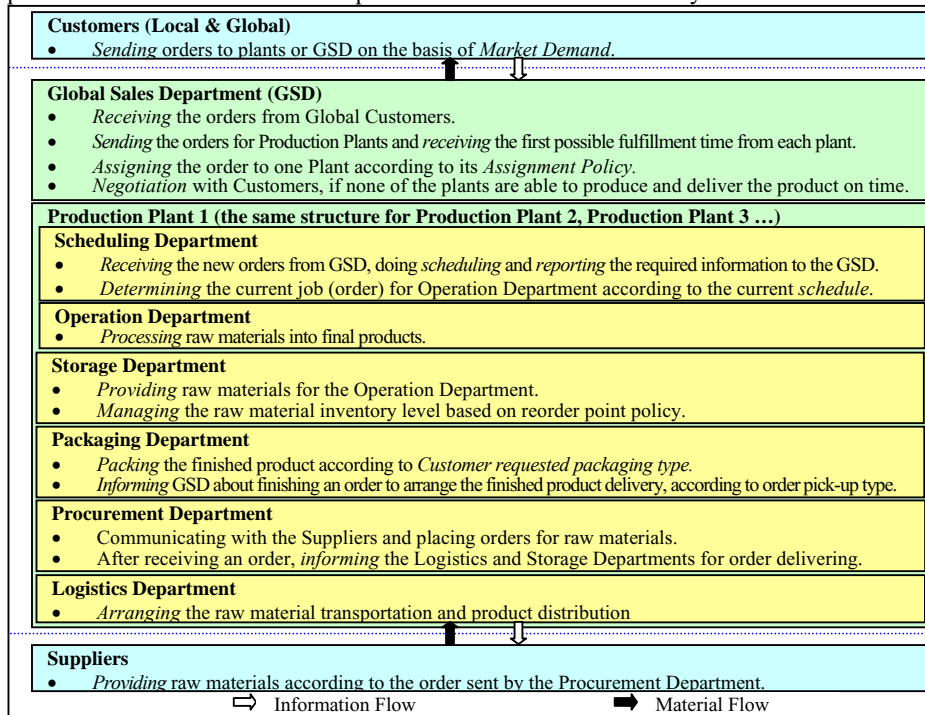
Figure 1. Main actors and their behaviour/interactions in a multi-plant enterprise

For analyzing the lube additive manufacturing enterprise behavior, an agent-based model is developed in Repast simulation platform [16] and Java programming environment. Each actor is modeled as an autonomous agent with specific roles and goals. The system behavior emerges from the agents' behavior and their interactions. Using this model, some experiments are defined by changing the behavior of agents to

study the performance of the enterprise during normal and abnormal situations and evaluate different policies for managing abnormalities.

3.1. Scenario: Raw material supplier disruption

In this experiment, the supplier for production plant 1 faces with a problem in production from day 300 to 330 and accordingly, it can provide the raw material with 3 to 7 days delay to the production plant. The first column of Table1 shows the system performance for this case. With the assumptions made, the enterprise can process 434 orders and there are 35 late orders with 143 tardy days in total during the time horizon (1 year). To handle this disruption, two policies are defined and their effects are studied with the model.

Case 1: Decentralized (Plant-level) disruption management

In this policy, the affected plant (plant 1) itself tries to manage the disruption in its supply chain by negotiation with other production plants for order fulfillment. The disruption management process is done in four main steps:

1. Detecting the disruption and its consequences for plant 1.
2. If due date violation is expected in plant 1, the information about affected orders (the orders that need this raw material) is sent to other plants.
3. Other production plants reschedule their orders: if fulfilling extra orders does not result in any delay in their previously-committed orders, they accept them; otherwise they determine their late orders and send this information to plant 1.
4. If none of production plants 2 or 3 accepts fulfilling the orders, plant 1 considers the orders received from plant 2 and 3 and reschedules its orders:
   a. Plant 1 checks if order exchange with plant 2 results in lower delay comparing with lateness because of raw material delivery delay. If not, the negotiation with plant 2 is terminated. If yes, it proceeds to step 4-2.
   b. Plant 1 checks if order exchange with plant 3 results in lower delay comparing with lateness because of raw material delivery delay. If not, the negotiation with plant 3 is terminated. If yes, it proceeds to step 4-3.
   c. If the total tardiness because of fulfilling the orders of plant 2 and 3 is less than the lateness because of raw material delivery delay, the plant 1 will exchange its orders with the plant in which fulfilling its orders causes less delay.

Case 2: Centralized (Enterprise-level) disruption management:

To handle the effects of disruption, Global Sales Department (GSD) collects the orders from all plants and re-assigns them to the plants considering the new raw material constraint for plant 1. The process has three main steps:

1. Same as above
2. If the raw material disruption causes delay in the committed orders by plant 1, GSD collects the information of unfulfilled orders from all production plants.
3. GSD arranges all collected orders and sends them, one-by-one, to all plants taking into account that orders requiring not-available raw materials can only be sent to plant 2 or 3. After receiving the order details, each plant replies with the earliest date when it can make the product and deliver it to the customer. Based on the replies, GSD re-assigns the order to the production plant with the first possible date.

The results of implementing these two policies (cases 2 and 3) are shown in Table 1. The decentralized policy improves the enterprise performance by reducing the number of late orders to 33 and total tardiness to 132. The centralized policy results in much more significant improvement with only 27 late orders and 101 total tardiness.

Similar to supplier disruption, the model can be used to study the effects of other possible abnormal situations and analyze different disruption management policies.

Table 1. Effect of disruption management on the enterprise performance

	Supplier Disruption	Decentralized Policy	Centralized Policy
Number of orders assigned to Plant 1	161	161	165
Number of orders assigned to Plant 2	130	133	131
Number of orders assigned to Plant 3	143	140	140
Number of orders assigned to All Plants	434	434	436
Number of late orders by Plant 1	11	12	5
Number of late orders by Plant 2	13	13	11
Number of late orders by Plant 3	11	8	11
Number of late orders by all Plants	35	33	27
Total tardiness for Plant 1 (days)	67	69	12
Total tardiness for Plant 2 (days)	48	36	32
Total tardiness for Plant 3 (days)	28	27	57
Total tardiness for all Plants (days)	143	132	101

4. Conclusion and future work

Application of agent-based modeling as a decision support tool for managing abnormal situations in a multi-plant chemical enterprise is discussed. This model allows evaluating different policies for different actors and their effect on overall system performance. Our next step concerns the coordination of plant activities through negotiation in a case where each plant maximizes its own explicit utility even during disruptions.

Acknowledgements

References

[1] G. Yu and Q. Xi, Disruption management: Framework, models and applications. World Sci. Pub. Co., 2004, Singapore.
[2] Y. Xia, J. Yang, B. Golany, S. Gilbert and G. Yu, IIE Trans., 36 (2004) 1.
[3] B.T. Tomlin, Manage. Sci. 52(2006) 639.
[4] M. C. Wilson, Trans. Res. 43(2007) 295.
[5] A. J. Schmitt and M. Singh, Win. Sim. Conf., 2009.
[6] K. H. van Dam, Grasping socio-technical systems with agent-based modeling, thesis (PhD), Delft University of Technology, 2009.
[7] M. Mishra, R. Srinivasan and I. Karimi, AIChE annual meeting, 2003.
[8] M. Mishra, R. Srinivasan and I. Karimi, AIChE annual meeting, 2004.
[9] N. Julka, R. Srinivasan and I. Karimi, Comp. & Chem. Eng., 26 (2002a) 1755.
[10] N. Julka, R. Srinivasan and I. Karimi, Comp. & Chem. Eng., 26 (2002b) 1771.
[11] Y. Sheffi, The resilient enterprise, MIT Press, 2005, Cambridge.
[12] J. Clausen, J. Hansen, J. Larsen and A. Larsen, OR/MS Today 28 (2001) 40.
[13] M. Christopher and H. Peck, Int. J. Log. Manage., 15 (2004) 1.
[14] C. Wong, A. Adhitya and R. Srinivasan, AIChE Annual Meeting, 2008.
[15] H. Zhang, A. Adhitya and R. Srinivasan, Int. Conf. on Infra., 2008.
[16] M. J. North, N. T. Collier, J. R .Vos, ACM Trans. Model. & Comp. Sim., 16(2006) 1.

A rigorous comparative study of temporal versus spatial Lagrangean decomposition in production planning problems

Sebastian Terrazas-Moreno,[a] Ignacio E. Grossmann,[a] Philipp Trotter[b]

[a]*Carnegie Mellon University, 5000 Forbes Ave, Pittsburgh PA 15232, USA,
grossmann@cmu.edu*
[b]*RWTH Aachen University, Templergraben 55, 52056 Aachen, Germany,
philipp.trotter@rwth-aachen.de*

Abstract

Temporal and spatial Lagrangean decompositions are alternatives for solving large-scale planning problems. In this paper we compare the strength of the bounds provided by both decompositions on the full space optimal solution. We also use the economic interpretation of the Lagrange multipliers to speed the convergence to the optimal dual solution.

Keywords: Production planning, Lagrangean relaxation, temporal decomposition, spatial decomposition, mixed-integer programming.

1. Introduction

The optimal planning of a network of manufacturing sites and markets is a complex problem. It involves assigning which products to manufacture in each site, how much to ship to each market and how much to keep in inventory to satisfy future demand. Each site has different production capacities and operating costs, while demand for products varies significantly across markets. Production and distribution planning is concerned with mid to long-term decisions usually involving several months, adding a temporal dimension to the spatial distribution given by the multi-site network. In addition, the production of each product can involve a setup or cleaning time that in some cases is dependent on the sequence of production. When setups and sequence-dependent transitions are included, the optimal planning problem becomes a mixed integer linear programming (MILP) problem. The computational expense of solving large-scale MILPs of this type can be decreased by using decomposition techniques. This paper presents temporal and spatial Lagrangean decompositions that allow the independent solution of time periods, production sites, and markets. The importance of choosing the best between alternative decomposition strategies such as temporal and spatial is discussed in Gupta and Maranas (1999). Although it has been reported that temporal decomposition provides a tighter bound on the full space solution and has faster dual convergence (Jackson and Grossmann, 2003), there is no rigorous generalization of the observed result. One objective of this paper is to compare the bounds obtained through Lagrangean temporal and spatial decompositions for a class of MILPs derived from the lot-sizing problem with setup times. The second objective is to use the economic interpretation of the Lagrange multipliers to provide a reduced dual search space and accelerate the convergence of the optimal multipliers.

2. Problem Statement

There is a set of products that are manufactured in several production sites and shipped to a set of markets where they are sold. Let I, S, and M be the sets of products, production sites, and markets, respectively. There is a finite time horizon divided into a set T of time periods of length L_t. Figure 1 shows the multi-period, multi-site network structure.
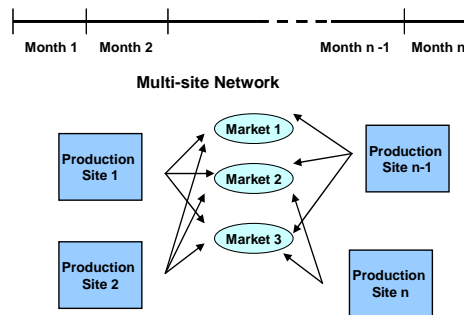

Figure 1. Network of production sites and markets for a multi-period planning problem

At each market $m \in M$ and time period $t \in T$ there is a forecasted demand $d_{i,t}^m$ for product $i \in I$. The production at each site $s \in S$ incurs in a manufacturing cost $\alpha_{i,t}^s$ and the shipment from any site s to any market m involves a transportation cost $\gamma_{i,t}^{s,m}$. Each site has a limited production rate (production per unit time) θ_i^s. Every product has a setup time ST_i^s and setup cost SC_i^s. The products have a sale price of $\beta_{i,t}^m$ at market m. The excess production can be stored as inventory at the production sites, where the inventory holding costs are $\delta_{i,t}^s$. All products that are sent to a market are sold, but the sales cannot be greater than the demand $d_{i,t}^m$. The planning problem is to determine the production in each site, the inventory levels, and the amounts of products shipped to each market during each time period, in order to maximize the profit. Profit is defined as sales minus production, shipment, inventory, and setup costs. We assume that the size of the problem prohibits its direct solution and we consider that temporal and spatial lagrangrean decomposition techniques are alternatives to overcome this challenge. The objective of this work is to rigorously compare the strength of both decompositions. We also illustrate the use of the economic interpretation of the Lagrange multipliers involved in the decomposition schemes to provide a reduced search space for the optimal Lagrangean dual. We illustrate our analysis using a planning problem with setup times but no sequence-dependent changeovers. The results will be extended to a more general model with sequence-dependent changeovers in a full length paper.

3. Illustrative Production planning model

The planning problem can be formulated as the following MILP model.

A rigorous comparative study of temporal versus spatial Lagrangean decomposition in production planning problems

$$\max \pi = \sum_{s \in S} \sum_{t \in T} \sum_{i \in I} \sum_{m \in M} \beta_{i,t}^{m} sl_{i,t}^{s,m}$$

$$- \sum_{t \in T} \sum_{i \in I} \sum_{s \in s} \left(\alpha_{i,t}^{s} x_{i,t}^{s} + \delta_{i,t}^{s} inv_{i,t}^{s} + SC_{i}^{s} stp_{i,t}^{s} \right) - \sum_{t \in T} \sum_{i \in I} \sum_{m \in M} \sum_{s \in S} \gamma_{i,t}^{s,m} f_{i,t}^{s,m} \qquad (1a)$$

$s.t.$

$$\overline{inv}_{i,t-1}^{s} + x_{i,t}^{s} = \sum_{m \in M} \overline{f}_{i,t}^{s,m} + inv_{i,t}^{s} \qquad \qquad \forall i \in I, t \in T, s \in S \qquad (1b)$$

$$x_{i,t}^{s} \le \left(\theta_{i,t}^{s} L_{t} \right) stp_{i,t}^{s} \qquad \qquad \forall i \in I, t \in T, s \in S \qquad (1c)$$

$$\sum_{i \in I} \theta_{i}^{s} x_{i,t}^{s} + \sum_{i \in I} ST_{i}^{s} stp_{i,t}^{s} \le L_{t} \qquad \qquad \forall t \in T, s \in S \qquad (1d)$$

$$f_{i,t}^{s,m} = sl_{i,t}^{s,m} \qquad \qquad \forall i \in I, t \in T, s \in S, m \in M \quad (1e)$$

$$\sum_{s \in S} sl_{i,t}^{s,m} \le d_{i,t}^{m} \qquad \qquad \forall i \in I, t \in T, m \in M \qquad (1f)$$

$$inv_{i,t}^{s} = \overline{inv}_{i,t}^{s} \qquad \qquad \forall i \in I, t \in T, s \in S \qquad (1g)$$

$$f_{i,t}^{s,m} = \overline{f}_{i,t}^{s,m} \qquad \qquad \forall i \in I, t \in T, s \in S, m \in M \quad (1h)$$

$$inv_{i}^{s} \le inv_{i,s}^{UP} \qquad \qquad \forall i \in I, t \in T, s \in S \qquad (1i)$$

$$\overline{inv}_{i}^{s} \le inv_{i,s}^{UP} \qquad \qquad \forall i \in I, t \in T, s \in S \qquad (1j)$$

$$f_{i,t}^{s,m} \le f_{i,s,m}^{UP} \qquad \qquad \forall i \in I, t \in T, m \in M, s \in S \quad (1k)$$

$$\overline{f}_{i,t}^{s,m} \le f_{i,s,m}^{UP} \qquad \qquad \forall i \in I, t \in T, m \in M, s \in S \quad (1\ell)$$

$$f, \overline{f}, sl \in \Re_{+}^{|I| \times |M| \times |S| x |T|}; \quad inv, \overline{inv}, x \in \Re_{+}^{|I| \times |S| \times |T|}; \quad stp \in \{0,1\}^{|I| \times |S| \times |T|} \qquad (1m)$$

The variable $sl_{i,t}^{s,m}$ represents the sales of product i in market m at time t; the superscript s indicates the production site of origin. The variable $inv_{i,t}^{s}$ is the level of inventory, $x_{i,t}^{s}$ is the amount produced of product i in site s during time t, and $f_{i,t}^{s,m}$ corresponds to the shipments between s and m. The setup variable $stp_{i,t}^{s}$ takes the value of one when a product i is manufactured in site s during time t.

Equation (1a) represents the maximization of profit. Equation (1b) is the mass balance of each product at each site and time period. Equation (1c) is a setup constraint where $\theta_{i,t}^{s} L_{t}$ is a valid upper bound for the production of each product. In (1d) the summation of production, setup, and transition times should be less than the length of the time period. Equations (1e) and (1f) are the market constraints. Constraints (1g) and (1h) set the duplicated variables $inv, \overline{inv}$ and $f, \overline{f}$ to be equal. Constraints (1i) - (1ℓ) contain upper bounds for inventory and transportation levels. Finally constraint (1m) indicates the domains of all variables.

4. Temporal and spatial Lagrangean duals

The problem can be made decomposable into time periods dualizing constraint (1g). The objective function takes the following form:

$$\max \pi^t = \sum_{s\in S}\sum_{t\in T}\sum_{i\in I}\sum_{m\in M}\beta_{i,t}^m sl_{i,t}^{s,m} - \sum_{t\in T}\sum_{i\in I}\sum_{s\in s}\left(\alpha_{i,t}^s x_{i,t}^s + \delta_{i,t}^s inv_{i,t}^s + SC_i^s stp_{i,t}^s\right) -$$

$$\sum_{t\in T}\sum_{i\in I}\sum_{m\in M}\sum_{s\in S}\gamma_{i,t}^{s,m} f_{i,t}^{s,m} + \sum_{s\in S}\sum_{i\in I}\sum_{t\in T}\lambda t_{i,t}^s (inv_{i,t}^s - \overline{inv}_{i,t}^s) \tag{2a}$$

On the other hand, the model can be decomposed into individual sites and markets by dualizing constraint (1h), and using the following objective function:

$$\max \pi^s = \sum_{s\in S}\sum_{t\in T}\sum_{i\in I}\sum_{m\in M}\beta_{i,t}^m sl_{i,t}^{s,m} - \sum_{t\in T}\sum_{i\in I}\sum_{s\in s}\left(\alpha_{i,t}^s x_{i,t}^s + \delta_{i,t}^s inv_{i,t}^s + SC_i^s stp_{i,t}^s\right) -$$

$$\sum_{t\in T}\sum_{i\in I}\sum_{m\in M}\sum_{s\in S}\gamma_{i,t}^{s,m} f_{i,t}^{s,m} + \sum_{t\in T}\sum_{i\in I}\sum_{m\in M}\sum_{s\in S}\lambda s_{i,t}^{s,m}(f_{i,t}^{s,m} - \overline{f}_{i,t}^{s,m}) \tag{3a}$$

Problems (2) and (3) can be completed by adding equation (2b)-(2f), (2h)-(2m) and (3b)-(3g), (3i)-(3m) identical to the corresponding constrains in problem (1). Let F, FT, and FS be the feasible regions of problems (1), (2), and (3). The temporal and spatial duals are defined as:

$$D^t = \min_{\lambda t}\ \max_{(sl,x,f,\overline{f},stp,inv,\overline{inv})\in FT} \pi^t \tag{4}$$

$$D^s = \min_{\lambda s}\ \max_{(sl,x,f,\overline{f},stp,inv,\overline{inv})\in FS} \pi^s \tag{5}$$

The well known result of Geoffrion (1974) establishes the following equivalences:

$$T = \left\{(sl,x,f,\overline{f},stp,inv,\overline{inv}):(sl,x,f,\overline{f},stp,inv,\overline{inv})\in Co(FT^{LP})\cap\left\{(inv,\overline{inv}):inv=\overline{inv}\right\}\right\} \tag{6}$$

$$S = \left\{(sl,x,f,\overline{f},stp,inv,\overline{inv}):(sl,x,f,\overline{f},stp,inv,\overline{inv})\in Co(FS^{LP})\cap\left\{(f,\overline{f}):f=\overline{f}\right\}\right\} \tag{7}$$

$$D^t = \left\{\max\pi:(sl,x,f,\overline{f},stp,inv,\overline{inv})\in T\right\} \tag{8}$$

$$D^s = \left\{\max\pi:(sl,x,f,\overline{f},stp,inv,\overline{inv})\in S\right\} \tag{9}$$

where F^{LP} refers to the linear programming relaxation of F. The main result of our analysis is that in general $D^t \le D^s$. The remainder of this section is devoted to presenting an outline of the derivation of this result. The complete proof will be available in a full length paper.

Definition

$$pF = proj_{sl,stp} F = \left\{(sl,stp):(sl,x,f,\overline{f},stp,inv,\overline{inv})\in F\right\}.$$

A rigurous comparative study of temporal versus spatial Lagrangean decomposition in production planning problems

Key Assumption

The larger the feasible region in the space of product sales (sl), the higher the attainable profit.

Theorem:

For the mixed integer planning problem presented in this work, temporal decomposition provides a tighter upper bound than spatial decomposition ($D^t \leq D^s$).

Outline of the proof:

By Fourier-Motzkin elimination, we can obtain the projection of FT^{LP} and FS^{LP} onto the (sl, stp) space. The projections show that $pFT^{LP} \subseteq pFS^{LP}$, which implies $p\left(Co(FT^{LP})\right) \subseteq p\left(Co(FS^{LP})\right)$. By showing that every point in $Co(FT^{LP})$ satisfies $f = \overline{f}$, we establish the following relationship: $p\left(Co(FT^{LP})\right) \subseteq p\left(Co(FS^{LP}) \cap \{(f, \overline{f}) : f = \overline{f}\}\right)$. Furthermore, we can conclude that $p\left(Co(FT^{LP}) \cap \{(inv, \overline{inv}) : inv = \overline{inv}\}\right) \subseteq p\left(Co(FT^{LP})\right) \subseteq p\left(Co(FS^{LP}) \cap \{(f, \overline{f}) : f = \overline{f}\}\right)$. This result and equations (6) and (7) can be used to show that $pT \subseteq pS$.

Given that a larger feasible region in the product sales space leads to a higher profit, $D^t \leq D^s$

□

5. Economic interpretation of Lagrange multipliers.

Solving the temporal and spatial Lagrangean duals (equations (4) and (5)) requires finding the optimal Lagrange multipliers λt and λs. We propose a method that exploits the economic interpretation of these multipliers to reduce their search space. For more details we refer the reader to Trotter (2009), available by request from the authors. The main idea is that the dual multipliers correspond to the transfer prices between time periods (for temporal decomposition) or between markets and sites (for spatial decomposition). The transfer prices depend on the set of active constraints. The sales are limited either by production capacity (equation 1d) or by forecast (equation 1f). Our approach for exploiting this interpretation can be outlined as follows. The first step is to solve the linear relaxation of the MILP problem (1) and obtain the values of the optimal dual variables of constraints (1g) for temporal decomposition or (1h) for spatial decomposition. The relaxed MILP solution can overestimate the production capacity since the setup variables can take fractional values. For this reason, constraint (1d) can become active in the MILP solution when it is inactive in the solution to the linear relaxation. However, if constraint (1d) is active in the relaxed solution it is also active in the MILP solution. In a second step, this property of the active set and the values of the dual variables are used to obtain rigorous bounds on the multipliers λt and λs.

6. Computational results

We ran an instance of the planning problem (1) that consists of three time periods, markets, production sites, and products. This example has two objectives. The first is to illustrate the difference in the bounds on the full space solution obtained by using temporal and spatial Lagrangean decompositions. The second is to show the effect of using the economic interpretation of the Lagrange multipliers on the computational effort required to find the dual solutions. We solve the Lagrangean dual using a generic implementation of the cutting plane algorithm (Kelley, 1960). We choose this algorithm since it has a rigorous convergence criterion. Other subgradient-based algorithms can be

much faster and useful in practice but lack the convergence properties of the cutting plane method. The full space model of this example consists of 27 binary variables, 298 continuous variables, and 253 constraints. The full space and decomposed problems were solved using the MILP and LP solver CPLEX 11.2.0 in GAMS 22.9. In Table 1 we show the optimal solutions of the full space problem and both duals. As shown in the table, the bound provided by temporal decomposition is tighter than the one provided by spatial decomposition. Also, the temporal dual requires fewer iterations to converge to its optimal solution than the spatial dual. Note that we are not concerned with comparing CPU times in this example; we assume that decomposition techniques are required as would be the case in larger problem instances. We leave such examples for a full length paper.

Table 1. Full space, temporal dual, and spatial dual solutions.

	Optimal Profit	Cutting plane iterations**
Full space solution*	41.6	-
Temporal dual solution	41.8	467
Temporal dual solution with economic interpretation of multipliers	41.8	28
Spatial dual solution	42.0	5000§
Spatial dual solution with economic interpretation of multipliers	42.3	1627

* 0% optimality gap in Branch and Bound algorithm
** 1% tolerance for convergence of cutting plane algorithm
§ Maximum number of iterations reached with 1.9% GAP

7. Conclusions

A multi-site, multi-period production planning problem based on lot-sizing models with setup times was presented in this paper. We assumed that the solution of the resulting mixed-integer linear programming (MILP) formulation requires decomposition techniques. Our work presents a rigorous comparison of the bounds on the optimal solution obtained by temporal and spatial Lagrangean decompositions. Temporal decomposition was shown to be the preferred scheme. We also proposed a procedure based on the economic interpretation of the multipliers for reducing their search space in order to accelerate the convergence of the Lagrangean dual. The results confirm that temporal decomposition provides a tighter dual bound and that less computational effort is required when the economic interpretation of the multipliers is exploited.

References

Maranas C. D. and Gupta A.,1999, A Hierachical Lagrangean Relaxation Procedure for Solving Midterm Planning Problems, Ind. Eng. Chem. Res., Vol 38, No 5 pp 1937 – 1947.

Jackson. J. R. and Grossmann I. E., 2003, Temporal Decomposition Scheme for Nonlinear Multisite Production Planning and Distribution Models, Ind. Eng. Chem. Res., Vol 42, No 13 pp 3045 – 3055.

Geoffrion A. M., 1974, Lagrangean Relaxation for Integer Programming, Math. Prog. Studies, Vol 2, pp 82 – 114.

Trotter P. A., 2009, Economic Interpretation of Lagrange Multipliers in Lagrangean Decomposition of a Planning Problem , Technical Report, RWTH Aachen University.

Kelley J. E., 1960, The Cutting-Plane Method for Solving Convex Programs, Journal of the SIAM, Vol 8, pp 703 – 712.

Integrated solvent and process design for the reactive separation of CO_2 from flue gas

N. Mac Dowell, A. Galindo, G. Jackson, C. S. Adjiman[*]

Department of Chemical Engineering, Centre for Process Systems Engineering, Imperial College London, London SW7 2AZ, U.K. c.adjiman@imperial.ac.uk

Abstract

A model-based platform for assessing alternative designs and solvent blends for reactive separation processes is presented and applied to the problem of chemisorption of CO_2 from flue gas with amine solvents. We combine state-of-the-art thermodynamics with rigorous process simulation techniques for this purpose. A rate-based model of chemisorption units for CO_2 capture is implemented in the gPROMS modelling environment, using the statistical associating fluid theory for potentials of variable range (SAFT-VR)[1] to represent mixture thermodynamics. Important features of our model are that both the reaction and phase equilibria are incorporated in the thermodynamic model, and as a result, enhancement factors are not required in the process model. As reaction products are accounted for at the level of the thermodynamic model (from a physical perspective), the generation of reaction products or heat is considered implicitly in the mass and energy balances, further simplifying the model. We validate our model using published pilot plant data, and subsequently apply this approach to the treatment of a typical industrial scale flue gas stream and demonstrate how it can be used to optimise simultaneously solvent composition and process operating conditions.

Keywords: solvent design, CO_2 capture, reactive separation, SAFT-VR

1. Introduction

It is well accepted that fossil fuels will continue to play an important role in the generation of heat and power. However, concerns surrounding anthropogenic emissions of carbon dioxide (CO_2)[2] have resulted in a number of initiatives to reduce CO_2 emissions. Carbon capture and storage (CCS) technologies are accepted as being a promising route to a near-term meaningful reduction in CO_2 emissions. The power generation sector, as the largest stationary point source of CO_2 emissions[2], is actively developing post-combustion CO_2 capture technology[3]. Currently, the most mature technology for large scale CO_2 capture is solvent scrubbing with monoethanolamine (MEA) based solutions. However, this process is very costly, primarily due to the energy required for solvent regeneration. Therefore, there exists a strong imperative to reduce the energy penalty associated with CO_2 capture by developing improved solvents and processes.

In this work, we develop an approach for integrated solvent and process design for the reactive separation of CO_2 from flue gas. We present a rate-based model of a CO_2 capture process which we validate with published pilot plant data.[4,5] The thermodynamic framework used is briefly discussed in the next section and the proposed process model and key assumptions are described. In Section 3, the model is validated against pilot plant data using MEA and 2-amino-2-methyl-1-propanol (AMP)

as solvents. We show in Section 4 how the model can be used to design an improved solvent blend and process for CO_2 capture.

2. Methodology

2.1. Thermophysical property model

The reactive nature of mixtures of amines in aqueous solution with CO_2 is well known; there is a large body of experimental and theoretical work in place detailing the mechanisms and rates of these reactions[6]. In this work, we use the SAFT-VR[1] equation of state to calculate the thermodynamic properties of the system as it has previously been shown to model complex fluids successfully for a wide range of conditions. In a previous paper[7], we have described in detail the development of the SAFT-VR[1] models of H_2O, MEA and CO_2. We incorporate a description of all the interactions between the components at the level of the thermodynamic model, using the SAFT[8] formalism to represent simultaneously the effect of the physical and chemical interactions in the fluid. Chemical equilibrium between the species is modelled through association sites. This approach facilitates a consistent, physically-based description of the numerous competing interactions in both the gas and liquid phases. Moreover, the SAFT approach allows the implicit consideration of reaction products: when CO_2 and MEA react, the products are described in the SAFT-VR model as associated complexes of CO_2 and MEA molecules (e.g., trimers formed from 2 MEA molecules bound to 1 CO_2 to describe the formation of carbamate). The thermophysical properties of the mixture of CO_2, MEA, H_2O and associated complexes are entirely described by the SAFT-VR model parameters for CO_2, MEA and water. Provided that the process operates in the mass-transfer limited regime, so that chemical equilibrium can be assumed, this simplifies the mass and energy balance equations in the process model. It also reduces the dependence on the experimental data: no kinetic data or rate equations are required to describe the effect of the chemical reactions on the mixture phase behaviour or thermophysical properties. This constitutes a major advantage in systems comprising novel, reacting solvents where these experimental data are not available.

If one were to rigorously apply a chemical theory to describe the phase behaviour of reacting mixtures, such as a mixture of ammonia+CO_2+H_2O, for example, there would be, 72 temperature-dependent, interaction parameters required for the description of this system[9]. All these parameters cannot be estimated reliably. Consequently one would need to perform a sensitivity analysis in order to determine the most significant parameters, which are then fitted by comparison to experimental data. The parameters which are deemed to be less important would be fixed to an arbitrary value. This approach is likely to reduce the accuracy and predictive ability of such models. In contrast, in developing the SAFT-VR models of the MEA+H_2O+CO_2 system considered in this work, it was necessary to estimate only three unlike binary interaction parameters.

2.2. Absorber and desorber models

A rigorous rate-based model of the CO_2 absorption process has been developed, based on the approach of Krishna and Taylor[10,11]. The packed column is represented as a cascade of non-adiabatic, non-equilibrium stages. The equations used to describe the packed column comprise mass and energy balances, equilibrium relations, and mass and

energy transfer rate equations. The two-film model[12] is used to describe the heat and mass transfer phenomena. This model is chosen because of its success in modelling systems similar to the one considered here[13-14]. We consider a model fluid comprising H_2O, MEA, CO_2 and N_2. For simplicity, we do not include oxygen in our model fluid, despite the fact that there would be a significant O_2 content in any real flue gas stream. Instead, we replace O_2 with an additional N_2 content. This assumption is justified as the physical solubility of O_2 in H_2O is similar to that of N_2 at processing conditions. The inclusion of O_2 would be necessary to account for the oxidative degradation of the solvent, a phenomenon not considered in our model due to the lack of reliable data. A desorber model has also been developed using the same principles as for the absorber model.

The key assumptions in our models are as follows:

1. The bulk phase of each stage is assumed to be well-mixed, i.e., there are no temperature or composition gradients in either the bulk gas or liquid phase.
2. Each stage is mechanically stable, i.e., there is no pressure gradient from the bulk gas phase to the bulk liquid phase.
3. Both phase and chemical equilibrium prevail at the vapour-liquid interface.
4. All chemical reactions are deemed to proceed to equilibrium. This is a reasonable assumption in systems which are mass transfer limited. As a result, we do not need enhancement factors in the rate equations for mass transfer.

The chemical equilibrium is taken into account when performing thermophysical property calculations with the SAFT-VR model at the interface and in the bulk phases. This is illustrated in Figure 1.

3. Model validation

The process model has been implemented in gPROMS[15] using a SAFT-VR Foreign Object[16] and tested against published pilot plant data. We have selected the data of Tontiwachwuthikul *et al.*[4] and that of Tobiesen *et al.*[5] for use in validating our absorber and desorber models respectively because both these data sets have been accurately predicted in other contributions[19,5]. The proposed model is used in a predictive manner: no parameters are fitted to the pilot plant data. For both MEA and AMP solvents, the proposed absorber model provides a quantitatively accurate prediction of the liquid phase temperature profiles, vapour phase CO_2 concentrations, and liquid phase CO_2 loading, which is defined as the number of moles of CO_2 per mole of amine. On average, the relative error in predictions of liquid phase temperature profiles, T, is = 0.17%, and the absolute errors in the mole fraction for the CO_2 gas phase concentration and liquid phase loading are 0.003 and 0.002, respectively Results for a typical run with MEA are shown in Figure 2.
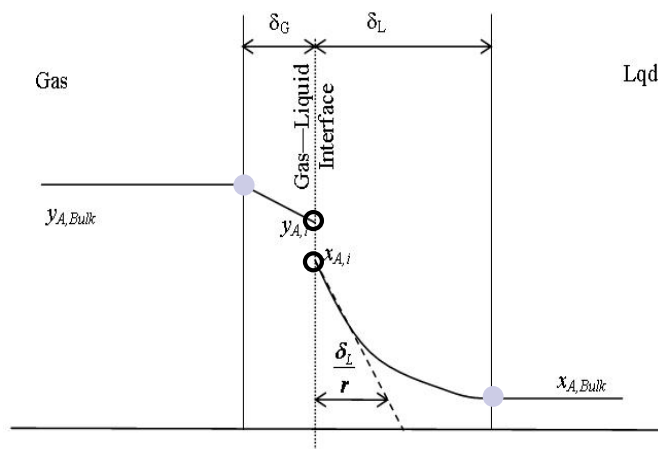
Figure 1. A schematic of the two-film model of a column stage. $y_{A,Bulk}$ and $x_{A,Bulk}$ denote the mole fraction of component A in the bulk vapour and liquid phases respectively, $y_{A,i}$ and $x_{A,i}$ the mole fraction of component A at the "interface" in the vapour and liquid phases respectively. The grey circles and empty circles denote where SAFT calculations are carried out, implying different equilibrium compositions for the reacting species. A phase equilibrium calculation between the phases denoted by the empty circles is performed at the vapour-liquid interface.

The pilot plant data of Tobiesen *et al.*[5] have been used to validate the desorber model. Due to the higher temperatures in the desorber, it is expected that the assumption of chemical equilibrium is more readily met. Indeed, the predictions of our model are in good agreement with the data: an isothermal temperature profile is observed in the packed column section, and the energy input to the reboiler and the associated extent of solvent regeneration are in good agreement with the experimentally observed values.

4. A solvent design case study

Solvent blends, particularly those including sterically hindered compounds, are considered to be an attractive approach to developing solvents for CO_2 capture. Recently, blends of AMP (a "slow" amine) and NH_3 have been shown to be particularly promising[17] when compared to a standard 30wt% MEA solvent. Specifically, this blend was presented as being as "fast" as the MEA solvent, but with a significantly superior capacity for absorbing CO_2. A 30wt% AMP + 5wt% NH_3 blend, denoted 30:5, has been suggested as promising. Our objective is to use the proposed model to identify an improved blend. To this end, we formulate an objective function as follows:
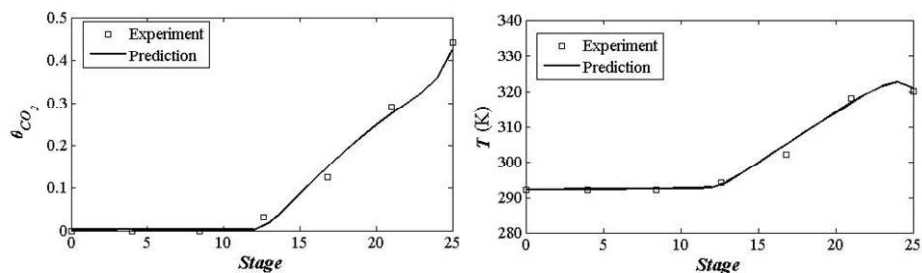
Figure 2. Sample model predictions and experimental data for CO_2 loading (θ_{CO2}) and temperature T as a function of stage number (0 denotes the top of the column and 25 the bottom). This corresponds to Run 22 in Ref [5]

$$f_{obj} = A\, y_{CO2} + B\, y_{RNH2} + L\, Cp$$

where A, B are weights, with $A=6,000$ and $B=8,500$; y_{CO2} is the outlet mole fraction of CO_2 in the flue gas, y_{RNH2} is the total outlet mole fraction of amine in the flue gas (i.e., the sum of the mole fractions of AMP and NH_3); L is the flow rate of the rich solvent stream leaving the absorption column (kmol.s^{-1}); C_p is the heat capacity of that stream (kmol.kJ^{-1}.K^{-1}). The first term is related to the emission of CO_2 to the atmosphere, the second to the emission of amine, and the third to the regeneration cost expressed in terms of the total heat capacity of the rich solvent stream.

A solvent design problem has been solved on the basis of the proposed objective function using the inlet temperature, T, and composition, x_{NH3} and x_{AMP} of the lean solvent stream and the flowrate L as design variables. The constraints include a minimum water content of 50wt% for the lean solvent, to avoid corrosion problems due to the presence of AMP. The results of the optimisation are summarised in Table 1 and compared to the best performance that can be achieved with the 30:5 blend. The optimal solvent blend has an increased AMP content of 38wt% and a decreased NH_3 content of 2wt%, a greater solvent flow rate, and a reduced inlet temperature. Its performance in terms of separation effectiveness is similar to that of MEA: the mole fraction of CO_2 in the flue gas (y_{CO2}) is 0.004 for MEA and 0.009 with the optimal blend. Given these promising results, a more realistic objective function must now be developed, to account in more detail for the compromise between absorption effectiveness and regeneration cost and to incorporate additional costs such as cooling the solvent to the desired inlet temperature.

	30:5 AMP-NH_3 Blend	Optimal AMP-NH_3 Blend
f_{obj}	30,578	10,251
L (kmol/s)	68.70	73.51
T (K)	303.15	295.01
x_{AMP} *(mol %)*	0.08	0.11
x_{NH3} *(mol %)*	0.07	0.03
y_{CO2} *(mol %)*	0.02	0.009
y_{RNH2} *(mol %)*	0.05	0.02

Table 1: Comparison of the performance of the 30:5 AMP-NH_3 blend and the optimised AMP-NH_3 blend.

5. Conclusions

A novel methodology for the design of solvents and solvent blends for reactive separation processes has been developed and applied to the case of CO_2 capture. This rests on a predictive model for the process and for the thermodynamics of the mixtures, including the complex chemical equilibrium. The process model is rate-based and does not require the use of enhancement factors to describe the effect of the chemical reaction on the absorption process. By using the SAFT approach, reaction products are implicitly accounted for at the physical level of the thermodynamic model of the fluid. This simplifies process modelling, since there are no generation terms in either the mass or energy balance equations. Thus, there is no need for kinetic information such as rate equations or rate constants, nor data on the enthalpy of absorption to model most new solvent mixtures. This significantly reduces the requirement for experimentation and physicochemical data for the solution of solvent design problems. This modelling approach has been validated on pilot plant data and has led to promising results in designing an optimised blend of AMP and NH_3, based on a simple objective function.

6. Acknowledgements

References

[1]   Gil-Villegas, A., Galindo, A., Whitehead, P. J., Mills, S. J., Jackson, G. and Burgess, A. N., J. Chem. Phys. 106, (1997), 4168-4186
[2]   Metz, B., Davidson, O. , de Coninck, H. C., Loos, M. and Meyer, L. A.IPCC, 2005: IPCC Special Report of Carbon Dioxide Capture and Storage. Prepared by Working Group III of the Intergovernmental Panel on Climate Change, Cambridge University Press, Cambridge, United Kingdom and New York, NY, USA, 2005
[3]   Rochelle, G. T., Science, 325, (2009), 1652-1654
[4]   Tontiwachwuthikul, P., Meisen, A. and Lim, C.,. Chem. Eng. Sci., 47, (1992), 381-390
[5]   Tobiesen, F. A., Juliussen, O. and Svendsen, H. F., Chem. Eng. Sci., 63, (2008), 2641-2656
[6]   Blauwhoff, P. M. M., Versteeg, G. F. and Van Swaaij, W. P. M., Chem. Eng. Sci., 39, (1984), 207-225
[7]   Mac Dowell, N., Llovell, F., Adjiman, C. S., Jackson, G. and Galindo, A., Ind. Eng. Chem. Res., (2009), In Press, DOI: 10.1021/ie901014t
[8]   Chapman, W.G., Gubbins, K. E., Jackson, G. and Radosz, M., Ind. Eng. Chem. Res., 29, (1990), 1709-1721
[9]   Maurer, G. Fluid Phase Equilib., 30, (1986), 337-352
[10]   Krishnamurthy, R. and Taylor, R., AIChE J., 31, (1985), 449-456
[11]   Krishnamurthy, R. and Taylor, R.,AIChE J., 31, (1985), 456-465
[12]   Lewis, W. K. and Whitman, W. G., Ind. Eng. Chem., 16, (1924), 1215-1220
[13]   Pandya, J. D., Chem. Eng. Commun., 19, (1982), 343-361
[14]   Kucka, L., Muller, I.,Kenig, E. Y. and Gorak, A., Chem. Eng. Sci., 58, (2003), 3571-3578
[15]   gPROMS v 3.1.5 Process Systems Enterprise (PSE) Ltd. http://www.psenterprise.com/index.html
[16]   Kakalis, N. M. P. , Kakhu, A. I. and Pantelides, C. C., Ind. Eng Chem. Res., 45, (2006), 6056-6062,
[17]   Choi, W.-J., Min, B., Y., Shon, B. H., Seo, J. B. and Oh, K. J , J. Ind. Eng. Chem., 15 (2009), 635-640

Optimization based Conceptual Design of Reactive Distillation for Selectivity Engineering

Vinay Amte,[a] Sriharsha Nistala,[b] Sanjay M Mahajani,[c] Ranjan K Malik[d]

[a,b,c,d]*Department of Chemical Engineering, Indian Institute of Technology, Bombay-400076, India*

Abstract

This work presents some representative reactive distillation (RD) models to obtain attainable region for multi-component reaction systems when feed contains significant amount of inerts in it. One can achieve maximum selectivity for a quantitative conversion unlike conventional reactors with semibatch reactive rectification (SRR) as the best RD configuration when the inert is more volatile than the reactant. The other plausible RD configurations were suggested for different volatility patterns of the components and the inert. The MINLP optimization technique is also applied to industrially important case of dimerization of isobutylene for maximizing the selectivity towards di-isobutylene and the results were in agreement to those obtained from conceptual design. This work highlights the potential of optimization as a tool to complement the conceptual design method for selectivity engineering in RD columns.

Keywords: reactive distillation, selectivity engineering, MINLP optimization, reactive rectification, dimerization.
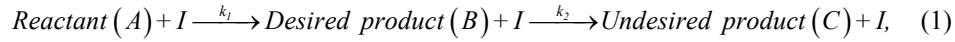
1. Introduction

Reactive Distillation (RD) which combines reaction and distillation is a promising multi-functional reactor which has a potential of giving desired selectivities. The underlying principle in application of RD for selectivity engineering is to facilitate the separation of selected components and favorably manipulate the concentration profiles in the reactive zone of the RD column to expedite the desired reaction. When designed astutely, RD can offer notably higher yields as compared to the conventional reactors.

Attainable region (AR) was first proposed by Horn (1964) to study reactor networks. It is a region in concentration space that can be obtained by a combination of conventional reactors and mixing (Glasser et al., 1987). Nisoli et al. (1997) extended the AR approach to reactor-separator synthesis by defining a reactor-separator vector. Agarwal et al. (2008) presented a conceptual design method for RD, based on the AR approach. They introduced batch reactor-separator models *viz.* reactive condenser (RC) and the reactive reboiler (RR). These models were capable of surpassing the performance of conventional reactors, depending on kinetics and vapor-liquid equilibria (VLE) of a given reacting system. These models were further modified to reactive rectification (Rrect) and reactive stripping (Rstrip) shown in Fig.1, which act as multistage surrogates of the RC and RR models respectively. The Rrect and Rstrip models operate with total reflux and reboil modes respectively. With the help of these models, a selectivity of almost 100% was realized for the entire range of conversion.

2. Reaction System

In this study, we focus our attention on a reaction system consisting an inert component. The series reaction scheme shown below is considered to study the effect of inert (*I*) in

the feed that does not take part in the reaction but has the potential to modify the composition profiles in the column depending on its volatility (α_I) .

$$Reactant\,(A)+I \xrightarrow{\ k_1\ } Desired\ product\,(B)+I \xrightarrow{\ k_2\ } Undesired\ product\,(C)+I, \quad (1)$$

Such systems are analyzed with the help of selectivity parameter *S*, defined as the ratio of the rate of desired reaction to that of the undesired reaction. For the given series reaction scheme with unity rate constants, it is required to maximize *S* by enhancing the concentration of *A* and efficient removal of *B* from the reactive stages to obtain higher selectivity in the RD column. For more volatile reactants, Rrect offers close to 100% selectivity whereas for less volatile reactant Rstrip gives similar performance by maintaining high value of selectivity parameter all over the reactive zone.

In systems with inert, the parameter *S* can still be studied to arrive at a RD configuration eventhough the concentration of the inert does not directly feature in the selectivity parameter. The presence of inert, either more or less volatile than the reactant limits the performance of the RD models. But, it is possible to remove the inert from the top or bottom of the column and maintain a high concentration of reactant in the column by employing semi-batch Rrect or Rstrip model.

3. Model Development

The multi-stage semi-batch model introduced by Agarwal et al. (2008) was a hybrid RD unit for a system wherein reactant acts as a saddle (Fig. 1). They showed that it is possible to remove the desired product from the top of the column and maintain a high concentration of reactant on the reactive stages when the desired product is lighter than the undesired product. Hence, the selectivity is enhanced to its maximum possible value. For a system containing a volatile inert, one possibly removes the inert from the top of the column and maintains a high concentration of reactant in the column by the use of SRR model. The assumptions made while formulating equations for RC and RR models (Agarwal et al., 2008) applies here as well.
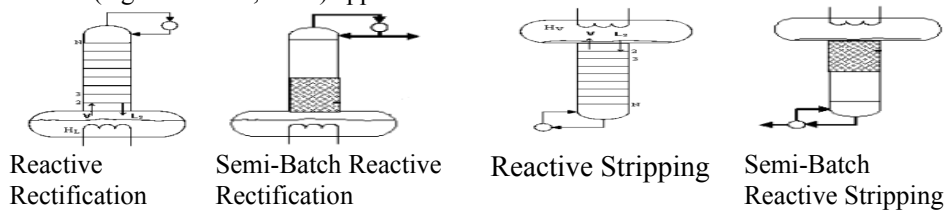

Reactive Rectification   Semi-Batch Reactive Rectification   Reactive Stripping   Semi-Batch Reactive Stripping

Figure. 1. Representative Multi-stage RD Models

The overall and component mass balances around the reboiler (*j* = 1) are given by:

$$\frac{d(H_j)}{d\xi} = H_j\left(-1 + l_{j+1}\right) \tag{2}$$

$$\frac{d(x_{i,j})}{d\xi} = (x_{i,j} - y_{i,j}) + l_{j+1}(x_{i,j+1} - x_{i,j})\ \ for\ i = 1,2...NC\text{-}1 \tag{3}$$

where, $d\xi = V_1 dt/H_j$ is dimensionless warp time and, $l_j = L_j/V$

The overall and component mass balances on the jth stage (*j* = 2 to *N*-1) gives:

$$-l_j + l_{j+1} + \frac{Da}{N-1}\sum_{k=1}^{p} v_{T,k} r_k^{'}(x_j) = 0 \tag{4}$$

$$-l_j x_{i,j} + l_{j+1} x_{i,j+1} + y_{i,j-1} - y_{i,j} + \frac{Da}{N-1}\sum_{k=1}^{p} v_{i,k} r_k^{'}(x_j) = 0\ \ for\ i = 1,2...NC\text{-}1 \tag{5}$$

Optimization based Conceptual Design of Reactive Distillation for Selectivity Engineering

where, Damkohler number is given by $Da = k_{ref} H_{cat}^T/V$, H_{cat}^T is total catalyst loading. Similarly, the overall and component mass balances around condenser ($j = N$) give:

$$-l_N + 1 - \frac{D}{V} + \frac{Da}{N-1}\sum_{k=1}^{p} v_{T,k} r_k'(x_N) = 0 \tag{6}$$

$$-l_N x_{i,N} + y_{i,N-1} - \frac{Dx_{i,D}}{V} + \frac{Da}{N-1}\sum_{k=1}^{p} v_{i,k} r_k'(x_N) = 0 \quad for\ i = 1,2...NC-1 \tag{7}$$

In addition, the VLE equations are also pertinent to all stages except the total condenser. Similarly, equations were developed for the semi-batch reactive stripping model (SRS).

4. Performance Analysis by AR approach

The conceptual design method is applied to obtain ARs of RD models which are then compared to select best RD configuration for a given reaction system. The volatility order is considered to be $\alpha_A > \alpha_B > \alpha_C$ unless mentioned otherwise.

4.1. Inert is more volatile than reactant

One can enhance the AR for RD models by decreasing the value of *Da* either by reducing the catalyst loading or increasing the vapor (in Rrect or SRR) or liquid (in Rstrip or SRS) flow rate. However, there exists a critical *Da* value (*Da_c*) below which no further improvement in performance is possible. So, for the critical *Da* which is evenly distributed over the total number of stages *N*, the ARs for both Rrect and SRR expand with an increase in number of stages for any percentage of inert in the feed as shown in Fig. 2. But there exists a bound on conversion, because of loss of reactant through distillate, in the case of the SRR model which is imposed by the reflux ratio (see Fig. 2b).
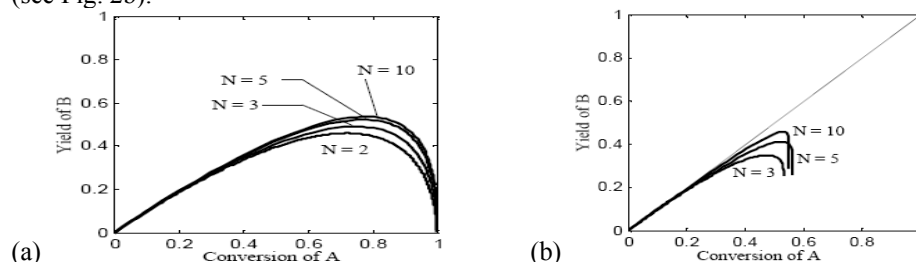

(a)   (b)

Figure. 2. Effect of number of reactive stages *N* on the AR for (a) Rrect, (b) SRR (reflux ratio = 99) with 25% inert in feed

When the reactant is less volatile than inert, its concentration on the reactive stages decreases. As a result, there is a drop in the column performance due to the presence of inert. The increase in the volatility of the inert is responsible for suppressing the concentration of the reactant in the Rrect column. However, efficient separation of the volatile inert facilitates its removal in the overhead stream of SRR model which leads to an improvement in selectivity towards the desired product.

The performance of Rrect decreases with an increase in the percentage of inert in the feed. The volatile inert suppresses the concentration of reactant on the reactive stages and hence, affects the selectivity at any possible conversion. On the other hand, in the case of SRR, the limit on maximum possible conversion increases with an increase in the percentage of inert in the feed. Efficient separation of the inert from the reactant in the column due to its higher volatility facilitates the removal of almost pure inert in the overhead stream and limits the loss of reactant.

The introduction of non-reactive stages further boosts the yield at any possible conversion (Agarwal et al., 2008). This is due to strengthened distillation effect which improves the separation of the desired product from the reactive zone as shown in Fig. 3a. The feasible region expands and the maximum possible conversion increases with
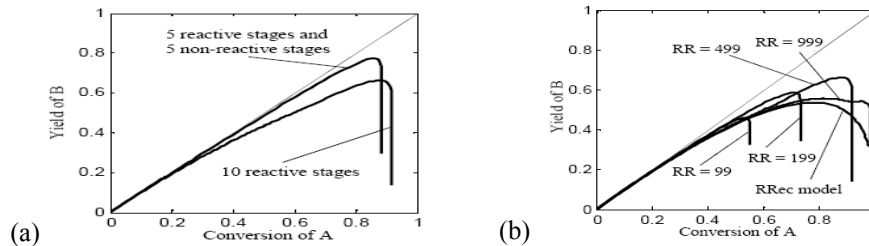

Figure. 3.(a) Effect of non-reactive stages, (b) Effect of reflux ratio on AR of 10 stage SRR with 50% inert in feed.

an increase in the reflux ratio (Fig. 3b). Increase in reflux ratio prevents the loss of reactant from the overhead stream and also maintains a high concentration of reactant on the reactive stages. As expected, the performance of the SRR model approaches that of the Rrect model as the reflux ratio is increased. The RRect model may thus be considered as a special case of SRR model at infinite reflux ratio. The other representative RD models for different volatility cases are shown in Table 1.

Table 1. Guidelines for selecting best RD configurations for various orders of volatility of the components and inert for the series reaction sequence

| Order of volatility	RD configuration	Remarks (*x* is composition)
$\alpha_I \ggg \alpha_A > \alpha_B > \alpha_C$		x_A is not effectively diluted by x_I as it gets removed in the overhead stream.
$\alpha_A > \alpha_B > \alpha_C$ and $\alpha_A > \alpha_I$		x_A is least affected by the presence of x_I.
$\alpha_C > \alpha_B > \alpha_A > \alpha_I$		Removal of least volatile inert from the bottom stream
$\alpha_C > \alpha_B > \alpha_A$ and $\alpha_I > \alpha_A$		Inert may not be able to dilute the reactant in the column due to its high volatility
Reactant is a saddle and desired product is the most volatile component		
$\alpha_I > \alpha_B > \alpha_A > \alpha_C$ and $\alpha_I = \alpha_B > \alpha_A > \alpha_C$		Removal of more volatile inert along with *B* in the overhead stream
$\alpha_B > \alpha_A > \alpha_C$ and $\alpha_B > \alpha_I$		Removal of *B* in the overhead stream. Inert may dilute the reactant in the column
Reactant is a saddle and undesired product is the most volatile component		
$\alpha_C > \alpha_A > \alpha_B$ and $\alpha_I > \alpha_B$		Removal of B in the bottom stream. Inert may dilute the reactant in the column
$\alpha_C > \alpha_A > \alpha_B = \alpha_I$ and $\alpha_C > \alpha_A > \alpha_B > \alpha_I$		Removal of less volatile inert along with *B* in the bottom stream

4.2. Inert is as volatile as reactant

The ARs for both Rrect and SRR models expand with an increase in number of stages for any percentage of inert in the feed. This is due to enhanced separation of desired
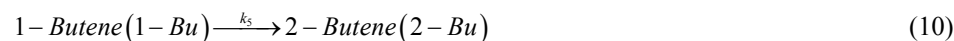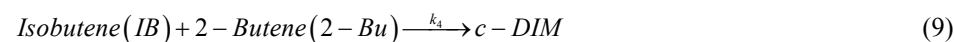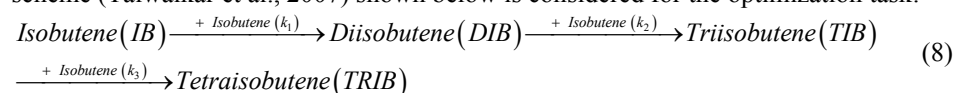
product and thereby suppressing undesired side reaction. With increase in inert in the feed, the concentration of reactant decreases on reactive stages. This leads to a drop in Rrect performance as compared to SRR wherein inert is simultaneously being removed. At low values of reflux ratio, the conversion in SRR is limited due to loss of reactant in the overhead stream; the upper bound on conversion improves with increase in reflux ratio. The hybrid SRR further enhances the yield at any conversion due to distillation effect which improves the separation of the desired product from the reactive zone. The other representative RD models are shown in Table 1 for different volatility pattern.

4.3. Inert is least volatile

Due to the low volatility of the inert which is not significant enough to alter the concentration of the reactant, the model performance is unaffected. In SRR model, an increase in the reflux ratio results in AR expansion by increasing the maximum conversion as in the previous case. The summary of the proposed RD configuration based on inert volatility is presented in Table 1.

5. Optimization of RD column for Dimerization of Isobutene

An industrially important reaction, dimerization of isobutene (IB) to di-isobutene (DIB) is considered as a case study. This reaction assumes significance due to occurrence of undesired side reactions and it becomes important to improve the selectivity towards the desired product DIB. In this study, optimization technique is applied to qualitatively validate the conceptual design results presented in the previous section. The reaction scheme (Talwalkar et al., 2007) shown below is considered for the optimization task:

$$Isobutene(IB) \xrightarrow{\text{+ }Isobutene\,(k_1)} Diisobutene(DIB) \xrightarrow{\text{+ }Isobutene\,(k_2)} Triisobutene(TIB)$$
$$\xrightarrow{\text{+ }Isobutene\,(k_3)} Tetraisobutene(TRIB) \tag{8}$$

$$Isobutene(IB) + 2-Butene(2-Bu) \xrightarrow{k_4} c-DIM \tag{9}$$

$$1-Butene(1-Bu) \xrightarrow{k_5} 2-Butene(2-Bu) \tag{10}$$

An ideal VLE model is applied by the use of relative volatilities which were determined using the T-x-y data generated using steady state simulator AspenPlus (Aspen Technology Inc., 2004). The decision variables were feed stage and reactive stage location, reflux ratio and distillate to bottoms rate ratio. The model was implemented in GAMS 22.8 and solved with the codes CONOPT and SBB (Brooke et al., 2009) on a 3 GHz Pentium 4 PC with 1 GB of main memory.

Three different cases of feed were considered (i) pure feed of IB, (ii) mixture of 80 mol% of IB and 20 mol% of 1-butene and, (iii) mixture of 60 mol% IB, 30 mol% 1-butene and 10 mol% isobutane. The optimal RD configurations that maximize the selectivity towards DIB for cases (i) and (ii) correspond to the continuous mode of the Rrect model with feed to bottom most stage and the product is also removed from the bottom. The catalyst is placed on the top most stage in the column. It is known that such a configuration maintains a low concentration of desired product on the reactive stages and inhibits the undesired reaction. The optimal configuration for case (iii) resembles the continuous SRR model. This is due to the fact that removal of 2-butene (more volatile than the IB) from the overhead stream facilitates the enhancement in selectivity. The optimal configurations were simulated in ASPEN using UNIQUAC model to represent the VLE of the system and it was found that the simulation results are in good agreement with those of optimization. The typical composition and temperature profiles for an optimal RD configuration are shown in Fig. 6.
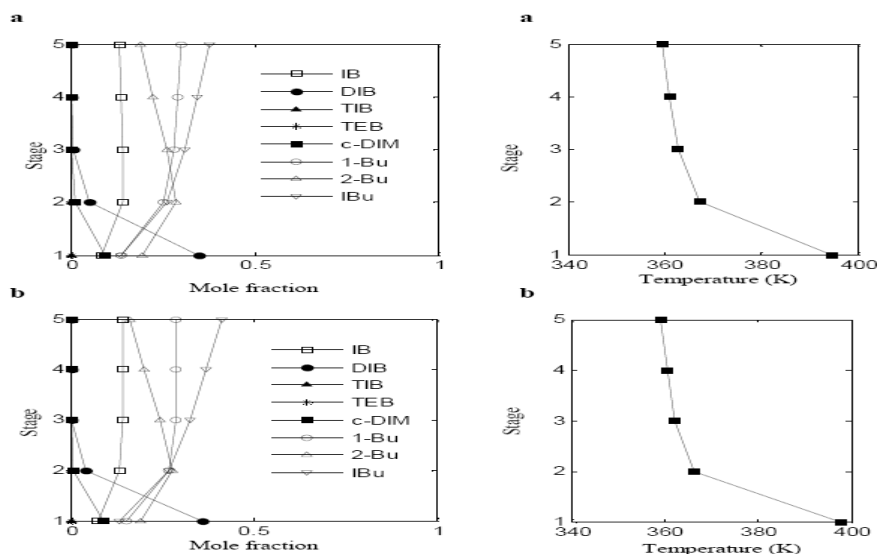
Figure 6. Composition and temperature profiles for an optimal configuration at 90% conversion, selectivity = 43.71% (maximum possible selectivity = 50%), isobutene system, feed: 60% IB, 30% 1-butene and 10% isobutane, (a) from GAMS, (b) from ASPEN

6. Conclusions

In the present work, AR approach is extended further to include few representative RD configurations and is applied to series reaction scheme with inert present in feed. When the reactant is more volatile than the products, it was found that the Rrect model offers maximum selectivity in the presence of an inert whose volatility is equal to or less than that of the reactant. Alternatively for an inert that is more volatile than the reactant, SRR is preferred. The other plausible RD configurations were suggested for different orders of volatility. The MINLP optimization technique was then applied to the dimerization of IB in the presence of an inert. The RD with bottom or top withdrawal was found to be the optimal configuration based on inert volatilites. These results qualitatively validate the conceptual design findings and hence, establish the potential of optimization as a tool to complement the conceptual design method for selectivity engineering in RD.

References

A. Brooke, D. Kendrick, A. Meerhaus, R. Raman, 2008, GAMS Release 22.8-A User's Guide, GAMS Development Corporation, Washington.

A. Nisoli, M. F. Malone, M. F. Doherty, 1997, Attainable Regions for Reaction with Separation, A.I.Ch.E.J., 43 (2), 374-386.

Aspen Plus User Manual, Aspen Plus version 2004.1, ASPEN Technologies Inc.: Cambridge, MA, 2004.

D. Glasser, D. Hildebrandt, C. Crowe, 1987, A geometric approach to steady flow reactors: The attainable region and optimization in concentration space, Ind. Eng. Chem. Res., 26, 1803-1810.

F. Horn, 1964, Attainable and non-attainable regions in Chemical reaction technique, In Proceedings of 3rd Euro. Symp. on Chemical Reaction Engg. , Pergamon Press, New York, 1-10.

S. Talwalkar, S. Mankar, A. Katariya, P. Aghalayam, M. Ivanova, K. Sundmacher, S. M. Mahajani, 2007, Selectivity engineering with reactive distillation for dimerization of C4 olefins: Experimental and theoretical studies, Ind. Eng. Chem. Res., 46 (10), 3024-3034.

V. Agarwal, S. Thotla, S. M. Mahajani, 2008, Attainable regions of reactive distillation - Part I. Single reactant non-azeotropic systems, Chem. Eng. Sci, 63, 2946-2965.

Optimization-based design of reactive distillation columns using a memetic algorithm

Maren Urselmann[a] and Sebastian Engell[a]

[a]*Process Dynamics and Operations Group, Technische Universität Dortmund,
44221 Dortmund, Germany,{maren.urselmann | sebastian.engell}@bci.tu-dortmund.de*

Abstract

This contribution deals with the optimization of the design of reactive distillation columns by using a memetic algorithm (MA) which is a combination of an evolution strategy (ES) and a mathematical programming (MP) solver. The focus of this paper is on the restriction on the number of feed trays that introduces discrete variables into the problem formulation. The optimization of the number and of the location of the feed streams is addressed by the EA which can deal with discontinuous cost functions and integrality constraints. A minimization of the number of feed streams is achieved by adding penalties to the fitness function of the EA. The results of the new approach are compared to the results of the optimization without restriction on and the minimization of the number of feed streams.
Keywords: Conceptual design, mathematical optimization, memetic algorithm, reactive distillation

1. Introduction

A current trend in process design is towards integrated processes, i.e. the integration of reactive and separating functionalities into a single apparatus as, e.g., a reactive distillation column. Compared with the classical serial arrangement of unit operations, this advanced concept has the potential to decrease the dimensions of the equipment and to increase the degree of heat integration. Furthermore, it provides the opportunity to overcome chemical and thermodynamical boundaries, such as chemical equilibria or distillation boundaries due to azeotropes. Separations of non-ideal mixtures with simultaneous chemical reactions belong to the most difficult design problems and should be solved in an integrated fashion. The design of a reactive distillation column constitutes a constrained combinatorial optimization problem which is amenable to MINLP techniques. In practice, such problems are often hard to solve due to non-linear and integrality constraints and the nonconvexity of the problem.

Recently, a memetic algorithm (MA) for the global solution of reactive distillation problems without restrictions on the number of feeds was introduced [1]. By the use of this method the computational effort needed for a local search of the continuous design optimization with a fixed number of trays could be reduced by 75% in comparison to the reference algorithm (OQNLP/CONOPT). The MA consists of an evolutionary algorithm (EA) and the mathematical NLP solver CONOPT. The EA generates initial points for the local solver. It works in the space of the design variables and the state variables of the column designs are computed by the same solver that performs the local optimization. The MA was able to identify a huge number of local solutions, but more than 80% of them represent column designs with more than three streams per feed. For the more complex model instances with $N > 40$ trays it was not possible to find good solutions with less than three feed streams in reasonable computation times.

In this work, the memetic algorithm is extended by a restriction of the number of feed streams. Three different approaches are tested and compared to the basic approach without restrictions ($MTBE_{NLP}$). The first formulation ($MTBE_{NLP1}$) is almost equivalent to the basic approach but if the MP solver finds local solutions with an unrealistic number of feed streams, they are considered to be infeasible by the EA. In the second formulation ($MTBE_{CF1}$) the EA addresses the optimization of the number and of the location of the feed streams which is done by the introduction of discrete variables. In the third formulation ($MTBE_{CF2}$), the minimization of the number of feed streams is additionally taken into account by adding a penalty term for each feed stream.

2. The case study

The kinetically controlled production of MTBE from isobutene and methanol (IB + MeOH $\leftrightarrow$ MTBE) in the presence of n-butane at a pressure of 8 bar is considered here. The desired purity of the product is 99 mole-%. The total amount of the feed streams is fixed ($F_{1,tot}$ = 6.375 mole/s MeOH, $F_{2,tot}$ = 8.625 mole/s IB/n-butane). The reboiler and the condenser of the distillation column are modeled as trays. The number of trays N is fixed. It is assumed that there can be a certain amount of catalyst on each tray of the column, restricted by the volume of a tray. No reaction is taking place in the reboiler nor in the condenser. The objective is to maximize the annual profit which is calculated by the annual revenues for the products minus the annualized investment cost, annual operating cost and annual cost for raw materials. The investment cost are calculated by heuristic functions for the cost of the column shell, the internals, the catalyst, the condenser and the reboiler; the operating cost are calculated by the heat loads for heating and cooling.

The model variables are related to the operating conditions inside the column, e.g. the pressure, the vapour velocities and the concentrations of the substances. The set of design variables $d \in D$ consists of the amounts of both feeds $i = 1, 2$ on each tray $k = 1, ..., N$ denoted by $F_i(k)$, the amounts of catalyst on each tray $k = 2, ...,N-1$ denoted by $E_{cat}(k)$ and two variables α_{top} and $\alpha_{bottom} \in (0, 1)$ for the reflux ratio at the top and the ratio of the evaporation rate to the product removal at the bottom of the column. In case of the model without restrictions on the feeds ($MTBE_{NLP}$), it is assumed that fractions of both feed streams can enter the column on each tray of the column including reboiler and condenser.

2.1. NLP Model

The optimization problem is modeled as an NLP optimization problem. All variables (model variables and design variables) are free decision variables. The model consists of a large number of algebraic equations formulated in the modeling language GAMS. These equations describe the mass transfer and the reaction on each tray of the column (the model equations) and the annualized cost. Furthermore it comprises a set of inequalities to fulfil the required purity of the product and the required throughput, and to restrict operating conditions to feasible choices. For a more detailed description of the model see [3].

2.2. Simulation Model

The simulation model comprises a subset of the equations and of the inequalities of the optimization model. In the simulation model, the design variables were removed from the set of free variables. The equations and the inequalities that restrict the feasible values of the design variables were also removed from the set of constraints (see [1] for more details).

3. Solution approach and previous work

Memetic algorithms [4] are hybrid evolutionary algorithms coupled with local refinement strategies. In this work, an evolution strategy (ES) which is a special form of an EA is used. ES are designed for continuous search spaces and have a special feature: the strategy parameters, e.g. the parameters to determine the mutation strength, are adapted during the search (*self-adaptation*). A detailed introduction to ES can be found in [5]. In Figure 1, the structure of the MA used here can be seen.
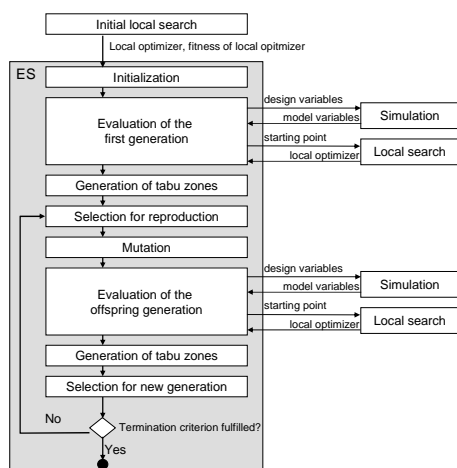

Fig. 1: Structure of the memetic algorithm

At the beginning of each optimization an initial local search is performed by CONOPT. This procedure is necessary to compute initial values for the model variables within the simulation model, because a good initialization is crucial for the solution of the model equations. During the subsequent local optimization, the values of the model variables of the nearest feasible point found so far (measured by the Euclidean norm) serve as initial values for the simulation. Then the ES is started to address the global optimization of the design variables. In order to evaluate the individuals of the population, the corresponding model variables are computed by CONOPT using the simulation model. The resulting point in the space of all variables represents a possible column design which is used as a starting point for the local optimization in the space of all variables that is performed also by CONOPT. For each localized local optimum, a region can be defined (a *tabu zone*) to approximate the basin of attraction of this optimum. These regions are excluded from the subsequent search procedure of the EA. It could be shown in [1] that the use of these zones can improve the efficiency of the search. A detailed description of all elements of the MA can be found in [1].

4. Restriction of the number of feeds

In order to obtain realistic solutions, a maximal number of three feed trays for each feed is assumed. Three new approaches were developed and tested: In the formulation $MTBE_{NLP1}$, solutions with more than three feed trays per feed found by the MP solver are not selected for the next generations by the EA.

In approach $MTBE_{CF1}$, the EA defines three possible feed tray locations per feed for each individual. 'Possible' means, that the amount of feed on these trays is nonnegative (but can be zero). The amounts of feeds on the other trays are fixed to zero, so they cannot be changed by the local solver. After the simulation model is solved for the remaining variables, the local solver optimizes the amounts of feeds on the fixed tray locations in the subsequent local search procedure. E.g., if the starting point proposed by the EA defines three possible feed trays for F_1 with two positive feed streams on these trays the local solver may remove one of the feed streams or add another stream on one of the three trays.

In formulation $MTBE_{CF2}$, the number of feeds not only is restricted but it is also minimized by the EA by means of adding a penalty term for each non-zero feed stream to the fitness function. These penalty terms make the fitness function jump whenever

the number of feeds changes. Since CONOPT cannot handle discontinuous functions, the objective function in the optimization model is not changed. In order to ensure, that the results of the local search are local solutions with respect to both functions (the fitness function and the objective function), the procedure in MTBE$_{CF1}$ is modified such that CONOPT is not able to add feed streams during the local search, whereas removing a feed stream is still allowed. This is because adding an additional feed stream would cause the fitness function jump to a lower (=worse) value although the objective function increases. That cannot be considered by the MP solver. Removing a feed stream on the other hand would cause the fitness function jump to a higher (=better) value. I.e. a solution with a smaller number of feeds that is a local optimum with respect to the objective function is also a local optimum with regard to the fitness function of the EA. Hence, in this formulation, the EA defines variable values from one to three to the numbers of possible feed trays with only positive amounts of feeds on these trays. So, the MP solver cannot add feed streams during the local search.

5. Extension of the MA

5.1. Representation

In evolutionary strategies, individuals are represented by a vector that represents the object variables of the optimization problem (here: the design variables) and a strategy parameter vector. In the formulations MTBE$_{NLP}$ and MTBE$_{NLP1}$, the genes of an individual are given by the continuous design variables described in Section 2. The strategy parameter vector contains one step size parameter for the mutation of the amounts of feeds for each feed, $N - 2$ parameters for the amount of catalyst on the trays and one for each parameter α_{top} and α_{bottom}. In formulations MTBE$_{CF1}$ and MTBE$_{CF2}$, the vector of object variables is extended by the discrete variables noF_i that represent the number of feed streams per feed and by two vectors $indF_i$ with noF_i distinct discrete elements $d \in \{1, ..., N\}$ that represent the indices of the feed trays for feed i. The extended strategy parameter vector contains four additional parameters: one parameter for each variable noF_i and one for each vector $indF_i$.

5.2. Mutation

In evolution strategies the mutation of the strategy parameters and of the object parameters is done consecutively. The strategy parameters are mutated first. In this work, reflection is used to avoid that a parameter exceeds its feasible domain. The continuous object parameters and the corresponding strategy parameters are mutated by the standard mutation operator for continuous search spaces [5]. The strategy parameters for the discrete variables are mutated as described in [6]. The mutation of the variables that correspond to the feed streams is done in a hierarchical fashion. At first the numbers of feed streams, noF_i are mutated independently by the standard mutation operator for integer variables [6]. If the number of feed streams noF_i is changed by the mutation, the corresponding index vector $indF_i$ is adjusted by removing or adding an index randomly. The mutation of the indices of the feed trays corresponds to a relocation of the feed streams along the column. At first, the total number of trays for which the feed streams should be relocated is determined. The indices of the feed trays are perturbed by a random number until the predefined number of relocations was performed. A repair procedure is applied to ensure that the vectors $indF_i$ do not contain duplicates. Before the variables corresponding to the amounts of feed F_i are mutated as described in [1], they are also relocated to the new feed locations. In case of the approach MTBE$_{CF2}$ the application of a repair procedure is necessary that ensures a positive amount of feed on each feed tray.

Optimization-based design of reactive distillation columns using a memetic algorithm

5.3. Fitness function

The fitness function f is equal to the objective function c of the case study described in Section 2. Only in approach $MTBE_{CF2}$ a penalty term for each positive feed stream is added to the cost function. This penalty cost is chosen to be an additional investment of *20,000€* to ensure that solutions with fewer feeds are preferred. It is multiplied by the cost index that is already used to scale the investment cost of the column and multiplied by the sum of the number of feed streams. Tabu zones were not used in this study (neither in the basic algorithm $MTBE_{NLP}$ nor in the new formulations).

6. Evaluation

The three approaches described in the last section were tested on a PC with 3.06GHz and 2GB RAM. 5 runs for each model instance with $N = 10, ..., 60$ trays were done. The strategy parameters for all runs of the MA are $\mu = 5$, $\kappa = 5$ and $\lambda = 10$. The termination criterion is a limit of 50 generations. The best selection operator was determined by preliminary tests. In case of the approaches $MTBE_{CF1}$ and $MTBE_{CF2}$ a (μ, κ, λ)-selection was applied, whereas a selection procedure that chooses randomly μ solutions without allowing duplicates out of the set of all parents and offspring individuals led to the best results for approaches $MTBE_{NLP}$ and $MTBE_{NLP1}$.

6.1. Analysis of the problem size

Without regarding restrictions on the number of feed streams, the problem space of the optimization model is connected and contains $149\,N + 14$ continuous model variables, whereas the search space of the EA is formed by $3\,N$ continuous design variables. The restrictions of the number of feeds divide the formerly single continuous problem into a huge number of sub-problems that have to be solved separately. This number corresponds to the product of the number of possible combinations to distribute noF_i feed trays on N trays. If noF_i can have variable values, then the sum of all possible combinations for these values has to be taken into account. The exact number of sub-problems for a model instance with N trays and $noF_i = 1,..., maxNoF_i$ feed trays can be determined by equation (1).

$$\sum_{noF_1=1}^{\max NoF_1} \sum_{noF_2=1}^{\max NoF_2} \binom{N}{noF_1}\binom{N}{noF_2} \tag{1}$$

The total number of sub-problems for $N = 10,..., 60$ trays is $\approx 1.04 \cdot 10^{10}$ in case of approach $MTBE_{CF1}$ and $\approx 1.18 \cdot 10^{10}$ in case of approach $MTBE_{CF2}$.

6.2. Numerical results

The two plots in Figure 2 show a comparison of the profit of the best solutions found by the application of the tested approaches. Formulation $MTBE_{CF1}$ leads to the best realistic solutions on all instances. In case of the more complex models with $N > 45$
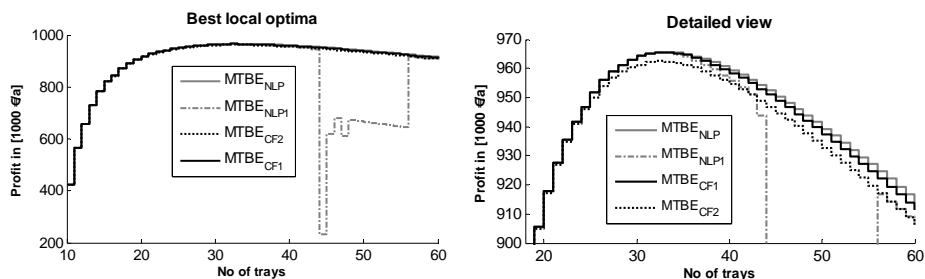

Fig. 2: Comparison of the quality of the best solutions

trays, an improvement of up to 75 % in comparison to the results of MTBE$_{NLP1}$ could be reached. On average the profit of the solutions found by the approach MTBE$_{CF2}$ is 0.3% lower than the profit (without the penalty) of the solutions found by the formulation MTBE$_{CF1}$ while the median number of feed streams is reduced from 6 to 3.
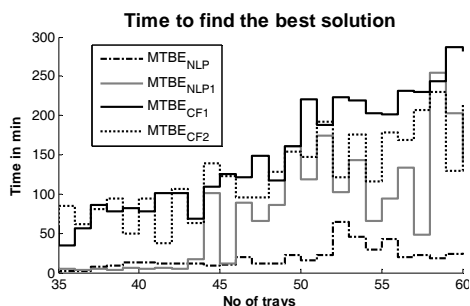

Fig. 3: Mean computation times needed to find the best solution

For model instances with N < 45 trays, the restrictions of the number of feeds cause a significant increase of the computational effort needed to find good realistic solutions (see Figure 3). In case of the more complex models, good solutions could only be found by regarding the restrictions.

In Figure 4 the globally optimal solutions found by the different approaches are shown. The solutions for the approaches MTBE$_{NLP}$, MTBE$_{NLP1}$, and MTBE$_{CF1}$ are identical, whereas the solution of approach MTBE$_{CF2}$ has three feed streams less, while the profit without regarding the penalty is only 3,150 €/a smaller.

7. Conclusions & Outlook

Four approaches to the global optimization of a reactive distillation column with and without restrictions and minimization of the number of feeds were compared. Including the restrictions divides the problem into more than 10^{10} continuous sub-problems, what makes the computational effort needed to find good solutions increase significantly. These additional expenses are not justified in case of model instances with up to N ≈ 40 trays, because the formulation without restrictions can find realistic solutions with almost similar objective values faster. For the more complex model instances with N > 45 trays, the quality of the results as well as the computational effort needed to find these solutions can be improved by explicitly modeling the restrictions. The use of an

Fig. 3: Global solution: MTBE$_{NLP}$, MTBE$_{NLP1}$, and MTBE$_{CF1}$ (left), MTBE$_{CF2}$ (right)

additional minimization procedure by adding a penalty term to the cost function leads to results with almost the same objective values and helps to decrease the number of feeds without a further increase in the computational effort.

In future work, the MA will be extended to the optimization of the number of trays.

References

[1]   M. Urselmann, G. Sand, S. Engell, Proc. IEEE CEC (2009), pp. 1721-1728.
[2]   S. Barkmann and G. Sand and S. Engell, CIT (80) (2008), pp. 107-117.
[3]   P. Moscato, Caltech Concurrent Computation Program, report 826, (1989).
[4]   H.G. Beyer and H.P. Schwefel, Natural Computing: An International Journal 1(1) (2002), pp. 3-52.
[5]   G. Rudolph, PPSN III LNCS, Berlin (1994), pp. 193-197.

Eco-Efficiency Analysis For Chemical Process Design

Adama Ouattara,[A] Azzaro-Pantel Catherine,[A] Pibouleau Luc,[A] Domenech Serge,[A] Baudet Philippe,[B] Yao Benjamin,[C]

[a] *LGC-CNRS-INPT, Université de Toulouse, 4, Allée Emile Monso, BP 84234, F-31432, Toulouse1, France, Catherine.AzzaroPantel@ensiacet.fr*
[b] *ProSim, Stratège Bâtiment A, BP 27210, F-31672 Labège Cedex, France, Philippe.Baudet@prosim.net*
[c] *Institut National Polytechnique Houphouët-Boigny, Département de génie chimique et agro-alimentaire, BP 1093 Yamoussoukro, Côte d'Ivoire, beyao@yahoo.fr*

Abstract

This paper addresses the problem of analyzing the various objectives involved in eco-efficient processes, meaning that ecological and economic considerations are taken into account simultaneously at the preliminary design phase of chemical processes. The environmental aspect is quantified at the preliminary design phase of chemical processes by using of a set of metrics or indicators following the guidelines of sustainability concepts proposed by IChemE [1]. The resulting multiobjective problem is solved by use of a genetic algorithm implemented in the so-called Multigen library. The trade-off between economic and environmental objectives is illustrated through Pareto curves. A key point of the methodology is the use of the package ARIANE[TM], a decision support tool dedicated to the management of plants utilities (steam, electricity, hot water...) and pollutants (CO_2, SO_2, NO_x, etc..), used here both to compute the primary energy requirements of the process and to quantify its pollutant emissions. The well-known benchmark process for hydrodealkylation (HDA) of toluene to produce benzene, revisited here in a multi-objective mode, is used to illustrate the usefulness of the approach in finding environmentally friendly and cost-effective designs.

Keywords: multiobjective optimization, genetic algorithm, eco-efficiency, economic criterion, environmental impact

1. Introduction

In traditional chemical process design, attention has been focused primarily upon the economic viability. Yet, chemical plants can no longer be designed on the unique basis of technico-economic concerns and the two other dimensions of sustainability – environmental and social – leading to the so-called "Triple Bottom line", must be part and parcel of the design phase. A major difficulty to tackle the problem is that there are many independent but often competing objectives that have to be considered simultaneously. Lots of ongoing research aim to develop a set of metrics or indicators (the amount of metrics may vary from 10 [2] to 134 [3]) to draw a quantitative profile of sustainability. This study aims at the development of a design framework for eco-efficient processes, following the guidelines of the environmentally conscious design (ECD) methodology proposed by Allen and Shonnard [4]. For this purpose, several indexes of environmental impact including ozone depletion, global warming potential, human and aquatic toxicity, photochemical oxidation as well as acid rain potentials have to be taken into account. The environmental aspect is quantified by using a set of

metrics or indicators as proposed by IChemE [1]. Such problems lead to multiple and most often conflicting goals and must be solved by means of efficient multiobjective optimization tools. In this study, optimization is performed by genetic algorithms implemented in the previously developed Multigen library [5], that is particularly well-suited for multiobjective optimization and mixed integer nonlinear problems. The trade-off between economic and environmental objectives is then illustrated through Pareto curves. The well-known benchmark process for hydrodealkylation (HDA) of toluene to produce benzene [6] is revisited here in a multi-objective mode and illustrates the usefulness of the approach in finding environmentally friendly and cost-effective designs. A key point of the methodology is to capture in the modelling approach both process and utility production units, since the environmental impact of a chemical process not only contains the material involved in the process but also the energy consumption, the effect of flow recycle, percent conversion and so on...

2. Eco-efficient chemical process design framework

2.1. General framework

The design of eco-efficient chemical processes (see Fig. 1) integrates a mathematical process model, coupled with an impact assessment model which is embedded in an outer multiobjective optimization loop. The proposed framework is an alternative design methodology for waste minimization to the so-called WAR algorithm [7], which has been extensively used in the literature: let us recall that this method is based on a potential environmental impact (PEI) balance for chemical processes. The PEI is a relative measure of the potential for a chemical to have an adverse effect on human health and the environment (e.g., aquatic ecotoxicology, global warming, etc.). The result of the PEI balance is an impact (pollution) index that provides a quantitative measure of the impact of the waste generated in the process.
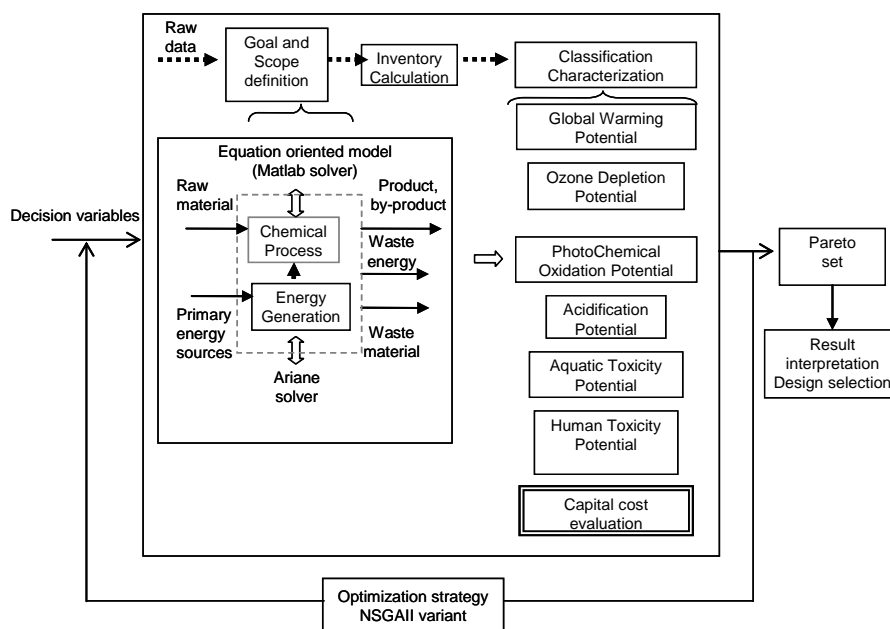

Fig. 1: General optimization and modelling framework

Eco-Efficiency Analysis For Chemical Process Design
.

Recently, several systematic methodologies are available for the detailed characterization of the environmental impacts of chemicals, products, and processes, within Life Cycle Assessment (LCA) concepts. In this work, the idea is to use their potential to develop a Life Cycle Analysis method dedicated to process development. Since our approach focuses on decreasing the environmental impacts of the manufacturing stage and associated utility systems, only a "cradle-to-gate" analysis is performed.

Concerning the multiobjective optimization strategy, the MULTIGEN library, offering a variant of the classical NSGAII algorithm [5, 8] has been used: it has been designed as a library of multiobjective genetic algorithms codes written in VBA, equipped with an Excel® interface.

A key point of the methodology is the use of ARIANE™ [9], a decision support tool dedicated to the management of plants that produce energy under the form of utilities (steam, electricity, hot water...) and of its add-in module Plessala™ developed by ProSim. ARIANE™ is used here both to compute the primary energy requirements of the process and to quantify the pollutant emissions due to energy production.

2.2. Economic and environmental models

The economic functions used for main equipment items follow the general correlations proposed by Guthrie [10]. For utility systems, capital cost estimation is performed by using expressions developed in [11].

The environmental impacts are quantified by the so-called environmental burdens proposed by IChemE [1] and obtained from the flowsheet through material and energy balances.

The Environmental Burdens (EB) that have been taken into account are the following ones: Global Warming Potential (GWP), Acidification Potential (AP), Ozone Depletion Potential (ODP), Photochemical Oxidation or smog formation Potential (PCOP), Human Toxicity Potential Effect (HTPE) and Aquatic Toxicity Potential (ATP). The Environmental Burden (EB) caused by the emission of a range of substances, is then calculated by adding up the weighted emission of each substance:

$$EB_i = \Sigma W_N \times PF_{i,N} \qquad \text{(Eqn. 1)}$$

EB_i = i[th] environmental burden
W_N = weight of substance N emitted, including accidental and unintentional emissions
$PF_{i,N}$ = potency factor of substance N for i[th] environmental burden.

Let us recall that the potency factor of each environmental burden is expressed relative to the contribution of a reference substance. For instance, the potential contribution to global warming from a given quantity of the gas is relative to the contribution for a corresponding quantity of CO_2. The weighting factor is known as the "potency factor". Note that because a single substance will contribute differently to different burdens, each substance will have a number of different potency factors.

3. Application to the HDA case

3.1. Presentation of the illustrative example

The hydrodealkylation (HDA) process for producing benzene, a benchmark example in chemical process synthesis, is used in this study. This process has been extensively studied by Douglas [6] using a hierarchical design/synthesis approach. In this process, benzene is formed by the reaction of toluene with hydrogen. The hydrogen feed stream has a purity of 95% and involves 5% of methane; this stream is mixed with a fresh inlet

stream of toluene, a recycled toluene, and a recycled hydrogen streams. The feed mixture is heated in a furnace before being fed to an adiabatic reactor. The reactor effluent contains unreacted hydrogen and toluene, benzene (the desired product), biphenyl, and methane; it is quenched and subsequently cooled in a flash separator to condense the aromatics from the non-condensable hydrogen and methane. The vapour steam from the flash unit contains hydrogen that is recycled. The liquid stream contains traces of hydrogen and methane that are separated from the aromatics in a stabilizer. The liquid stream from the stabilizer consisting of benzene, biphenyl and toluene is separated in two distillation columns. The first column separates the product, benzene, from biphenyl and toluene, while the second one separates biphenyl from toluene, which is recycled back into the reactor.

To model the HDA process, design and flowsheeting packages could be used to model the process in order to compute the objective functions: yet, to identify more easily the failure cause identification of simulation, the equations proposed by Douglas [6] have been directly implemented and solved by the Matlab solver *fsolve* based on a classical Newton-Raphson method. The HDA case, considered here as a test bench, refers to the standard example proposed by Douglas [6]. As abovementioned, ARIANE^{TM} [9] is used here to model the utility production unit: a typical steam generating facility is considered where steam is produced from a conventional mono-fuel fired boiler (40 bar, 376 °C) and let down to the lower pressure levels (respectively, 10 bar, 336°C and 5 bar, 268 °C) through turbines which produce electricity used in the plant. It must be also highlighted that the fired heater (furnace) of the process is considered as a bi-fuel fired heater (mix of natural gas and fuel oil).

The environmental impact contributions for the components in the HDA process have been taken from [12]. Biphenyl has been considered as a pollutant.

Using the previous economic and environmental objective functions, the multiobjective nonlinear problem is formulated as follows:

Determine decision variables (process operating conditions) in order to simultaneously:

Min (Investment)   (Eqn. 2)

Min (EB_i), i=1, 8   (Eqn. 3)

> *s.t.*
> *Mass and energy balances (Matlab solver and ARIANE^{TM})*
> *Bounds on decision variables*

Table 1 shows the decision variables, their initial values and associated bounds.

The additional following constraints have been considered in the Multigen interface:
- The purity of the benzene product is at least 99.97 %
- The benzene production rate is maintained at 265 kmol/h
- The hydrogen feed must have a purity of 95 %
- The reactor outlet temperature is less than 704.50 °C
- The quencher outlet temperature is less than 621.16 °C
- All pollutants, CO_2, NO_x, CO, SO_2, dusts flowrate (kg/h) must take positive values.

The choice of the decision variables that have been taken into account will not be discussed here. The lower and upper bounds as well as the initial value have been set from the analysis of Douglas [6] and Turton et al. [13].

Eco-Efficiency Analysis For Chemical Process Design

.

Table 1: Decision variables for the HDA process

Decision variables	Lower bound	Initial value	Upper bound
Adiabatic reactor temperature (°C)	600	634.5	704.50
hydrogen purge flow rate (kmol/h)	31	372.50	1612
Flash pressure (bar)	30	32.04	34
Stabilizer pressure (bar)	4	9.30	10
Column 1 pressure (bar)	2	2.80	4
Column 2 pressure (bar)	1	1.03	2
Fuel oil flowrate (bi-fuel fired boiler) (t/h)	1	11.10	50
Ratio (bi-fuel furnace)	0.1	0.50	0.9
Vapor steam output flowrate (mono-fuel boiler) (t/h)	10	219.22	400

4. Results and Discussion

4.1. Simulation of a reference HDA case (Douglas [6])

The standard case of the HDA process has served as a reference case to demonstrate the interest of the approach. Although a detailed presentation concerning the various contributions of both units can be performed component by component, item by item only significant results relative to the impacts of both facilities are displayed in Fig. 2. Not surprisingly, the utility facility contributes strongly to GWP, PCOP and AP whereas HDA process has a strong impact on HTPE and ATP and contributes moderately to GWP and PCOP.
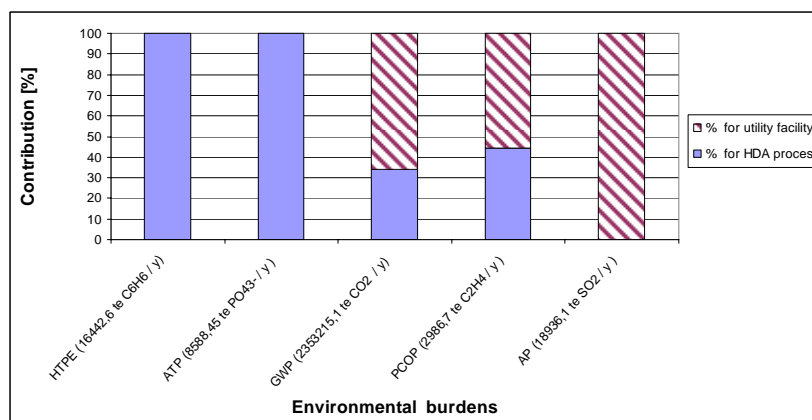
Fig. 2: Environmental burdens of the standard case

4.2. Multiobjective results

All environmental and investment criteria have been considered simultaneously for optimization purposes. The results obtained from the general framework of Fig. 1 are presented in 2D-graphs (Fig. 3a to 3e), with the economic criterion and one of the environmental impacts. Very interesting solutions are found at the so-called "knee" (S1, S2, S3), which all lead to a strong improvement of all objectives and constitute good candidates for design options, since the decision variables lead to acceptable values for the practitioner (the natural gas/fuel ratio being a sensitive variable for GWP decrease).
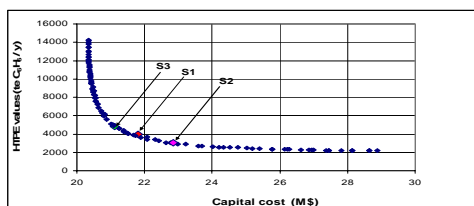
Fig. 3a
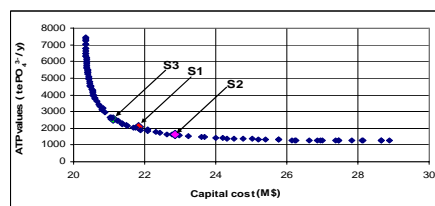
Fig. 3b
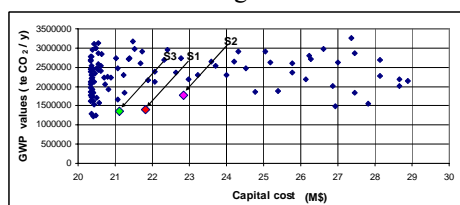
Fig. 3c
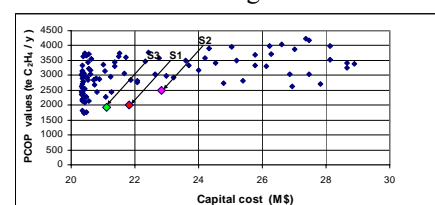
Fig. 3d
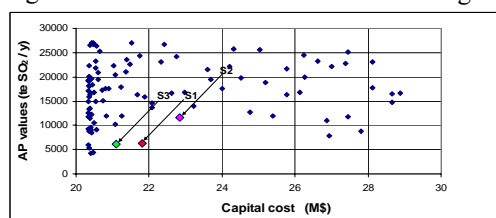
Fig. 3e

Fig. 3: Environmental burdens of the standard case

5. Conclusions

This paper has presented a methodology for ecodesign and optimization of chemical process taking into account the contribution of utility generation. The methodology incorporates environmental factors into the chemical process synthesis at the beginning design stage, which is totally different with the traditional end-of-pipe treatment method. The interest of NSGAII is that no preference between criteria is required and the Pareto-optimal solutions can be directly generated. The next step is now to develop a decision-aid technique from this large set of compromise solutions.

References

[1] IChemE, The Sustainability metrics, http://www.icheme.org/sustainability/
[2] AICHE, Sustainability Metrics, 1998. http://www.aiche.org/cwrt
[3] CSD, Indicators of Sustainable Development, New York, 1996.
[4] D.T. Allen, D. R. Shonnard,, AIChE J., NJ 07458 (2002)
[5] A. Gomez, C. A.-Pantel, L. Pibouleau, S. Domenech, , ESCAPE 18, Elsevier, 2008
[6] J. M. Douglas, Conceptual design of Chemical Process, McGraw Hill, New York, 1988
[7] D. Young, R. Scharp and H., Cabezas, J. Hazard. Mater, (95) (2002)
[8] K. Deb, IEEE Trans. Evol. Comput., 6 (2002) 182-197
[9] ARIANE^{TM}, http://www.prosim.net/fr/energy/ariane.html
[10] K. M. Guthrie, Capital cost estimating, Chem. Eng, 76(6) 1969,
[11] J. C Bruno, F. Fernandez, F Castells, I. E Grossmann, Chem. Eng. Res. Des, 79(A) (1998)
[12] S. K. Sikdar, M.M. El-Halwagi, Process design tools for the environment, Taylor and Francis, New York, 2001
[13] R. Turton, R. C. Bailie, W. B. Whiting, J. A. Shaeiwitz, Analysis, Synthesis, and Design of Chemical Processes, Prentice Hall PTR, New York, 1998

Optimal Design of Combined LNG/LPG Regasification System

Hosoo Kim, Ik Hyun Kim and En Sup Yoon

School of Chemical and Biological Engineering, Seoul National University
Shillim-dong, Gwanak-gu, Seoul, Republic of Korea, lake@pslab.snu.ac.kr

Abstract

In compatibility of liquefied natural gas (LNG) calorific values between importing countries has become one of the important technical and commercial issues in the LNG industry. Until now, heavy hydrocarbon removal (HHR) system, or nitrogen dilution system has been used in controlling the calorific value; however, these methods can make the process more complicated and less profitable. This is especially the case for offshore plants. A new concept of "Rich LNG" may bring the fundamental and economic solution for the calorific compatibility problem by combining the existing chains of LNG industry. But, in "Rich LNG", the natural gas liquids (NGL) removal must be efficiently performed by utilizing abundant available cold energy during the LNG regasification process in offshore or onshore. In this study, a multi-objective and mixed-integer nonlinear programming technique is applied for generating the complete Pareto optimal sets for a new calorific value adjustment process.

Keywords: regasification, liquefied natural gas, calorific value, mixed-integer nonlinear programming

1. Introduction

In traditional heavy hydrocarbon removal (HHR) systems, natural gas liquids are removed from the LNG in a distillation process by vaporizing the LNG in a distillation column known as a demethanizer. Natural gas liquids (NGLs) typically are comprised of C2+ hydrocarbons which not only may increase the calorific values of the natural gas beyond specification limits, but also may have a greater market value in their own right. McCartney [1] describes such regasification process and configurations. While these configurations and methods typically operate satisfactorily for onshore facilities, these configurations and methods for offshore facilities would be unacceptable as these configurations require relatively substantial space. [2,3] Also, it requires a large electric power supply for compressors and a high heat load for the reboiler of the distillation column.

In current offshore LNG regasification terminals, LNG is typically heated to pipeline specification in offshore vaporizers using seawater or submerged combustion vaporizers and sent out through a riser and by subsea pipeline to the customer. However, because of the significant cost for refrigeration and recompression, adjustment of calorific values after vaporizing is generally not economical. The traditional HHR system using distillation column is not appropriate due to the space limitation in an offshore environment. Thus, many methods and configurations are being studied to determine a simple and cost-effective method and configuration to adjust calorific values in an offshore regasification terminal.

2. Problem Statement

Figure 1 illustrates a possible superstructure for a new calorific adjustment process. LNG is first pressurized from the storage tanks and divided into two portions. One portion of the pressurized LNG is then heated through several heat exchangers and supplied to the 1st separation system. In the 1st separation system, a simple 2-phase separator or distillation column can be used for LNG processing. For the distillation column, column overhead condensers, which are plate-and-fin or shell-and-tube exchangers, use low-temperature LNG as the cooling medium. The pressure of the liquid stream from the 1st separation system is reduced to a lower pressure using a valve and fed to a 2nd separation system which has three processing options. One is a simple 2-phase separator, another is the distillation column, and the other is just a by-pass. As in the 1st distillation column, low-temperature LNG is used as the cooling medium for the condenser of the 2nd column. The overhead stream generated from the 2nd separation system is compressed and combined with the overhead stream of the 1st separation system.
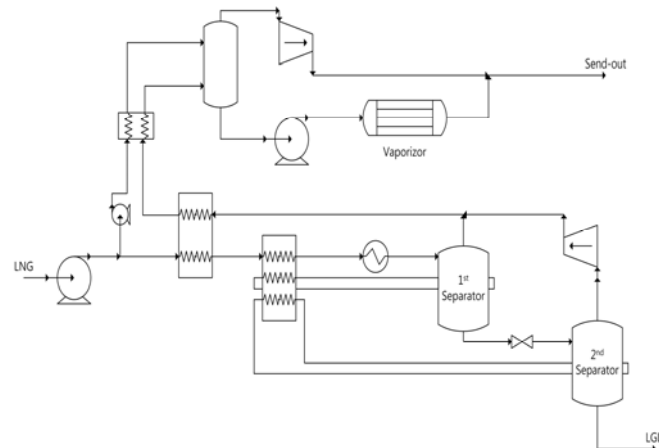
Figure 1. Superstructure for a new calorific adjustment process

The combined overhead stream is finally recovered as product to meet the specification limit of calorific values. Then, this lean and cold natural gas is re-condensed to a liquid using low-temperature LNG entering the process that allows the liquid to be pumped (rather than vapor compression) at about 100 barg, thus substantially lowering energy consumption, capital, and operational costs. The second portion of the pressurized LNG is used as cold energy source for re-condensing the product gas and then is combined with the condensed lean LNG. In this case, additional equipment (heat exchanger and pump for the split portion) and a process constraint like a minimum flow for the pump are necessary. If the initial pressurized LNG was not divided, the process could be simpler but larger. The compressed small quantity of vapor residue is mixed with vaporized lean LNG and sent-out through the pipeline.

3. Solution Procedure

This design problem has been defined in two parts, generalized disjunctive programming (GDP) problem with one objective and multiobjective problem for minimizing the operating costs and the performance of NGLs. First, the GDP problem has been mathematically reformulated as the MINLP problems [4] and the MINLP technique incorporated into the process simulator using its own optimization

capabilities [5] has been suggested. For solving the resulting bi-criterion problem with MINLP problem, we have suggested the heuristic procedure that reduces the number of discrete solutions which are necessary for complete Pareto optimal sets.

To find an effective discrete solution, we applied the concept of an 'Ideal Point' (IP) which was originally used to scalarize problems having multiple objectives and minimize the Euclidean distance between IP and the set of Pareto optimal solutions. [6] A discrete solution is obtained by solving the problem whose objective is the minimization of the Euclidean distance between the ideal point and the set of objectives. This problem is described as follows.

$$\min f(distance) = \min \left\{ (f_2^{ideal} - f_2)^2 + (f_1^{ideal} - f_1)^2 \right\}^{1/2}$$

$$s.t.$$

$$f_1(x, x_d, u) \le q$$
$$f_2(x, x_d, u) \le p$$
$$g_c(x, x_d, u) \le 0$$
$$h_c(x, x_d, u) = 0 \tag{1}$$
$$\underset{i \in D_j}{\vee} \begin{bmatrix} Y_{i,j} \\ h_{sij}(x, x_d, u) = 0 \end{bmatrix} \quad \forall j \in J$$
$$\Omega(Y) = True$$
$$Y \in \{True, False\}$$
$$x \in \Re^n$$

where x_d corresponds to the design variables which are the degrees of freedom in the process simulator. x corresponds to all the other variables of the process which are calculated by the process simulator. u corresponds to a vector of fixed parameters for existing equipment determined by disjunctions. The equations h_{sij} are implicit equations solved by process simulator that cannot be accessed by the user. These equations include all the mass and energy balance equations, thermodynamic equations, physical properties, etc. Equations h_c and g_c are overall process constraints declared in the optimization tool of the process simulator and solved by its own optimizer. J is a set of disjunctions and i is the index of components of the disjunction D_j.

After the discrete solution is determined, the heuristic procedure reducing the search region is applied and the ideal point is moved. In Eq. (1), Euclidean distance is used for objective function and f_1 and f_2 are transferred to parametric constraint for reducing the search region in the criterion search space. Figure 2 demonstrates the search region reduction process. In Figure 2a, z* is the initial ideal point, which components f_1^0 and f_2^0 are values, determined by the original MINLP problem with respect to each objective function. The minimum point z^1 and discrete solution y^1 are obtained by solving the Eq. (1) with an infinite bound of parametric constraints. The unmarked regions are infeasible and region Z_1 is dominated by z^1. Therefore, they can be neglected and the search regions are reduced to R_1 and R_2. We denote the two regions as follows.

$$R_1 = \{ z \in \Re^2 \mid \exists (x, y) \in \Omega : z = f(x, y) \& f_2 \le f_2^1 \} \tag{2}$$

$$R_2 = \{z \in \Re^2 \mid \exists (x,y) \in \Omega : \ z = f(x,y) \ \& \ f_1 \leq f_1^1 \} \tag{3}$$

In figure 2b if R_1 is chosen, the ideal point is moved to the left lower point of the region. This time, however, the additional integer cut must be included to exclude the previously found discrete solutions. The next minimum point z^2 and discrete solution y^2 are obtained by solving the Eq. (1) within region R_1. The region Z_2 is dominated by z^2 and R_1 is reduced again.

(a)   (b)
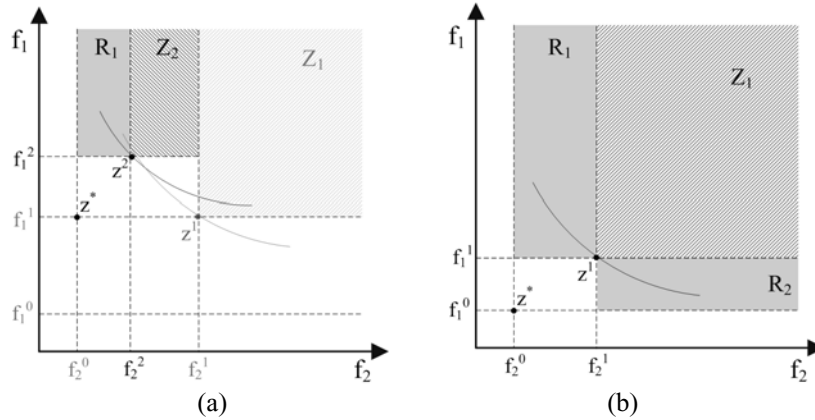
Figure 2. Reducing the search regions

4. Numerical Result

Figure 3 shows the Pareto curve and intermediate points obtained by the previously proposed procedure. Also, the cross-shaped points which are not connected with the Pareto curve are a locally-obtained Pareto optimal set by the enumeration. All these points are dominated by Pareto curve obtained by the proposed procedure.

Before the selection of discrete solution, the objective values were normalized with two MINLP problems in the initialization step. After the first discrete solution y^1 and point z^1 are determined, y^2 and y^3 were obtained reducing the region R_1. In the direction of region R_2, y^4, y^5, y^6 and y^7 were found in sequence. These discrete solutions are the minimal candidate set and the complete Pareto optimal set was obtained by solving the parametric NLPs and comparing the local sets. Although the point z^5 is not included in the complete Pareto optimal set in this case, the figure 3 shows that the proposed heuristic procedure is a simple and effective rule in this problem.

Configurations for each discrete solution are indicated in Table 1. In terms of NGL performance (f_2), it is more favorable to choose a distillation column instead of a 2-phase Separator. If a distillation column is selected, it is to a greater advantage to be used in the 1st separation process rather than the 2nd separation process; however, these choices show contrary characteriscs with respect to operating cost (f_1).

Table 1. Configurations for each discrete solution

	Split System	1st Separation	2nd Separation
y^1	Used	Distillation	2-Phase Separator
y^2	Used	2-Phase Separator	Distillation
y^4	Used	2-Phase Separator	2-Phase Separator

Optimal Design of Combined LNG/LPG Regasification System
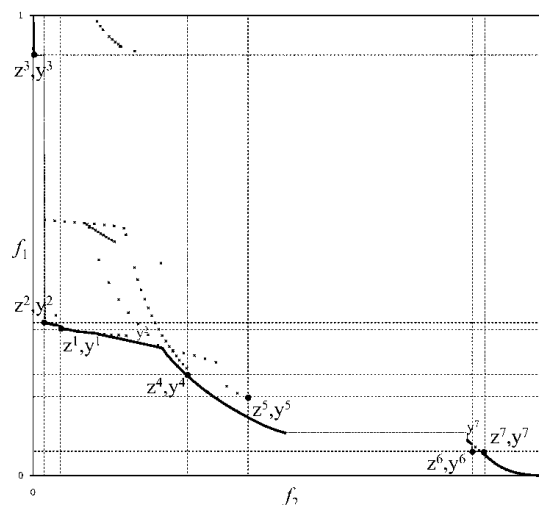
Figure 3. The attainable set of the design problem. Pareto curve obtained by the proposed procedure is emphasized

5. Conclusion

This paper has addressed the optimal design of a chemical process for simple and cost-effective adjusting calorific values in an offshore regasification terminal. This problem has been mathematically formulated as the multiobjective and mixed-integer nonlinear problems for minimizing the operating costs and the performance of NGLs. Process modeling has been performed in the commercial simulator for using rigorous models and thermodynamics under severe conditions. This design problem has been defined in two parts, generalized disjunctive programming (GDP) problem with one objective and multiobjective problem for minimizing the operating costs and the performance of NGLs. The Pareto curves obtained by our method provide quantitative relationships between conflicting objectives and thus can be used as a decision supporter for optimal design with robustness under varying situations.

6. Acknowledgements

References

[1] O.P.K. Daniel G. McCartney, Method for Vaporizing and Recovery of Natural Gas Liquids from Liquefied Natural Gas, U.S. Patent US 6564579, 2003.

[2] J. Mak, G. Curt, Configurations and Methods of Integrated NGL Recovery and LNG Liquefaction, U.S. Patent WO 2007/008525 A3, 2007.

[3] J. Mak, High Ethane Recovery Configurations and Methods in LNG Regasification Facility, U.S. Patent WO 2007/139876 A3, 2007.

[4] J.A. Caballero, A. Odjo, I.E. Grossmann, AIChE J. 53 (2007) 2351-2366.

[5] Aspen Technology, Inc., (2009).

[6] R. Steuer, Multiple Criteria Optimization : Theory, Computation, and Application, Wiley, New York, 1986.

Global Optimal Design of Mechanical Vapor Compression (MVC) Desalination Process

Marian Marcovecchio[a], Pio Aguirre[a], Nicolas Scenna[a], Sergio Mussati[a]

[a] INGAR Instituto de desarrollo y diseño, Avellaneda 3657, S3002GJC Santa Fe, Argentin, {mariangm; paguir; nscenna; mussati}@santafe-conicet.gob.ar

Abstract

This paper deals with the optimal design of Mechanical Vapor Compression (MVC) desalination process. Precisely, a detailed mathematical model of the process and a deterministic global optimization algorithm are applied to determine the optimal design and operating conditions for the system. The resulting model involves the real-physical constraints for the evaporation process. Nonlinear equations in terms of chemical-physical properties and design equations (efficiencies, Non-Allowance Equilibrium, Boiling Point Elevation, heat transfer coefficients, momentum balances, among others) are used to model the process. The model has been solved by using a deterministic global optimization algorithm previously developed by the authors [7] and implemented in a General Algebraic Modeling System GAMS [1]. The generalized reduced gradient algorithm CONOPT 2.041 [2] is used as NLP local solver.
The model was successfully solved for different seawater conditions (salinity and temperature) and fresh water production levels. The influence of the production requirements on the process efficiency as well as the algorithm's performance is presented.

Keywords: Vapor Compression (MVC) desalination, global optimization method.

1. Introduction

Desalination of seawater is one of the main alternatives to overcome the problem of fresh water supply. Various types of desalination systems are known and are in use. Typically, such systems include a water pre-treatment system, a desalination unit and a post-treatment system. The desalination of seawater in such systems is achieved through thermal or membrane processes. The thermal processes for seawater desalination include multistage-flash distillation (MSF), multi effect distillation (MED) and vapor compression (VC). Further, the main membrane systems are reverse osmosis (RO) and electrodialysis (ED) processes.

The product of a Vapor Compression system enjoys similar qualities to the other distillation processes. Its source of driving force is rotating mechanical energy, generally obtained from a motor. VC units tend to be small plants in isolated locations while the other processes are usually used for large fresh-water productions.

Furthermore, certain desalination systems can employ renewable energy sources for powering the desalination system. In fact, a mechanical vapor compression (MVC) desalination system may be powered by a wind turbine. Typically, wind powering of a MVC desalination system may be achieved either by direct mechanical coupling of the turbine shaft to the compressor axle of the desalination system, or by generating electrical power that is utilized to drive the electrical compressor drive. However, the

mechanical coupling does not provide any means for power regulation or speed control of the compressor drive.

A brief outline of the articles focusing on mathematical modeling and analysis of single effect VC desalination units can be found in [3, 4, 5].

In the present work, a mathematical model recently developed by the authors [6] which was solved for local optimality is properly extended in order to get a more realistic design of the process. Mass, energy and momentum balances are included in the mathematical model. Rigorous correlations to compute Boiling Point Elevation, total heat transfer coefficient and physical-chemical properties of all streams, among others, are also considered. The model involves non-convex constraints like bilinear terms and logarithms which can lead to local optimal solutions. For this reason, a deterministic global optimization algorithm previously developed by the authors [7] is implemented in this paper to find the global optimum for the resulting mathematical model.

The paper is outlined as follows. Section 2 briefly describes the process. Section 3 introduces the problem formulation. Section 4 summarizes the mathematical model. Section 5 presents applications of the developed model and results analysis. Finally, Section 6 presents the conclusions and future work.

2. Process description

Figure 1 shows a Single-Effect Mechanical Vapor Compression (MVC) desalination process. As is shown, the energy input is entirely as mechanical power to drive the compressor. No live steam is required except for preliminary heating to raise the plant to working temperature. Vapor formed is recompressed and introduced to the evaporator. Two primary methods can be used for the compression of the vapor: 1) Thermal Vapor Compression (TVC) and, 2) Mechanical Vapor Compression (MVC).

The main equipments used in the MVC desalting process are the evaporator, the compressor, pumps and pre-heaters.
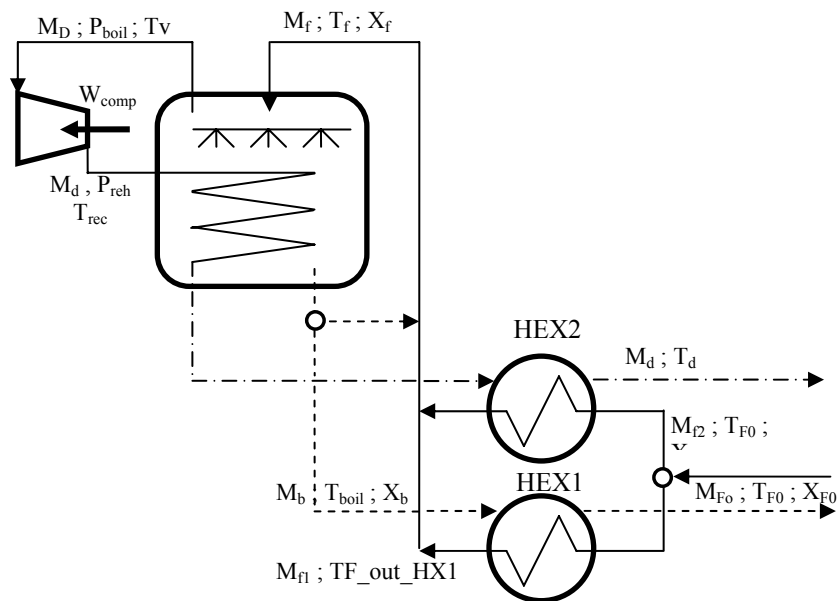

Fig. 1 Single-effect Mechanical Compression Vapor (MVC) desalination

As is shown in Fig. 1, the incoming seawater [Mf] is passed through two heat exchangers [HEX1 and HEX2] where it is preheated by the heat transferred from the discharged brine [Mb] and product [Md] streams. The sea water is then recycled and sprayed on the outside of a bundle where it boils and partially evaporates. Then, the produced steam is drawn through the demister to the centrifugal compressor [VC] which increases the pressure and temperature of the steam by compression. This steam is then discharged into the inside of the heat transfer tube bundle where it condenses into distillate [Md]. The compressor provides, through its suction, a pressure lower than the equilibrium of the brine facilitating the evaporation of the seawater. The energy performance of the system depends on the pressure increment in the mechanical compressor, on the thermodynamic efficiency of polytropic process and on the efficiency of the electric motor.

MVC plants are in service with energy consumptions around 11 kWh/m3, and designs have been developed with power consumptions as low as 8 kWh/m3. This is the lowest energy consumption of any distillation so far developed and is competitive with seawater reverse osmosis including energy recovery. Capital and energy costs are significant factors on the total water production cost. The main energetic consumption of the MVC distillation unit is represented by the electricity which is mainly required to drive the compressor motor, while there is no steam requirement. The operation and maintenance of the compressor motor may be half of the total operating cost. The process includes pumps for the seawater, brine and product. In some designs, part of the discharge brine is recycled by using a recirculation pump. MVC unit ratings have so far been limited to about 1500 m3/d. The most disadvantages of MVC system are the maximum allowable tip velocity of the compressor blades and mechanical compressor which limit the fresh water production.

3. Problem formulation

The optimization problem can be stated as follows: Given the water demand and seawater conditions, the goal is to determine the optimal operating conditions in order to minimize the electricity used by the compressor and total heat transfer area of the process. Another optimization problem could be formulated as follows: Given the total heat transfer area (evaporator and pre-heaters), the goal of the problem is to maximize the ratio of fresh water production to the electricity used by the compressor. Finally, other objective function may be the total annual cost of the process (minimization).

4. Assumptions and Mathematical Model

The resulting mathematical model is based on the following assumptions: a) Product is pure water, b) Heat losses from the evaporator surface are negligible, c) No recycle is considered, and d) equal overall heat transfer coefficients in the two heat exchangers are assumed. Basically, the model involves real-physical constraints for the evaporation process and is derived from the rigorous energy, mass and momentum balances. Physical and chemicals properties of the seawater and fresh water are computed by detailed correlations.

Next, the main equations of the model are presented.

Total mass and energy balance:

$$M_f = M_b + M_d \tag{1}$$

$$M_f \, X_f = M_b \, X_b + M_d \, X_d \tag{2}$$

The outside temperature at the compressor is computed as:

$$\frac{T_{sobrec}}{T_v} = Rp^{\frac{\gamma-1}{\gamma}} \tag{3}$$

where

$$Rp \ P_{boil} = P_{reh} \tag{4}$$

The mechanical power consumed by the compressor is given by:

$$W_{comp} \ \eta \ MW \ 3600 = M_d \ \frac{\gamma}{\gamma-1} R \ T_v \ \frac{T_{sobrec}}{T_v - 1} \tag{5}$$

The energy balance on the evaporator is as follows:

$$M_d \left(H_{sobrec} - H_{vap\_sat}\right) + M_d \ \lambda_{TD} = M_f \left(H_{Tboil} - H_{Tf}\right) + M_d \ \lambda_{Tv} \tag{6}$$

The energy balances on the two pre-heaters are given by:

$$Q_{HEX1} = M_b \left(H_{boil} - H_{b\_disch}\right) \tag{7}$$

$$Q_{HEX1} = M_{f1} \left(HTF_{out\_HEX1} - H_0\right) \tag{8}$$

The mass balance on the seawater spliter is:

$$M_f = M_{f1} + M_{f2} \tag{9}$$

$$Q_{HEX2} = M_d \left(H_{Td} - H_{Td\_prod}\right) \tag{10}$$

$$Q_{HEX2} = M_{f2} \left(H_{TF\_out\_HX2} - T_{SW}\right) \tag{11}$$

$$M_f \ H_{TF} = M_{f1} \ H_{TF\_out\_HX1} + M_{f2} \ H_{TF\_out\_HX2} \tag{12}$$

The temperature differences on hot/cold sides of pre-heaters are computed as follows:

$$\Delta t1_{HX1} = T_{boil} - TF\_out\_HEX1 \tag{13}$$

$$\Delta t2_{HX1} = Tb_{disch} - T_{SW} \tag{14}$$

$$\Delta t1_{HX2} = Td - TF\_out\_HEX2 \tag{15}$$

$$\Delta t2_{HEX2} = T_{d\_prod} - T_{SW} \tag{16}$$

The logarithmic mean temperature differences to compute the heat transfer area of pre-heaters (LMTD) are given by:

$$LMTD_j = \frac{\Delta t1_j - \Delta t2_j}{\ln \dfrac{\Delta t1_j}{\Delta t2_j}} \qquad j = HEX1, HEX2 \tag{17, 18}$$

Then, the heat transfer areas are given by:

$$Q_j = U_j \ A_j \ LMTD_j \qquad j = HEX1, HEX2 \tag{19, 20}$$

On the other hand, the evaporator area is calculated by:

$$A_{evap} = \frac{M_d \ \lambda_{TD}}{U_{evap} \ (T_d - T_{boil})} + \frac{M_d \ Cp_v (T_{rec} - T_{boil})}{U \ LMTD} \tag{21}$$

In total, the proposed mathematical model involves 41 variables and 36 constrains, most of them being non-convex.

The model was solved for global optimization by applying the ME-D (Minimization of the Error and Discard) algorithm [7] implemented in the software GAMS. The tolerance for the global optimality was set at 0.001 which represents about 0.09% of the global optimal objective value.

5. Results and discussion

In this section, the optimal solutions corresponding to three representative case studies are presented due to space limitation.

Table 1 shows the optimal solutions obtained for different fresh water demand. As is mentioned in Section 3, the goal is to determine the optimal operating conditions to minimize the electricity consumption in order to satisfy the given fresh water demand. The seawater salinity and temperature are 45000 ppm and 298 K, respectively.

Table I. Optimal values corresponding to different fresh water demands.

	Md = 1000.00 [Kg/h]	Md = 800 [Kg/h]	Md = 700.00 [Kg/h]
W_comp [kW]	58.059	44.969	39.348
Rp **	2.10 ***	2.10 ***	2.10 ***
T_d [K]	340.41	328.98	328.907
Tboil [K]	325.32	316.12	315.00 ##
T_f [K]	317.30	306.59	306.59
Mb[Kg/h]	900.000	720.00	630.000
Mf [Kg/h]	1900.00	1520.000	1330.000
LMTD_HX1 [K]	12.287	8.30	5.156
LMTD_HX2 [K]	36.935	30.54	13.940

*** Variables reaching their upper bounds
Variables reaching their lower bounds

As is expected, the obtained results show that the electricity consumption as well as the flow-rate of streams increase with the fresh water demands.

From a mathematical point of view, it should be mentioned that the same optimization problem was solved by using a local optimization algorithm (CONOPT) instead of a global optimization method. Despite that the simplicity of the proposed model, authors conclude that the non-convex constraints involved by the mathematical model such as logarithms to compute the logarithmic mean temperature differences (LMTD) and bilinear terms lead to local optimal solutions when a local optimization algorithm is used and the solutions depend strongly on the initial values.

Consequently, authors remark the advantages and the importance of employing global optimization algorithm.

In order to verify the accuracy of the proposed model, output results are compared to those reported previously by [4]. For comparison purposes, it was necessary to fix some optimization variables at the same values as in [4], in such way that the model had the functionality of a simulator. The obtained results agree satisfactorily to the reported in [4]. Then, the presented model represents adequately the process and can be used to optimize the design and operating condition, meanwhile the resolution method provides the best solution (global optimum).

6. Conclusions

A mathematical model for the optimal design of MVC desalination process was presented. The proposed model was successfully solved by using a deterministic global optimization method. Precisely, the ME-D algorithm developed by [7], within a very tight global optimality tolerance, was used to solve the model and the results agree satisfactorily with those reported by other authors. Different objective functions have been implemented and studied by the model and methodology proposed.

As future work, the effect of the non-condensable gases on the process, the velocity of steam inside the evaporator and dimensions, among others, will be included into the model in order to get more realistic process designs.

7. Acknowledgements

References

[1] A. Brooke, , D. Kendrick, A. Meeraus, A. A. 1992, GAMS- A User's Guide (Release 2.25). The Scientific Press. San Francisco, CA,.

[2] A. Drud, 1996, CONOPT 2.041: A system for large scale nonlinear optimization. Reference manual for CONOPT Subroutine Library, 69p, ARKI consulting and development A/S, Bagsvaerd; Denmark.

[3] Ettouney Hisham. 2006, Design of single effect Mechanical Vapor Compression. Desalination J. (190) 1-15.

[4] Helal A.M., Al-Malek S.A., 2006, Design of a solar-assited mechanical vapor compression (MVC) desalination unit for remote areas in the UAE, Desalination J. (197) 273-300.

[5] Aly Narmine H., El-Fiqi Adel, 2003, Mechanical vapor compression deslalination systems. A case study. Desalination J. (158) 143 – 150.

[6] S. Mussati, N. Scenna, E. Tarifa, S. Franco, J. Hernandez, 2009, Optimization of the mechanical vapor compression (MVC) desalination process by using mathematical programming, Desalination and Water Treatment, (5), 124-131.

[7] M. G. Marcovecchio, "Modelado de procesos y Métodos de optimización global aplicados a la síntesis de procesos de desalinización", Tesis doctoral, Facultad de Ingeniería y Ciencias Hídricas. Universidad Nacional del Litoral, Santa Fe, 2007.

Molecular Signature Descriptors for Integrated Flowsheet and Molecular Design

Nishanth G. Chemmangattuvalappil, Charles C. Solvason,

Susilpa Bommareddy, Mario R. Eden

Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA

Abstract

The reverse problem formulation is a technique for solution of integrated process and product design problems from a properties perspective. In this work, an algorithm is introduced for reverse problem formulations using property operators based on molecular signature descriptors. A general framework has been developed for the integration of flowsheet design techniques with the solution of combined process and molecular design problems.

Keywords: Reverse problem formulation, Molecular signatures, Flowsheet design

1. Theoretical background

1.1. Reverse problem formulation (RPF) technique and property operators

Reverse problem formulation is a technique used to reduce the complexity of integrated process and product design problems (Eden *et al*., 2004). In RPF, the integrated problem is broken down into two reverse problems. The first reverse problem identifies the property targets to achieve optimum process performance and the second reverse problem generates the molecular structures that meet the property targets identified in the first problem. To track the properties, property operators are used which are functions of the original properties tailored to obey linear mixing rules. The normalized property operator, Ω_{js} is obtained by dividing it by a reference operator value (Shelley and El-Halwagi, 2000). If $\psi_j\,(P_{js})$ is the property operator of the jth property P_{js} of stream s, x_s is the fractional contribution, and N_s is the number of streams then:

$$\Omega_{js} = \sum_{s=1}^{N_S} x_s \frac{\psi_j\left(P_{js}\right)}{\psi_j^{ref}} \tag{1}$$

1.2. Molecular signatures

Molecular signature is a descriptor used for representing the atoms in a molecule using extended valancies of atoms to a pre-defined height (Visco *et al.,* 2002). If G is a molecular graph and x is an atom of the molecule, the atomic signature of height h is a canonical representation of the sub graph of G containing all atoms that are at a distance h from x. Faulon *et al.* (2003a) identified the relationship between topological indices (TIs) and signatures, which is given as Eq. (1). In Eq. (1), k is a constant, $^h\alpha_G$ is the vector of occurrences of atomic signatures of height h. TI (root ($^h\Sigma$)) is the vector of TI values calculated for each root of atomic signature. TI (root ($^h\Sigma$)) can be estimated by summing up the TI values of all the atomic signatures that constitute the molecule.

$$TI(G) = k \cdot {}^h\alpha_G \cdot TI\left(root\left({}^h\Sigma\right)\right) \tag{2}$$

1.3. Flowsheet property model

A flowsheet property model can quantify the efficiency of different processing routes from raw materials to products. In a recently published work, a flowsheet property model has been proposed to estimate the energy consumption of a unit operation (d'Anterroches and Gani, 2005). If NG is the total number of process groups, d_{ij}^{k} is the maximum driving force of process group k, a_k is the contribution of process group k, A is a constant, which is different for different unit operations, p_k is a topology factor, nt is the number of separation tasks that should be performed before task k and $\overline{D_i}$ is the maximum driving force of task i, the energy index E can be defined as:

$$E = \sum_{k=1}^{n=NG} \left(\frac{1+p_k}{d_{ij}^{k}} \times a_k + A \right) \tag{3}$$

$$p_k = \sum_{i=1}^{nt} \overline{D_i} \tag{4}$$

2. Integrated flowsheet and molecular design

In RPF, property targets of the input molecules into a process will be identified corresponding to the optimal performance. Existing algorithms for the identification of molecular structures corresponding to the target properties are based on group contribution (GC) models. However, the properties are limited for which GC models are available. In this work, an algorithm has been developed based on molecular signature descriptors for the molecular design part. Since many existing QSPR expressions can be reformulated in terms of molecular signatures, algorithms based on signatures will be able to track a wide variety of property targets. The flowsheet property corresponding to the designed molecule will be estimated to reject unsuitable candidates.

2.1. General problem statement

Design the molecules with the best dominant property which also satisfy the set of property constraints identified during the process design. Identify the molecules that result in low energy consumption for the process.

2.2. Stepwise procedure
- The property targets for the input molecules to provide the optimum process performance will be calculated using Eq. (1).
- Identify the QSAR/QSPR/GC models that predict the properties corresponding to the optimum performance.
- Identify heights of molecular signatures corresponding to the TIs used in QSPR
- Based on the structural constraints and transformation techniques explained in sections 2.3-2.5, identify the signatures and generate candidate molecules from signatures based on the algorithm by Faulon *et al.* (2003b).
- Calculate the energy index of the unit operations involving the shortlisted candidates from the previous step. The final pool of molecules for rigorous simulation will be formed from the candidates with low energy index.

2.3. Problem formulation

If θ is the property function of property P, the property operator corresponding to P is estimated as follows:

$$\theta = f(TI) \tag{5}$$

$$TI = \sum_{i=1}^{N} {}^{h}\alpha_i \cdot TI(root({}^{h}\Sigma)) = \sum_{i=1}^{N} {}^{h}\alpha_i L_i \tag{6}$$

$$\Omega(P) = \sum_{i=1}^{N} x_i L_i \tag{7}$$

The dominant property, which is expressed in terms of the occurrences of atomic signatures, can be optimized subject to the property constraints. If Ω_j is the property operator corresponding to the dominant property and Ω_{ij} is the normalized property operator of molecule i, an optimization problem can be formulated as follows:

$$Max/Min \ \Omega_j \tag{8}$$

$$\Omega_j^{min} \leq \Omega_j \leq \Omega_j^{max} \tag{9}$$

To ensure that the collection of signatures will form a molecule without any free bonds, the *handshaking lemma* from graph theory is used (Trinajstic, 1992). If $D(i)$ is the number of degrees of each vertex i and M is the total number of edges, then:

$$\sum_{i=1}^{N} D(i) = 2M \tag{10}$$

If N_i is the total number of signatures, the number of rings in the final structure is N_r, N_{Di}, N_{Mi} and N_{Ti} are the signatures with one double bond, two double bonds and one triple bond in the parent level, respectively, then the following expression can be used to describe the relationship between the number of signatures and edges:

$$M = \left[\sum_{i=1}^{N} x_i + \frac{1}{2} \sum_{i=0}^{N_{Di}} x_i + \sum_{i=0}^{N_{Mi}} x_i + \sum_{i=1}^{N_{Ti}} x_i \right] - 1 + N_r \tag{11}$$

In order to differentiate between different types of atoms, vertex coloring has been used. The coloring function has to be selected based on the types of atoms, nature of the final molecule and type of bonds. The coloring should start from the root atom to all atoms up to level *h-1*. In a complete molecule, each bond is shared by two atoms, therefore the colors involved in the edge that joins the two vertices must be the same for both vertices. However, the color order will be different for both vertices since the reading of the color has to start from the root. In addition, the bond type should be consistent in the joined signatures. For instance, if there is an edge between the vertices A and B, there will be an A→B color sequence as well as a B→A color sequence. This consistency should be followed by all such color sequences. If $(l_i \rightarrow l_j)_h$ is one coloring sequence at a level *h* then:

$$\sum (l_i \rightarrow l_j)_h = \sum (l_j \rightarrow l_i)_h \tag{12}$$

Equation (12) has to be satisfied even for those edges that join the same vertex colors (for instance, a B-B bond). In other words, the number of such signatures should be an even number.

In some signatures, there will be more than one child with a specific color (say *m*) for a single parent. In such cases, it must be ensured that the number of complementary signatures with the previous parent in the child level is more than *m*. If n_i is the number of child vertices with a higher degree than the parent vertex, *i* and *j* represents the child and parent colors:

$$\sum x_i n_i \le \sum x_j \tag{13}$$

The dominant property function can be optimized subject to the problem constraints. Integer cuts can be used to form other feasible solutions. An algorithm is available (Faulon *et al.,* 2003b) to enumerate candidate molecules from the identified signatures.

2.4. Expression of group contribution (GC) models with signatures

In GC, the property function of a compound is estimated as the sum of property contributions of all the molecular groups present in the molecule. Molecular signatures of sufficient height can be used to re-write GC expressions. This transformation will allow us to solve the property models based on TIs and GC models simultaneously. Vertex coloring is used to classify the signatures to different molecular groups. The root atom on every signature can be colored with three numbers/letters - the first color is the type of atom, the second color is the number of neighboring atoms and third is the bond type. The effects of higher order molecular groups (Marrero and Gani, 2001), which are combinations of molecular groups can also be successfully estimated by this approach.

2.5. Property models with different signature heights

In a molecular design problem with multiple property constraints, different TIs may be describing different properties of interest. It is also possible to have one QSAR/QSPR containing different TIs. If the signature heights of the TIs are different, the signatures corresponding to the largest height have to be enumerated first and the number of signatures of smaller height can be obtained from the higher signatures. Consider the situation where QSPR is based on signatures of heights one and two respectively. The property operator for property *Y* can be written in terms of signatures as follows:

$$f(Y) = \sum_i L_i^1 \alpha_i + \sum_j L_j^2 \alpha_j \tag{14}$$

If the height of the largest signature of interest is *h* and the height of the lower signature is *h-m,* then the total number of each *h-m* level signatures can be obtained by adding the *h* level signatures under the same color at level *h-m*. For example, assume the signatures of interest are N1(C), N2(CC) and N3(CCC) which are signatures with height one. Now, signatures of height two with root vertex N can be divided into three sets: S_1, S_2 and S_3. The elements in these sets are signatures of height two with N vertex and vertex color 1, 2 and 3 respectively. The occurrences of signature N1(C) can be obtained as follows:

$$N(C): \; {}^1\alpha_{N(C)} = \sum_{S_1} {}^2\alpha_i \tag{15}$$

The number of occurrences of other signatures of height one is similarly estimated.

3. Case Study: Acid gas removal

A gas treatment process uses MDEA ($OH(CH_2)_4N(CH_3)OH$) to remove acid gases from an alkane rich feed. Two recycled process streams (S_1 and S_2), which mainly consist of

2,5-dimethyl-hexane, are also fed to the process and will be separated from the amine after the acid gas absorption. The objective of this case study is to identify a solvent that will reduce the consumption of MDEA by 50% and will utilize all available recycle streams. Apart from the process constraints, the designed molecule should have minimum soil sorption coefficient (log K_{oc}) to avoid accumulation of the escaping solvent in one place and a high toxic limit concentration (TLC) value. The energy index for the separation of the alkane from the final molecule must be low so that the alkane can be easily separated after the absorption.

Table 1. Property data for acid gas removal problem

Property	Feed Properties		MDEA Properties	Lower Bound for Sink	Upper Bound for Sink
	S_1	S_2			
VP (mm of Hg)	63.2	43.1	0.26	-	10
H_v (kJ/mol)	39	41	89	60	90
V_m (cm^3/mol)	178	189	114	110	140
Flow (kmol/h)	50	70	180	300	350

The property operators corresponding to the three properties of interest are:

$$\psi_{VP} = \sum_{s=1}^{NS} x_s VP^{1.44} \qquad \psi_{V_m} = \sum_{s=1}^{NS} x_s V_m \qquad \psi_{H_v} = \sum_{s=1}^{NS} x_s H_v \tag{16}$$

The property targets for molecular design have been obtained by solving Eq. (17). The limit for TLC is kept as 10 ppm. The following property models have been used:

Table 2. Property models

Property	Property Model	Reference
log K_{oc}	$\log(K_{oc}) = 0.53\left(^1\chi\right) - 1.25\left(\Delta^1\chi^v\right)$ $- 0.72\left(\Delta^0\chi^v\right) + 0.66$	Bahnick and Doucette, 1988
H_v	$\Delta H_v = h_{v0} + \sum n_i h_{vi}$	Marrero and Gani, 2001
V_m	$V_m = 33.52\varepsilon + 30.67$	Dian *et al.*, 1998
TLC	$\log(TLC) = 4.204 - 1.385\left(^2\chi^v\right)$	Koch, 1982

Here, $^n\chi$ is the connectivity index of order n and ε is the edge connectivity index. There is no reliable group contribution or QSPR model for vapor pressure. So, an empirical relationship using boiling point is employed (Sinha and Achenie, 2001):

$$\log VP = 5.58 - 2.7\left(\frac{T_{bp}}{T}\right)^{1.7} \tag{17}$$

The GC model (Marrero and Gani, 2001) is used to obtain the boiling point corresponding to the target VP. Lower limit for molecular design is calculated as 425K.

.

The molecular property operators corresponding to the target properties have been obtained from Eqs. (5-7). The signature height required to re-write all the topological indices involved is three. They along with the targets are shown in table 3.
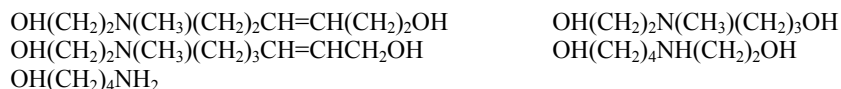
Table 3. Property operators for acid gas removal problem

Property	Ω_j	Lower Bound	Upper Bound
VP	$\exp(T_b/t_{b0})$	6.75	-
H_v	$H_v - h_{v0}$	45.3	145.3
V_m	$(V_m - 30.67)/33.52$	0.28	5.75
TLC	$(4.204 - \log(TLC))/1.385$	2.21	-
$\log(K_{oc})$	$\log(K_{oc}) - 0.66$	Minimum	

Based on the structures of common acid gas absorbents, the following molecular groups have been used for design:

OH   CH$_3$ CH$_2$ CH$_3$N   CH$_2$NH$_2$ CH$_2$NH   NHCH   CH=CH

The signatures of height 3 have been formulated corresponding to these molecular groups and structural constraints have been formulated from Eqs. (11-13). An optimization problem has been set up and solved for the minimum value of soil sorption coefficient. The best five candidates are:

OH(CH$_2$)$_2$N(CH$_3$)(CH$_2$)$_2$CH=CH(CH$_2$)$_2$OH   OH(CH$_2$)$_2$N(CH$_3$)(CH$_2$)$_3$OH
OH(CH$_2$)$_2$N(CH$_3$)(CH$_2$)$_3$CH=CHCH$_2$OH   OH(CH$_2$)$_4$NH(CH$_2$)$_2$OH
OH(CH$_2$)$_4$NH$_2$

In the final step, the energy index value for the separation of 2,5-dimethyl-hexane from the mixture of MDEA and these five candidates is to be compared using the maximum driving force approach. However, due to the lack of VLE data available for these mixtures, this step is omitted in this paper.

4. Conclusions

A general framework has been proposed for integrating flowsheet design techniques with reverse problem formulations. The newly introduced concept of molecular signature descriptors has been used as tool for the molecular design part of the general reverse problem formulation framework. The ability of signatures to represent a wide variety of topological indices and group contribution models allowed the application of RPF for a variety of property targets.

References

D.A. Bahnick, W.J. Doucette, (1988), *Chemosphere*, 17(9), 1703-1715
L. d'Anterroches, R. Gani, (2005), *Fluid Phase Equilibria*, 228/229, 141-146
J. Dian, L. Jin, L. Wang, (1998), *Progress in Natural Science*, 8(6), 754-761
M.R. Eden, S.B. Jorgensen, R. Gani, M.M. El-Halwagi, (2004), *Chem. Eng. & Proc.* 43, 595-608
J.L. Faulon, D. P. Visco Jr., R. S. Pophale, (2003a),*J. Chem. Inf. Compt. Sci.* 43, 707-720
J.L. Faulon, C.J. Churchwell, D.P. Visco Jr., (2003a), *J.Chem. Inf. Compt. Sci.*, 43, 721-734
R. Koch, (1982), *Chemosphere*, 11(9), 925-931
J. Marrero, R. Gani, (2001), *Fluid Phase Equilibria*, 183/184, 183-208.
M.D. Shelley, M.M. El-Halwagi, (2000), *Computers & Chemical Engineering*, 24
M. Sinha, L.E.K. Achenie, (2001), *Adv. Env. Res.*, 5, 239-249.
N. Trinajstic, Chemical Graph Theory, 2nd Edition, CRC press, Boca Raton, FL, 1992
D.P. Visco Jr., R.S. Pophale, M.D. Rintoul, J.L. Faulon, (2002), *Mol. Graph. Mod.* 20 429-438

Micro reactor design for Distributed Fuel Generation

"Pádraig Mac Suibhne,a Toshko Zhelev,a Erzeng Xue, Reg Mannb†"

[a] *University of Limerick, Limerick, Ireland*
[b]*University of Manchester, Manchester, England*

Abstract

This paper presents a new concept of waste CO_2 utilisation for fuel re-synthesis. It extends this challenging approach to GHG emission mitigation into distributed fuel generation concept based on micro-reactor technology. Further the paper outlines the steps taken to begin the fabrication of a series of micro reactors intended for the reverse water gas shift and Fischer-Tropsch synthesis. The design methodology for the micro channels within the reactors is also outlined. The designs theorized for the micro channels were simulated using two approaches; manual, macro-scale calculations and CFD micro-fluidic simulation. The results from these simulations allow a comparison to be made between the two methods. They also provide a solid foundation upon which to base the chosen micro reactor designs. The CFD simulation results provide insight into the expected gas flow rates and hence, the required micro channel arrangement. The micro reactors to be fabricated are described and were chosen based upon the results presented.

Keywords: Micro reactor, r-WGS, Fischer-Tropsch, micro channel, gas flow.

1. Introduction

This project has developed in response to the desire to tackle the constant build-up of green-house gases, particularly CO_2. With some exceptions, it is clear that the CO_2 storage is not a sustainable alternative. Our understanding is that the most promising approach to CO2 mitigation would be its conversion in useful products. The best solution would be to utilise the valuable carbon content by closing the loop of Hydrocarbon fuel to Carbon Dioxide back to synthetic fuel. We plan to take advantage of existing reaction pathways and to develop this promising concept for fuel re-synthesis identifying obstacles and challenges and finding engineering solutions to tackle them.

2. Background

The existing reaction pathways to generate fuel from CO2 are the Reverse Water Gas Shift (r-WGS) and the Fischer-Tropsch Synthesis (FTS). The r-WGS reaction has been known to chemistry since the 1800's. It is a catalytic reversible reaction and its progression towards the generation of CO raises technical challenges. CO is the main substance of so-called syngas, which is the required starting point for hydro-carbon synthesis.

The Fischer-Tropsch reaction converts syngas into hydrocarbons from C_1 to C_{100+} [i]. The fuels derived from Fischer-Tropsch products are of high quality and are more environmentally sound because of their lower aromatic and zero sulphur content[ii]. The middle range carbon number fraction has a high cetane number, better combustion and

lower emissions. FTS is currently carried out profitably at the industrial scale by many petro- chemical companies, e.g. Exxon-Mobil, Shell, Sasol and Syntroleum[iii].

The challenges to be met when performing CO_2 conversion to hydrocarbons following r-WGS and FTS pathway can be summarised as follows: 1. Technical concept; 2. Feasible raw materials supply; 3. Catalyst choice; 4. Mathematical modelling; 5. Reactors design; 6. System's synthesis; 7.Economic evaluation; 8. Prototyping. This paper focuses on the concept development, reactor design and corresponding portion of mathematical modelling.

3. The Concept

The novel concept in our case is the intent to perform sequentially r-WGS and FTS processes in micro-reactors. There are two main expected advantages of this novel concept: (a) Development of fundamentals for possible distributed fuel generation, and (b) Utilisation of micro-processing advantages, i.e. better heat and mass transfer, higher reaction rates, more controllable product quality, lower temperatures and pressures, better resources management, reduced missions and cost, enhanced safety and reduced environmental impact. There are reports of carrying out successfully, but separately r-WGS and FTS in micro-reactors[iv, v]. The new concept proposed in this paper will provide alternative of fossil fuels and save the design and manufacturing infrastructure of internal combustion engines/vehicles beyond the fossil fuel completion era.

The fuel re-synthesis requires pure H_2 and CO_2. Intense research is ongoing presently focusing on CO_2 capturing and storage[vi]. CO_2 re-capturing from the exhaust flue gas within power plants is still in the early stages, with some currently in operation[vii]. It has also been proposed that Oxi-fueled power generation (A power plant using pure O2) could be a better approach for the future[viii].
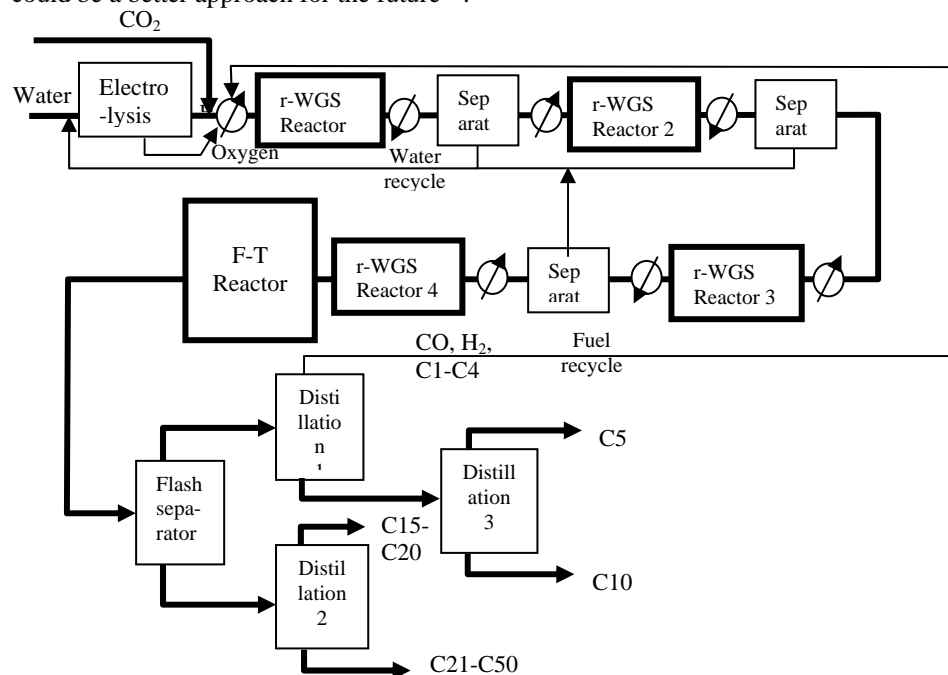

Figure 1: Flowsheet of the proposed fuel re-synthesis process

Micro reactor design for Distributed Fuel Generation

Since the exhaust from such a power plant would consist of highly pure CO_2 and H_2O this proposed method of power generation would be an ideal feedstock source for the r-WGS. The practice of CO_2 absorption from atmospheric air [vi] has also been thoroughly examined using various approaches. Such a source of CO_2 would enable a distributed fuel generation approach, as all of the necessary raw materials could be present in any convenient location.

The other and necessary raw material for this process is of course hydrogen. Through the electrolysis of water hydrogen can be conveniently sourced. Electrolysis as a source of pure hydrogen would drastically increase the energy demand of any fuel re-synthesis system. Therefore other feedstock options that could provide the necessary building blocks are also considered. Syngas and hydrogen derived form bio-waste gasification products is one of these and FTS has already been successfully carried out using such feedstock[ix][x]. Another option is to take advantage of wind energy or other forms of renewable energy. With costs of wind-generators rapidly falling down, this can be a source of sustainable hydrogen generation. The general flow-sheet of the proposed micro-reactor system is given in Figure 1.

The performance of the r-WGS and FTS within micro channels is expected to be very effective. For this to be achieved we need to reach the threshold of reactor sizes, where the beneficial micro-effects start to take place. Therefore we need to specify the sizes, lengths and topology of reactor channels.

4. Micro-channel Design

There has been much attention focused on micro reactor technology[xi]. Certain studies even specifically look at the reactions that we intend to perform[v]. The term "micro reactor" refers to a chemical reactor whose characteristic internal dimensions of the internal structures are below the 1mm scale. The well known advantages of micro reactors have been summarised by Fichtner et.al. [xii]. Improved heat and mass transfer is noticed as well as the optimisation of slower reactions getting closer to thermodynamic equilibrium[xiii].

The modelling and the performance of fuel re-synthesis and the application of the reactions to micro reactor technology is the ultimate goal of this study. In the design of the micro-channel reactors as much care as possible needed to be taken to ensure their performance will be as expected. We must ensure that the pressure loss, flow rate and residence times are acceptable. By accurately controlling and manipulating these reaction conditions we intend to determine the optimum process conditions for the system. Apart from negotiating the link between the individual system's components, i.e. r-WGS, FTS, Micro-reactor Technology (Design and Fabrication) and simulation of the proposed process, the project will involve ensuring that the reactions perform as required. To properly evaluate the potential of such a micro reactor system, a variety of unique micro reactor designs are to be fabricated and experimentally evaluated.

It is known that traditional fluid flow equations begin to become invalid at the micro scale.[xiv] The gas flow behaves differently and care needed to be taken when attempting to model the micro channel flow. The scale of the micro channels was chosen to take advantage of the benefits which come at the micro scale, i.e. improved heat transfer & increased surface to volume ratio. Although as the size of the channels gets smaller, accurate prediction of the flow, through conventional means, becomes difficult.

Manual calculations were carried out in order to predict the performance of various micro-channel designs. For the flow rate be sufficient to allow analysis through the process of Gas Chromatography, a gas flow rate of at least 10[μL/min] (at reactor

temperature and pressure) was desired. Naturally with the inevitable energy losses in the fluid flow through the micro reactor, this flow rate, in some cases is not achievable through the use of a single micro-channel.

While conducting these calculations certain assumptions needed to be made about the nature of the fluid and the flow for calculation purposes. The flow is assumed to be incompressible. This is an incorrect assumption which would cause an overestimation of the flow rate. A second assumption that the flow is in the no-slip regime (velocity = 0 at the channel wall) was made. This assumption has been shown in previous studied to begin to become invalid at this micro scale[xiv]. Choosing the correct flow regime has a significant effect of the accuracy of the calculations. The flow regime can be determined by the Knudsen number (Kn). The flow within the micro channels is within the slip-flow regime. (0.001<Kn<0.1). A no-slip assumption should result in an underestimation of the flow rate. This result of this assumption opposes the effect of the compressible flow assumption. In the manual calculations various energy losses in the channel flow were accounted for and calculated using standard channel flow equations. These were derived from the Navier-Stokes Equations for fluid flow. The contributing factors to the pressure losses considered were; Entrance Losses, Friction from flow in each part of micro reactor & Losses in Bends. In an effort to investigate the reactions within narrow micro channels, we decided upon the range of micro channel sizes to evaluate, (6, 16, 60 &160μm).

Due to the incorrect assumptions made in the initial flow rate calculations, it was necessary to carry out CFD simulations on the micro channel flows to verify or correct the manual calculations. A thorough CFD simulation of the gas flow was also conducted using Conventorware micro-fluidic simulation software. Conventorware is a powerful and versatile software package capable of analysing gaseous flows at the micro scale. With these fluid flow simulations, together with the manual calculations the micro reactor designs can be finalised. For both of the flow prediction methods used the necessary number of channels in parallel to maintain the desired flow rate was found. The number of bends and micro channel size was taken into account as well as the residence time expected within the reactors being calculated. The Driving Pressure was limited to 10 BAR as this is the limit of performance of the silicon micro reactor. A flow rate of 10 [μL/min] (@ reactor temperature and pressure) was deemed sufficient as the Gas Chromatography to be performed on the reactor product requires a sample of gas of 5 [μL] in volume. Four micro channel widths were considered; 6μm, 16μm, 60μm & 160μm. The depth of each of the micro channels is fixed at 12 [μm]. This shallow depth of the micro channels is to encourage contact with the catalyst at the channel wall. The average flow rate expected was calculated for each of these channel sizes. The length of the channel was varied by incorporating 180^0 bends, meanders or zigzags into the designs.

5. Results

The results of both the simulation approaches are presented below. They clearly illustrate that there will be a need for several parallel micro channels within each reactor. This is necessary to ensure an adequate flow rate.

Manual Calculations

# Channels needed		P=10 Bar, T= 600 K.	
Q=0.01ml/min			
Width of Channels [μm]	0 Bends	2 Bends	4 Bends
6[μm]	**50.3**	**150.2**	**250.3**
16[μm]	**6.5**	**19.4**	**32.5**

60[μm]	0.6	1.9	3.2
160[um]	0.4	0.7	1.1

$e_f = \dfrac{4fLV^2}{2d}$ Energy loss in flow.	$e_f = \dfrac{K_f V^2}{2}$ Energy loss in channel bend
$Re = \dfrac{\rho U D_h}{\mu}$. Reynolds Number.	$\Delta P = \rho . e_f$ Pressure Drop.
$K_f = 2$ Bend loss coefficient.	$f . Re = 16$ Poiseuille number.

The results from Conventorware simulations

# Channels Needed		P=10 Bar, T= 600 K	
Q=0.01ml/min			
Width of Channels [μm]	0 Bends	2 Bends	4 Bends
6[μm]	17.9	53.8	89.6
16[μm]	2.1	6.4	10.7
60[μm]	0.4	1.1	1.8
160[μm]	0.2	0.7	1.2
Predicted residence times within micro-channels			
P=10 Bar, T= 600 K. Residence times predicted [s]			
Width of Channels [μm]	0 Bends	2 Bends	4 Bends
6[μm]	0.12	1.05	2.90
16[μm]	0.04	0.33	0.92
60[μm]	0.02	0.20	0.57
160[μm]	0.04	0.38	1.04

The micro fluidic simulations allowed a clear visualisation of the nature of the flow within the micro channels. This Conventorware micro fluidic simulation software produced results which are far more reliable than the initial manual calculations. The assumptions made in the manual calculations (for calculation purposes), were not needed for the Conventorware simulations, i.e. no slip & incompressible flow. This comparable simulation data allowed the reactor designs to be finalised and the fabrication to begin. The Conventorware simulations visually allowed more insight than the original manual calculations.

6. Discussion/Conclusion

The micro-fluidic CFD simulations of the channel flow are a far more solid foundation upon which to base the micro channel designs, particularly for the smaller micro-channels. The two sets of simulation data agree with each other more so when considering the larger micro channels. The manual calculations predict a much lower flow rate than the CFD simulations in the narrower micro channels. This discrepancy makes sense as the manual calculations assume no slip at the channel walls. This "no-slip" assumption becomes invalid as the characteristic dimensions of the flow, (i.e. channel diameter) tend to the lower end of the micro scale. Knowing the effect of the channel geometry on the flow rate allows us to vary the dimension of micro channels (while maintaining an adequate flow rate) and hence vary the residence time within the micro reactors. This allows us to study the affect of residence time on the reactor performance in terms of yield and conversion rate.

It is evident that the flow rate of the gas within the reactors is proportional to the length of the micro channel. The CFD simulations show that the effect of the micro channel bends is negligible and the increase in the pressure loss is almost entirely due to the resulting increase in channel length. The results are promising and illustrate that an acceptable gas flow can be maintained through the introduction of as many micro channels in parallel as possible. Micro second residence times may be achieved; this has been reported to be a necessary operating condition for a r-WGS micro-reactor [ii].

The four micro channel widths will be fabricated (6,16,60 &160um) with the predicted number of parallel micro channels necessary to maintain an adequate flow. The depth of all the channels will be 12[um]. Each channel size will have 3 length variations. This will total 12 unique micro reactor designs. Each unique design will be fabricated 3 times. All together 36 micro reactors will be made.

7. Future work

The future work for this study will involve experimental work in the deposition of catalyst within the reactor micro channels and investigating the performance of each of the reactions. A process optimisation of the fuel re-synthesis system is planned with the aid of a further, more detailed simulation of the reactor's performance. A final feasibility analysis for the proposed process will also need to be carried out. Such a critical review of the fuel re-synthesis system along with the experimental data should provide a clear picture of the viability of such a method of fuel production. We expect that the scale of our proposed micro reactors will improve the yield and selectivity of the chemical reactions taking place. Working with devices at this micro-scale should prove to be more difficult for certain aspects of the study (e.g. Catalyst deposition), although the potential improvement to the performance could be significant.

8. Acknowledgements

References

[i] C. Maretto, R. Krishna, Catalysis Today 52 (1999) 279-289.

[ii] G.P. V der Laan, A. Beenackers, Applied Catalysis A: General 193 (2000) p39-53.

[iii] A.P. Steynberg, M.E. Dry, B.H. Davis, B.B. Breman, S. André, D. Mark, Studies in Surface Science and Catalysis, Elsevier, 2004, pp.64-195.

[iv] J.D.Holladay, K.P.Brooks, R.Wegeng, J.Hu, J. Sanders, S.Baird, Catalysis Today 120 (2007) pp. 35-44.

[v] S. Zhao, V.S. Nagineni, N.V. Seetala, D. Kuila, Industrial & Engineering Chemistry Research 47 (2008) pp.1684-1688.

[vi] V.Nikulshina, C.Gebald, A.Steinfeld, Chem. Engineering J. 146 (2009) 244-248.

[vii] S. Kishimoto, T. Hirata, M. Iijima, T. Ohishi, K. Higaki, R. Mitchell, Energy Procedia 1 (2009) pp.1091-1098.

[viii] K. Jordal, M. Anheden, J. Yan, L. Strömberg, E.S. Rubin, D.W. Keith, C.F. Gilboy, M. Wilson, T. Morris, J. Gale, K. Thambimuthu, Greenhouse Gas Control Technologies 7, Elsevier Science Ltd, Oxford, 2005, pp. 201-209.

[ix] K.-W. Jun, H.-S. Roh, K.-S. Kim, J.-S. Ryu, K.-W. Lee, Applied Catalysis A: General 259 (2004) pp.221-226.

[x] S. Srinivas, R.K. Malik, S.M. Mahajani, Energy for Sustainable Development 11 (2007) 66-71.

[xi] K.F. Jensen, Chemical Engineering Science 56 (2001) pp.293-303.

[xii] T.L. M. Fichtner, R. Wunsch, K. Schubert, Scientific Reports, F. Forschungszentrum Karlsruhe, Karlsruhe, 1998, p. 121.

[xiii] D.L.K. W.E. TeGrotenhuis, K.P. Brooks, B.J. Golladay, R.S. Wegeng, 6th Int. Conf. on Microreaction Technology, 2002, New York, AIChE (2002) p. 18.

[xiv] Gad-el-Hak, M. "The Fluid Mechanics of Microdevices," J. of Fluids Engng, Vol. 121 (1999) pp. 5-33.

Integration of Design and Control for a large scale flowsheet

Seon B. Kim and Andreas A. Linninger*

Laboratory for Product and Process Design
Departments of Chemical and Bioengineering, University of Illinois at Chicago
Chicago, IL 60607, USA, linninge@uic.edu

Abstract

In order to design a process that operates close to tighter boundaries safely, much attention has been devoted to the integration of design and control, in which the design decisions, dynamics, and controlled performance are considered simultaneously in optimal fashion. However, rigorous methods for solving design and control simultaneously meet challenging mathematical formulations which become computationally intractable. In an earlier paper of our group, a new mathematical formulation to reduce the combinatorial complexity of integrating design and control was introduced. We showed that substantial reduction in the problem size can be achieved by embedded control for specific process designs. In this paper, we extend the embedded control to the plantwide process. This case will demonstrate the current capabilities of the methodology with integrated design and control under uncertainty.

Keywords: Integrated design and control, embedded optimization, uncertainty, dynamic feasibility

1. Introduction

Rigorous incorporation of the process dynamics is important for operational safety and efficiency. By considering dynamic controllability and operability for uncertain condition in early design stage, we can achieve better overall performance of the system than classical approaches which only consider steady state, thus dynamic constraint violations cannot be detected. Integration of process design and control pursues minimizing cost, while guaranteeing smooth process operation in spite of dynamic disturbances and process uncertainty. Integration of design and control has received attention in the scientific community for the last 30 years. And several remarkable methodologies have been developed. Excellent reviews of integrated design and control methodologies can be found elsewhere (Sakizlis et al., 2004; Seferlis & Georgiadis, 2004).

One main difficulty of integration of design and control for large-scale processes stems from the large computational time requirement which makes it impossible to apply current optimization algorithms. Recently, we proposed a new method entitled embedded control optimization (Malcolm et al., 2007).

In this paper, the embedded control optimization methodology is enhanced to incorporate moving horizon estimator for achieving better dynamic process performance compared to our earlier fixed horizon state estimation. Also a plantwide scale of process-isomerization flowsheet is solved to provide a realistic example for the capability of this methodology.

This paper is organized as follows. Section 2 briefly reviews methodology of embedded control optimization. Section 3 demonstrates the application of embedded control

optimization for integrated design and control of an isomerization process flowsheet. This paper closes with conclusions.

2. Methodology

Mathematical problem decomposition for design under uncertainty

The conceptual problem of the integration of process design and control under uncertainty is a stochastic infinite dimensional mixed integer dynamic optimization problem. To solve these integrated design and control problems requires expensive computational time, integer decisions, and non-convex equations introduced by feedback pose an extreme challenge to existing mathematical programming techniques. To overcome the intractability of the original problem, Pistikopoulos and co-workers proposed a problem decomposition algorithm as shown in **Figure 1** (Mohideen et al., 1996). In this decomposition technique, the optimal design choices are solved stochastically in a discrete sampling space. Because the discrete sampling space may not contain all critical scenarios, a separate search for critical constraint violations needs to be performed. Accordingly the rigorous feasibility test explores whether the current design is feasible in the entire uncertain space. If a new critical scenario is identified, this critical situation is added to the discrete sample spaces. Thus, this decomposition technique is composed of three steps- *sampling step* (A), *main optimization step* (B) and *dynamic feasibility test step* (C). For this feasibility analysis, several methods can be used. (Dimitriadis & Pistikopoulos, 1995; Grossmann & Floudas, 1987; Moon et al., 2008; Swaney & Grossmann, 1985).
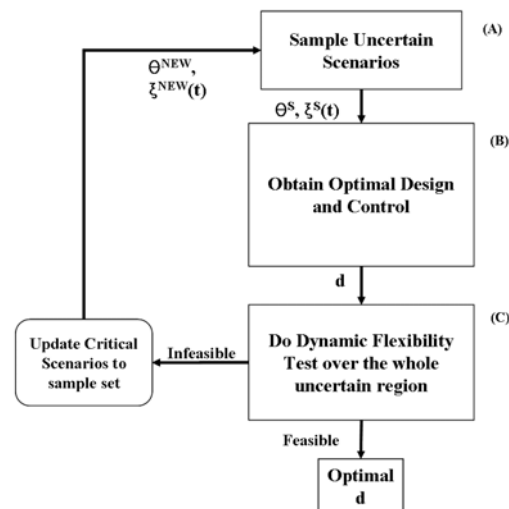
Figure 1. Decomposition algorithm for integrated design and control under uncertainty. Main optimization problem (B) is separated from flexibility test. (Mohideen et al., 1996)

Embedded control optimization

Even though the problem decomposition substantially reduces the problem size, it still remains a challenge due to combinatorial complexity of the NP-hard search space. Specifically, the control decisions such as the insertion of feedback loops, or pairing of manipulated and control variables cause a combinatorial explosion in the possible process design and control realizations. We therefore propose to separate the design

decisions from the control decisions. The master level fixes design decisions such as reactor dimensions, residence time, reactor length and diameter that govern dynamic process performance. No control decisions are made at this level. For a given design, we assess its dynamic process performance by solving the process dynamics problems rigorously. Moreover, the optimal control action is calculated with relatively ease way because it operates on linear state space models which are updated dynamically in each time step. In our algorithm, a Linear Quadratic Regulator (LQR) computes the best control action to minimize a cost function. We map the nonlinear dynamic process model to linear state space model using linear identification methods. This identification is executed in every step of the discretized time horizon. The required input-output data sets are obtained by sampling the dynamic system model with suitably chosen sampling intervals. The adaptive identification involves the solution of a least square fitting problem. For solving this problem, the sequential least squares method was used in previous paper (Malcolm et al., 2007). In this paper, we used a *Moving Horizon Estimation* (Haseltine & Rawlings, 2005). It estimates the state and parameters using a moving data window of fixed size. When new observation becomes available, new data are added to the data window and the same amount of oldest data is removed from the window. It provides a generic approach to the state and parameter estimation which can be applied both linear and nonlinear processes.

Next section, we will show effectiveness of embedded control optimization by designing isomerization process flowsheet.

3. Case study-Integrated Design and Control of isomerization process flowsheet

This section will demonstrate the effectiveness of embedded control optimization for an entire flowsheet. We show the large scale case study- for integrated design and control of an isomerization process under uncertainty (Luyben et al., 1998).

Isomerization Process description
Isomerization process converts normal butane to isobutene, as shown in Eqn. (1).

$$nC_4 \rightarrow iC_4 \qquad (1)$$

The isomerization flowsheet consists of a reactor, a heat exchanger, and two distillative separation columns as shown in Figure 2. An input feed is the mixture of nC_4 and iC_4. It also has small amount of propane (nC_3) and isopentane (iC_5). Since the input feed already has some amount of iC_4, it does not enter reactor directly. It enters Deisobutanizer column (DIB) and some of iC_4 is separated from input feed. The propane, the lightest component also comes out in the distillate stream. Because of similar volatilities of iso/normal butane, it is hard to separate. Thus, relatively higher number of tray and reflux ratio is required. The bottom stream of DIB goes into the purge column in which, most of iC_5 is purged to bottom stream. The upper stream of the purge column is vaporized and goes into reactor by passing the heat exchanger. In the reactor, some of normal butane is converted to isobutene in the vapor phase shown in (2).

$$R = kC_{nC4}, \quad k = k_o \exp\left(\frac{\lambda}{RT}\right) \qquad (2)$$

where R is reaction rate, k is temperature-dependent reaction rate constant, k_0 is pre-exponential factor, λ is activation energy , T is the reactor temperature in Rankin.
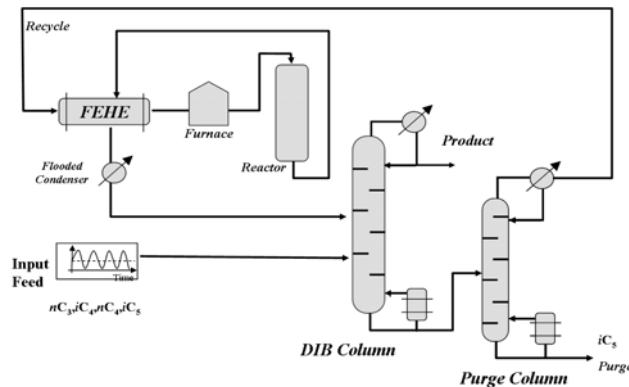
Figure 2. Isomerization process (Luyben et al., 1998). This process converts normal butane to isobutane. It consists of a reactor, a feed-effluent heat exchanger, and two distillation columns.

In the heat exchanger, heat transfer between input stream of reactor and output stream of reactor is occurred. The input feed of the reactor is heated with heat of output stream. However since this heat exchange does not heat input feed sufficiently, a furnace is used to heat input stream to the desired temperature. The effluent of reactor is fluidized in the condenser and goes into DIB. It is fed at a higher tray than the fresh feed because the concentration of iC_4 in the reactor effluent is higher than in the fresh feed.

Operational constraints and control action strategy
We set two quality constraints and one safety constraint for operating this process. These constraints are adopted from Luyben et al. (1998). For the product quality, the mole fraction of nC_4 (x_1) of final product should be less than 2%. For complete elimination of heaviest inert iC_5, the mole fraction of iC_5 (x_2) at top stream of purge column should not exceed 0.1%. For safety, the reactor pressure (x_3) should never exceed 700 psia to prevent explosion. To satisfy quality and safety constraints, we use reflux ratio (u_1) and vapor ratio of DIB column (u_2), reflux ratio of purge (u_3), and input temperature of reactor (u_4) as manipulated variables.

$$x_1 \leq 0.02 \tag{3}$$

$$x_2 \leq 0.001 \tag{4}$$

$$x_3 \leq 700 \tag{5}$$

Among these constraints, violation of (5) is unacceptable at any times under any circumstances. On the other hand, it is impossible to keep the product specifications exactly at the set point target in a real operation. Therefore, the quality constraints for bottom and product streams are soft. However, the optimal process should deviate as little as possible for this target.

Design variables
To optimize this problem we consider four main design variables -Reactor volume size (D_1), Total tray number of DIB (D_2), Total tray of purge column (D_3), Heat exchanger size (D_4). While larger values of these design variables ease the separation task, it increases the capital cost. Therefore we should consider the trade off between controllability and the capital cost.

Uncertainty scenarios

To account for uncertainty in the process operations, several expected uncertainty scenarios are considered. We assume two dynamic scenarios: varying compositions of nC_4 and iC_4 in input feed, and the increase of the feed rate at certain time (t=1,500) by 15 percent. For time-invariant uncertainty, we wish to investigate the impact of two main uncertain parameters associated with chemical reactions. The first parameter is the pre-exponential factor, k_0, and the second is the heat energy of reaction, λ. We assume that 10% of variation exists from nominal points.

Optimal design process

To maximize the performance, while at the same time planning flexible operation, we performed the design optimization under uncertainty as follows. First we collected uncertain samples (step 1) and performed embedded control optimization (step 2). Finally we checked feasibility of design obtained from step 2 (step 3).

First attempt: As a first attempt to perform the stochastic optimization, we chose 10 samples in the uncertain space of reaction conditions and evaluated the probabilities of each parameter set to calculate expected cost. Then we rigorously determined the design and manipulated variables such that the process does not violate the constraints in desired purity limit in every realization of the reaction conditions and design variability in the dynamic performance due to uncertainty. For capital cost, we considered the reactor, the columns, and the heat exchangers. For operating cost, we considered energy consumption in the furnace, condensers, and reboilers. The master level of this problem is to maximize total profit. To solve optimal design problem, a Nelder-Mead simplex method is used in master level of our methodology, and embedded control was used to adjust control decisions. To handle the constraints, penalty function is used. We found the best optimal design after 24 iterations. (D_1=156ft^3, D_2=59, D_3=32, D_4=1,034ft^2).

Next we rigorously tested the steady and dynamic state feasibility of the design specifications obtained previous section. A fully dynamic feasibility test is very challenging, so we used following strategy. At first, we performed a steady state feasibility test. If the current design is feasible in steady state, then rigorous dynamic feasibility test follows steady state feasibility test. We used active constraint strategy for this test (Dimitriadis & Pistikopoulos, 1995; Grossmann & Floudas, 1987). In steady state feasibility test, we found the critical point (k_0=2.66e7, λ=-3,900), which disobeys constraints with the design spec. This critical point is located at the vertex of the uncertain region.

Second attempt: We solved the main optimization problem again with ten uncertain parameter set and the critical point. In this time we obtained these optimal values - D_1=432ft^3, D_2=52, D_3=33, D_4=1,011ft^2. We tested feasibility tests again in both steady state and dynamic state. This time no violations were found. Thus we concluded obtained result is the best solution. We presented the simulation result of design obtained in Figure 3. It shows a tight control. x_1 and x_2 go under the set point after t=1,500, x_3 never goes higher than the set point.

4. Conclusions

This paper describes a conceptual framework for design and control integration. This paper enhanced and refined the embedded optimization approach for integrated design and control, originally suggested by our group (Malcolm et al., 2007). Our methodology recasts the problem of design and control integration into a solvable mathematical programming formulation. Moving horizon estimation is used to convert the nonlinear

dynamics into the linear state space model adaptively. With the linear process model; Linear Quadratic Regulator could easily find optimal control analytically. The case study of designing an isomerization process- described control and design integration for large scale flowsheet. We apply our novel methodology to plantwide process successfully.
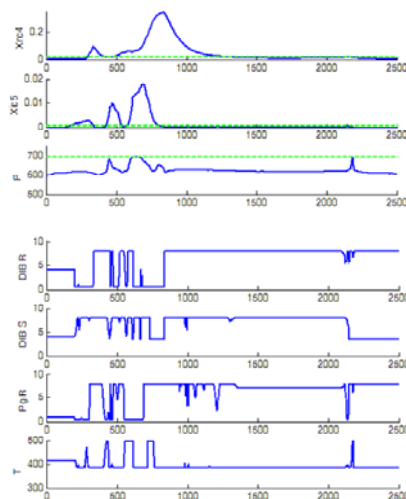

Figure 3. Simulation result of final optimal design. It shows tight control of all process variables. The pressure never exceeded more than 700 psia for all simulation time. The constraints x_{iC5} and x_{nC4} were stabilized under the set points after t=1,500(sec). These constraints and are soft constraints, they are not required to be satisfied for all time period but should be satisfied from specific point of time through the end-operation time. So we set them point constraints $t=[t_1, t_2]$ and for this case, t_1 is 2,250 and t_2 is 2,500 (sec). Here, t_1 means the stabilized operation time after control. However, The pressure constraint should be satisfied through all simulation periods ($t= [0, t_2]$).

References

Dimitriadis, V. D., & Pistikopoulos, E. N. (1995). Flexibility Analysis of Dynamic Systems. Ind. Eng. Chem. Res., 34, 4451.

Grossmann, I. E., & Floudas, C. A. (1987). Active constraint strategy for flexibility analysis in chemical processes. Computers & Chemical Engineering, 11, 675-693.

Haseltine, E. L., & Rawlings, J. B. (2005). Critical evaluation of extended Kalman filtering and moving-horizon estimation. Industrial & Engineering Chemistry Research, 44, 2451-2460.

Luyben, W. L., Tyréus, B. D., & Luyben, M. L. (1998). Plantwide process control. New York: McGraw-Hill.

Malcolm, A., Polan, J., Zhang, L., Ogunnaike, B. A., & Linninger, A. A. (2007). Integrating systems design and control using dynamic flexibility analysis. AIChE Journal, 53, 2048-2061.

Mohideen, M. J., Perkins, J. D., & Pistikopoulos, E. N. (1996a). Optimal design of dynamic systems under uncertainty. AIChE Journal, 42, 2251.

Moon, J., Kulkarni, K., Zhang, L., & Linninger, A. A. (2008). Parallel Hybrid Algorithm for Process Flexibility Analysis. Industrial & Engineering Chemistry Research, 47, 8324-8336.

Sakizlis, V., Perkins, J. D., & Pistikopoulos, E. N. (2004). Recent advances in optimization-based simultaneous process and control design. Computers & Chemical Engineering, 28, 2069-2086.

Seferlis, P., & Georgiadis, M. C. (2004). The integration of process design and control. Amsterdam ; Boston: Elsevier.

Swaney, R. E., & Grossmann, I. E. (1985). An index for operational flexibility in chemical, process design .1. Formulation and theory. AIChE Journal, 31, 621-630.

Multi-Scale Chemical Product Design using the Reverse Problem Formulation

Charles C. Solvason, Nishanth G. Chemmangattuvalappil, Mario R. Eden

Department of Chemical Engineering, Auburn University, Auburn, AL 36849, USA

Abstract

The main objective for this research is to extend the reverse problem formulation algorithm to include aspects of multiple length scales, thereby creating a framework where product synthesis, design, and optimization can be achieved with significantly reduced computational time. In order to achieve this objective a centralized framework was developed that combines property clustering with chemometric techniques like principal component analysis (PCA) and partial linear regression onto latent surfaces (PLS) to solve the design problem in a property descriptor sub-domain. Information from the molecular scale on short range order, such as group structure, conformation, and stereoregularity, is combined with information from the mesoscale on long range order, such as the particle size, to design a set of alternative excipients that exhibit superior properties for use in an acetaminophen tablet without the use of parallel, grid, or supercomputing techniques.

Keywords: Reverse problem formulation, Property clustering, Multi-scale design

1. Background

The U.S. National Research Council has recently published a report on an emerging paradigm in product design focused on integrated computational materials engineering (CICME, 2008). In the report, the committee noted that in order to alleviate the strain put on U.S. manufacturers from the swiftly changing and increasingly global marketplace, *integrated* design closely coupling computational models with manufacturing processes would be required. The term 'integrated' recognizes that the properties of products are controlled by a multitude of separate and often competing mechanisms that operate over a wide range of length and time scales. It is the linkage of the scales that remains the 'Grand Challenge' (CICME, 2008). The traditional method of linking the various length-scales has been to compute information at smaller scales and pass it to models at larger scales by removing degrees of freedom (coarse-graining) with the objective being to predict macroscopic properties from molecular information (Fermeglia and Pricl, 2009). While often the most accurate method for predicting properties, simulation utilizes *a priori* knowledge of the molecular architecture and the computational cost of such a hierarchical nesting method typically prevents an accurate modeling of mesoscopic structure such as the morphology of polymers (Fermeglia and Pricl, 2009). Furthermore, when this method is integrated within the product-process design framework (Hill, 2004), the computational intensiveness exponentially increases since each projected molecular architecture must be simulated to determine its physico-chemical properties. To minimize the computational cost in these types of problems, property prediction simulations are typically approximated with structure descriptor models based on the property-molecular architecture relationships. These descriptor based models can be of the group contribution type (GCM) that directly relate the presence of functional groups to

thermodynamic intensive properties, GCM that describe a set of core properties from which other thermodynamic properties are calculated, topological indices from which QSPR/QSAR models can be used to describe properties, case-specific characterizations that describe molecular architecture which can be empirically related to macro-scale properties, and/or many other types.

Since it is often the non-linear nature of these property models that complicates design problems, resulting in MINLP formulations which require tedious solution strategies, one way to reduce the complexity of the problem is to use the reverse problem formulation (Gani and Pistikopoulos, 2002; Eden *et al*., 2004). In this solution strategy, the complexity of the problem is reduced by using the duality of linear programming to reformulate the design problem as a series of reverse problems solved in the property domain. Essentially, the target properties can be determined from the solution of the reverse simulation problem and, as long as these targets are matched, any number of property models may be used to generate the molecular architectures at the various scales to ensure a solution (Eden *et al.*, 2004). The advantage of using reverse problem formulation in multi-scale product design is that the problem complexity is significantly reduced by decoupling the constitutive property models from the design.

2. Concept

Chemical product design is a relatively new concept in the systems community which has shifted focus from developing only the physical form, function, and aesthetics of assembled products to the design of chemically *formulated* products (Hill, 2004). Chemically formulated products are products designed at the molecular, quantum-, nano-, meso-, macro-, and mega-scale levels to deliver a specific desired attribute. Examples of formulated products include personal care products (Hill, 2004), nano-structured materials (Fermeglia and Pricl, 2009) and many other product types. Hill (2004) notes that the objective of computational product design is to guide and focus experimentation; and CICME (2008) notes that because of a lack of robust design models, complimentary experimental and theoretical approaches are of profound importance. Combining these observations with the reflection that the requisite properties will be consumer attributes (since consumer preference drives the value of a product) it follows that the relationship between the underlying fundamental physical-chemical properties and the consumer attributes will most likely involve empirical relationships. Based on these observations, the objective of this work is to utilize chemometrics to demonstrate how to deconstruct the computational difficulty of a multi-scale product design problem by solving it in a latent property sub-domain under reduced dimensionality.

3. Method

Chemometrics are used to illustrate the solution concept shown in Fig. 1 where a set of user defined attributes are mapped to the latent sub-domain so that simultaneous molecular and meso-scale design can take place. The solution method is as follows:

1. Characterization is used to describe the molecular & micro-structures
2. Decomposition (e.g. Principle Component Analysis), is used to find the latent property sub-domain and its relationship with the descriptor properties
3. Design of Experiments (DOE) & Partial Linear Regression onto Latent Surfaces (PLS) are used to determine the attribute – latent property relationship

4. Reverse problem formulation (RPF) and property clustering are used to map the molecular, meso, and macro-scale domains into the latent sub-domain
5. Simultaneous mixture, molecular, and meso-scale particle size design is accomplished via characterization based group contribution methods and mixture analysis.
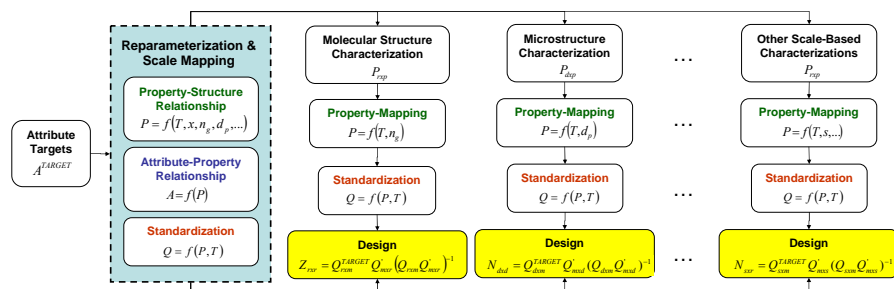

Figure 1. Solution concept for simultaneous multi-scale product design

3.1. Characterization

Characterization is a class of tools associated with the determination of not only chemical constituency or molecular structure, but also of larger structural characteristics describing the orientation and alignment of these molecules (often called microstructure at the meso-scale). Some examples of characterization techniques include nuclear magnetic resonance (NMR), x-ray diffraction (XRD), and infrared spectroscopy (IR). The techniques are applied to a training set of molecules defined by an experimental design used to explore the interesting facets of a set of property attributes A_{nxq}. A more detailed description of the types of characterization techniques utilized in this method can be found in Solvason *et al.* (2009a) and Gabrielsson *et al.* (2002).

3.2. Decomposition

The most common decomposition technique is principal component analysis (PCA). By definition, PCA uses the variance-covariance structure to compress the property data to principal component data that contains much of the system variability. This result also improves the interpretation of the data structure by consolidating multiple property descriptors P_{nxp} into single, underlying latent variables which are devoid of co-linearity. Using the inverse definition of the latent property substructure, the loadings L_{pxm} represent the underlying latent properties, the scores T_{nxm} are mixtures of those properties and the property descriptors are the weights of the underlying latent variables:

$$T_{nxm} = P_{nxp}L'_{pxm} \tag{1}$$

This structure can be normalized through the use of a standardization matrix to ensure that the property descriptors are weighted between 0 and 1. A detailed description of this analysis can be found in Solvason *et al.* (2009b,c).

$$Q_{nxm} = R_{nxp}L_{pxm} \tag{2}$$

3.3. DOE & PLS

The relationship between the principal component scores T_{nxm} and the attribute properties A_{nxq} is then developed using a PLS model of a new DOE factorial design

where the scores are varied between their high (+1) and low (-1) levels. A detailed description of the analysis can be found in Gabrielsson *et al.* (2002).

$$A_{nxq} = T_{nxm}B_{mxq} \tag{3}$$

Where B_{mxq} are the regressed coefficients found using PLS. It should be noted that the PLS model uses a separate set of scores and loadings to develop the relationship between A_{nxq} and T_{nxm}.

3.4. Property Clustering

To visually represent the latent property subspace, property clustering is used to deconstruct the design problem into a Euclidean vector in the cluster domain and a scalar called the Augmented Property Index (AUP). The clusters themselves are conserved surrogate properties described by property operators, which have linear mixing rules, even if the operators themselves are nonlinear. Methods for the application of group contribution methods for molecular design have previously been developed using property clustering by Eljack *et al*. (2007) and Solvason *et al.* (2009b). To utilize the latent variables in the property clustering algorithm, it is important to recognize that the data structure of Eq. 2 follows a linear mixing rule. By extension, other complete molecules, molecular groups, or microstructure subspace properties Q_{nxm}, Q_{rxm}, and Q_{sxm} can be found by multiplying the latent variables L_{pxm} by the associated fractions R_{nxp}, R_{rxp}, and R_{rxp}, respectively.

3.5. Simultaneous Design

Since the molecular group and particle subspace property relationships in Eqs. 1-3 were derived using a decomposition technique, the constraints imposed by decomposition should also be observed for any new molecular architectures created.

$$Q_{jxm} = Z_{jxr}Q_{rxm} \qquad j = n,r,s,...,x \tag{4}$$

For example, the molecular design of Eq. 4 is a representation of a linear mixture of the underlying latent variable subspace properties, all of which are linear in nature; it is essentially a *linear mixture of linear mixtures*. This observation assumes that any nonlinearity in the attribute system is handled by the attribute-latent property relationship and not the molecule-group subspace property relationship (Muteki and MacGregor, 2006).

4. Case Study: Pharmaceutical Tablet Design

The case study discussed in this work is an extension of the review published by Solvason *et al.* (2009a), but specifically focused on simultaneous molecular and particle size design of pharmaceutical excipients. Three attributes that are important to direct compression tablet manufacturing are disintegration time, crushing strength, and ejection force. These attributes have been notoriously difficult to analyze based on traditional mixing design because of the complex and highly nonlinear nature of pharmaceutical excipients. In order to characterize the molecular architecture that contributes to these attributes, they are mapped down to a latent domain subspace where they can be approximated as linear combinations of molecular group and particle size parameters. The domain subspace was found to be characterized by three latent properties using a training set of 24 excipients P_{nxp}, which primarily consisted of mannitol, maltitol, xylitol, maltodextrin, dextrin, isomalt, and microcrystalline cellulose from various suppliers. To reduce the number of parameters in the subspace,

decomposition was performed using PCA. It was found that the 1st latent property explained 96% of the data, although, it was decided to keep all properties for illustrative purposes. Using PLS models developed from the training set, a consumer specific set of targets A_{nxq} was mapped to the domain subspace Q_{nxm} as shown in Table 1.

Table 1. Attribute targets mapped to the descriptor sub-domain

Subspace Targets	Q1	Q2	Q3
UL	2.00	2.00	1.00
LL	1.00	0.00	0.00

The molecular groups present in the training set and identifiable by the characterization were CH, CH_2, OH, CHO, O, CH_2-O, CHOH, CH_2OH, $CHCH_2OH$, CHCHO, α-pyranose, O_{cyc}, β-pyranose, and cellulose. The last 2 groups were only present in the micro-crystalline cellulose excipient and, as such, were excluded from the short chain molecular design. In the previous molecular design using these molecular groups, the selection of the α stereoisomer of pyranose compounds as candidates was preferred, while the microcrystalline cellulose (MCC), which relies on the β stereoisomer was notably outside the target range (Solvason *et al.*, 2009a). This was an unexpected result since MCC is not only one of the most common excipients used in acetaminophen tablets, but also contained the most information on long-range order when characterized. In order to better understand this conclusion, a simple investigation of the meso-scale micro-structure was conducted by analyzing the influence of particle size (50-700 μm) on the attribute properties.

There are two ways to illustrate the influence of particle size on the design problem. The first is to illustrate an average IR/NIR characterization for each particle size. Although a mathematically valid method for visualizing the particle size effect, a particle size 'IR/NIR spectra' has no physical meaning. A more appropriate method is to observe the effect of particle size on the individual chemical constituents in the model. Since increasing particle size results in larger absorbances (and smaller transmittances), the primary effect was an increase in the AUP values of the training set molecules and individual groups. A smaller, secondary effect due to a disproportionally strong effect observed in the 4500-6000 cm[-1] range for polyols (Storz and Steffens, 2004) caused a slight alteration of the location of the training set and molecular group clusters (Fig. 2). Using lever arm analysis on the average and individual particle sizes adjusted for positive AUPs (not shown), suggests that the final particle size should be in the 180-200 μm domain. Reparameterizing the molecular design for this particle size results in the updated candidates shown in Table 2. In comparison to the candidate list published by Solvason *et al.* (2009), some of the α-pyranose candidates are no longer preferred, indicating that as the particle size is increased, the orientation of the glycosydic linkages has a diminishing effect on the product attributes.

Table 2. Designed molecules

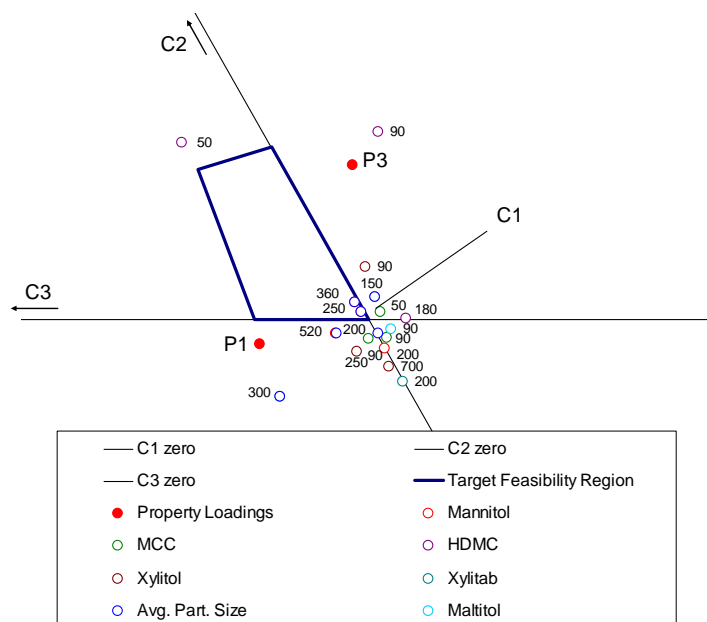| Candidate Molecules	Q1	Q2	Q3
CH2OH-CH2OH	1.25	1.94	0.91
CH2OH-CH2-O-CH2OH	1.78	1.97	0.54
CH2OH-O-CH2OH	1.78	1.93	0.52
OH-CH2OH	1.19	0.94	0.32
OH-(α)pyranose-CH2OH	1.82	0.97	0.66
CH2OH-CH(OH)-OH	1.76	0.92	0.19
OH-CH(CH2OH)-CH2OH	1.77	1.30	0.33

Figure 2. Design cluster diagram featuring particle sizes in μm.

Conclusions

In summary, visualizing the reverse problem formulation using the combination of property clustering and chemometrics provides a framework to solve property driven processes without commitment to components and/or structures *a priori.* In particular, it has been shown that using multivariate characterization techniques to describe both the molecular and micro-structure of excipient particles combined with PCA and PLS to map the attribute data into a latent property subspace provides insight into the structure of pharmaceutical tablets. Planned further analysis will show that this interdependency can be utilized for simultaneous molecular and microstructure design (including polymorphism) to occur, circumventing the heavy combinatorial expense typically associated with solving problems in the meso-scale.

References

Committee on Integrated Computational Materials Engineering, National Research Council (2008). *The National Academies Press*, USA.

M.R. Eden, S. B. Jørgensen, R. Gani and, M. El-Halwagi (2004). *Chem. Eng. Proc.* 43.

F.T. Eljack, M.R. Eden, V. Kazantzi, M.M. El-Halwagi (2007), *AIChE Journal*, 53(5)

M. Fermeglia and S. Pricl (2009). *Comp. & Chem. Eng,* 33(10)

J. Gabrielsson, N. Lindberg, and T. Lundstedt (2002). *J. Chemometrics.* 16.

R. Gani. and E.N. Pistikopoulos (2002). *Fluid Phase Equilibria,* 194-197.

M. Hill (2004), *AIChE Journal*, 50(8), 1656-1661.

K. Muteki and J.F. MacGregor (2006), *Chemom. & Intell. Lab.* Sys., 85, 186-194

C.C. Solvason, N.G. Chemmangattuvalappil, M.R. Eden (2009a), *Comp. Aided. Chem. Eng.,* 26,

C.C. Solvason, N.G. Chemmangattuvalappil, M.R. Eden (2009b), *Comp. & Chem. Eng.*, 33(5).

C.C. Solvason, N.G. Chemmangattuvalappil, F.T. Eljack, M.R. Eden (2009c), *I&EC Res.* 48(4).

E. Storz and K. Steffens (2004). *Starch/Starke*. 56.

Solvent design for a Menschutkin reaction by using CAMD and DFT calculations

Heiko Strübing, Panagiotis G. Karamertzanis, Efstratios N. Pistikopoulos, Amparo Galindo, Claire S. Adjiman,

Department of Chemical Engineering, Centre for Process Systems Engineering, Imperial College London, London SW7 2AZ, U.K., c.adjiman@imperial.ac.uk

Abstract

The design of solvents that optimise reaction performance is challenging due to the complexity of solvent effects on reactions. We propose an *ab initio* computer-aided methodology for identifying optimal solvents based on a combination of quantum mechanical calculations and empirical models. In the first step of the approach, a number of quantum mechanical calculations are performed to predict the relative rate constant for the reaction of interest in several solvents. A continuum solvation model is used to model solvent effects on the relative thermodynamic stability of reactants and the transition state. This is then used to develop a surrogate model based on a linear free energy relation, the solvatochromic equation, and on group contribution techniques. This simplified model allows the fast identification of alternative solvents via an optimisation-based computer-aided molecular design approach due to its low computational cost. We apply the proposed approach in the case of the Menschutkin reaction between pyridine and phenacyl bromide to identify solvents that maximise the rate constant. The results are compared to experimental reaction rate data and show that the method is capable of designing promising solvents.

Keywords: Solvent design, CAMD, solvent effects, quantum mechanics rate constants

1. Introduction

Solvents are widely used in the chemical industry to fulfill a number of roles, such as supplying heat to endothermic reactions, absorbing heat from exothermic reactions, transporting reactants to a location in the correct concentrations and promoting or demoting reaction rates [1]. Hence, the optimal choice of solvent impacts process performance in many ways as it can alter the rate of a given reaction by several orders of magnitude[2]. Thus, when optimising a solvent-mediated reactive process, it is essential to consider the choice of solvent as part of the key design variables.

In spite of the large impact of solvents on reactive process performance, there are relatively few design methodologies that consider the effect of solvents on reaction kinetics. In most approaches, the focus has been on the effect of the solvent on phase equilibrium (e.g. solubility) and aspects other than the effect of the solvent on the rate constant [1][3]. To date, two design methods that explicitly incorporate kinetics:

- the approach of Folic *et al.* [4][5], who developed an empirical model using group contribution (GC) techniques that relate solvent structure and rate constant and incorporated it within a Computer-Aided Molecular Design (CAMD) algorithm, and
- the approach by Stanescu and Achenie [6], who used quantum mechanical (QM) calculations to compute rate constants for the reaction of interest in

promising candidate solvents, as identified by a CAMD algorithm based on complementary criteria.

In this work, we propose an *ab initio* method for designing reaction solvents by explicitly maximising reaction rate. We apply this approach to a Menschutkin reaction that has been used as a model for solvent effects. We compare the results of the proposed approach to that of Folic *et al.* [4], in which some experimental data are required for model building.

2. Methodology

2.1. *Overall methodology*

It is assumed throughout that the solvent is inert during the reaction. The proposed design methodology for a given reaction is as follows:

1. Compute reaction rate constant for a few diverse solvents using Transition State Theory (TST) and QM calculations with a continuum solvation model [7].
2. Build a simplified reaction model by regressing an empirical model (solvatochromic equation [8][9]), which, when combined with group contribution methods [10], can be used to predict the rate constant in new solvents.
3. Use optimisation-based CAMD [4] to design solvent molecules based on the proposed model.

This approach differs from that of Folic *et al.* [4] in the first step, as in our approach reaction rate constants are predicted from *ab initio* calculations rather than being measured experimentally. The use of *ab initio* calculations is made possible by focusing on the prediction of relative rate constants, rather than absolute values, and by translating the information extracted from expensive QM calculations into an inexpensive model in Step 2, which can then be used to systematically assess the performance of a very large number of solvents.

2.2. *Prediction of reaction rate constants using QM and continuum solvation*

In the first step, the reaction rate constant k^{QM} is predicted in several solvents of varying polarity and containing a range of functional groups. We used conventional transition state theory [11][12] that is only briefly summarised. Eyring [11] postulated that for an elementary reaction $A + B \rightarrow C + D$, a quasi-equilibrium exists, so that,

$$A + B \underset{k_{-1}}{\overset{k_1}{\rightleftarrows}} (AB)^{\ddagger} \xrightarrow{k_2} C + D , \tag{1}$$

where A and B are the reactants, C and D are the products and $(AB)^{\ddagger}$ is the activated complex or transition state in quasi-equilibrium with the reactants. The rate-determining step occurs when the transition state decomposes into the products and is described by:

$$k^{\text{QM}} = k_2 = \frac{k_B T}{h} \frac{\gamma_A \gamma_B}{\gamma_{(AB)^{\ddagger}}} \frac{C^{\circ}_{(AB)^{\ddagger}}}{C^{\circ}_A C^{\circ}_B} \exp\left(-\frac{\Delta_r G^{\circ,\ddagger}}{RT}\right) , \tag{2}$$

where k_B is the Boltzmann constant, T is absolute temperature, h is the Planck constant, γ_i is the activity coefficient of species i (assumed here to be equal to 1, although in

general it can be different from 1 and to depend on the solvent), C_i° is the standard state concentration of species *i*, *R* is the universal gas constant and $\Delta_r G^{\circ,\ddagger}$ is the Gibbs free energy barrier (activation energy) between the reactants and the transition state.

The Gibbs free energy of activation is computed using GAUSSIAN [13] and the PCM model [14]. Such a computation is demanding: it requires the energy minimisation of the reactant molecules and the identification of the transition state in a dielectric continuum and the vibrational frequencies of the reactants and transition state in gas phase.

In order to relate the calculated k^{QM} to experimentally-measured rate constants, it is important to use the same standard states. This is normally done by finding a scaling between the computed and measured rate constants. Thus, in the absence of any experimental data, it is not possible to compute absolute rate constants (k^{QM}) that predict the experimentally-derived rate constants (k^{exp}). Instead, relative rate constants, which relate the rate constant in one solvent to that in a reference solvent, can be readily obtained and used in solvent design.

2.3. Construction of an empirical model for predicting reaction rate constants

In Step 2 of the design methodology, a simplified model of solvent effects on the reaction rate constant is derived from *ab initio* computed rate constants for a small set of solvents. We used the solvatochromic equation [8][9], a linear solvation energy relationship that links the reaction rate constant to solvent properties. An important feature of the solvatochromic equation is that reaction parameters and solvent properties are decoupled, making the prediction of solvent effects straightforward. The solvatochromic equation is given by:

$$\log\left(k^{solv}\right) = a_0 + a_S S + a_\delta \delta + a_A A + a_B B + a_H \delta_H^2 / 100, \tag{3}$$

where k^{solv} is the reaction rate constant, *S, A, B* are solvatochromic parameters (*S* measures dipolarity and polarisability, *A* the hydrogen bond acidity, *B* the hydrogen bond basicity), δ is the polarisation correction factor (0.0 for non-halogenated, 0.5 for polyhalogenated and for 1.0 aromatic solvents), δ_H^2 is the cohesive energy density which quantifies solvent-solvent interactions. The parameters S, A, B, δ and δ_H^2 depend only on the solvent. On the other hand, $a_0, a_S, a_\delta, a_A, a_B, a_H$ are parameters that depend on the reaction and can thus be regressed from knowledge of the solvent properties and (relative or absolute) reaction rate constants for a statistically sufficiently large set of solvents.

For the illustrative example in this paper we used QM to predict the rate constants of the chosen Menschutkin reaction in seven solvents that were subsequently used to regress the reaction-specific parameters. The solvent properties S, A, B, δ and δ_H^2 are obtained via group contribution methods [15][4] that have been developed for a set of 43 atom groups.

2.4. Formulation of the CAMD problem

The solvent design is formulated as a mixed-integer minimization problem, following the approach of Folic *et al* [4][5]. Any number of reactions can be considered within the problem formulation, provided each is represented by a solvatochromic equation. Binary variables in the problem formulation include variables that constrain the number of groups of each type to integer values and the type of molecule that can be designed (acyclic, monocyclic and bicyclic molecules were allowed). For the illustrative example

we chose to maximize the rate constant, k^{solv}, for a single reaction. The solvent design formulation is given by:

$$\max \log\left(k^{\text{solv}}\right)$$

$$s.t. \log\left(k^{\text{solv}}\right) = a_0 + a_S S + a_\delta \delta + a_A A + a_B B + a_H \, \delta_H^2 \big/ 100$$

    property estimation techniques   (4)
    design constraints
    molecular complexity constraints
    chemical feasibility constraints

The property estimation techniques include GC expressions and other equations that relate the structure of the solvent molecule to its properties. The design constraints are imposed to ensure that the solvent is within a desired operating range, e.g., that it is liquid under operating conditions. The molecular complexity constraints limit the size and number of groups in the candidate solvents. The chemical feasibility constraints ensure all molecules designed are structurally meaningful (e.g., there are no free bonds in the molecule), and relate binary variables to the number of groups of each type. A detailed formulation can be found in Strübing *et al.* [16].

3. Results

The proposed *ab initio* approach is applied to the Menschutkin reaction [17] between phenacyl bromide and pyridine:

$$C_6H_5C(O)CH_2Br + C_5H_5N \rightarrow C_6H_5C(O)CH_2N^+C_5H_5 + Br^- \tag{5}$$

3.1. Ab initio prediction of rate constant

The rate constant for the reaction in seven solvents (nitromethane, methanol, acetone, dichloromethane, tetrahydrofuran, chlorobenzene and toluene) is computed at the B3LYP/6-31+G* level of theory. The predicted rate constants relative to the reaction rate in nitromethane are reported in Table 1. Any solvent can be used as a reference to calculate relative rate constants. In this work, nitromethane was chosen because it has the largest rate constant value, and thus the impact of uncertainty in that value on the relative rates is likely to be the smaller than when using rate constants much smaller than 1.

Table 1. Rate constants relative to nitromethane for the Menschutkin reaction at 35$^{\text{o}}$C, as calculated using QM with PCM.

Solvent	nitro-methane	methanol	acetone	dichloro-methane	tetra-hydrofuran	chloro-benzene	toluene
Relative rate constant	1.0000	0.6063	0.6251	0.4433	0.1547	0.0123	0.0051

3.2. Simplified reaction model

The solvatochromic equation is regressed with rate constants determined in the first step and solvent properties from GC. The resulting equation is given by

$$\log\left(k^{\text{solv}}\right) = -6.42 + 6.47S + 0.36\delta + 14.90A + 3.27B - 3.67\delta_H^2 \tag{6}$$

Solvent design on a Menschutkin reaction

This equation can then be used to predict the rate constant for the reaction in other solvents. An alternative solvatochromic equation is also obtained based on experimentally-measured rate constants [18] for comparison. This is given by

$$\log\left(k^{GC}\right) = -3.49 + 3.32S + 1.31\delta + 5.84A + 4.21B - 1.24\delta_H^2 \tag{7}$$

3.3. Solvent design results

The solvent design problem is formulated by considering the simplified reaction model Eq. (6) and a design space of solvents. These are obtained from a list of 43 groups [4] after removing unsuitable groups such as those that can react with the reactants (e.g., halogenated groups, COOH). Integer cuts are added to the problem, which is then solved repeatedly to obtain a ranked list of solvent candidates.

The best solvent identified is 3-methyl-2-methylene-1,3-butanediol, with chemical formula $C_6H_{12}O_2$. This solvent is depicted in Figure 1. The predicted rate relative to that of nitromethane is 4.877.
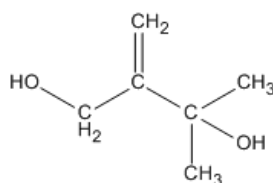

Figure 1. The best solvent candidate for the Menschutkin reaction found using the proposed *ab initio* approach.

The best solvent molecules identified by the CAMD algorithm contain OH groups. A significant proportion of the best performing solvent molecules also contain the $CHCl_2$, $CH_2=C$ or CH_2NO_2 group. This is consistent with the fact that nitromethane and methanol are predicted to perform well as solvents (see Table 1).

In Table 2, these results are compared with the results obtained when using the solvatochromic equation derived from experimentally measured rate constants in Eq. (7). The results of both approaches are consistent, although the QM-based model indicates that the $CHCl_2$ group is favourable, while the experimentally-derived model does not.

Table 2. Functional groups in the best solvent candidates using two simplified reaction models.

Solvatochromic model	Groups in most solvent candidates	Groups in many solvent candidates
Eq. (6) Based on *QM* calculations	OH	$CHCl_2$ $CH_2=C$ CH_2NO_2
Eq. (7) Based on experimental data	OH	$CH_2=C$ CH_2NO_2

4. Conclusions

An *ab initio* methodology for designing solvents has been developed. It is based on the use of quantum mechanical calculations with a continuum solvation model to predict rate constants in several solvents relative to a reference solvent. This is then used to develop a simplified yet predictive reaction rate model, based on the linear

solvatochromic equation and group contribution techniques. This computationally inexpensive model is used in an extensive and systematic exploration of the solvent design space through a CAMD algorithm. This approach has been applied successfully to a Menschutkin reaction, suggesting a family of optimal solvents. The results have been shown to be consistent with the results obtained by regressing the simplified model using experimentally measured reaction rates. Further development of the approach will require closing the loop, by predicting the reaction rate of the CAMD generated solvents using QM and generating an improved empirical reaction rate model until self-consistency is achieved.

5. Acknowledgements

References

[1] R. Gani, J-G. Concepción, D.J.C. Constable, Comput. Chem. Eng., 29 (2005), 1661.

[2] C. Reichardt, 3rd Ed., Solvents and Solvent Effects in Organic Chemistry, WILEY-VCH, Weinheim, 2005.

[3] A.I. Papadopoulos, P. Linke, Chemical Engineering Research and Design, 83 (2005), 674.

[4] M. Folić, C.S. Adjiman, E.N. Pistikopoulos, AICHE Journal, 53 (2007), 1240.

[5] M. Folić, C.S. Adjiman, E.N. Pistikopoulos, Ind. Eng. Chem. Res., 47 (2008), 5190.

[6] I. Stanescu, L.E.K. Achenie, Chemical Engineering Science, 61(2006), 6199.

[7] C.J. Cramer, Essentials of Computational Chemistrz – Theories and Models, John Wiley and Sons Ltd., Chichester, UK.

[8] M.H. Abraham, R.M. Doherty, M.J. Kamlet, J.M. Harris, R.W. Taft, J. Org. Chem., 46 (1987), 3053.

[9] M.H. Abraham, R.M. Doherty, M.J. Kamlet, J.M. Harris, R.W. Taft, J. Chem. Soc., Perkin Trans. 2, (1987), 913.

[10] A. Fredenslund, R. Jones, J. Prausnitz, AICHE Journal, 21 (1975), 1086.

[11] H. Eyring, Journal of Chemical Physics, 3 (1935), 107.

[12] M.G. Evans, M. Polanyi, Transactions of the Faraday Society, 31 (1935), 875.

[13] M.J. Frisch, G.W. Trucks, H.B. Schlegel, G.E. Scuseria, M.A. Robb, J.R. Cheeseman, J.A. Montgomery, Jr., T. Vreven, K.N. Kudin, J.C. Burant, J.M. Millam, S.S. Iyengar, J. Tomasi, V. Barone, B. Mennucci, M. Cossi, G. Scalmani, N. Rega, G.A. Petersson, H. Nakatsuji, M. Hada, M. Ehara, K. Toyota, R. Fukuda, J. Hasegawa, M. Ishida, T. Nakajima, Y. Honda, O. Kitao, H. Nakai, M. Klene, X. Li, J.E. Knox, H.P. Hratchian, J.B. Cross, V. Bakken, C. Adamo, J. Jaramillo, R. Gomperts, R.E. Stratmann, O. Yazyev, A.J. Austin, R. Cammi, C. Pomelli, J.W. Ochterski, P.Y. Ayala, K. Morokuma, G.A. Voth, P. Salvador, J.J. Dannenberg, V.G. Zakrzewski, S. Dapprich, A.D. Daniels, M.C. Strain, O. Farkas, D.K. Malick, A.D. Rabuck, K. Raghavachari, J.B. Foresman, J.V. Ortiz, Q. Cui, A.G. Baboul, S. Clifford, J. Cioslowski, B.B. Stefanov, G. Liu, A. Liashenko, P. Piskorz, I. Komaromi, R.L. Martin, D.J. Fox, T. Keith, M.A. Al-Laham, C.Y. Peng, A. Nanayakkara, M. Challacombe, P.M.W. Gill, B. Johnson, W. Chen, M.W. Wong, C. Gonzalez, J.A. Pople, 'Gaussian 03, Revision E.01', Gaussian, Inc., Wallingford, CT, 2004.

[14] J. Tomasi, B. Mennucci, E. Cancès, Journal of Molecular Structure (Theochem), 464 (1999), 211.

[15] T.J. Sheldon, C.S. Adjiman, J.L. Cordiner, Fluid Phase Equilibria, 231 (2005), 27

[16] H. Strübing, S. Konstantinidis, P.G. Karamertzanis, E.N. Pistikopoulos, A. Galindo (eds.), C.S. Adjiman (eds.), Molecular Systems Engineering, Chapter: Computer-Aided methodologies for the design of reaction solvents, In Press.

[17] P. Barnard, B. Smith, Journal of Chemical Education, 58 (1981), 282.

[18] D. Willitts, Solvent Effects in a Menschutkin Reaction, Internal report, Imperial College London, 2009.

The Virtual Product-Process Laboratory applied to personal care formulations

Elisa Conte,[a] Rafiqul Gani,[a] Tahir. I. Malik[b]

[a]CAPEC-Dep. of Chem. and Biochem. Eng., Tech. Univ. of Denmark, 2800 Kgs. Lyngby, Denmark, elc@kt.dtu.dk
[b]AkzoNobel RD&I, Wilton Centre, Redcar TS10 4RF, U.K.

Abstract

The objective of this paper is to illustrate the new advances in the virtual product-process design laboratory (virtual PPD-lab) which has been previously presented [1, 2]. Through the virtual PPD-lab, it is possible to generate and/or analyze alternatives for products and processes matching a set of predefined targets by performing virtual experiments in order to verify product feasibility, identify problems and validate the product. The significance of the virtual laboratory is that the experimental effort during the development of new products and processes can be reduced, sparing therefore, valuable time and resources. This paper highlights the new advances in the capabilities of the virtual PPD-lab and illustrates them through an industrial case study involving the analysis of a hair spray.

Keywords: product design, hair spray, polymer-solvent system

1. Overview of the virtual laboratory

Chemical product-process design involves the identification of product candidates (molecules and/or mixtures) that exhibit specific desirable or targeted behavior and the design of a process that can manufacture them with the same qualities. The common practice for the design and development is experiment-based trial and error approach, where past and expert knowledge play an important role. The virtual PPD-lab is an innovative and alternative approach to product-process design. It contains methods, tools, work-flows/templates for different types of design problems corresponding to different activities within the work-flow. As a number of the needed methods, tools and models are already available in ICAS [3], they are accessed from the virtual PPD-lab through an established interface. The current version of the virtual PPD-lab consists of templates for the design of microcapsules for the controlled release of active ingredients; for the design of pesticide formulations; and the newly added template for the design of formulation [1,2].

2. Computer-Aided design and analysis of formulation

Many chemicals based products used by millions of people everyday are formulations where Active Ingredients (AIs), solvents and additives are blended together in order to obtain a product with specific properties and a particular form (spray, cream,…) which meet the targets defined by the consumers/market/regulatory. The AIs are responsible for the main function of a product, for example, the activity of an insect repellent is to be effective against mosquitoes. The solvent mixture is the carrier for the AIs and it is therefore present in high concentrations. Also, it needs to be carefully selected (designed) in order to completely dissolve the AIs and to give the formulated product

the desired (target) behavior. Additives such as wetting agents, perfumes, preservatives, etc., are added to enhance the end-use properties of the product, while other additives such as emulsifiers in a cream product are responsible for conferring the right form to the product. The formulations considered in this work are lotions, that is, formulation with a liquid form and a work-flow for their design has been developed [2] and consists of four main tasks (see Fig. 1). At first, a market analysis is performed to understand the consumer needs, which are then translated into target properties (rule based translation [2,4]) on which constraints are set. If several target properties affect one consumer need, all of them should be taken into consideration in the design. These actions contribute to the definition of the design problem (task 1). At task 2 database alternatives are screened in order to identify the most suitable AI/AIs. If the AI/AIs chemical properties are not available in the database they have to be calculated. In the solvent mixture design task, candidate one phase liquid mixtures matching the predefined property targets are identified. Additives are then selected in task 4. This methodology has been applied for the design of a paint [1], a liquid-spray insect repellent and a liquid-spray sunscreen [2].

While the above methodology (work-flow) is sufficient for the design of new products, modifications are needed for analysis, verification and possible improvements of existing formulated products. The aim of this paper is therefore to extend the current methodology to allow these new features and to highlight the application of this extended version through an industrial case study involving a hair-spray product.

2.1. Computer-Aided analysis of formulations

Unlike the design of formulations where the identities of the chemicals present are not known, in the verification and analysis of formulations, the identities of most of the chemicals are known, avoiding thereby the step of screening numerous alternatives. The aim in analysis of liquid formulated products is to verify the (liquid) phase stability of the actual formula and to calculate its chemical and physical properties to check the product performance. In some cases, the exact formula may not be available and a small number of candidates may need to be generated from an available short list of AIs, solvents or additives. The main issue is how to predict the phase stability and liquid mixture properties for very complex mixtures consisting of AIs (pigments, polymers), solvents (organic chemicals and water), additives (aroma chemicals, surfactants, preservatives, etc.). Models to predict properties of such mixtures do not exist while experimental verification is a reliable but time consuming and expensive option. The objective, therefore, is to manage this complexity by breaking down the problem into a number of sub-problems where available data and/or models can be used. At first the binary solvent mixtures, for which models and data are available, are studied. Then, binary mixtures of AIs and solvents are considered. Here, if models or data are not available, they can easily be adopted through a few data-points
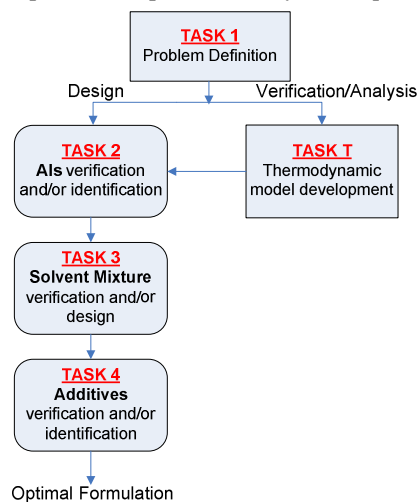

Figure 1. The modified flowchart.

(measured or generated through other models). Next, only those binary mixtures that satisfy the requirements are combined to form ternary mixtures and the models verified for binary mixtures are now tested for ternary mixtures. That is, the binary parameters are being extrapolated to the ternary composition space. Few experimental data are needed to verify the model performance. Finally, systems with more than two solvents are considered, if necessary, with the same models. It is worth to note that verification and analysis of the specified product is made indirectly without using the actual multicomponent complex mixture.

Based on the above, the previously developed work-flow [2] for the design of formulated products has been modified to introduce the option verification/analysis (see Fig. 1). The problem definition task (task 1) is common to both the options of design and verification. If the verification/analysis option is chosen, an appropriate set of thermodynamic models is developed, if necessary, in task T (T stands for 'Thermo'). The chemical systems under consideration are complex because of the nature of formulation, therefore the problem is decomposed into sub-problems. The actions to be performed in task T are the following:

Sub-task T.1: The solvent mixture is analyzed. At first, a search of the open literature is performed to look for availability of experimental (solubility) data/information. Then, available property model (such as UNIFAC, NRTL, UNIQUAC) usually able to handle the prediction of liquid equilibria for the systems under consideration are used to generate the phase diagrams.

Sub-task T.2: Analysis of binary systems AI-solvent. A search of the open literature is also performed to look for solubility information for the involved AIs. Here, an appropriate property model, capable of prediction of solubility and miscibility needs to be used. At first, the ability of available models (group/segment contribution based) to handle the systems under consideration is tested, that is, the necessary interaction parameters should be available. If not, new parameters need to be regressed or the existing parameters may be fine-tuned on experimental data. If experimental data are not available, pseudo-experimental data could be generated. Once the interaction parameters are available, binary phase equilibrium diagrams can be generated.

Sub-task T.3: Analysis of ternary systems AI-solvents. Literature search is performed, as in the previous steps. If new binary or ternary interaction parameters are needed, the same procedure used in T.2 is employed.

Sub-task T.4: Analysis of systems with more than two solvents. The procedure is the same as above.

If more than one AI needs to be considered, the procedure outlined above is repeated for each AI. As a last sub-task, the multicomponent system including the AIs and solvents should be considered. Additives are not included in task T since their concentration is usually very low and it can be assumed they will not influence the phase behavior of the system. If a choice between several AIs needs to be performed, further modeling may be required in task 2 in order to choose the most promising AI. If a solvent mixture needs to be designed, further modeling may be needed in task 3, in order to compare the properties of different candidates and select the most promising. In task 4, the influence of the additives on the properties of designed formula can be evaluated. The methodology is highlighted through an industrial case study involving the analysis of a hair spray product.

3. Case study

A hair spray is usually constituted by a polymer blend, a solvent mixture, a propellant and some additives such as neutralizing agents, plasticizers and aromas [5]. The polymer blend constitutes the active ingredient (AI) with the function of holding the curls. The solvent mixture is usually a water-based mixture of organic chemicals and it constitutes the delivery system together with the propellant. The neutralizing agents are usually alkaline and are needed to rend the product water soluble (rinsable with water).

In this case study, the AI and the neutralizing agent are known, while the solvent mixture has not been defined yet, but a short list of candidate solvents is available. Because of confidentiality reasons, details such as chemical identity can not be given. However, sufficient explanations are provided for the reader to understand the main concepts. The AI is the co-polymer $M1_xM2_yM3_z$ where $M1$, $M2$ and $M3$ are three polymer repeat-units and x, y, z are the weight fractions of each repeat unit (note that $x \gg y > z$). Molecular weight and polydispersity index are also known. The list of solvents includes 5 candidates: A, B, C, D and E. The objective of the case study is to investigate the stability of the formulation in terms of single liquid phase solvent mixtures that satisfy the performance criteria set by the market and that can at the same time dissolve the AI. The temperature of interest is room temperature (300 K).

According to task 1 of the methodology (see Fig. 1), consumer preferences have to be identified. Market survey has indicated that consumers want a hair spray to give good curl retention and holding power, without giving a harsh and brittle feeling to the hair [6]. Shine and luster are also required attributes [7]. It must be possible to easily rinse the product with water [6] and with short drying time. In addition, the spray does not have to retain electric charge, which is responsible of the 'electric look' [7]. Flammability and toxicity concerns must also be taken into considerations [7] and environmental friendly products are preferred therefore VOCs emissions must be below the regulated limit [5,7].

The curl retention/holding power, shine and luster are provided by the AI. The tacky or gummy feeling also depends on the AI. The solvent mixture is responsible for the product drying time, flammability and toxicity. The overall mixture should contain less than 80% (by weight) of alcohol in order to limit the VOCs emissions. In addition, the static charge of the solvent mixture should be kept low. Numerical values for the constraints are also set: drying time is set between 480 and 960 seconds, toxicity parameter LC_{50} has to be higher than 0.1 mol/l, dielectric constant ε is fixed close to the one of the water (between 50 and 70) to avoid the dielectric behavior, the flash point has to be at least higher than the room temperature. The drying time constraints have been chosen considering the information given in [8]. For the problem defined above (task 1), the appropriate property models need now to be developed.

3.1. Development of the property models

The decomposition of the problem into sub-problems as outlined in section 2.1 is illustrated in Figure 2 (additional sub-tasks are needed because the AI is a copolymer). sub-tasks T.2, T.3 and T.4 have to be performed for each polymer $P1$, $P2$ and $P3$. At the end of sub-task T.4, the polymer repeat units can be combined and the whole picture of the complex formulation can be obtained. At this moment, binary and ternary solvent mixtures have been analyzed, and the first polymer $P1$ constituted by the repeat unit $M1$ has been considered. UNIFAC GC-method has been employed and it has been found that total miscibility cannot be guaranteed for the following solvent binary mixtures: (A, D), (A, E), (C, D) and (D, E). Since solvent D is causing immiscibility in many of the

binary solvent mixtures, it is eliminated from further consideration. Ternary phase diagrams have also been investigated for the remaining solvents. In sub-task T.2, binary systems composed of a polymer *P*1 and one solvent have been analyzed. Experimental data for the following systems have been found: (*P*1, *A*), (*P*1, *B*) and (*P*1, *C*). The FV-UNIQUAC model [9,10], which is suitable for polymer-solvent solubility calculations, has been selected from the model library to perform the phase stability checks. Even though the model is available in the model library, the needed interaction parameters are not available. The existing experimental data has been used to regress the parameters for the systems (*P*1, *A*), (*P*1, *B*) and (*P*1, *C*) while for the remaining two systems (*P*1, *D*) and (*P*1, *E*), pseudo-data have been generated based on the assumption that the energetic interactions between segments are not significantly dependent on molar mass, therefore, pseudo-experimental data generated with low molecular weight compounds can be used for the regression of the polymer-solvent parameters [9,10]. The UNIFAC GC-method has been employed for the pseudo-data generation. Systems (*P*1, *A*) and (*P*1, *B*) are found to show a two phase region with an Upper Critical Solution Temperature behaviour. In sub-task T.3, ternary mixtures are considered. No information on the ternary systems have been found. Models used in sub-task T.3 do not require ternary interaction parameters, but additional binary interaction parameters have to be regressed (interaction solvent-solvent).
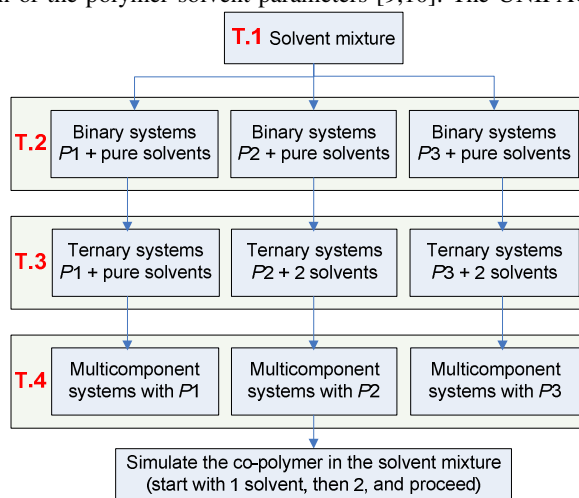

Figure 2. The thermodynamic task T.

3.2. Design of the solvent mixture

In task 1 the performance criteria were fixed and constraints on the target properties were defined. According to task 3 a solvent mixture matching the targets has to be identified from the combinations of solvents *A*, *B*, *C*, *D* and *E*. The mixture property models employed for this screening are linear mixing rules for the toxicity parameter and the dielectric constant, while rigorous models based on group contributions are selected to check the evaporation rate [11] and the flash point of the mixtures [12]. A constraint on the composition of the mixtures is also applied to account for the VOCs emissions. Table 1 shows the results: the mixtures matching the constraints, their composition in terms of weight fraction w_1, and the values of the properties LC_{50} (mol/l), ε (-), T_{90} (s), T_{flash} (°C).

Table 1: feasible solvent mixtures for the hair spray formulation.

Solvent 1	Solvent 2	w_1	LC_{50}	ε	T_{90}	T_{flash}
A	B	0.645	0.706	70.0	945.95	30.35
A	C	0.597	0.637	70.0	849.85	28.95

The last step is to evaluate if the identified solvent mixtures form a single liquid phase with the polymer *P*1. Figure 3 shows the ternary phase diagrams for the systems (*P*1, *A*, *B*) and (*P*1, *A*, *C*). The composition of interest is also reported (the polymer weight fraction is fixed at 5%). The ternary diagrams show that the first formula falls in the single phase region, while the second formula is in the two phase region. Therefore the only feasible formula, in the case the AI is constituted by the only polymer *P*1, is the formula (*P*1, *A*, *B*) with the corresponding composition (0.05, 0.613, 0.337).
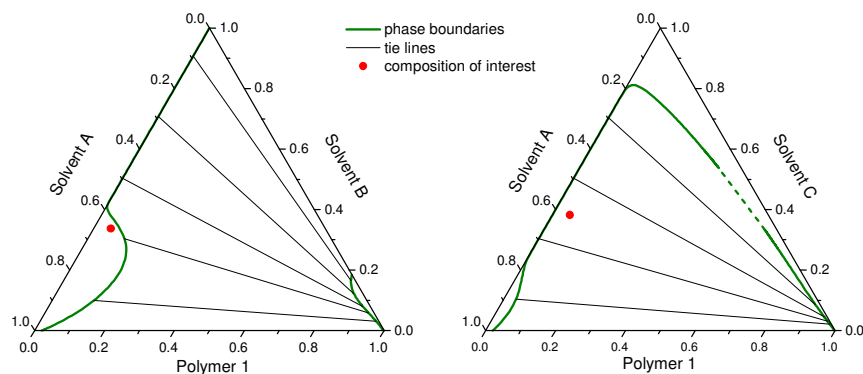

Figure 3. Weight based ternary phase diagrams for the systems (*P*1, *A*, *B*) and (*P*1, *A*, *C*).

4. Conclusions and future work

The computer-aided methodology for the design of formulated products has been extended to include the verification and analysis of specified formulations. A procedure for the development of models needed for the new options has been suggested. Through an industrial case study, the application of the new developments has been highlighted. The case study has also shown how to face the problem of non-availability of experimental data for model parameter regression, and how the complexity of a problem can be managed with computer-aided tools. Further work is necessary to incorporate all the options into the Virtual PPD-lab software.

References

[1] Conte, E., Morales-Rodriguez, R., Gani, R., 2009, Proceedings of 19th European Symposium on Computer Aided Process Engineering, 249-254.
[2] Conte, E., Morales-Rodriguez, R., Gani, R., 2009, Proceedings of 10th International Symposium on Process System Engineering, 825-830.
[3] ICAS Documentation, 2003, Internal report, CAPEC, KT-DTU, Lyngby, Denmark.
[4] Harper, P. M., Gani, R., 2000, *Comp. and Chem. Eng.*, 24, 677-683.
[5] Shah, S. M., Fernandez, C. A.,1992, US Patent 5294437.
[6] Varco, J., Williams, C. E., 1985, US Patent 4567040.
[7] Shernv, S., 1989, US Patent 5053218.
[8] Morawsky, N., Martino, G. T., 1997, US Patent 5599524.
[9] Bogdanic, G., Vidal, J., 2000, *Fluid Phase Equlib.*, 173, 241-252.
[10] Bogdanic, G., 2001, *Fluid Phase Equlib.*, 191, 49-57.
[11] Klein, J. A., Wu, D. T., Gani, R., 1992, *Comp. and Chem. Eng.*, 16(5), 229-236.
[12] Liaw, H.-J., Lee, Y.-H., Tang, C.-L., Hsu, H.-H., Liu, J.-H., 2002, *J. of Loss Prev. in the Proc. Ind.*, 15, 429-438.

Systematic Analysis and Design of the Poly-Production Process for Power and Olefin from Natural Gas

Yu Qian*, Xiuxi Li, Zhixian Huang, Yun Chen

School of Chemical Engineering, South China University of Technology, Guangzhou 510640, China, *ceyuqian@scut.edu.cn*

Abstract

A novel polygeneration system for olefin and power production from natural gas is proposed. Technological-economical evaluation is based on the internal rate of return (IRR) and exergy efficiency analysis. The proposed polygeneration system for simultaneous production of ethylene and propylene and electricity is energy integrated and thus more thermodynamically efficient and economically viable. IRR and exergy efficiency of the proposed polygeneration system are higher than that of GSMTO system 18.9% and 49.9%, respectively. The biggest exergy destruction segments and possible measures for improvement are investigated through simulation and thermodynamic analysis.

Keywords: poly-production. exergy efficiency, natural gas, olefin

1. Introduction

Industrial energy use is facing dual pressure from access to energy resources and environmental considerations. The development and deployment of integrated energy conversion and chemical systems - poly-generation system- could be a promising approach to meeting increasingly stringent criteria and demand for reliable power supplies. There are many such integrated systems, particularly in the chemical and energy industries where processes are flexible and output products could be cascaded and recycled to minimize environmental impacts [1].

Although the investment cost of poly-generation systems is higher, in the long run they are more economical than systems where power, heat and cooling are generated individually [2]. Poly-production has been highlighted in the literature as a promising alternative for the simultaneous production of electricity, hydrogen, synthetic liquid fuels, heat, and chemicals [3], and CO_2 capture and storage [4]. Using coal gasification, co-production of electricity and C1 chemicals is currently the focus of poly-generation research [5-9]. To our knowledge, however, using natural gas as feedstock, the combined production of power and olefin on integrated plant has still been lack of discussion. There are abundant natural gas resources in western China to be exploited and developed, but they are remote from markets or pipe lines. Thus it is a promising option to monetize such ''stranded" gas by converting it into easily transported chemical products to bring it to market. With crude oil prices rising sharply and the public's increasing awareness of environmental issues unremitting efforts are being made to enhance the diversity of energy supply and efficient use of energy. Therefore, it is clearly necessary to develop and deploy integrated technologies to permit the conversion of natural gas into higher quality and more convenient energy carriers and/or

chemical products in an efficient and clean manner with minimal environmental impacts.

2. The integrated olefin with power generation process

Using coal-based polygeneration as a model approach, methanol and olefin are synthesized simultaneously through different chemical conversions using natural gas as the feed stock. Steam produced from the synthesization processes and fuel gas released from the chemical conversion process is sent to a power plant to generate electricity. Combined olefin with power generation is schematically presented in Fig.1. In the gasification process, syngas is produced by steam reforming in a multi-tubular fixed-bed reactor with a nickel catalyst at about 860-850   and 18-20 bar. Energy for the strongly endothermic reactor is provided by heating the reactor tubes in a furnace by combustion of natural gas or some fuel gas [10]. The main equilibrium reactions are:

$$CH_4 + H_2O = CO + 3H_2O \tag{1}$$
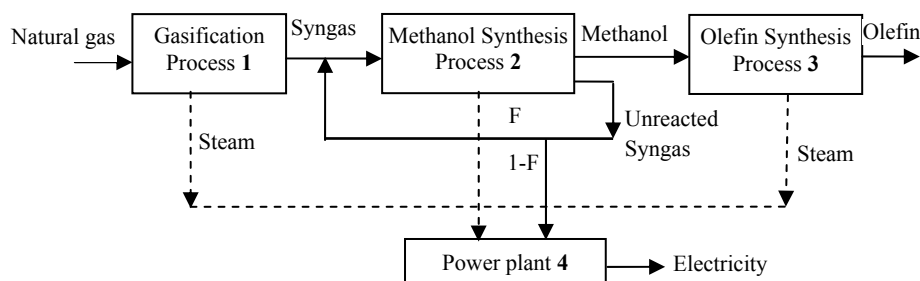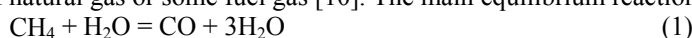$$CO + H_2O = CO_2 + H_2 \tag{2}$$

Fig.1 Superstructure block diagram of co-production of chemicals and energy.

The syngas produced from the gasification process is compressed and sent to the methanol process as the feed material. The methanol synthesis reactor uses a fixed bed of copper-based catalyst at 200 - 280   and 80 - 100 bar [11]. The main reactions are:

$$CO + 2H_2 = CH_3OH \tag{3}$$
$$CO_2 + 3H_2 = CH_3OH + H_2O \tag{4}$$
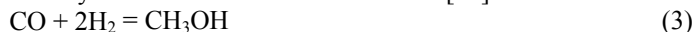
Both reactions are exothermic, so the temperature of the reactor needs to be kept at about 250   by controlling the evaporation pressure of the cooling water in order to maximize the equilibrium conversion. The reactor effluent is cooled to 40-45   to condense the crude methanol product and the unreacted gas is separated. The unreacted gas is divided into two streams in a separator: one is compressed and recycled by mixing with fresh gas as a recycle steam in order to increase the production of chemical products, the other is sent to the power side as a fuel feedstock.

In the superstructure both producers and consumers of thermal energy are given. The output stream (CO, H_2 and H_2O) has a higher temperature than needed in the methanol synthesis reactor. Cool water cools the synthesis gas and heat is recovered to generate steam. In the production of methanol, the synthesis reactions are exothermic. Hence, product streams have large amounts of thermal energy and can likewise be used elsewhere in the structure.

Several process integrations are performed in this proposed poly-generation system. Instead of being discharged to the environment, the purge stream from the methanol synthesis unit is recovered and utilized as fuel feed. In addition, on the basis of energy

integration, the usable heat released from the gasification process, the methanol process and the olefin process is recovered to heat the water for the steam from the power side.

3. Flowsheeting simulation

Aspen Plus is applied to simulate the whole polygeneration system. Models of the cogeneration plant consist of four main sections: (1) the gasification process, (2) the methanol synthesis process, (3) the olefin synthesis process, and (4) the power plant. The proposed polygeneration system is shown in Fig.2. A series of specified parameters for the process are selected based on the literature [10-13], while the thermodynamic properties (RK-SOVE, ELECNRTL, PR-BM) are selected specifically for each process.
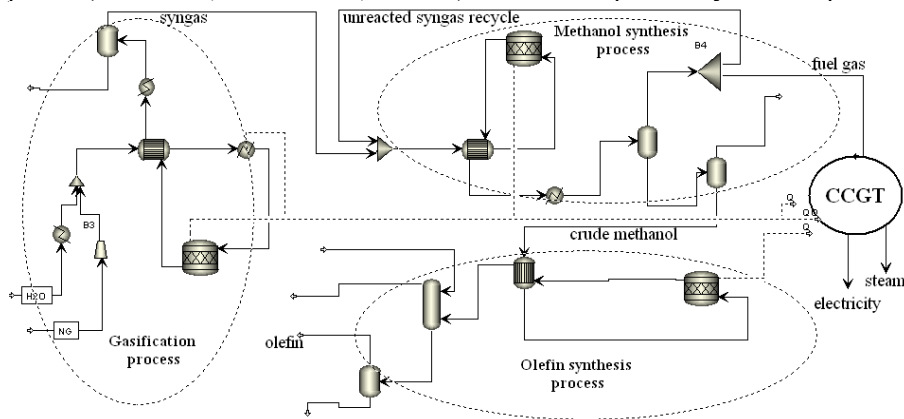

Fig.2 Flowsheet for simulation of the polygeneration system

In natural gas-based poly-generation of olefin and power, the effect of temperature and pressure of each reactor on their chemical conversion is limited. The recycled unreacted syngas in the methanol synthesis process, however impacts not only the olefin production but the output electricity, so it is a key process parameter. The ratio of the recycled unreacted syngas to the whole output is defined as fraction, F. (shown at point B4 in Fig.2). The value of F ranges from 0 to 1: F=1 denotes that all unreacted syngas recycles back to the reactor in the methanol synthesis unit, and F=0 denotes that syngas passes once through the reactor and all unreacted syngas serves as the feed of power generation.

4. Thermoeconomic optimization of the polygeneration system

The main exergy losses are determined through exergy analysis, especially internal exergy destruction in chemical reaction and heat transfer process. Increasing the recycle rate of unreacted syngas in the methanol synthesis process can improve the exergy efficiency of the system, but it also increases the load on equipment in the methanol synthesis process and follow-up processes, which leads to greater equipment investment and operating costs. High exergy efficiency does not mean the optimal economic performance. Thermoeconomic analysis has been used in this work to discuss the influence of the recycled unreacted syngas on the economic performance and exergy efficiency of the system, to seek the optimal process parameters. All the exergies entering the system are defined as "fuel" and the output exergies are defined as "product". The thermoeconomic cost of products is the sum of the thermoeconomic cost of "fuel", investment and O&M costs [14-15]. For a steady flow opening system of one

product, it is easy to list the exergy cost equation.

$$c_f E_X^f + Z_{cap+OM} = c_p E_X^p \qquad (5)$$

A number of methods have been recommended for cost analysis of polygeneration systems. Some are based on rigorous accounting procedures in which the cost of each involved energy/exergy stream is determined. Others are based on direct cost accounting, which allocates all cost components between chemical products and electricity according to certain rules of thumb such as exergy pro-rating or cost allocation based on functional considerations [16]. A thermoeconomic optimization model of the polygeneration system is built. The trade-off between exergetic efficiency and economic benefit of the system can be determined by solving the optimization model, which minimizes energy consumption and unit cost and maximizes the profit. The thermoeconomic model of the polygeneration system is presented in Fig. 3.
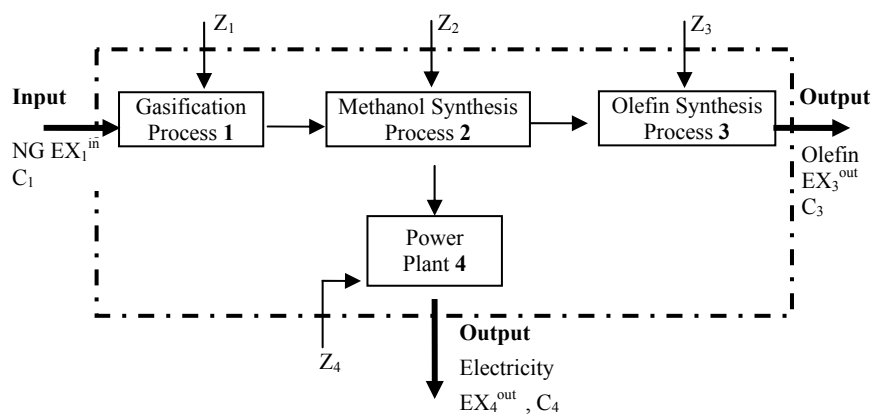
Fig.3 Thermoeconomic optimization model of the proposed polygeneration system

In order to quantify the influence of process parameters (F value) on different units through the whole polygeneration system, the system is simulated under different F values. When the amount of recycled unreacted syngas increases, other operating conditions remain unchanged, the reaction equipment's volume in this unit increases correspondingly. In the methanol synthesis process, when changing the value of the fraction, the production of methanol increases. So the exergy of the feed syngas conversion into methanol also increases, the relationship of the increased rate with F are shown in Fig. 4.
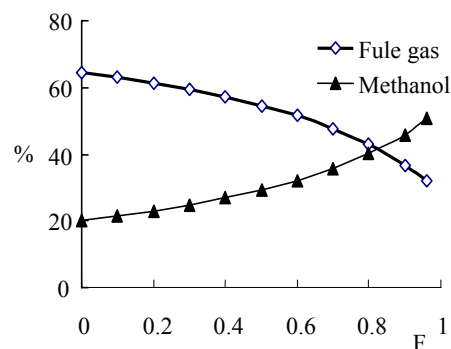
Fig.4 Percentage of the exergy of methanol/fuel gas in syngas feed at different F

Systematic Analysis and Design of the Poly-Production Process for Power and Olefin from Natural Gas

In contrast to methanol, amount of the purge fuel gas decreases with the conversion rate of syngas, and exergy of the syngas conversion into the fuel gas decreases correspondingly. Since the feed of the olefin synthesis process is the product of the former process, the production of olefin grows with the increase of *F*, as shown in Fig.5.

Fig.5 Effect of *F* on the production of olefin and power

The thermoeconomic optimization model of the polygeneration system is built based on the exergy cost equation, the exergy equilibrium equation, and the polynomial regression functions.

$$Object: \quad \max \; profit = \sum Ex^{out}c - \sum Ex^{in}c - \sum Z_i \qquad (6)$$
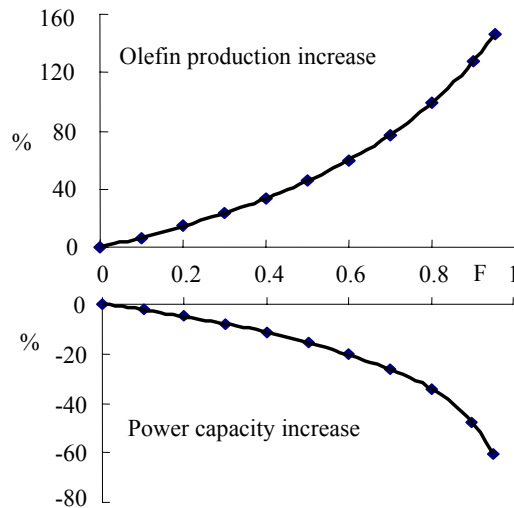
where Ex^{out} and Ex^{in} denote the output and input exergy, respectively; c is the unit exergetic cost; $\sum Z$ denotes the sum of equipment depreciation, operation management fees, maintenance costs, etc in the system.

$Constrain$: $\quad EX_i^{out}=f_i \cdot EX_i^{in} \quad (i=1, 2, 3, 4)$
$\qquad\qquad\quad EX_i^{in}=EX_{(i-1)}^{out} \quad (i=2, 3, 4)$
$\qquad\qquad\quad Y_i=\lambda_i(F) \qquad (i=2, 3, 4)$
$\qquad\qquad\quad Z_i=Z_i^0 \cdot Y_i^{0.62} \quad (i=2, 3, 4)$

where i denotes the number of the subsystem; f_i denotes the function of the input-output exergy of the i th process; λ_i denotes the function relationship between the increased rate of equipment volume and F, which can be obtained by polynomial regression; Z_i^0 denotes the sum of equipment depreciation, operation management fees, maintenance costs, etc. in the i the subsystem at F=0; and 0.62 is the scale index.

The optimal value of *F* and the maximum economic benefit are obtained by solving the thermoeconomic optimization model of the proposed polygeneration system. Fig.6 shows the relationship between economic benefit and olefin production when changing the value of *F*. Olefin production and exergy efficiency increase with the increase of *F*, but the profit of the system decrease significantly when the value of *F* exceeds 0.85. The main reason is that increasing *F* results in greater demands regarding the volume of equipments, investment, and operating costs. To achieve the best overall system

performance, a trade off between exergy efficiency and economic profit, and F=0.78 is obtained as the optimal point of the system.
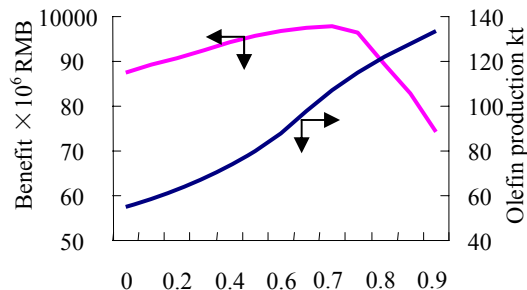

Fig.6 Relationship between system benefit and olefin production

5. Conclusions

This paper proposes a concept for efficient use of natural gas in the cogeneration of electricity power and chemical products (methanol, ethylene, and propylene) in one integrated facility combining CCGT technology with GSMTO synthesis. Systematic analysis indicates that the proposed polygeneration system is more thermodynamically efficient and economically viable than single purpose power generation and chemical products plants. Exergy destruction occurs mainly in the gasification process and the power plant, which account for 36.4% and 42.1% of the whole exergy loss, respectively. Physical exergy destruction in heat transfer and chemical exergy destruction in the combustion process are the main reasons for the exergy loss of the polygeneration system. A thermoeconomic optimization model of the polygeneration system is built, and a conceptual design of the system is conducted..

References

[1] K. Yamashita and L. Barreto, Energy, 30(2005)2453.
[2] P..M. Heteu and L. Bolle, International Journal of Thermal Sciences, 41(2002)1151.
[3] W. D. Ni, Z. Li, and X. Yuan, Paper presented to the workshop on East Asia energy futures, Beijing, 2000.
[4] H. Jens, Z. Li, S. S. Xu, Appl Energy, 86(2009)219.
[5] P. Chiesa, S. Consonni, T. Kreutz, et al., International Journal of Hydrogen Energy, 30(2005)747.
[6] T. Kreutz, R. Williams, S. Consonni, et al., International Journal of Hydrogen Energy, 30(2005)769.
[7] B. Chen, L. Gao, and H. G. Jin, Journal of Engineering Thermophysics (Chin.), 25(2004)741.
[8] H. G. Jin, L. Gao, Z. Liu, and D. X. Zheng, Proceedings of the First Joint China/Japanese Chemical Engineering symposium, (2004)403.
[9] L. W. Ma, W. D. Wei, Z. Li, Power Engineering (Chin.), 24(2004)603.
[10] H. I. Wong, Bachelor of Engineering Thesis, University of Queesland, 2004.
[11] U.S.Department of Energy, 2001, http://www.netl.doe.gov/publications/factsheets/program/ prog049.pdf.
[12] F. J. Keil, Microporous and Mesoporous Materials, 29(1999)49.
[13] L. G. Zheng, and E. Furimsky, Energy Conversion and Management, 44(2003)1845.
[14] G. Temir, and D. Bilge, Applied Thermal Engineering, 24(2004)2689.
[15] A. S. Nafey, and H. E. Fath, and A. Mabrouk, Desalination, 201(2006)224.
[16] O. A. Hamed, A. A. Hamed, and A. A. Holayil, Energy, 31(2006)2699.

Multi-Objective Design for the Consequential Life Cycle Assessment of Corn Ethanol Production

Arinola Abiola[a], Eric S. Fraga[a] & Paola Lettieri[a]

[a]*Centre for CO$_2$ Technology, Department of Chemical Engineering, University College London, Torrington Place, London WC1E 7JE.*

Abstract

Sustainable process design involves the simultaneous consideration of economic, environmental, and social factors. This approach is demonstrated in the production of corn ethanol by using a life cycle assessment (LCA) methodology. The system boundaries considered are corn farming and transportation, and ethanol production and distribution. The analysis is extended to studying the impact of alternative processing technologies on sustainability performance. The goals of this work are achieved by mixed integer nonlinear programming (MINLP) using the General Algebraic Modelling System (GAMS). Multi-objective optimisation techniques are used in assessing the trade-offs between the sustainability aspects.

Keywords: Process design, bioethanol, life cycle assessment (LCA), multi-objective optimisation

1. Introduction

The depletion of fossil fuel reserves and the environmental concerns relating to their use are some of the reasons for the renewed interest in biofuels. Some of the arguments in favour of biofuels, relative to fossil fuels are lower greenhouse gas emissions and renewability. Despite these proposed benefits, the last few years have witnessed an intense debate regarding their efficacy.

Bioethanol is the most common biofuel in the World today and has long been recognised as a cleaner-burning alternative to gasoline. Over 50% of global bioethanol production was produced from corn last year, making it the most widely used feedstock [3]. The viability of corn-ethanol as an alternative to gasoline has been studied independently on the basis of different criteria - energy balance, greenhouse gas emissions, and economic viability [1]. Independent consideration of the different aspects, however, often leads to designs which have destructive effects on process sustainability. We therefore propose a methodology which can be used to concurrently evaluate the efficacy of the corn-ethanol life cycle using multiple criteria.

A life cycle assessment (LCA) methodology is applied in evaluating the sustainability impact of the corn-ethanol process. LCA studies can be attributional or consequential [2]. Attributional LCA studies describe the environmental properties of a process and its subsystems while consequential studies describe the effect of changes within the process life cycle. The LCA study of biofuels has, however, been limited to the attributional aspect [5]. The attributional LCA method is therefore extended to assessing the economic and social performance indices of the process within a consequential LCA framework. Consequently, the novelty of the methodology adopted herein is

the assessment of the consequential aspect of the process life cycle in a context which incorporates sustainability issues. This is achieved by studying the impact of alternative technologies for the transformation, separation, and co-product recovery stages of the process.

This article therefore presents, for the first time, a model suitable for use within a multi-objective design optimisation framework to assess the trade-offs between the economic, social, and environmental performance of corn-ethanol use as fuel, while simultaneously studying the impact of alternative technologies. The goal is to determine *a-priori* the possible optimal designs which can be operated for the corn ethanol plant.

2. Methodology

The process is modeled using the GAMS/LindoGlobal software [4] and designed for an annual ethanol production rate of 40M U.S gallons. A simplified conventional corn-ethanol process is depicted in Figure 1. The first stage is the washing and milling (feedstock handling) of the feedstock. Corn is hydrolysed to glucose (hydrolysis) which is fermented to hydrous ethanol (fermentation). Hydrous ethanol is separated in the fermenter output stream from unused feedstock components (distillation). The hydrous ethanol stream is then separated into anhydrous ethanol and water (dehydration). The bottoms product from the distillation column is further processed as coproduct or DDGS (coproduct recovery).

Alternatives are introduced for the fermentation, distillation, and co-product recovery stages. For the transformation, the impact of converting the feedstock to ethanol by using an integrated continuous fermentation-pervaporation system (CF-scheme) is studied as an alternative to the conventional batch fermentation process (BF-scheme). For the separation stage, the conventional distillation-molecular sieve dehydration process (D-scheme) may be substituted by an integrated distillation-pervaporation system (P-scheme). Anaerobic digestion of coproduct (AD-scheme) is studied as an alternative to DDGS production. Consequently, a BF-D-DDGS scheme can be used to describe the conventional corn ethanol process which involves batch fermentation for transformation, distillation for the separation, and DDGS production for the coproduct recovery.

LCA, being a multidisciplinary tool can be used to estimate numerous environmental impacts including eutrophication, global warming, etc. The global warming potential (GWP) has, however, been selected as the environmental performance index because an attributional LCA study indicates it is the major environmental impact associated with the corn ethanol life cycle. The total cost is selected as the economic index, and is estimated by summing up the annualised equipment and operating costs. Social issues, in principle, may include job creation, the use of child labour, protection of local citizens, provision of friendly working conditions and so on but many of these are difficult to quantify. This partly explains why the International Organisation for Standardisation (ISO) have not published the metrics for quantifying social sustainability. One important social issue in this area, however, is the diversion of land from food crops production to the production of biomass specifically for liquid biofuel generation. The social aspect has consequently been modelled as the land area required

Multi-Objective Design for the Consequential Life Cycle Assessment of Corn Ethanol Production

to meet feedstock demands. This is estimated as a linear function of the amount of corn given the yield of corn.
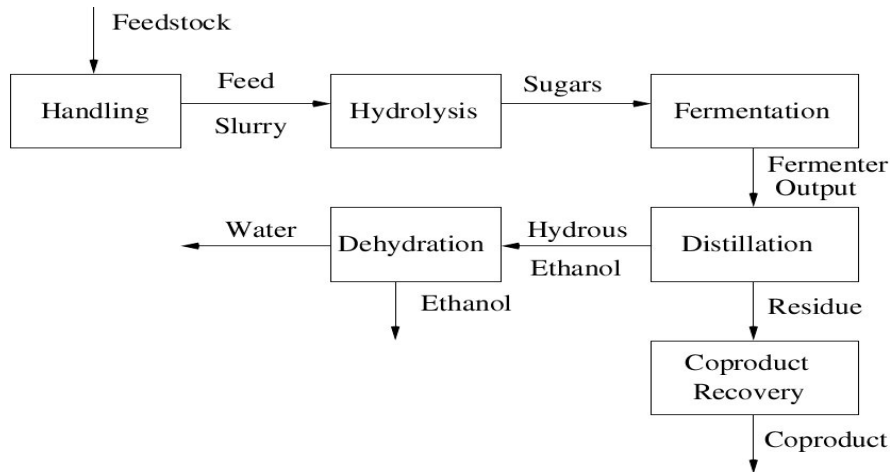
Figure 1: The corn-ethanol process

The process life cycle is implemented as a mixed integer nonlinear problem (MINLP), with a 1238 by 1274 matrix size. Multiobjective optimisation techniques are used to evaluate the trade-offs between the three aspects to obtain the Pareto optimal solutions. The degrees of freedom emanate from the design variable bounds specified in the model, and examples of this include the process operating residence times and temperatures. The CPU time expended in converging the model to a feasible solution for different combinations of α and β weighting factors ranges from several minutes to less than an hour.

The economic, environmental, and social sustainability indicators have been selected as the total cost, C_T (M\$a^{-1}), global warming impact, G (g CO_2 kg$_{ethanol}^{-1}$), and the land area, L_a (ha) required to meet feedstock demands respectively. The multi-objective optimisation problem is presented in equation 1:

$$min \ Z = \quad \alpha G + \beta C_T + (1 - \alpha - \beta)L_a$$

$$s.t \quad \begin{aligned} h(x,y) &= 0 \\ g(x,y) &\leq 0 \end{aligned} \qquad \text{(Eqn. 1)}$$

where h(x,y) refers to equality constraints (mass and energy balances), g(x,y) the inequality constraints (design variable bounds like operating time and temperature) in the model where $\alpha, \beta \in [0,1]$.

3. Results

The objective function values corresponding to the optimal economic, environmental, and social sustainability designs are presented in Table 1. (M\$ refers to millions of dollars, or 10^6\$).

Table 1: The values in each row are the criteria values for the structure identified as optimum for the single objective in the first column

Objective	L_a(kha)	Criterion C_T (M\$/a)	G (gCO$_2$/ kg$_{ethanol}$)	Structure
min L_a	29.87	84.12	664.95	*CF-P-AD*
min C_T	34.69	44.36	1025.02	*BF-D-DDGS*
min G	29.87	85.47	658.03	*CF-P-AD*

The CF-P-AD structural configuration minimises both land area and GWP, as also shown in Figure 2. The CF-designs require less feedstock than their BF-alternates because of the reduced impact of product and substrate inhibition during fermentation. The CF-scheme involves yeast recycle to the fermenter, leading to a higher input volume and lower glucose concentration. The resulting reduction in glucose concentration minimises substrate inhibition effects but this is not possible when the batch scheme is operated because of closed system operation. The effect of product inhibition is also reduced by the in-situ withdrawal of ethanol from the unit by the integration of a membrane system.

The minimisation of the GWP by the CF-P-AD scheme is the result of a lower feedstock requirement and structural configuration of the separation stage. The integrated membrane system reduces the input feed material into the separation stage, thus leading to a lower energy duty. The pervaporation scheme operated for the separation stage also requires a lower energy duty because, unlike distillation, the energy requirement is limited to the enthalpy of the permeate stream. The AD-scheme, unlike the DDGS alternate, does not contribute to the life cycle fossil energy requirement because the energy required is offset by the natural gas product stream of the process.

Although the operation of a similar structural configuration minimises G and L_a, the actual designs generated are different in each case as the objective function values show. The difference in design performance indices is due to the difference in coproduct recovery and separation inlet flowrates. More specifically, the separation block inlet flowrate for the optimal L_a exceeds that of the G design. This gave reboiler duties of 46.22gJh^{-1} and 42.92gJh^{-1} respectively and this explains the observed difference in the G value of both designs. The other factor responsible for the difference in G values is the higher retentate flowrate into the anaerobic digester unit. In the optimal G design, the natural gas production rate was able to offset all of the plant energy requirement, and vice versa for the optimal area design. The higher inlet volumetric rate increased the anaerobic digester cost from 3.26M\$/a in the L_a design to 4.15M\$/a in the G design. The digester residence time also changed from a value of 10 days to 18 days in the latter case. The observed constancy in the land area index can be attributed to the fact that neither of the separation or coproduct recovery stages contributes to the actual transformation stage.

Multi-Objective Design for the Consequential Life Cycle Assessment of Corn Ethanol Production

The BF-D-DDGS design minimises the total cost as shown in Table 1. The BF-schemes are generally cheaper than their CF-alternates because of the integrated pervaporation system. The modular nature of the membranes in the pervaporation unit results in lower economies of scale, and the same reason explains the economic advantage of distillation over pervaporation. The AD-scheme is more expensive than the DDGS design because the slow rates of waste metabolism by methanogens results in a higher capital outlay because of the residence times involved. It was found, in addition, the operating cost savings offered by the AD-scheme in terms of natural gas provision is not outweighed by the equipment cost increase.

Figure 2 shows the different structural configurations that trade-off the different criteria.
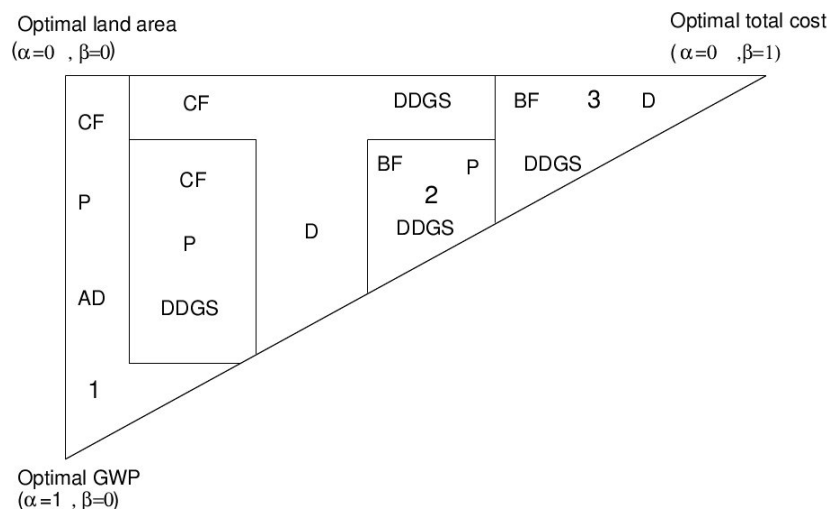
Figure 2: This diagram represents the domain of the multi-objective optimisation search space with the weighting factors α and β as independent variables. The points at the three vertices are representative of mono-criteria solutions where just one sustainability issue contributes towards the objective. The lines along the vertices represent bicriteria points where two of the three objectives count. In addition, any point inside the ternary diagram is representative of a design which simultaneously takes all three criteria into account, albeit at differing degrees. Furthermore, the lines within the rectangle, although depicting regions where the three criteria count towards the performance objective, also signify changes between the generated designs.

The batch transformation designs dominate the design space at regions where the total cost contributes more than the environmental and social sustainability objectives to the overall performance index. The prominence of the continuous transformation scheme in the objective space increases as the land area and GWP contribute more to the performance objective for earlier described reasons. The choice of operating region or design might depend on a number of factors such as the geographical location and decision maker preferences. Point 1 in Figure 2 for example, might be of interest to a design engineer who values environmental and social sustainability issues over economic performance. Point 2, on the contrary, might be the design of choice in cases where social concerns are not as important as environmental and economic issues. Finally, a point 3 design might be taken further for implementation when a decision maker places a higher priority on economic performance.

On a more general note, a change in design from the CF-P-AD design to the CF-P-DDGS is witnessed as α increases contrary to the switch to the CF-D-DDGS design when α=0. This is explained by the resulting increase in the contribution of the GWP to the objective function as the CF-P-DDGS design offers a better environmental performance than the CF-D-DDGS scheme for afore-described reasons. The switch to the conventional BF-D-DDGS design as β approaches 1 is explained by the higher contribution of the total cost to the performance index. As α increases, the design switches to a BF-P-DDGS design, which is more expensive than the BF-D-DDGS scheme, as β approaches 1. The greater contribution of the GWP to the objective function as α increases results in a switch to a design which is more environmentally friendly as α+β approach unity. It can also be observed from Figure 2 that the CF-P-AD design completely dominates the design space as α approaches 1. In this region, the GWP contributes more to the objective function than the other indicators. The CF-P-AD scheme offers the best environmental performance for earlier mentioned reasons, thus explaining the switch.

4. Conclusions

The efficacy of corn ethanol use as an alternative to gasoline has been a subject of intense debate. The viability of the process has been evaluated on the basis of the conventional design (BF-D-DDGS) adopted in industry. Furthermore, the LCA studies which have been performed on this process have been limited to attributional analysis.

This paper has presented the optimal design of the processes for generating corn ethanol through the assessment of sustainability with multiple performance indicators. The performance indicators include the economic, environmental, and social aspects. The process models include alternative technologies for key process aspects - the transformation, separation, and coproduct recovery.

The results obtained indicate that the major advantage of the conventional BF-D-DDGS design is a lower cost of operation. In addition, it has been shown that trade-offs exist between the different sustainability aspects, so other designs can offer better benefits than the conventional industrial designs. Alternative designs have been identified which can enable a decision maker find the optimal path to process improvement.

References

[1] Roel Hammerschlag. Ethanol's Energy Return on Investment: A Survey of the Literature 1990-Present. *Environmental Science Technology*, 40(6):1744-1750, February 2006.
[2] P.Miettinen and R.P Harmalainen. How to benefit from decision making in environmental life cycle assessment (LCA). *European Journal of Operational Research*, 102(2):279-294, October 2007.
[3] Renewable Fuels Association. Available at www.ethanolrfa.org/index.shtml
[4] Richard E. Rosenthal. GAMS – A User's Guide. Technical report, GAMS Development Corporation, Washington, D.C, USA., 2008.
[5] Michael Wang. Lifecycle Energy and Greenhouse Gas Emission Impacts of Different Corn Ethanol Plant Types. *Environmental Research Letters*,. 2:1-13, May 2007.

A Meta-Heuristics Approach for the Design and Scheduling of Multipurpose Batch Plants

Nelson Chibeles-Martins[a,c], Tânia Pinto-Varela[§,a,b], Ana Paula Barbósa-Póvoa[b], A. Q. Novais[a]

[a]*Unidade de Modelação e Optimização de Sistemas Energéticos (DMS- INETI), LNEG, Lisboa, Portugal*
[b]*Centro de Estudos de Gestão, IST, UTL, Av. Rovisco Pais, 1049-101 Lisboa, Portugal*
[c]*Centro de Matemática e Aplicações, CMA, FCT-UNL, Qta da Torre, 28259-516 Caparica, Portugal*

Abstract

The optimal design and scheduling of multipurpose batch plants is a complex task, since the design of the plant resources and the associated schedule should be undertaken together. Due to the nature and dimension of these problems, they often result into large Mixed Integer Linear Program (MILP) formulations that come associated with a high computational burden. In order to overcome this difficulty, a meta-heuristic approach, based on the Simulated Annealing (SA) methodology, is developed and tested along this paper. Sensitivity analysis is performed to the main meta-heuristic parameters. Several examples proposed by Pinto *et al.*(2008) are solved using the SA and the results are compared to an exact optimization approach.

Keywords: Design and Scheduling, Simulated Annealing, MILP, Batch Plants

1. Introduction

In its most general form, batch production involves general purpose facilities where a variety of products is produced by sharing the available resources (e.g. equipment, man-power, utilities, etc.). Such plants are defined as multipurpose batch facilities (Barbosa-Póvoa, 2007). The design of such facilities considers the selection of the number, type and capacities of the resources involved, as well as the definition of its operability (periodic or non-periodic), so as to produce a set of products while guaranteeing a set of pre-defined conditions and optimizing a given objective. Due to the inherent flexibility of the multipurpose resources utilization, where the same resource can be used to perform different tasks, operational scheduling considerations need to be taken into account at the design stage. Thus, in order to guarantee optimal solutions, most design formulations based on mathematical programming approaches have to consider a large number of resource items, out of which are selected the ones that are incorporated into the final plant design. This fact, together with the complexity of the recipes, leads to large Mixed Integer Linear Programming, MILP, problems, often associated with a high computational burden that grows together with the problem dimension. Despite the fact

---

§ Author to whom correspondence should be addressed, e-mail: tania.pinto@ineti.pt

that nowadays is possible to solve problems and models which were impossible to handle a few years ago, the development of efficient tools is still required, which thus constitutes an open area of research.

The research in this area has been mainly focused on the use of mathematical programming models (MILP and MINLP). These models when applied to real problems often become intractable. For this reason new approaches need to be explored to overcome this drawback as stated by Barbosa-Povoa (2007). These approaches may span from exact algorithms where decomposition techniques are explored till problem oriented heuristics, evolutionary algorithms, meta-heuristics, as well as hybrid methods.

However, few works have been published where such methods have been explored. For the case of Simulated Annealing some cases where reported. Xia and Macchietto (1997) using an evolutionary algorithm with simulated annealing also addressed the generic design of multipurpose batch plants, but where no detail on the plant topology was contemplated. The method appears as quite efficient and good results were reported. Later on Raaymakes *et al.*(2000) explored the scheduling in multipurpose batch plants using Simulated Annealing, but their models handled the scheduling while ignoring the plant design.

In this paper a Simulated Annealing algorithm is developed for the simultaneous design and scheduling of multipurpose batch plants where detailed plant topology is considered. A number of instances are tested and the statistical results presented and compared to results of a previous work where the use of mathematical programming methods was explored (Pinto et al., (2008). In all cases, the plant topology, scheduling, process equipment design and storage profiles are determined.

2. Design Problem

The optimal plant design can be obtained by solving the following problem:
Given:
- Process description, through a RTN representation;
- The maximal amount of each type of resource available, its characteristics and unit cost;
- Time horizon of planning;
- Demand over the time horizon (production range);
- Task and resources operating cost data;
- Equipment and connection suitability;

Determine:
- The amount of each resource used;
- The process scheduling;
- The optimal plant topology as well as the associated design for all equipment and connectivity required.

while maximizing the plant profit.

A non-periodic plant operating mode defined over a given time horizon is considered. Mixed storage policies, shared intermediated states, material recycles and multipurpose batch plant equipment units with continuous sizes, are allowed.

3. Modelling framework

The meta-heuristic approach developed in this paper uses the Simulated Annealing algorithm as proposed by Kirkpatrick *et al.* (1983) and Cerny *et al.* (1985), but adds up

several adaptations so as to improve the algorithm's efficiency and effectiveness. These adaptations explore the characteristics of the problem in study.

Simulated Annealing can be classified as a Local Search Meta-Heuristic. It can be initialized with a random or with a constructive heuristic and improved iteratively. For each iteration the algorithm chooses a solution from the neighborhood of the current one. In order to prevent an early stop of the algorithm on a local optimum, a mechanism based on the Metropolis Algorithm was incorporated, as well as the tuning of some parameters was undertaken, to guarantee efficiency and effectiveness. In particular a sensitivity analysis is performed in the following parameters:

- Initial temperature
- Cooling schedule
- Stop criterion

Other features are tailored based, such as:

- Objective function
- Initial solution generation
- Neighborhood function

The following *pseudocode* represents the developed algorithm where the s_0 represents the initial solution, s_i the current solution on iteration i, s'_i the neighbor solution generated on iteration i and $f(s)$ is the objective function. P_{ac} is the probability of accepting a worse neighbor solution and T_i stands for the temperature at iteration i.

```
Parameters initialization
i = 1, s₁ = s₀
while not Stop Criterion:
    Generate(s'ᵢ);
    If f(s'ᵢ) > f(sᵢ) then
        sᵢ = s'ᵢ
    else
        Pac = exp((f(s'i) - f(si)) / Tᵢ)
        u = random[0,1]
        if Pac > u then
            sᵢ = s'ᵢ
    Tᵢ₊₁ = f(Tᵢ)
    i = i + 1
Stop
```

The algorithm stops after a defined number of iterations without improvements in the objective function. The parameters were empirically adjusted.

Solutions are treated as matrices $[n \times H]$, where n represents the number of equipment units and H the time horizon. For each cell, the algorithm controls if a task starts in that moment and how much it is going to process. A neighbor solution s' is then generated from the current solution by randomly selecting a small increase or decrease in the batch size, the use of another equipment or both alternatives.

The algorithm developed for this work uses a Geometric Cooling Scheduling ($T_{i+1} = \alpha T_i$). However, the proposed algorithm presents several modifications, such as:

- The best solution found in all iterations is recorded and presented as the solution proposed by the algorithm;
- On the neighbor generation phase, if a task ended processing a batch below a pre-fixed operating level that task is removed from the solution;

- On the neighbor generation phase, if the selected equipment unit is not in use, the algorithm tries to assign it a task that will process a small random quantity of product, but sufficiently above the pre-fixed operating level.

The proposed algorithm is explored and compared with two more methodologies, which are the Resource-Task-Network and the proposed formulation for the design of multipurpose batch plants, i.e. the designated Resource-Task-Network adapted by Pinto *et al.* (2008), both use a MILP approach.

4. Examples

All the examples presented were solved in a Pentium 2 Duo, T7300, 2 GHz, 2 GB RAM. The MILP approach used GAMS 22.6/CPLEX 11.

4.1. Example 1

A plant is to be designed at a maximum profit to produce 80 ton of a single product (S3) from two raw materials, (S1) and (S2). Two different processing Tasks are considered. Task T1 transforms S1 into S3 after 1 hour, and Task T2 processes S2 during 2 hours and generates S3. A single campaign non-periodic mode of operation is assumed over a time horizon of 5 hours. In terms of equipment, three storage tanks are available (V1, V2 and V3) to store respectively, S1, S2 and S3, and a multipurpose reactor is suitable to process Task T1 and T2. Vessels V1 and V2 are connected to reactor R1 (connections c1 and c2), while the latter is also connected to vessel V3 (connection c3). The plant design provides in detail the plant structure, as well as the operational schedule.

The SA algorithm was applied and some statistic analysis obtained after a thousand runs, as shown in Table 1.

Table 1- Statistic analysis from the SA.

	OF x10^3 (m.u.)	Gap to F*	CPU (s)
Minimum	1247.60	0.075%	0.109
Median	1248.06	0.039%	0.180
Maximum	1248.54	0.000%	0.359
Average	1248.11	0.034%	0.186
Standard Deviation	0.242	0.019%	0.045

m.u: monetary units

The same example was solved using the MILP approaches and the computational results are presented in Table 2.

Table 2- Computational results.

Methodology	Tot. Var.	Bin. Var.	Const.	CPU (s)	OF x10^3 (m.u.)
RTN	504	223	651	0.171	1248.5
RTN-adapted	129	35	186	0.062	1248.5

m.u: monetary units

When analyzing the solutions obtained through the SA algorithm, it can be noticed that the objective function value is very near the optimal value (very small Gap to F*

column in Table 1). However, the CPU time used by the SA was slightly higher than MILP approaches. This is however expected due to the reduced dimension of the example.

4.2. Example 2

The example proposed by Pinto *et al*. (2008) was adapted to satisfy a higher demand and an increased time horizon.

A plant is to be designed at a maximum profit so as to produce three final products, S4, S5 and S6, with production capacities between [0:800] tons for S4 and S5, and [0:600] tons for S6, from two raw materials, S1 and S2. The process operates in a non-periodic mode over a time horizon of 120 hours. In terms of equipment suitability, reactors R1, R2 and R3 are multipurpose in nature. Task T1 may process S1 during two hours in R1 or R2 producing the unstable material S3; Task T2 may process S2 during two hours in R2 or one hour in R1 to produce S4, which is both an intermediate material and a final product; Task T3 processes 0.5 of S3 and 0.5 of S4 in R3 during three hours producing S5, which, like S4, is both an intermediate and a final product; Task T4 may process 0.5 of S3 and 0.5 of S5 in R3 during two hours producing the final product S6. Each vessel is suitable to store only one material State. For each equipment unit the capacity is defined in a continuous range with a minimum and a pre-specified maximum.

To apply the SA algorithm, the neighborhood function, previously defined, had to be adjusted in order to accommodate the unstable material, S3. The constraints characterizing the production demand of final materials S4, S5 and S6 were transformed into flexible constraints and were allowed to be temporarily violated during the local search phase.

Similarly to the previous example, some statistical analysis was made and presented in Table 3, after one hundred runs; the computational results from the MILP approaches are shown in Table 4.

Table 3- Statistic analysis from the SA.

	OF x10^3 (m.u)	Gap to F[1]	CPU time (s)
Minimum	2263.89	3.08%	2 187.2
Median	2294.00	4.76%	2 351.1
Maximum	2322.17	6.37%	2 567.0
Average	2291.51	4.79%	2 354.1
Standard Deviation	13.64	0.603%	83.05

m.u: monetary units; value of incumbent solution from BB, F[1] = 2400.5 x 10^3

Table 4 - Computational results.

Methodology	Tot. Var.	Bin. Var.	Const.	CPU (s)	Gap	OF x10^3(m.u.)
RTN	24 071	9 262	34 568	220 000	6.6 %	2400.5
RTN-adapted	8 694	2 182	14 209	220 000	4.6%	2400.5

m.u: monetary units

The RTN methodology obtained the value of 2400.5 x 10^3 m.u. for the objective function (F^1) in 220 000 s with a 6.6 % optimality gap (Table 4). The maximum productions for S4, S5 and S6 were satisfied. The RTN-adapted methodology using the same cpu time as RTN methodology reached the same value of 2400.5 x 10^3 m.u. for objective function with a 4.6% optimality gap. Like the RTN, the maximum production for all the final products is satisfied.

Comparing the results between the proposed algorithm and the MILP approaches, it can be seen in Table 3, that the SA obtains solutions very close to the incumbent solution produced by the Branch and Bound Algorithm (Table 4), requiring only 1% of the CPU time. Therefore, it can be concluded that the SA approach appears as a very good alternative to deal with complex design and scheduling problems of multipurpose batch plants, being able to produce solutions within a reasonable margin to the optimum in a much more efficient way.

5. Conclusions

In this paper, the design of multipurpose batch plants is explored, where different aspects are considered: plant topology, equipment design, scheduling and storage policies of multipurpose batch plants. A meta-heuristics approach, i.e. Simulated Annealing, is proposed and compared to other MILP approaches. The SA proposed converges to optimal solutions or to very good sub-optimal solutions, in a small amount of time.
By comparison with the MILP approaches, the scope offered by Meta-Heuristics is clearly demonstrated in dealing with complex problems, where MILP approaches are not efficient due to problem intractability. A good indicator of the viability of Meta-Heuristics is also illustrated by the possibility of constructing a Neighborhood Function, which enabled the algorithm to converge to sub-optimal solutions.
On-going research addresses the further tuning of the proposed SA approach and its application to more complex examples, in order to more firmly establish its potential as an alternative optimization approach to MILPs, when the application of the latter may lead to problem intractability.

References

A. P. Barbosa-Povoa (2007). "A critical review on the design and retrofit of batch plants." Computers & Chemical Engineering, 31(7): 833-855.

V. Cerny (1985). "A thermodynamical approach to the travelling salesman problem: an efficient simulation algorithm." Journal of Optimization Theory and Applications, 45: 41-51.

T. Pinto, A. Barbosa-Povoa and A. Q. Novais (2008). "Design of multipurpose batch plants: A comparative analysis between the STN, m-STN, and RTN representations and formulations." Industrial & Engineering Chemistry Research, 47(16): 6025-6044.

C. D. G. J. S. Kirkpatrick, and M. P. Vecchi (1983). "Optimization by Simulated Annealing " 220 (4598): pp. 671 - 680.

J. A. H. M. Raaymakes (2000). "Scheduling multipurpose batch process industries with no-wait restrictions by simulated annealing." European Journal of Operational Research, 126: 131-151.

Q.S. Xia and S. Macchietto (1997). Design and synthesis of batch plants MINLP solution based on a stochastic method. Computers & Chemical Engineering, 21, S697-S702.

Personal micro-blogging for workflow support in conceptual design

Michael Wiedau[a], Frank Alsmeyer[a], Andreas Harwardt[b], Wolfgang Marquardt[b]

[a] AixCAPE e.V., Intzestr. 1, 52072 Aachen, Germany,
{wiedau / alsmeyer}@aixcape.org
[b] AVT – Process Systems Engineering, RWTH Aachen University, Turmstr. 46, 52072
Aachen, Germany,   {andreas.harwardt / wolfgang.marquardt}@avt.rwth-aachen.de

Abstract

This contribution reports on how the conceptual chemical process design can benefit from workflow support using tracing and micro-blogging as part of a process design software environment. Micro-blogging allows users to write short messages on what they are doing, feeling or thinking. Micro-blogging is suggested to be integrated into the user interface of a conceptual process design software environment to support the engineer during his/her work. By adding features to the user interface, micro-blogging can informally extend the documentation of the design process beyond what the machine can log by itself. During the development of a new chemical process in the conceptual design phase, the engineer will not only be informed on what the machine logged in earlier design steps but will also be provided with former corresponding messages other engineers manually added to the system during their design activity. The important difference between the commonly known micro-blogging systems and the process design software concept proposed here is that the messages are stored locally on the client and are hence only accessible from within the company but not by the public. Another difference is the analysis, filtering and reuse of former messages. This way, methods and tools used during the conceptual design of a chemical plant are acquired from the engineers who use micro-blogging as an extension of a design software system. An exploratory case study has been carried out to show how corresponding blog-entries can be retrieved and what benefits can be achieved for the design engineering team using such a system. A larger case study with industrial users is in progress to help assess practical benefits and the still existing potential for improvement of the concept and its implementation.

Keywords: workflow support, conceptual design, micro-blogging, shortcut-tools

1. Introduction

The competitiveness of the chemical product in the production lifecycle of a chemical plant is largely predetermined during the early development and reengineering phases of a chemical process. The workflow in the early phases is largely ill-defined and lacks determinism because of its creative nature. Every activity in the design process is different and every design project is dominated by different requirements, specifications and constraints.

One possibility for the support of designers in the early phases of the design process by some kind of workflow support system is to record all activities during the work process. The recorded workflow can be used in following projects or during the design

of an alternative flowsheet by analyzing and recalling relevant former design steps. After an appropriate analysis, similarities between the former and the current design process can be detected to guide the user by providing references to related activities of the design process of former projects or to alternative flowsheets of the current design.

2. Conceptual design follows uncertain workflows

In chemical engineering, the conceptual design phase is a highly creative process which is not following a predetermined, concrete workflow, as it might be true for later phases of the project lifecycle, e.g. during detail engineering and construction. Many design steps have to be handled by experienced domain experts. Each activity and the overall design process can be improved by making use of the engineer's experience. Prominent workflow support systems are suitable for domains with complete, rigid and well structured process patterns that can be accurately predetermined in advance, such as business processes, and has been widely adapted by workflow management systems [9]. Still, problems appearing in a design project have appeared before in a similar way. However, the way these problems have been solved and the experience gained during the design process are rarely documented explicitly because of a lack of time and appropriate tool support. Further, the design experience on similar problems is often not shared between designers and design teams. As a consequence, any improvement of the design process, facilitated by a targeted sharing of past experience may have a significant economic potential. Fast access to past design practice may *speed up* the design process and *save cost* in the conceptual design phase. In addition, enhanced knowledge transfer may result in *better designs*.

Several concepts to reuse past design experience have been suggested in the past. For example, relations between design alternatives have been tracked in [11] or classification schemes have been suggested to organize information from different designers along multiple dimensions by capturing the evolution and the reuse of models [12]. Other attempts reuse the experience inside specific design support tools and make use of the *Case-Base Reasoning* paradigm for design cases from the past [13]. Most of these approaches implement external experience repositories decoupled from the design support system during run-time. Typically, the user has to manually provide a problem description to receive an answer to his/her query. In recent research projects on *design process tracing* [1] or *traces based reasoning* [14], software systems have been developed to provide guidance in the form of a-priori software process integration to achieve an automated capture of process traces. While this idea is attractive, automatic tracing and trace interpretation in creative work processes is challenging and has not yet been validated in chemical process engineering in an industrial context. In this contribution we therefore want to investigate an alternative to automatic tracing, i.e. we attempt to capture experience in previous projects in a less formal and automated way by extending a process design software environment by personal micro-blogging functionality. Micro-blogging has become a widely accepted form of socialization on the internet. The most popular system for micro-blogging nowadays is twitter.com.

3. Personal micro-blogging

3.1. Background

We are currently developing a software toolbox that supports conceptual design with shortcut calculation methods. With shortcut methods, feasibility and economic potential of process alternatives can be quickly assessed. For example, determining the minimum energy demand of a separation has been found to be a good way of estimating the cost

of a distillation system, because operating costs are dominated by the energy demand and investment costs are closely related to the vapor flow rate in the column. The minimum energy demand can be calculated using shortcut algorithms like the RBM (Rectification Body Method) [2] or infinity/infinity analysis [7] which are both part of a conceptual design software which is currently developed at AixCAPE in collaboration with AVT.PT. To learn how engineers use these innovative methods in various contexts and to improve their creative workflow, the micro-blogging functionality described in this paper has been integrated into the shortcut toolbox

3.2. Methodology

Figure 1 shows a screenshot of the micro-blogging user interface prototype. In the upper text box, the user can enter his comments (*human entries*). Additional to these entries, machine actions are saved as *machine entries*. The workflow trace described by these machine entries is not yet analyzed but only saved as entries to the database. In the second box, the users' last entries are displayed in a historical way. In the third box at the bottom, the *corresponding entries* are displayed, sorted by relevance with respect to the last inserted entry (at 10:34:35).
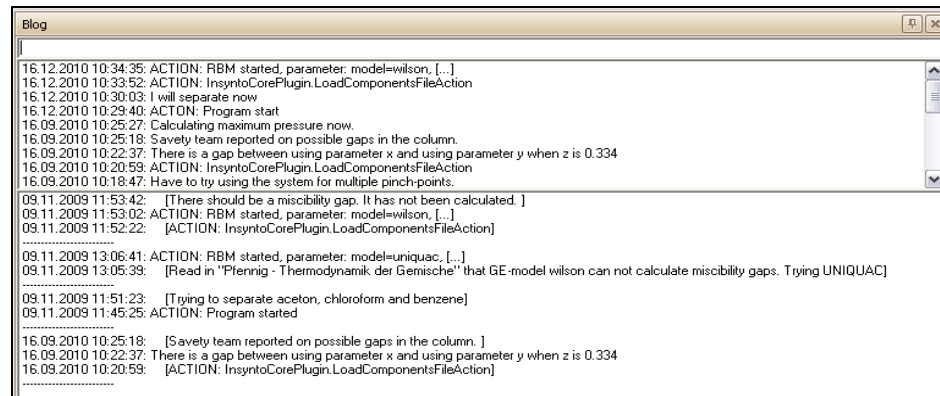

Figure 1 - Screenshot of the personal-micro-blogging software prototype. The window consists of 3 layers: 1. input-line, 2. last entries, 3. corresponding entries

All entries are saved to an embedded database which is used as a storage system in our software prototype. To find corresponding blog entries, the Vector Space Model (VSM) [4] is used to retrieve information and to filter all blog-entries according to the actually entered blog entry. VSM builds a document space of all entry messages by collecting all occurring words (called terms). Afterwards, for each entry i in the database, a vector v_i is calculated with the number of words that exist in the entry. Accordingly, a vector v_0 is calculated for the last blog entry. Similarity is measured by the cosine of the angle θ_i between two vectors, i.e.

$$\sigma_i = \cos \theta_i = \frac{v_i \cdot v_0}{\|v_i\| \cdot \|v_0\|}$$

The value σ_i represents the similarity between a blog-entry from the database and the current blog entry. The best matching entries according to this similarity are filtered and presented to the user as a list of corresponding entries. The calculated value represents the degree of correspondence of both entries. The best matching entries are presented to the designer. Due to the fact that each micro-blog-entry stays in correspondence to his previous and following entry, and this correspondence is important for the designer to

see the context of the entries, these are additionally presented to the user (shown in square brackets).

The major objective of this approach is that designers are supported with reusable knowledge from former design projects. In principle, the micro-blogging functionality can also be used as a feedback system for software developers. Whenever an error or malfunction in the underlying software system appears, a software-developer could see what actions and blog-entries the user has written while the malfunction appeared. That makes it easier to find specific software lacks for particular use cases and improves debugging procedures. A more advanced goal would be to provide project managers with a trace of the design process as a basis for its systematic improvement. For example, with the automatic tracing and analysis methods in [1], a previous workflow can be traced and analyzed by experts or project managers.

3.3. Case study

In a simple case study, we describe how the integration of micro-blogging in a conceptual design software might improve the design process by re-using knowledge. Suppose a chemical engineer uses the RBM [2] to identify the minimum energy demand of a separation. She/he uses the G^E-model parameters for the Wilson model, which are the only ones reported in the company's property database. Later, the designer recognizes that a reported miscibility gap is not calculated by the software. She/he consults with more experienced colleagues (or checks a thermodynamics textbook [3]) to find out that the problem is not related to the algorithm but rather to the Wilson model which is unable to describe two-phase behavior. With the UNIQUAC model, meaningful results could finally be obtained. Fig. 2 shows the micro-blogging history:

```
09.11.2009 11:45:25: ACTION: Program started
09.11.2009 11:51:23: Trying to separate aceton, chloroform and benzene
09.11.2009 11:52:22: ACTION: InsyntoCorePlugin.LoadComponentsFileAction
09.11.2009 11:53:02: ACTION: RBM started, parameter: model=wilson, [...]
09.11.2009 11:53:42: There should be a miscibility gap. It has not been calculated.
09.11.2009 13:05:39: Read in "Pfennig - Thermodynamik der Gemische" that GE-model wilson can not calculate miscibility gaps. Trying UNIQUAC
09.11.2009 13:06:41: ACTION: RBM started, parameter: model=uniquac, [...]
09.11.2009 13:07:36: Now, Miscibility gap calculated correctly.
```

Figure 2 - Possible history for the usage of a synthesis toolbox supporting micro-blogging. Machine entries are annotated with the word "ACTION".

One year later, another designer in the same company encounters the same problem, but for a process involving other chemical components. This time, the user is provided with corresponding entries from earlier work (see Fig. 1). One of them (09.11.2009, 13:05:39) shows that a colleague has read in [3] that a correct calculation can be done by means of the UNIQUAC model.

In this simple case study, the database did not yet contain a large number of entries from various design projects. However, it demonstrates in principle how micro-blogging can improve design processes by supporting designers with additional information.

3.4. Results and discussion

The case study illustrates the high potential for improving design processes by personal micro-blogging. The reuse of knowledge saves engineering time, it helps the designer to focus on his/her work to avoid pitfalls and finally helps to achieve better designs. There are many possible improvements to the system to increase its power and its usability. The system currently uses VSM to find corresponding entries. There are other information retrieval methods (like stop-word filtering [15]) that might improve the quality of the corresponding entry list. It is also possible to let the user rate the

corresponding-entry results such that the system can "learn" from the users' experience. Help functions of modern software systems are static and most often neither interactive nor personalized. If help entries would additionally appear in the list of corresponding entries, the benefit of the system for the user could be improved.

The system does not only support reuse of design knowledge, but may also be used to document scientific work by tracing the workflow in different case studies. The micro-blogging approach facilitates the logging and documentation on case studies in parallel.

A further scope of application for this micro-blogging plug-in is the identification of strengths and weaknesses of a design software environment or of the embedded calculation methods.

We see two main problems concerning the suggested micro-blogging approach. Firstly, designers have to formulate their micro-blogging entries manually which is time-consuming and does not result frequencies in an immediate benefit. Discussions in the community reveal that users having to enter blog entries in an input box via the computers' keyboard result in an acceptance problem, because different user types use their computer keyboard in differently. Since the quality of the workflow support obviously depends on the quality of the users' micro-blogging entries, acceptance is a crucial precondition. This problem has to be investigated in the future.

Another issue is that one word might have different meanings in different contexts. Also, different words might have the same meaning in the same context. The retrieval of corresponding blog entries might suggest entries that do not correspond with the user's intentional meaning. This problem could be avoided by means of semantic technologies like Latent Semantic Analysis [5] based on a domain ontology such as OntoCAPE [16] or the use of a lexical database like WordNet [6].

4. Conclusions and future work

We have suggested a new software concept to the improve design processes in chemical engineering by means of personal micro-blogging plug-ins into design software environments. Our prototype system analyses the activities in a design process by means of natural language log-entries provided by the user. Corresponding entries of the past are filtered according to the currently entered entry. A critical assessment clearly shows the potential of the suggested approach but also some shortcomings which need to be addressed in future work to effectively support the documentation and analysis of the design process and the design results in order to achieve the reusability of design knowledge

This research is part of a project where the software prototype is evaluated by designers in industry. We will not only evaluate if the blogging system is useful, where there are still shortcomings to be improved. Their entries will also teach us how the industrial colleagues use the shortcut design methods and their software implementation to provide feedback to the methods and system developers.

The ranking and the presentation of corresponding micro-blog entries is the main bottleneck of the current implementation of our idea. The additional use and improvement of methods of information retrieval and the rating of entry suggestions will contribute to removing the existing bottleneck.

Currently, all logged data are stored in a simple database in the software client. A central storage system within a company or a design team would facilitate interaction and collaboration. The micro-blog entries and consequently the design process can be analyzed more profoundly if domain knowledge e.g. encoded by means of ontologies [15,16] is linked to the blog-entries to result in enriched information for the designer.

To improve the detection of methodological or programming errors, the software system could send information about the workflow to the software vendor, whenever the user needs live support. The presented micro-blogging approach is currently bound to a specific software framework in the shortcut-based design environment and the data analysis software Alanda [8]. A tool-independent approach and integration into the operating system with an interface to different software tools would be preferable. In addition to the types *human entry* and *machine entry* it is also possible to store general *help-entries* as additional suggestions to the designer to refer the designer to help texts related to current design activities and blog-entries.

5. Acknowledgements

References

[1] Miatidis, M., Jarke M. and Weidenhaupt K. (2008): *Using Developers' Experience in Cooperative Design Processes*, Results of the IMPROVE Project, Springer

[2] Bausa J., Watzdorf, R. v and Marquardt, W.. (1998): *Shortcut Methods for Nonideal Multicomponent Distillation*: 1. Simple Columns, AIChE J., 44(10), 2181-2198

[3] Pfennig, A. (2004), *Thermodynamik der Gemische*, Springer Verlag, Berlin

[4] Salton G., Wong A., Yang C.S., *A vector space model for automatic indexing*, Communications of the ACM, v.18 n.11, p.613-620, Nov. 1975

[5] Deerwester S., Dumais S., Furnas G., Landauer T., Harshman T., *Indexing by Latent Semantic Analysis*, Journal of the American society for information science, 1990

[6] Fellbaum C., *WordNet - An Electronic Lexical Database*, the MIT Press, 1998

[7] Ryll, O., Blagov, S., Hasse, H.: *Rechnergestützter konzeptioneller Entwurf von Destillations-/Reaktionsprozessen*, Chemie Ingenieur Technik, 80 (2008) 207-213

[8] Soemers M., Alsmeyer, F., *ALANDA for Configuration-Free Analysis of Process Data,* 15th IFAC Symposium on System Identification, July 6 - 8, 2009, Saint-Malo, France

[9] van der Aalst, W., van Hee, K.: *Workflow Management*. MIT Press (2002)

[10] Pohl, K., Dömges, R., Jarke, M.: *Towards method-driven trace capture*. In: Proceedings of the 9th International Conference on Advanced Information System Engineering (CAiSE'97), Barcelona, Spain. (1997)

[10] Lim, W.C.: *Effects of reuse on quality, productivity, and economics*. IEEE Software 11:3 (1994) 23-30

[11] Banares-Alcantara, R., Lababidi, H.M.S.: *Design support systems for process Engineering, II. KBDS: An experimental prototype*. Computers & Chemical Engineering 19:3 (1995) 279-301

[12] Subrahmanian, E., Konda, S.L., Dutoit, A., Reich, Y., Cunningham, D., Patrick, R., Thomas, M., Westerberg, A.W.: *The n-dim approach to creative design support systems*. In: Proceedings of the ASME Conference. (1997)

[13] Leake, D., Hammond, K., Birnbaum, L., Marlow, C., Yang, H.: *An integrated interface for proactive, experience-based design support*. In: Proceedings of the 6th International Conference on Intelligent User Interfaces. (2001)

[14] Mille, A.: *From case-based reasoning to traces based reasoning*. In: 9th IFAC Symposium, Nancy, France. (2006)

[15] Kuropka, D., *Modelle zur Repräsentation natürlichsprachlicher Dokumente. Ontologie-basiertes Information-Filtering und -Retrieval mit relationalen Datenbanken*, Logos Verlag Berlin 2004

[16] Morbach, J, Wiesner, A., Marquardt, W, *OntoCAPE – A (re)usable ontology for computer-aided process engineering*. Computers and Chemical engineering, 2009

Computer Aided Design of Plate Heat Exchangers

Olga Arsenyeva[a], Leonid Tovazhnyansky[a], Petro Kapustenko[a], Gena Khavin[b]

[a]National Technical University "Kharkiv Polytechnic Institute",
21 Frunze Str., 61002, Kharkiv, Ukraine, kap@kpi.kharkov.ua
[b]AO SODRUGESTVO-T, 2 Krsnoznamenny Per., 61002, Kharkiv, Ukraine,
sodrut@gmail.com

Abstract

The computer aided design of plate heat exchanger with mixed grouping of plates is considered. It is formulated as the mathematical problem of finding the minimal value for implicit nonlinear discrete/continues objective function with inequality constraints. The optimizing variables include the number of passes for both streams, the numbers of plates with different corrugation geometries in each pass, the plate type and its size. To estimate the value of objective function in a space of optimizing variables the mathematical model of plate heat exchanger is developed. To account for thermal and hydraulic performance of channels between plates with different geometrical forms of corrugations, the exponents and coefficients in formulas for heat transfer coefficients and friction factors calculation are used as model parameters. The procedure and software for numerical experiment to identify model parameters by comparing the calculation results with those obtained with free available in web computer programs of plate manufacturers is developed. The sets of such parameters are obtained for a number of industrially manufactured plates. The described approach is implemented as software for plate heat exchangers calculation.

Keywords: Plate Heat Exchanger, Design, Mathematical Model, Model Parameters

1. Introduction

Plate heat exchangers (PHEs) are one of the most efficient types of heat transfer equipment. The principles of their construction and design methods are sufficiently well described elsewhere, see e.g. Hesselgreaves (2001), Wang, Sunden and Manglik (2007), Tovazshnyansky et al (2004). This equipment is much more compact and requires much less material for heat transfer surface production, much smaller footprint, than conventional shell and tubes units. PHEs have a number of advantages over shell and tube heat exchangers, such as compactness, low total cost, less fouling, flexibility in changing the heat transfer surface area, accessibility. Due to differences in construction principles from conventional shell and tube heat exchangers, PHEs require substantially different methods of thermal and hydraulic design. Such methods should be based on accurate enough mathematical models.

2. Mathematical model of PHE

The PHE consists of a set of corrugated heat transfer plates clamped together between fixed and moving frame plates and tightened by tightening bolts, see e.g Hesselgreaves (2001). The plates are equipped with the system of sealing gaskets, which are also separate the streams and organizing their distribution over the inter plate channels. In multi-pass PHE plates are arranged in such way, that they are forming groups of parallel channels. Such group is corresponding to one pass and the stream is going consequently

through the passes. The temperature distributions in passes are different, and in different groups of channels both counter-current and co-current flows may occur.

The mathematical model of PHE was presented by Arsenyeva et al.(2009). The PHE is regarded as a system of one-pass blocks of plates. The conditions for all channels in one such block are equal. The total number of blocks is $n_b = X_1 X_2$ and the number of heat transfer units in one block, counted for hot stream:

$$NTU_b = U_b \cdot F_b \cdot X_2 / (G_1 c_1) \tag{1}$$

where U_b – overall heat transfer coefficient in block, W/(m^{2}K); G_1- mass flow rate of hot stream, kg/s; c_1 – specific heat of hot stream, J/(kg·K); X_1 and X_2 – the number of passes for hot and cold stream, respectively.

When $G_1 c_1 / X_2 < G_2 c_2 / X_1$ block heat exchange effectiveness ε_b for counter current flow:

$$\varepsilon_b = \left[1 - \exp(NTU_b \cdot R_b - NTU_b)\right] / \left[1 - R_b \cdot \exp(NTU_b \cdot R_b - NTU_b)\right] \tag{2}$$

where $R_b = G_1 \cdot c_1 \cdot X_1 / (G_2 \cdot c_2 \cdot X_2)$ - the ratio of heat capacities of streams going trough block; G_2 and c_2 mass flow rate [kg/s] and specific heat [J/(kg·K)] of cold stream.

If $R_b = 1$, then $\varepsilon_b = NTU_b / (1 + NTU_b)$.

In case of co-current flow directions

$$\varepsilon_b = \left[1 - \exp(-NTU_b \cdot R_b - NTU_b)\right] / (1 + R_b) \tag{3}$$

On the other hand the heat exchange effectiveness of block i : $\varepsilon_{bi} = \delta t_{1i} / \Delta t_i$,

where δt_{1i} - temperature drop in block i; Δt_i - the temperature difference of streams entering block i. The temperature change of cold stream: $\delta t_{2i} = \delta t_{1i} \cdot R_b$

These relations can be regarded as mathematical model of block, which describes the dependence of temperature changes from characterising block values of F_b and U_b.

For every block we can write the equation which describes the link of temperature change in this block to temperature changes in all other blocks of PHE.

For any number of passes such system presented in matrix form:

$$[Z][\delta t_i] = [\varepsilon_{bi} \Delta], \tag{4}$$

where $[\delta t_i]$ - vector-column of temperature drops in blocks; $[\varepsilon_{bi} \Delta]$ - vector-column of the right hand parts of the system; $[Z]$ - matrix of system coefficients.

The numerical solution of linear algebraic equations system (7) easily made on PC. After that the total temperatures change in PHE calculated as:

$$\delta t_{\Sigma 1} = \sum_{i=1}^{X_1} \left(\frac{1}{X_1} \sum_{ii=1}^{X_2} \delta t_{(i-1)X_2 + ii} \right); \quad \delta t_{\Sigma 2} = \frac{(G_1 c_1)}{(G_2 c_2)} \delta t_{\Sigma 1}. \tag{5}$$

The total heat load of PHE:

$$Q = \delta t_{\Sigma 1} \cdot G_1 \cdot c_1, \tag{6}$$

This system should be accompanied by equations for calculation of overall heat transfer coefficient U, W/(m^{2}K), as below.

$$U = 1 / \left(1/\alpha_1 + 1/\alpha_2 + \delta_w / \lambda_w + R_f \right) \tag{7}$$

Computer Aided Design of Plate Heat Exchangers

where α_1, α_2 - film heat transfer coefficients for hot and cold streams, respectively, W/(m²K); δ_w - the wall thickness, m; λ_w - heat conductivity of the wall material, W/(mK); $R_f = R_{f1} + R_{f2}$ - the sum of fouling thermal resistances for streams, m²K/W.

For plate heat exchangers the film heat transfer coefficients are usually calculated by empirical correlations:

$$Nu = f(\text{Re}, \text{Pr}) = A * \text{Re}^n \text{Pr}^{0.4} (\mu / \mu_w)^{0.14} \tag{8}$$

Here μ and μ_w dynamic viscosity at stream and at wall temperatures, respectively;

Nusselt number is: $\qquad\qquad Nu = \alpha \cdot d_e / \lambda$,

The streams velocities are calculated as $\qquad w = g / (f_{ch} \rho)$

Where g is flowrate of stream through one channel, kg/s.

The pressure drop in one PHE channel

$$\Delta p = \zeta \cdot \left(L_p / d_e \right) \cdot \rho \cdot w^2 / 2, \tag{9}$$

where L_p -effective plate length; ζ – friction factor, which is usually determined by empirical correlations of following form:

$$\zeta = B / \text{Re}^m \tag{10}$$

For multi pass PHE the pressure drop in one pass multiplied by number of passes X.

In modern PHEs plates of one type are usually made with two different corrugation angles, that can form three different channels, when assembled in PHE. Plates of H type have corrugations with bigger angles (about 60^0) that form the H channels with higher efficiency of heat transfer and hydraulic resistance. Plates of L type have a lower angle (about 30^0) and form the L channels with lower heat transfer and hydraulic resistance. Combined, these plates form channels MH or ML with intermediate characteristics. Such principle of design enables to change thermal and hydraulic performance of plates pack with the level of discreteness equal to one plate in a pack.

In one PHE two groups of channels are usually used. One is of higher hydraulic resistance and heat transfer (x-channel), another of lower characteristics (y-channel). When the stream is going through set of such channels, the temperature changes in different channels are differ. After mixing in collector part of PHE block, the temperature is determined by heat balance. Then the heat exchange effectiveness of plates block with different channels:

$$\varepsilon_b = \left(g_x \cdot n_x \cdot \varepsilon_x + g_y \cdot n_y \cdot \varepsilon_y \right) / \left(g_x \cdot n_x + g_y \cdot n_y \right), \tag{11}$$

where n_x and n_y are the numbers of x and y channels in a block of plates, respectively;

$g_{x,y} = w_{x,y} \cdot \rho \cdot f_{ch}$ - the mass flow rates through one channel of type x or y. These flow rates should satisfy equation $\Delta p_x = \Delta p_y$ and material balance:

$$g_x \cdot n_x + g_y \cdot n_y = G_b, \tag{12}$$

where G_b - flow rate of the stream through the block of plates.

The principle of mixing plates in one heat exchanger gives the best results with symmetrical arrangement of passes ($X_1 = X_2$) and G_b is equal to total flow rate of respective stream. The unsymmetrical arrangement $X_1 \neq X_2$ is usually used when all channels are the same (any of the three available types).

When the numbers of channels are determined, the numbers of plates calculated as:

$$N_{pl} = \sum_{i=1}^{X_1} \left(n_{x1i} + n_{y1i} \right) + \sum_{j=1}^{X_2} \left(n_{x2i} + n_{y2i} \right) + 1 \tag{13}$$

The total heat transfer area of PHE (two end plates not included), m²:

$$F_{PHE} = (N_{pl} - 2) \cdot F_{pl}, \tag{14}$$

where F_{pl} - heat transfer area of one plate, m².

The above algebraic equations (1)-(14) describe the relationship between variables which characterize PHE and heat transfer process in it. They can be regarded as a mathematical model of PHE, which solution enables to calculate pressure and temperature change of streams entering the heat exchanger. It is a problem of PHE rating (analysis).

The problem of PHE design (synthesis) require to find its characteristics (such as plate type, numbers of passes, numbers of plates with different corrugations) which will in the best way satisfy the required process conditions. Here the optimal design with pressure drop specification is considered, in a sense as described by Wang and Sunden (2003). The objective function is total heat transfer area of PHE with conditions that specified heat load Q^0 and allowable pressure drops for both streams $\Delta p_1^{\,0}, \Delta p_2^{\,0}$ must be satisfied. These conditions can be regarded as partial inequality constraints:

$$Q \geq Q^0; \qquad \Delta p_1 \leq \Delta p_1^{\,0}; \qquad \Delta p_2 \leq \Delta p_2^{\,0} \tag{15}$$

Analysis of relations (1) – (15) lets to conclude, that we have the mathematical problem of finding the minimal value for implicit nonlinear discrete/continues objective function with inequality constraints. It does not permit analytical solution without considerable simplifications. To solve it by numerical methods, the software is developed for IBM compatible PC. The mathematical model contains some parameters, namely coefficients and powers in empirical correlations that are not easily available.

3. Identification of mathematical model parameters

As a rule the empirical correlations for design of industrially manufactured PHEs are obtained during tests on such heat exchangers at specially developed test rigs. Such tests are made for every type of new developed plates and inter plate channels. The results are property of manufacturing company and usually not published.

Based on described above mathematical model it was developed the technique of numerical experiment that enables to identify model parameters by comparison with results obtained for the same conditions with the use of PHE calculation software, which is now available in internet for most of PHE manufacturers. The results for some plates manufactured by Alfa Laval presented in Table 1. The geometrical parameters of plates, for which correlations were acquired are given in Table 2.

The comparison of results obtained with our software to those of Alfa Laval free available software has shown good agreement (discrepancies not more then 4% on surface area). We should note, that obtained correlations and developed software can be used only for preliminary calculations, when optimizing PHEs or heat exchanger network. The final calculations when ordering the PHE must be performed by its manufacturer.

Computer Aided Design of Plate Heat Exchangers

Table 1. Parameters in correlations for some Alfa Laval PHEs (Re>250)

Plate type	Channel type	A	n	Re	B	m	Re	B	m
M3	H	0.265	0.7	<520	33.0	0.25	≥520	10.7	0.07
	L	0.12	0.7	<1000	18.8	0.33	≥1000	8.8	0.22
	ML/MH[1]	0.18	0.7	<1000	44.0	0.4	≥1000	5.1	0.10
M6	H	0.25	0.7	<1250	10.0	0.2	≥1250	2.4	0.0
	L	0.12	0.7	<1500	5.1	0.3	≥1500	1.7	0.15
	ML/MH	0.165	0.7	<930	9.3	0.3	≥930	2.72	0.12
M6M	H	0.27	0.7	<1300	11.7	0.13	≥1300	4.55	0.0
	L	0.11	0.71	<2200	4.23	0.23	≥2200	1.88	0.12
	ML/MH	0.14	0.73	<2100	5.61	0.16	≥2100	1.41	0.0
M10B	H	0.24	0.7	<2000	11.1	0.15	≥2000	3.5	0.0
	L	0.11	0.7	<1500	12	0.36	≥1500	2.42	0.14
	ML/MH	0.12	0.74	<2700	6.2	0.2	≥2700	1.9	0.05
M15B	H	0.26	0.7	<2000	5.84	0.05	≥2000	5.84	0.05
	L	0.085	0.74	<2900	5.2	0.28	≥2900	1.57	0.13
	ML/MH	0.13	0.74	<3500	4.3	0.15	≥3500	1.25	0.0

Table 2. Geometrical parameters for some Alfa Laval PHE plates

Plate type	δ, mm	d_e, mm	b, mm	F_{pl}, m^2	$D_{connection}$, mm	$f_{ch} \cdot 10^3$, m^2	L_p, mm
M3	2.4	4.8	100	0.032	36	0.240	320
M6	2.0	4.0	216	0.15	50	0.432	694
M6M	3.0	6.0	210	0.14	50	0.630	666
M10B	2.5	5.0	334	0.24	100	0.835	719
M15B	2.5	5.0	449	0.62	150	1.123	1381

4. Case study

It is required to heat 5 m^3/h of distillery wash fluid from 28 to 90 ^{o}C by hot water coming with temperature 95 ^{o}C and flow rate 15 m^3/h. The pressure of both fluids is 5 bar. Allowable pressure drop for hot stream 1.5 bar. For cold stream 1 bar. The properties of wash fluid are taken constant as follows: density – 978,4 kg/m^3; heat capacity – 3,18 kJ/(kg·K); cconductivity – 0.66 W/(kg·m). Dynamic viscosity at temperatures $t = 25$; 60; $90^\circ C$ is taken as $\mu = 19,5; 16,6; 9\ cP$.

The results of calculations for different passes numbers X_1 and X_2 and optimal for those passes plates arrangements are presented in Table 3. The analysis show that the global optimum (38 plates) is achieved at X_1=2 and X_2=4 with all medium channels (19 H and 19 L plates in PHE). The closest other option (41 plates) is at X_1=X_2=2 with mixed

---

[1] In our study we neglected by small differences in ML and MH channels parameters

Table 3. The influence of passes and plate arrangement on number of plates in M6M PHE

X_2	X_1			
	1	2	3	4
1	56 plates 1*28H / 1*27H	235 plates 2*59H/1*116H	157 plates 3*26H/1*78H	184 plates 4*23H/1*91H
2	72 plates 1*35H / 2*18H	41 plates 2*(7H+3*ML)/ 2*(7H+3*MH)	60 plates 3*10MH / 2*14ML	64 plates 4*8L / 1*15L+1*16L
3	44 plates 1*21H / 2*7H+1*8H	49 plates 2*12MH / 3*8ML	43 plates 3*(4MH+3*L)/ 3*(4ML+3*L)	50 plates 3*6L+1*7L / 3*8L
4	63 plates 1*31MH / 1*7ML+3*8ML	38 plates 2*9MH / 1*4ML+3*5ML	44 plates 3*7L / 2*5L+2*6L	50 plates 3*6L+1*7L / 4*6L

channel arrangement in one pass. If we would have only one plate type in PHE, the minimal number of plates would be 44 for both H and L plates, or 15% higher than with mixed channels.

5. Conclusions

The algorithm and software for computer aided design of multi-pass PHE assembled with plates of different corrugation patterns is developed. The model parameters corresponding to some industrially manufactured plates are obtained. The examples of calculation results for case study show the possibility with such method to obtain optimal solutions with exact satisfaction of constraints for total heat load and pressure drop of one stream. It gives the considerable reduction in heat transfer surface area of PHE.

6. Acknowledgements

References

O. Arsenyeva, L. Tovazhnyansky, P. Kapustenko, G. Khavin, 2009, Mathematical Modelling and Optimal Design of Plate-and-Frame Heat Exchangers. Chemical Engineering Transactions, 18: 791-796.

A. Gogenko, O. Anipko, O. Arsenyeva, P. Kapustenko, 2007, Accounting for fouling in plate heat exchanger design. Chemical Engineering Transactions, 12: 207-213.

J. Hesselgreaves, 2001, Compact Heat Exchangers. Selection, Design and Operation. Elsevier, Amsterdam.

L. Tovazshnyansky, P. Kapustenko, G. Khavin, O. Arsenyeva., 2004, PHEs in Industry. NTU KhPI, Kharkiv (in Russian).

L. Wang, B. Sunden, R. Manglik , 2007, PHEs. Design, Applications and Performance. WIT Press, Southhampton.

L. Wang, B. Sunden, 2003, Optimal design of PHEs with and without pressure drop specifications. Applied Thermal Engineering. 23, 295-311.