Hybrid strategy for parameter estimation and PID tuning

Ling Wang*, Da-Zhong Zheng, De-Xian Huang

Department of Automation, Tsinghua University, Beijing 10084, P. R. China

Abstract

Parameter estimation and PID tuning are two crucial issues in control engineering. Classical methods either require some prior information or depend on some rules, especially they are short of generality and their performances are not satisfied in many engineering fields. Although genetic algorithm and simulated annealing approaches have gained much attention and applications during the past decades, it may cause the premature convergence of genetic algorithm and prohibitive time-consumption required for simulated annealing if executing them alone. In this paper, reasonably combining the parallel structure of genetic algorithm with the controllable jumping property of simulated annealing, a class of effective and general hybrid optimization strategy is proposed for parameter estimation and PID tuning. The proposed strategy is easy to be understood and implemented, and only a little pre-needed information is required. Numerical simulation results demonstrate that the hybrid strategy is of effectiveness, robustness on initial states, and adaptability on models or plants, and comparisons show that the hybrid strategy can achieve performances greatly better than those of pure genetic algorithm and classical methods.

Keywords: Hybrid strategy; parameter estimation; PID tuning

1. Introduction

Parameter estimation and PID tuning are two crucial issues in control engineering, which are of important theoretical value and engineering significance and have been widely studied so far. Traditional estimation methods, such as Least Square Method (LSM) and their generalizations, gradient estimation algorithms, and maximum likelihood algorithms require some prior information and model structure, e.g. time-delay, order etc, which greatly limit their applications, especially in the field of nonlinear systems. Moreover, most classic and improved methods are intrinsically dependent on the gradient information of the error index so as to be prone to be trapped in local optima. It is known that the control performances of PID are completely dependent on PID parameters, but classical tuning methods, such as Ziegler-Nichols method, Cohen-Coon etc (Astrom and Hagglund, 1995), are based on experiments and strongly depend on the plant model and the tuning results are not satisfied in many fields, which leads to the limitation of their applications. In the past decade, genetic algorithm (GA) gained much attention (Michalewicz, 1994) and was widely applied in many engineering fields, including control engineering (Szczerbicka and Becker, 1998). Versek, Urbancic and Filipic (1993) developed a three-stage framework based on genetic algorithms for learning control. Lima and Ruano (2000) proposed neural network models of tuning criteria together with the use of GA to achieve PID autotuning. Li and Shieh (2000) designed a GA-based fuzzy PID controller for non-minimum phase systems. Teo and Marzuki (1999) presented a neuro-fuzzy controller based on neural network with all the parameters tuned by GA. Kristinsson and Dumont (1992) used GA for model identification and then used the model parameters in a certainty equivalence control law based on pole-placement method. Lennon and Passino (1999) used a different method for fitness evaluation and employed a model reference approach. Wu and Yu (2000) proposed a GA based learning algorithm for the identification of a class of fuzzy models. But there still existed two main drawbacks when using pure GA alone, that is, difficult to determine operating parameters and pre-mature convergence.

In this paper, GA and Simulated Annealing (SA) are reasonably combined to construct an effective hybrid strategy (HS) for parameter estimation and PID tuning, which utilizes the population parallel search structure of GA and the controllably probabilistic jumping of SA. With some operations specific designed, the HS can be applied to various kinds of models and controlled plants. The computation procedure is simple and easy to understand and accomplish, and only a little pre-needed information is required. Numerical simulation results based on some classical problems demonstrate that the HS is of effectiveness, robustness on initial states, and adaptability on models or plants, and it can achieve better performances than classical approaches.

The organization of remain contents is as follows. Firstly, the problems to be studied are described, secondly the HS is proposed, then the strategy is implemented in detail for parameter estimation and PID tuning, and some numerical simulations and comparisons are carried out, lastly we end with some conclusions.

2. Problem statement

Generally, the representation style of the estimated model is known in advance, and it is supposed that

* Corresponding author. Email: wangling@mail.tsinghua.edu.cn. This paper was not presented at any IFAC meeting.

the system output can be measured and the ratio of signal/noise should be large enough, as well as the parameters to be estimated should be specified. Usually, a system model can be generally described as follows.

$$y(t) = f(r, \theta) \quad (1)$$

where, $y(t)$ is the system output, r is the system input, $\theta = (\theta_1, \theta_2, \cdots, \theta_k)$ is the parameters to be estimated, f is the model representation which can be expressed by transfer function, state space or ARMA model etc.

Parameter estimation means to obtain the estimated parameters using certain algorithm according to certain error index based on the model output and actual sampling data $y_0(t), t = 1,2,\cdots,n$, with certain model input. In this paper, hybrid strategy will be proposed to estimate the model parameters. The principle can be illustrated by Fig. 1, which will be explained in detail later.

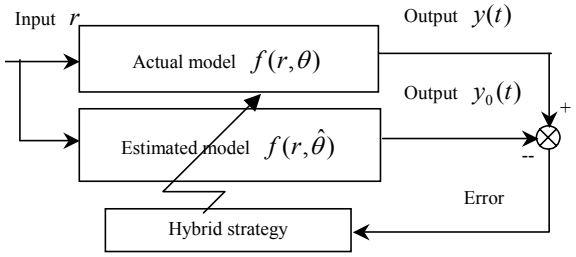


Fig. 1. Illustration of model parameter estimation

PID controllers are well known to engineers for their simple structure, easy implementation, good performances and strong robustness, so that they are widely used in various fields of industry, especially in chemical process industry. More than 90% of the controllers used in real applications are of the PID types. Generally, the formula of conventional PID and its discretized formula can be written as follow.

$$u(t) = K_p[e(t) + \frac{1}{T_i}\int_0^t e(\tau)d\tau + T_d\frac{de(t)}{dt}] \quad (2)$$

$$\Delta u(k) = K_p\{\Delta e(k) + \frac{T_0}{T_i}e(k) + \frac{T_d}{T_0}[\Delta e(k) - \Delta e(k-1)]\} \quad (3)$$

where K_p, T_i and T_d are proportional (P), integral (I) and derivative (D) parameters respectively, T_0 is sampling period, $e(t)$ and $u(t)$ are error variable and plant input respectively. Let $K_i = K_p T_0 / T_i$ and $K_d = K_p T_d / T_0$, then the formula (3) can be rewritten as follows.

$$\Delta u(k) = K_p \Delta e(k) + K_i e(k) + K_d[\Delta e(k) - \Delta e(k-1)] \quad (4)$$

PID tuning means to determine the above three parameters by certain algorithm to achieve the optimal control performances. In this paper, PID will be tuned by hybrid strategy, whose principle can be illustrated by Fig. 2 and interpreted later.

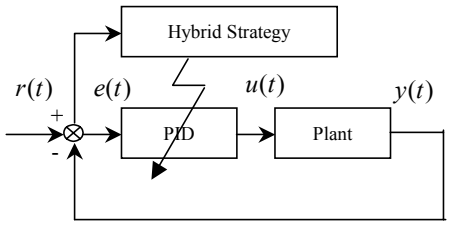


Fig. 2. Illustration of PID tuning

Intrinsically, both parameter estimation and PID tuning are the problems to search optimal parameters according to certain objective functions, which can be regarded as complex functional optimization problems with high dimensions and many local minima. To solve the problems effectively and achieve good performances, in next section a class of HS by combining genetic algorithm and simulated annealing will be proposed.

3. Hybrid optimization strategy

Based on the mechanics of natural selection and genetics, genetic algorithm combines the concept of *survival of the fittest* among solutions with a structured yet randomized information exchange and offspring creation. GA is naturally parallel and is able to exhibit *implicit parallelism*, which does not evaluate and improve a single solution but analyses and modifies a set of solutions simultaneously. There are three basic operators in pure GA, i.e. selection, crossover and mutation. The ability of GA, i.e., operating on many solutions simultaneously and gathering information from all current solutions to direct search, reduces the possibility of being trapped in a local optimum.

Originated from the similarity between statistical mechanics and combinatorial optimization, simulated annealing provides a framework for optimization of properties of very large complex system and can be viewed as an enhanced version of local optimization or iterative improvement algorithms (Kirkpatrick et al, 1983). SA attempts to avoid entrapment in a local optimum by sometimes accepting a move that deteriorates the value of the objective function. With the help of the distribution scheme, SA can provide a reasonable control over the initial temperature and cooling schedule so that it performs effective exploration of solution space and good confidence in the solution quality.

GA is a highly parallel procedure, which contains certain redundancy and historical information of past solutions. However, GA may lose solutions and substructures due to the disruptive effects of genetic operators. In addition, it is not easy to regulate GA's convergence and tune global parameters. Consequently, GA is easy to be premature and results in poor solution (Leung et al, 1997). On the other hand, SA maintains only one solution at a time, whenever they accept a new solution, the old one

must be discarded, which often causes low search efficiency. But, SA has the ability to escape from local optima that can be controlled by cooling schedule (Hajek, 1988). Reasonably combining these two approaches from mechanism to structure, it will develop novel hybrid strategy (HS) with more powerful search efficiency. So, utilizing the parallel searching framework of GA and incorporating SA to avoid individual being trapped in local minima with controllable probability, an efficient HS is proposed as follows.

Step1: initialize population, and determine the initial temperature t_0, and set $k = 0$.

Step2: if stop criterion has been satisfied, then output the results; else go on below steps.

Step3: implement selection and crossover operators.

Step4: implement mutation operator.

Step 5: perform simulated annealing for each individual i in parallel mode, then back to Step 2:

Step 5.1: if equilibrium condition has been reached, then decrease temperature $t_{k+1} = update(t_k)$ and set $k = k + 1$, and go to step2; otherwise go to step 5.2.

Step 5.2: generate a neighbor solution j from solution i randomly and calculate the difference of the objective values $\Delta c_{ij} = c_j - c_i$;

Step 5.3: if $\min\{1, \exp(-\Delta c_{ij} / t_k)\} > random[0,1]$, then let $i = j$, and update the best solution found so far if possible; else keep the old solution.

It can be seen that during the hybrid search process GA provides a set of initial solutions for SA at each temperature to perform Metropolis sample for each solution until equilibrium condition is reached, and GA uses the solutions found by SA to continue parallel evolution. Temperature is adjusted to control the behavior of SA, i.e., at a high temperature, SA performs a "course" search with high escaping probability from current solution; while at a low temperature, SA performs a "fine" search among the neighbor solutions of current solution. In addition, the optimization operators, such as mutation operator and the new solution generator of SA, can be different or hybrid used to yield a large neighborhood and efficiently explore better solutions among the solution space. Theoretically, Wang and Zheng (1998) analyzed the convergence behavior of such HS and a sufficient condition for global convergence was provided, and the job-shop scheduling was solved (Wang and Zheng, 2001).

Moreover, such HS reserves the generality of GA, SA and can be easily implemented and applied to any optimization problems by suitably modifying the encoding scheme, optimization operators, algorithm criteria and parameters. In next sections, the HS will be designed in detail for parameter estimation and PID tuning.

4. Implementation of the strategy

Encoding scheme: The real value encoding is used, i.e., all the parameters are specified by real values except that the model order is encoded by an integer.

Objective function or fitness function: The Integral of Time multiplied by Absolute Error (ITAE) index, i.e., $\int_\tau^\infty t|e(t)|dt$ is employed as objective function in PID tuning, which is able to restrain the overshoot and settling-time to certain extent. While $f = 1/[\sqrt{\sum_t (y(t) - y_0(t))^T (y(t) - y_0(t))} + 0.01]$ is used as fitness function in parameter estimation, where $y_0(t)$ is the actual output and $y(t)$ is the output of the estimated system.

Initialization of population and temperature: After generating initial population with P_{size} individuals randomly at the beginning of the procedure, the best and worst individuals with the objective index c_{best} and c_{worst} are determined. Then, at the initial temperature t_0 the probability to accept the worst individual with respect to the best individual is set as $p_r \in (0,1)$, i.e., $p_r = \exp[-(c_{worst} - c_{best}) / t_0]$. Hence, t_0 can be determined by $t_0 = -(c_{worst} - c_{best}) / \ln(p_r)$. Obviously, such process can be implemented easily and the relative performance of the initial population is used, so such method is of handleability and reasonability.

Selection: In parameter estimation, classical proportional selection based on fitness value is applied, i.e., individual i would be selected with probability $f_i / \sum f_j$, where f_i is the fitness value of i. While in PID tuning, rank-based selection is applied, i.e. all the individuals are arranged with decently order according to the objective value firstly, then the kth individual would be selected with probability $2k / [P_{size}(1 + P_{size})]$.

Crossover: Based on real-value-encoding scheme, crossover operator is designed as Equation 5 to generate two new individuals after selection operator. And such procedure is repeated $P_{size} / 2$ times (P_{size} is population size) to generate the new population. Then, the top P_{size} solutions with better objective values from the old population and new solutions are reserved for the next optimization operator.

$$\begin{cases} w_1' = a \cdot w_1 + (1-a) \cdot w_2 \\ w_2' = a \cdot w_2 + (1-a) \cdot w_1 \end{cases} \tag{5}$$

where $\alpha \in (0,1)$ is a random variable, w_1 and w_2 are parents, w_1' and w_2' are children.

Mutation, SA state generator: Due to the merit of HS, here mutation rate is set to one to perform a "fine" local neighbor search and all these operators can be conducted by appending random noise for each parameter.

$$w' = w + \eta \cdot \xi \tag{6}$$

where ξ is a random variable subjected to Gaussian distribution N(0,1), η is a scale parameter.

Moreover, during the evolution process the best solution found so far should be updated if possible to avoid the lost of good solution, i.e., "elitist" scheme.

Annealing function: Exponential cooling schedule is used to adjust the temperature, i.e. $t_k = \lambda \cdot t_{k-1}$, where $\lambda \in (0,1)$ is decrease rate.

Equilibrium condition and stop criterion: since theoretical convergence conditions may lead to huge computation and are not practicable, two approximate conditions are simply designed to provide a rather good compromise between quality and search efficiency. The Metropolis sample process is set to L_1 iterations fixed, and if the best solution found so far keeps fixed at L_2 consecutive temperatures, the algorithm will stop.

5. Numerical simulation

5.1 Simulations on parameter estimation

Based on three kinds of models (Jiang and Wang, 2000) described as follows, the performances of the HS are tested and some comparisons with simple GA are carried out.

Model 1 (Transfer function of 2-ordered system with time-delay): parameters to be estimated are k, T_1, T_2 and time-delay τ.

$$y(s)/u(s) = ke^{-\tau s}/(T_1 s^2 + T_2 s + 1) \tag{7}$$

Model 2 (Nonlinear state space model): parameters to be estimated are θ_1, θ_2, θ_3 and θ_4.

$$\begin{cases} \begin{bmatrix} x_1(t+1) \\ x_2(t+1) \end{bmatrix} = \begin{bmatrix} \theta_1 x_1(t) x_2(t) \\ \theta_2 x_1^2(t) \end{bmatrix} + \begin{bmatrix} 0 \\ u(t) \end{bmatrix} \\ y(t) = \theta_3 x_2(t) - \theta_4 x_1^2(t) \\ x_1(0) = x_2(0) = 1 \\ t = 0,1,...,50 \end{cases} \tag{8}$$

Model 3 (Hammerstein Model): parameters to be estimated are a_1, a_2, b_0, b_1 and d.

$$\begin{cases} A(q^{-1})y(k) = q^{-d}B(q^{-1})\phi[u(k)] \\ A(q^{-1}) = 1 + a_1 q^{-1} + a_2 q^{-2} \\ B(q^{-1}) = b_0 + b_1 q^{-1} \\ \phi[u] = \begin{cases} \sqrt{u + 1/2} - \sqrt{1/2} & 5 \ge u \ge -1/2 \\ -\sqrt{|u + 1/2|} - \sqrt{1/2} & -5 \le u < -1/2 \end{cases} \end{cases} \tag{9}$$

Combining Matlab and C++ simulation environment, and setting P_{size}=20, p_r=0.1, λ=0.85, η=0.1, L_1=30, L_2=20, sampling time T_0=0.1, numerical simulations are carried out on PIII/550 PC and the average results of 20 random simulation are summarized as follows.

Table 1. Estimation results of Model 1

Parameter	k	T_1	T_2	τ
Actual	1	1	2	1
Estimated	1	1	1.9997	1

Table 2. Estimation results of Model 2

Parameter	θ_1	θ_2	θ_3	θ_4
Actual	0.5	0.3	1.8	0.9
Estimated	0.5069	0.3048	1.8095	0.9077
Results of Huang and Wang (1996)	0.4916	0.3014	1.8432	0.9267

Table 3. Estimation results of Model 3

Parameter	a_1	a_2	b_0	b_1	d
Actual	-1.5	0.7	1	0.5	2
Estimated	-1.5004	0.6984	0.9861	0.4516	2
Results of Huang and Wang (1996)	-1.4982	0.6970	1.3654	-0.0371	2

The results demonstrate the effectiveness of the HS, which is competent for the models with different styles and properties, including nonlinear systems. In addition, in the Hammerstein model there are some parameters may affect the system output much less than others, e.g. b_1, which are hard to estimate. The estimation process of pure GA would lead to these parameters far deviate from the true value, but by incorporating SA into GA the hybrid strategy can achieve good results. The system output of actual model and the estimated one are shown on Fig. 3, from which it can be seen that the two curves are so adjacent each other. Compared with the results of Jiang and Wang (2000), the results from hybrid strategy are much better than those from pure GA. The reason is that the hybrid strategy takes the advantages of both GA and SA and improves the potential of global optimization.

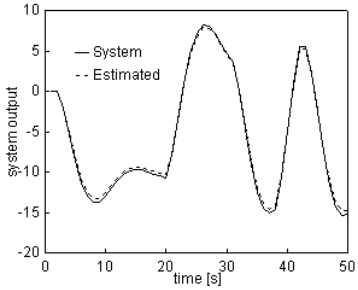


Fig. 3. Output of the actual and estimated Model 3

5.2 Simulations on PID tuning

Firstly based on the controlled plant $e^{-0.5s}/(2s+1)$, the statistical performances of PID controller tuned by hybrid strategy (HS) are investigated, and some comparisons with GA and Z-N methods are carried out. The parameter η is set to 0.6 and the others are the same as before. The statistical results of 20 random simulations are shown in Table 4 (comparatively, the ITAE object value by

Z-N method is 12.6951). The closed loop step output response using PID tuned by the three methods are shown in Fig. 4, and the decreasing curves of objective value are shown in Fig. 5.

Table 4. Average performances gained by the HS and GA

Algorithm	Average ITAE	ITAE Variance	Average overshoot	Average generation
HS	4.1625	1.5974	1.2%	39.80
GA	7.3840	2.3001	18.5%	49.15

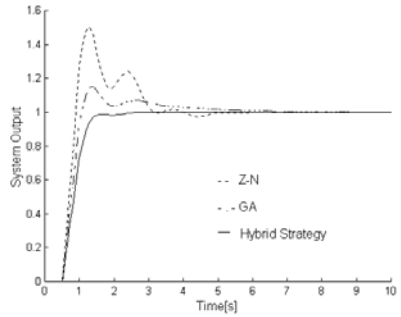


Fig. 4. The closed loop step output response

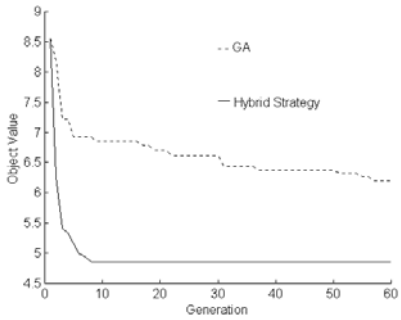


Fig. 5. Decent curves of objective value

To further test the tuning performances of the hybrid strategy, we study the plants as follows. The average performances of 20 random simulations are shown in Table 5, and the corresponding closed loop step output response curves are shown in Fig. 6~10.

$$G_1(s) = 1/(s+1)^2 \tag{10}$$

$$G_2(s) = 1/[s^2 + 2\xi s + 1], \xi = 0.2 \text{ or } 0.8 \tag{11}$$

$$G_3(s) = 1/(1+s)^3 \tag{12}$$

$$G_4(s) = e^{-0.1s}/[(1+s)(1+0.5s)(1+0.25s)(1+0.125s)] \tag{13}$$

Table 5. Average control performances of PID

Plant	Overshoot(%)		Average ITAE	
	GA	HS	GA	HS
$G_1(s)$	26.1	4.8	6.83	3.17
$G_2(s), \tau = 0.2$	15.1	3.5	5.18	3.13
$G_2(s), \tau = 0.8$	6.2	2.4	4.86	2.77
$G_3(s)$	23	0	21.29	10.92
$G_4(s)$	20.1	4.3	10.15	6.92

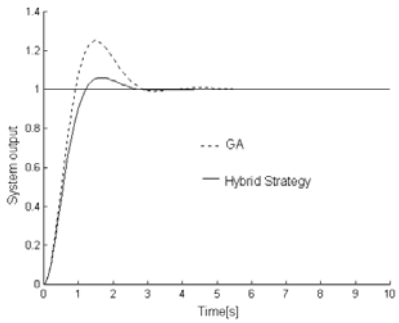


Fig. 6. The closed loop step output response of $G_1(s)$

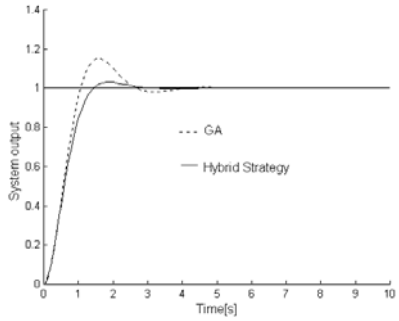


Fig. 7. The closed loop step output response of $G_2(s), \xi = 0.2$

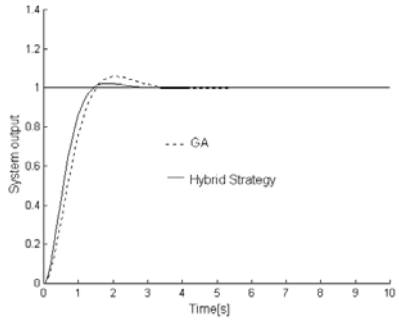


Fig. 8. The closed loop step output response of $G_2(s), \xi = 0.8$

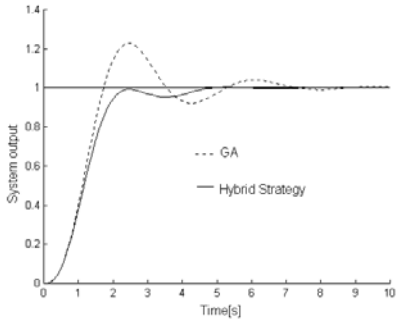


Fig. 9. The closed loop step output response of $G_3(s)$

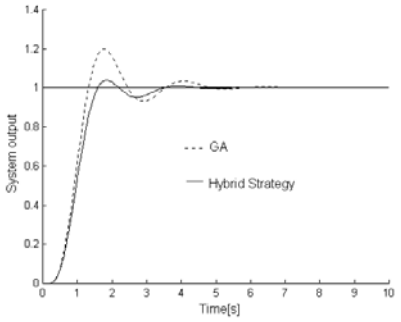


Fig. 10. The closed loop step output response of $G_4(s)$

Lastly, considering the water turbines plant $\frac{1}{1+0.2s}\cdot\frac{1-0.8s}{1+0.4s}\cdot\frac{1}{0.2+0.96s}$ with a non-minimum phase zero and taking η=0.1, T_0=0.04, the closed loop step output response curves using PID tuned by HS, GA and the simplex method (Liu and Mao, 1997) are show in Fig. 11.

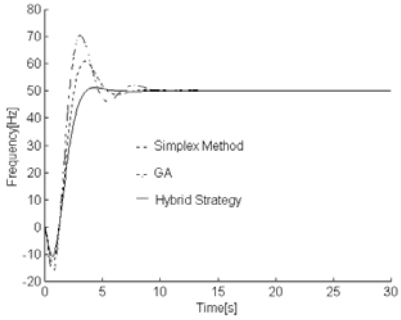


Fig. 11. The closed loop step output response

From the simulation results, it can be concluded that the HS can achieve good optimization performances, such as high quality, rapid speed and robustness on initial values. Secondly the controllers tuned by the HS can achieve better control performances than those of the controllers tuned by GA and Z-N methods, in particular, the overshoot, settling-time and error index are very small. In addition, the HS is independent of plant and control objective. So the HS is well fit for PID tuning.

In brief, the features of the proposed method can be summarized as followed. (1) The computation of parameters is easy and simple. (2) The estimation and tuning procedure is easy to understand and accomplish. (3) Only a little pre-needed information is required. (4) Hybrid global search can achieve very satisfied and better performances than traditional or pure GA methods. (5) The method is general and has a wide range of applications.

6. Conclusion

This paper proposed an effective hybrid strategy by combining SA and GA for parameter estimation and PID tuning. Numerical simulation results demonstrated the effectiveness, robustness on initial states, and adaptability on models or plants. The comparisons showed that the HS could achieve performances greatly better than those obtained by pure GA and traditional methods. The future work is to apply the proposed HS or combining fuzzy logic and neural networks in time-varying systems, some advanced controllers, such as fuzzy controller and neural network controllers, and online estimation and tuning, especially in actual industry environments.

Acknowledgements

This paper was supported by the NSFC under the Grant 60204008, and 973 Program under the Grant 2002CB312200.

References

Astrom, K., Hagglund, T. (1995). *PID controllers: theory, design and tuning*. New York: ISA.

Hajek, B. (1988). Cooling schedules for optimal annealing. *Mathematics of Operations Research*, 13(2): 311-329.

Huang, J., Wu, Y. 1996. On-line system identification based on genetic algorithms. *Information and Control* 25(3): 171-176.

Jiang, B., B. Wang. (2000). Parameter estimation of nonlinear system based on genetic algorithms (in Chinese). *Control Theory and Application*, 17(1): 150-153.

Kirkpatrick, S., Gelatt, C., Vecchi, M. (1983). Optimization by simulated annealing. *Science*, 220(4598): 671-680.

Kristinsson, K., Dumont, G. (1992). System identification and control using genetic algorithms. *IEEE Trans. Systems, Man and Cybernetics*, 22(5): 1033-1046.

Lennon, W. K., Passino, K. M. (1999). Genetic adaptive identification and control. *Engineering Applications of Artificial Intelligence*, 12: 185-200.

Leung, Y., Gao, Y., Xu, Z. B. (1997). Degree of population diversity- a perspective on premature convergence in genetic algorithms and its Markov-chain analysis. *IEEE Trans. Neural Networks*, 8(5): 1165-1176.

Li, T. S., Shieh, M. (2000). Design of a GA-based fuzzy PID controller for non-minimum phase systems. *Fuzzy Sets and Systems*, 111: 183-197.

Lima, J., Ruano, A. E. (2000). Neuro-genetic PID autotuning: time invariant case. *Mathematics and Computers in Simulation*, 51: 287-300.

Liu, L., Mao, Z. (1997). Water turbines PID controller based on genetic algorithm. *Automation of Electric Power Systems*, 21(12): 41-43.

Michalewicz, Z. (1994). *Genetic algorithms + data structures = evolution programs*. Berlin: Springer-Verlag.

Szczerbicka, H., Becker, M. (1998). Genetic algorithms: a tool for modeling, simulation, and optimization of complex systems. *Cybernetics and Systems*, 29: 639-659.

Teo, L., Marzuki, B. (1999). Tuning of a neuro-fuzzy controller by genetic algorithm. *IEEE Trans. Systems, Man, and Cybernetics*, 29(2): 226-236.

Varsek, A., Urbancic, T., Filipic, B. (1993). Genetic algorithms in controller design and tuning. *IEEE Trans. Systems, Man and Cybernetics*, 23(5): 1330-1339.

Voda, A. A., Landau, I D. (1995). A method for the auto-calibration of PID controllers. *Automatic*, 31(1): 41-53.

Wang, L., Zheng, D. Z. (1998). Study on a class of GASA hybrid strategy and its convergence behavior. *Control and Decision*, 13(6): 672-675.

Wang, L., Zheng, D. Z. (2001). An effective hybrid strategy for job-shop scheduling problems. *Computers and Operations Research*, 28(6): 595-596.

Wu, B., Yu, X. (2000). Fuzzy modeling and identification with genetic algorithm based learning. *Fuzzy Sets and Systems*, 113: 351-365.

INTEGRATION OF PRODUCT QUALITY ESTIMATION AND OPERATING CONDITION MONITORING FOR EFFICIENT OPERATION OF INDUSTRIAL ETHYLENE FRACTIONATOR

Hirokazu Kamohara * **Akitoshi Takinami** *
Makoto Takeda * **Manabu Kano** ** **Shinji Hasebe** **
Iori Hashimoto **

* *Showa Denko K.K., Oita 870-0189, Japan*
** *Kyoto University, Kyoto 606-8501, Japan*

Abstract: In this industry-university collaboration, a soft sensor for measuring a key product quality and a monitoring system for testing the validity of the soft sensor were developed to realize highly efficient operation of the ethylene production plant. To estimate impurity concentrations in ethylene products from online measured process variables, dynamic partial least squares (PLS) models were developed. The developed soft sensor can estimate the product quality very well, but it does not function well when the process is operated under unexperienced conditions. Therefore, a monitoring system was developed to judge whether the soft sensor is reliable based on the dynamic PLS model. In addition, simple rules were established for checking the performance of a process gas chromatograph by combining the soft sensor and the monitoring system. The soft sensor and the monitoring system have functioned successfully.

Keywords: soft sensor, partial least squares, statistical process control, monitoring, inferential control, ethylene fractionator

1. INTRODUCTION

Soft sensors are key technologies for producing high quality products when hard sensors of product quality are either not available or too expensive to install. Soft sensors are based on a first principle model, a black-box model, or their combination. Currently, a huge amount of process data is stored in computers, and the effective use of such data is anxiously expected. This situation motivates us to develop a black-box model rather than a first principle model.

Much research has been conducted to develop data-based soft sensors for various processes. A data-based soft sensor, however, does not always function well, because a black-box model is not valid when a process is operated outside certain condition where operation data used for modeling were obtained. The product quality and process performance will deteriorate if the estimates of the soft sensor are blindly believed by operators and used in a control system. To avoid such a situation, the validity of the soft sensor should be monitored online. When a soft sensor is judged to be invalid, the control system and operators should not use the estimates for any purpose.

In this industry-university collaboration, soft sensors for measuring key product qualities and monitoring systems for testing the validity of the soft sensors were developed to realize highly efficient operation of two ethylene fractionators at the SDK (Showa Denko K.K.) Oita plant in Japan. To estimate ethane concentration in ethylene products from online measured

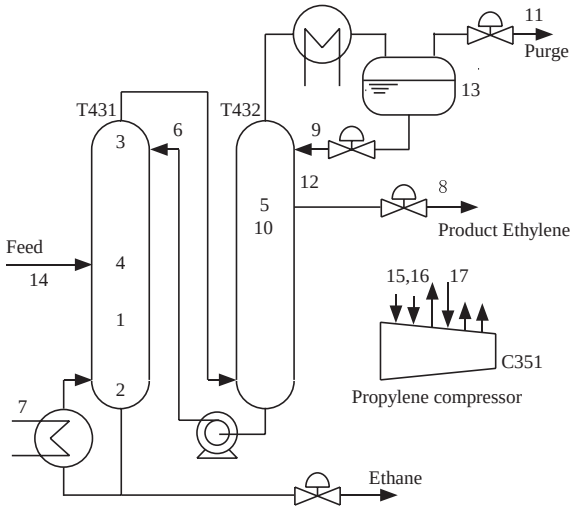


Fig. 1. Schematic diagram of the ethylene fractionator T431/2 at SDK Oita plant.

process variables, dynamic partial least squares (PLS) models were developed. Chemometric techniques, such as principal component analysis (PCA) and PLS, have been widely applied for process modeling, monitoring, and control (for example, Wise and Gallagher, 1996; Nomikos and MacGregor, 1994; Lakshminarayanan, 1997). The main advantage of those methods is that they can cope with correlated input variables. This characteristic is suitable for analyzing data from chemical processes, because chemical processes are multivariable systems and a great number of variables are mutually correlated. The soft sensors developed in the present work are based on PLS. Therefore, in order to test the validity of the soft sensor and judge whether the current operating condition is normal or not, a multivariate statistical process control (MSPC) technique based on PLS is used. In addition, a method is proposed for checking the performance of process gas chromatographs by using the developed soft sensors and monitoring systems. A process gas chromatograph sometimes gives a disturbed and incorrect measurement, which may disturb the operation and deteriorate the process performance. Therefore, it is very important to detect the incorrect measurement and report it to operators. In this article, application results of the PLS-based soft sensor and monitoring system are presented with real industrial data.

2. ETHYLENE FRACTIONATOR

The schematic diagram of the ethylene fractionator, referred to as T431/2, at the SDK Oita plant is shown in Fig. 1. This ethylene fractionator consists of two columns: the bottom column T431 and the top column T432. The feed stream enters the bottom column, and the product ethylene is drawn from the top column. The main specification is ethane concentration in the ethylene product. The ethane concentration must not exceed its upper bound. In order to keep the operation cost as low as possible, the ethane concentration should be kept as high as possible. This fractionator is controlled by using multivariable model predictive control. The number of controlled variables, manipulated variables, and disturbance variables is seven, four, and three, respectively. The controlled variables are ethane concentration and methane concentration in the ethylene product, T431 tray #29 temperature (1), T431 differential pressure, T432 differential pressure, condenser pot level, and reboiler pot level. Manipulated variables are T431 reboiler flow rate (7), T432 internal reflux flow rate (10), T432 purge flow rate (11), and T432 top pressure (13). The disturbance variables are T431 feed flow rate, T431 feed ethane concentration (14), and C351 #4 suction pressure (17). Here, C351 is a propylene compressor. Its #4 suction pressure affects propylene refrigerant temperature and then reboiler heat duty. The numbers in parentheses correspond to those shown in Fig. 1.

3. SOFT SENSOR

Soft sensors are key technologies for producing high quality products when quality hard-sensors are either not available or too expensive to install. Distillation compositions are such quality variables. The compositions can be measured by using, for example, gas chromatographs and near-infrared analyzers, but gas chromatographs suffer from large measurement delays, and most analyzers suffer from high investment and maintenance costs. Therefore, many researchers have investigated soft sensors and inferential control of distillation compositions.

3.1 *Overview*

In order to build a soft sensor by using past operation data stored automatically in a computer, linear or nonlinear black-box models have been widely used. When many process variables are used as input variables, the highly correlated nature of process data must be taken into account. In distillation processes, for example, tray temperatures close to each other change in nearly the same way. Applying a statistical modeling method to such highly correlated data causes a collinearity problem.

The simplest approach for tackling the collinearity problem is to select a few variables, which are mutually independent, from all process variables. Many articles have been published on this matter,

for example, Weber and Brosilow (1972), Joseph and Brosilow (1978), Morari and Stephanopoulos (1980), and Moore et al. (1987). However, this simple approach would not be optimal, because additional measurements may improve the performance of an estimator.

To solve the collinearity problem, composition estimators using PLS have been widely used (Kresta et al., 1994; Mejdell and Skogestad, 1991a, b). In their work, steady-state inferential models of product compositions were built. Mejdell and Skogestad (1993) compared three different estimators using a linear model of a binary distillation column. They concluded that good control performance could be achieved with the steady-state PCR (Principal Component Regression) estimator, which was almost as good as the dynamic Kalman filter, because the steady-state estimator has an inherent feedforward effect. The inherent feedforward effect was investigated in more detail by Kano et al. (2002). They suggested using predictive inferential control with a dynamic inferential model within the cascade control configuration to achieve good performance without demanding the iterative modeling approach. The proposed control system is a feedback control system with a feedforward control effect. An application of a composition estimator to an industrial packed-bed column was reported by Fujii et al. (1997). Their inferential model is a static PLS model based on pressure, flow rate, and temperature measurements. Kano et al. (2000) further investigated PLS-based inferential models, which can estimate the product compositions of the multicomponent distillation column from on-line measured process variables. They compared steady-state, static, and dynamic inferential models and found that the estimation accuracy could be greatly improved by using dynamic models.

In the present work, dynamic PLS is used for estimating the output variable, i.e., the ethane concentration in the ethylene product, from correlated process variables.

3.2 *Dynamic PLS Model*

Kano et al. (2000) thoroughly investigated the selection of input variables and sampling intervals. They concluded that the estimation accuracy was improved by using not only tray temperatures but also other process variables such as reflux flow rate, reboiler heat duty, and pressure. In addition, they strongly recommended using a dynamic inferential model to improve estimation accuracy and control performance. Based on these results, almost all measured process variables are considered as candidates for input variables. Then, optimal selection of input variables and sampling intervals was carried out by trial and error. The selection was mainly based on the correlation analysis, engineers' knowledge, and validation tests.

Table 1. Input variables of PLS model.

No.	Variable
1.	T431 tray #29 temperature
2.	T431 bottom temperature
3.	T431 top temperature
4.	T431 tray #37 temperature
5.	T432 tray #129 temperature
6.	Flow rate from T432 to T431
7.	T431 reboiler flow rate
8.	Product ethylene flow rate
9.	T432 reflux flow rate
10.	T432 internal reflux flow rate (=9-8)
11.	T432 purge flow rate
12.	T432 reflux ratio
13.	T432 top pressure
14.	T431 feed ethane concentration
15.	C351 #2 discharge pressure
16.	C351 #2 discharge temperature
17.	C351 #4 suction pressure

C351 is a propylene compressor. Propylene is used for heating and cooling at the reboiler and the condenser.

For building a dynamic PLS model, the operation data obtained in the period from December 2001 to February 2002 were used. A part of the operation data, which represents abnormal operation or sensor malfunction, was excluded. Although measurements of process variables are stored in the computer every minute, moving averages of five points are used for modeling to reduce measurement noise. Therefore, the minimum sampling interval, which is available for modeling, is five minutes. All variables were mean-centered and scaled to have unit variances. The selected input variables are listed in Table 1. A total of 17 process variables, including temperatures, pressures, flows, and reflux ratio, were selected. In addition, measurements at the current sampling instant were used together with those at 5, 10, 15, 20, 25, 30, 35, 40, 45, 50, 75, 100, 125, 150, 175, 200, 225, 250, 275 minutes before when a dynamic PLS model was built. Therefore, the total number of input variables is 340 ($= 17 \times 20$). The developed estimator has the form of

$$\hat{y}(t) = \sum_{i=1}^{17} \sum_{j=1}^{20} \alpha_{ij} x_i(t - s_j) \qquad (1)$$

where $\hat{y}(t)$ and $x_i(t)$ denote the estimated product concentration and the input variables at the time t, respectively. s_j is sampling instants and α_{ij} is regression coefficients.

The number of latent variables was determined on the basis of cross-validation tests. Twenty latent variables are used in the final model. Furthermore,

no nonlinear transformation is used for dealing with nonlinearity between input variables and the output, i.e., the ethane concentration, because a well-known logarithmic transformation of the product quality did not improve the estimation accuracy in this application.

3.3 *Estimation Results*

The developed soft sensor has been applied to the ethylene fractionator T431/2. The estimation results in three periods are shown in Fig. 2: (A) Nov. 10 2001 through Nov. 30 2001, (B) May 1 2002 through June 13 2002, and (C) Aug. 1 2002 through Aug. 31 2002. Here, measurements and estimates of ethane concentration (Ethane conc.) and estimation errors (Error) are scaled. The soft sensor functions very well. In Fig. 2(top), it is difficult to distinguish between measurements and estimates because estimation errors are so small. The relative estimation error is less than 10 % at almost all times except in the period from 1100 to 1200 hours. In this period, a trouble occurred and thus operators changed the operating condition considerably.

The product quality will deteriorate if estimates of the soft sensor are blindly believed by the control system and operators, while the soft sensor does not function well. To avoid such a situation, an operating condition should be monitored online and the validity of the soft sensor should be tested. The control system and operators should not use the estimates for any purpose when the soft sensor is judged to be invalid. In the next section, a monitoring system for testing the validity of the soft sensor is proposed.

4. MONITORING SYSTEM

A black-box model does not always function well. Since it is a data-based model, a black-box model is valid only when a process is operated within a certain condition where operation data used for modeling were obtained. Therefore, to successfully apply a data-based soft sensor to an industrial process, the validity of the soft sensor should be tested. In other words, the operating condition should be monitored online to judge whether the estimated value is reliable or not.

The soft sensor developed in the present work is based on PLS. Therefore, the operating condition, where operation data used for modeling were obtained, can be easily defined in the subspace spanned by major latent variables retained in the PLS model and its orthogonal complement space. In order to test the validity of the soft sensor, it is necessary to judge whether the current operating condition is inside the defined operating condition or outside. Such a monitoring system can be realized by using a multivariate statistical process control (MSPC) technique based on PLS.

4.1 *Overview*

Chemical processes are multivariable systems consisting of a large number of mutually correlated variables. MSPC was developed to monitor such multivariable processes. The original Shewhart-type control chart for correlated variables is the Hotelling T^2 control chart. Jackson (1959) used principal component analysis (PCA) and proposed a T^2 control chart for principal components. Later, Jackson and Mudholkar (1979) investigated PCA as a tool of MSPC and introduced a residual analysis. The control chart was introduced for the sum of squared residuals Q as well as T^2.

$$T^2 = \sum_{r=1}^{R} \frac{t_r^2}{\sigma_{t_r}^2} \tag{2}$$

$$Q = \sum_{p=1}^{P} (x_p - \hat{x}_p)^2 \tag{3}$$

where t_r is the r-th principal component score and $\sigma_{t_r}^2$ is the variance of t_r. x_p and $\hat{x}_p$ are a measurement of the p-th variable and its predicted (reconstructed) value, respectively. R and P denote the number of principal components retained in the PCA model and the number of process variables, respectively. The T^2 statistic is a measure of the variation within the PCA model, and the Q statistic is a measure of the amount of variation not captured by the PCA model. These two statistics, T^2 and Q, can be used for PLS-based MSPC by substituting latent variables for principal components. Many successful applications of PLS-based MSPC to industrial data have shown its practicability (Kourti et al., 1995; Macgregor and Kourti, 1995; Kourti and MacGregor, 1995).

4.2 *Monitoring System Design and Results*

The monitoring system is based on the dynamic PLS model designed for the soft sensor. The number of input variables is 340, and the number of latent variables is 20. That is, $R = 20$ and $P = 340$ in Eqs. (2) and (3). The developed monitoring system integrated with the soft sensor has been applied to the ethylene fractionator T431/2. The control limits of T^2 and Q are 200 and 100, respectively. These control limits were determined so that they represent 99% confidence limits. The monitoring results in three periods

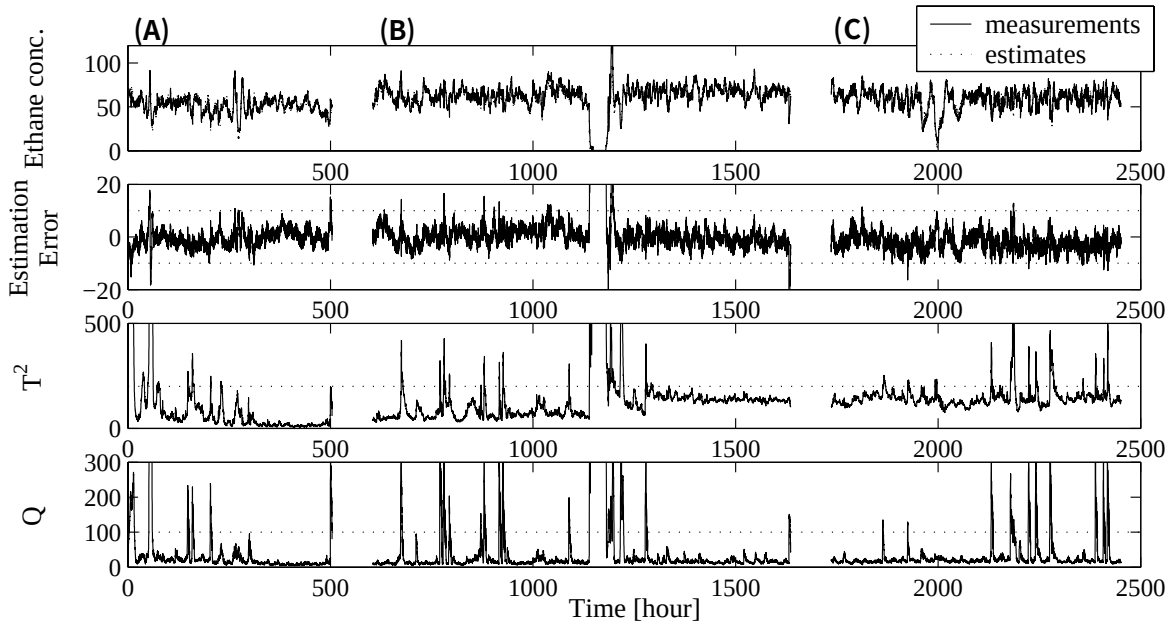


Fig. 2. Estimation and monitoring results of ethane concentration in the ethylene product.

are shown in Fig. 2. In the period from 1100 to 1200 hours, when the process was operated under the abnormal operating condition and the estimation error was crucial, both T^2 and Q statistics considerably exceeded their control limits. That is, the monitoring system indicates that the operating condition is abnormal and the soft sensor is not reliable in this period. It should be noted here that "abnormality" in this context does not necessarily mean the occurrence of faults. It means that the current operating condition is different from the past operating condition in which data used for modeling were obtained. In Fig. 2, estimation errors tend to become large and exceed ± 10 when T^2 or Q exceeds its control limit. Spikes of T^2 or Q are observed when load on the tower increases. These results demonstrate that the developed monitoring system can judge whether the soft sensor is reliable in the current operating condition or not. The soft sensor and the monitoring system have functioned successfully.

Furthermore, the T^2 statistic indicates that the operating condition in the period (C) differs from that in (A). The operating condition in (A) is quite similar to the normal operating condition, where reference data were obtained. On the other hand, an improved multivariable control system was installed and the process has been operated near the optimal condition in (C). Therefore, T^2 in (C) is larger than that in (A). Estimation errors are quite small even when T^2 indicates the significant operating condition change. That is, the developed soft sensor is valid under a sufficiently wide range of operating conditions.

Table 2. Rules for checking the performance of a process gas chromatograph.

Error of GC and soft sensor	Operating condition (Reliability of soft sensor)	Performance of GC
small	good	good
small	bad	good (unclear)
large	good	bad
large	bad	good

4.3 *Validation of GC*

The developed monitoring system can be used for another purpose. A process gas chromatograph (GC) sometimes gives a disturbed and incorrect measurement, and such an incorrect measurement disturbs the operation and deteriorates the process performance. In addition, a GC needs to be repaired if its measurements are not reliable. Therefore, it is important to detect the incorrect measurement and report it to operators. The performance of a GC can be checked by comparing its measured value with an estimated value of the soft sensor. The rules summarized in Table 2 were established in the present work. Here, an error of GC and soft sensor is a difference between a measured value of GC and an estimated value of the soft sensor. The operating condition is judged by the developed PLS-based monitoring system. Estimates of the soft sensor are reliable when operating conditions are good. Therefore, the soft sensor and the GC function well, when an error is small and the operating condition is good. On the other hand, a GC measurement is judged to be incorrect if an error is large while the operating condition is good, because the soft sensor must be reliable when the operating condition is good. In

such a situation, a large estimation error is caused by malfunction of GC.

5. CONCLUSION

In this industry-university collaboration, a soft sensor for measuring ethane concentrations in ethylene products and a monitoring system for testing the validity of the soft sensor were developed to realize highly efficient operation of two ethylene fractionators at the SDK Oita plant.

To estimate the ethane concentration from online measured process variables, dynamic PLS models were developed. The developed soft sensor can estimate the product quality very well. The relative estimation error is less than 10 % at T431/2; however, the soft sensor does not function well when the process is operated under abnormal conditions. To test the validity of the soft sensor, a monitoring system was developed based on the dynamic PLS model designed for the soft sensor. The monitoring system can judge whether the soft sensor is reliable or not. The usefulness of the developed monitoring system was demonstrated with real operation data. In addition, the performance of a process gas chromatograph can be checked by using the soft sensor and the monitoring system. Simple rules were established for this purpose. The soft sensor and the monitoring system have functioned successfully in the SDK Oita plant.

REFERENCES

Fujii, H., S. Lakshminarayanan, and S. L. Shah (1997). Application of the PLS technique to the estimation of distillation tower top composition. *Prep. of ADCHEM*, Banff, Canada.

Jackson, J. E. (1959). Quality Control Methods for Several Related Variables. *Technometrics*, **1**, 359-377.

Jackson, J. E. and G. S. Mudholkar (1979). Control Procedures for Residuals Associated with Principal Component Analysis. *Technometrics*, **21**, 341-349.

Joseph, B. and C. B. Brosilow (1978). Inferential Control of Processes. *AIChE J.*, **24**, 485-509.

Kano, M., K. Miyazaki, S. Hasebe, and I. Hashimoto (2000). Inferential control system of distillation compositions using dynamic partial least squares regression. *J. Proc. Cont.*, **10**, 157-166.

Kano, M., N. Showchaiya, S. Hasebe, and I. Hashimoto (2002). Inferential control of distillation compositions: selection of model and control configuration. *Control Engineering Practice* (in press)

Kourti, T. and J. F. MacGregor (1995). Process analysis, monitoring and diagnosis, using multivariate projection methods. *Chemometrics and intelligent laboratory systems*, **28**, 3-21.

Kourti, T., P. Nomikos, and J. F. MacGregor (1995). Analysis, monitoring and fault diagnosis of batch processes using multiblock and multiway PLS. *J. Proc. Cont*, **5**, 277-284.

Kresta, J. V., T. E. Marlin, and J. F. MacGregor (1994). Development of inferential process models using PLS. *Comp. Chem. Engng*, **18**, 597-611.

Lakshminarayanan, S., S. L. Shah, and K. Nandakumar (1997). Modeling and control of multivariable processes: dynamic PLS approach. *AIChE J.*, **43**, 2307-2322.

MacGregor, J. F. and T. Kourti (1995). Statistical process control of multivariate processes. *Control Eng. Practice*, **3**, 403-414.

Mejdell, T. and S. Skogestad (1991a). Estimation of distillation compositions from multiple temperature measurements using Partial-Least-Squares regression. *Ind. Eng. Chem. Res.*, **30**, 2543-2555.

Mejdell, T. and S. Skogestad (1991b). Composition estimator in a pilot-plant distillation column using multiple temperatures. *Ind. Eng. Chem. Res.*, **30**, 2555-2564.

Mejdell, T. and S. Skogestad (1993). Output estimation using multiple secondary measurements: high-purity distillation. *AIChE J.*, **39**, 1641-1653.

Moore, C., J. Hackney, and D. Canter (1987). Selecting sensor location and type for multivariable processes In: *Shell Process Control Workshop* (D. M. Prett and M. Morari, Ed.). pp.291-308, Butterworth, Boston.

Morari, M. and G. Stephanopoulos (1980). Optimal selection of secondary measurements within the framework of state estimation in the presence of persistent unknown disturbances. *AIChE J.*, **26**, 247-259.

Nomikos, P. and J. F. MacGregor (1994). Monitoring batch processes using multiway principal component analysis. *AIChE J.*, **40**, 1361-1375.

Weber, R. and C. B. Brosilow (1972). The use of secondary measurements to improve control. *AIChE J.*, **18**, 614-623.

Wise, B. M. and N. B. Gallagher (1996). The process chemometrics approach to process monitoring and fault detection. *J. Proc. Cont.*, **6**, 329-348.

ON LINE LOWER-ORDER MODELING USING FUZZY SYSTEMS

H. F. Ho, A. B. Rad*, Y. K. Wong, and W. L. Lo

The Hong Kong Polytechnic University
Department of Electrical Engineering
Hung Hom, Kowloon
Hong Kong SAR, China
E-mail: eeabrad@polyu.edu.hk

Abstract: In this paper, we present a novel on-line approximation technique to find the parameters of a First-Order plus Time Delay (FOPTD) model of higher-order systems using fuzzy reasoning. Based on the information obtained from the model, the parameters of a PID controller can be adjusted on-line. The performance of this algorithm is verified by simulation studies. The simulated examples demonstrate the feasibility and adaptive property of the proposed algorithm.

Keywords: Adaptive Control, fuzzy Logic, gradient descent

1. INTRODUCTION

Many systems are represented mathematically by high order dynamics. However, a lower-order model is sufficient for controller tuning (Ashworth, 1982). It is widely accepted that for the purpose of controller design, a First Order Plus Time Delay (FOPTD) model can approximate such systems adequately and hence facilitate controller design. In general, the parameters of this model, namely system gain, apparent time constant and apparent time delay can be used to tune a PID controller. There are many techniques to determine the parameters of FOPTD (Ziegler, 1942, Smith, 1967, 1997, Sundaresan, 1978). However, most of them are off-line approximating methods, for which the parameters are obtained from process reaction curve. In such cases, it is difficult to apply these methods to describe adequately the time-varying characteristic of the plant.

In recent years, there has been an unprecedented increase in applications of the so-called Soft-computing methodologies in identification and control of dynamic systems (Bha, 1990, Czogala, 1981, Lu, 1997, Narendra, 1990). Soft-computing methods are referred to techniques that employ fuzzy systems, neural networks and genetic algorithms either alone or in hybrid form.

In particular, fuzzy logic theory (Zadeh, 1965) has been the focus of much research in the areas of control and identification. Its integration with model-based systems theory has produced a unique approach entailing the human knowledge and heuristic methods with rigourous mathematical methods for stability and convergence analysis and several successful applications in control and identification have been reported (Chen, 1998, Wang, 1996). Whereas majority of earlier efforts was focused in fuzzy controllers, the emerging area of fuzzy identification has become very important in fuzzy system theory in the last decade (Sugeno, 1986, Babuska 1996). Fuzzy identification methods fall into three categories, linguistic fuzzy model (Wang, 1996), fuzzy relational modelling (Wang, 1997) and Takagi and Sugeno (TS) modeling (Sugeno, 1986). It is interesting to note that not much attention has been paid to reduced order modeling. This

may be due to the fact that fuzzy logic systems are essentially model free approaches. This has motivated the authors to develop an on-line approximation method to determine the parameters of FOPTD using fuzzy systems.

This paper presents a simple and new approach to the on line lower-order model identification for unknown processes using fuzzy system. The idea is to integrate a fuzzy system with a model generator with known structure. The parameters learning task is performed using the gradient descent algorithm (Wang, 1997).

The rest of this paper is organized as follow: Section 2 is devoted to the idea of approximating a high-order system with a FOPTD model using fuzzy system. The proposed method combined with PID controller is derived in Section 3. In Section 4 simulations studies are presented. Finally, the paper is concluded in Section 5.

2. LOWER ORDER APPROXIMATION OF HIGHER-ORDER SYSTEMS WITH FUZZY SYSTEM

2.1 The On-line Approximating Approach

It is well known that high-order processes dynamic can be described with sufficient accuracy by a first order plus time delay model (Sundersan, 1978). Consider:

$$\frac{Y(s)}{U(s)} = e^{-sq}\,\frac{b_0 s^n + b_1 s^{n-1} + ... + b_m}{s^n + a_1 s^{n-1} + ... + 1} \approx \frac{K \cdot e^{-t \cdot s}}{Ts + 1} \qquad (1)$$

where K is the system gain, T is the dominant time constant, τ is the apparent dead time and $Y(s)$ and $U(s)$ are the Laplace transformed output and input signals respectively. The proposed approach is conceptually simple and is realized by cascading a fuzzy system and model generator in parallel with the process to be identified as shown in Figure 1. The input signal $u(t)$ is applied to the high-order system, the fuzzy system, and the FOPTD model generator at the same time. The fuzzy system has three parameters, namely, the gain K, the time constant T, the dead time τ. These three parameters are fed to the first-order plus dead time model generator to get the output of the model. The error between the output of the plant and the output of the model is used to train the consequent part of the fuzzy system. The training process tends to force the output of the FOPTD model generator to approximate the output of the system. Thus, the inputs of the FOPTD model generator are the approximating parameters of the first-order representation of the high-order system. The output of the FOPTD model is expected to match the output of the high-order system after the model converges.

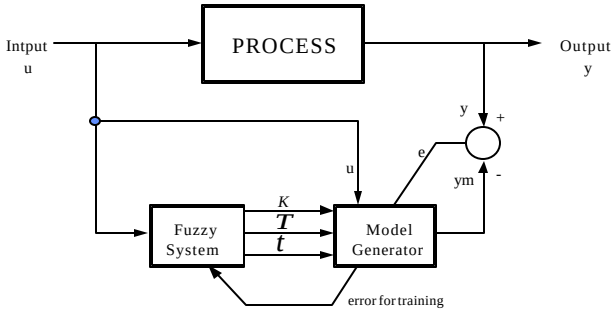


Fig. 1 Block Diagram of the proposed method

The transfer function of the FOPTD model generator is rewritten below:

$$\frac{Y_m(s)}{U(s)} = \frac{K \cdot e^{-t \cdot s}}{Ts + 1} \qquad (2)$$

2.2 The Fuzzy System Structure and Training Algorithm

In this paper, we apply the fuzzy identification techniques to obtain the process model directly. The ith rule of the fuzzy model is of the following form:

Rule i: *If* x_1 *is* A_{i1} *and* $\cdots$ *and* x_n *is* A_{in} *then* $\hat{p}_i$ *is* c_i

where $x \in R^n$ and $p_i \in R$ are the input vector (process input $u(t)$) and output value (estimated model parameters) of the fuzzy system respectively, A_{ij} ,$i=1,2.,m$, $j=1\ldots n$, are the(1) fuzzy sets. Given the input data x, by using a singleton fuzzifier, product fuzzy inference and weighted average defuzzifier, the output value of the fuzzy system is inferred as follows:

$$\hat{p}_i(x) = \frac{\sum_{i=1}^{m} c_i \prod_{j=1}^{n} m_{ij}(x_j)}{\sum_{i=1}^{m} \prod_{j=1}^{n} m_{ij}(x_j)} \qquad (3)$$

if we fix the m_{ij} *(Membership value of* x *for* A_{ij}*)* and view the c_i (*Consequence of Rule* i) as adjustable parameters, then equation (3) can be rewritten as:

$$\hat{p}_i(x) = \theta^T \xi(x) \qquad (4)$$

where $\theta = (c_1, \cdots\cdots, c_m)^T, \xi(x) = (\xi_1(x), \cdots\cdots, \xi_m(x))^T$ is a regression vector defined as

$$\xi_i(x) = \frac{\prod_{j=1}^{n} m_{ij}(x_j)}{\sum_{i=1}^{m} \prod_{j=1}^{n} m_{ij}(x_j)} \qquad (5)$$

To train the above fuzzy system, a direct learning the gradient descent algorithm (Wang, 1997) is employed. The consequent parameters are adjusted in each iteration is derived below. The error function E is define as:

$$E = \frac{1}{2}e^2 = \frac{1}{2}[y_m(t) - y(t)]^2 \quad (6)$$

where y and y_m are the output of the plant and the output at the FOPDT model at any time instant t. Within each time interval from t to $t+1$, the gradient descent algorithm is used to update the consequent parameters according to the following relationship:

$$c_i(t+1) = c_i(t) - h_i \cdot \frac{\partial E}{\partial c_i} \quad (7)$$

where η_i is the learning rate. Using the chain rule, one has

$$\begin{aligned}\frac{\partial E}{\partial c_i} &= \frac{\partial E}{\partial y_m(t)} \cdot \frac{\partial y_m(t)}{\partial p_i(t)} \cdot \frac{\partial p_i(t)}{\partial c_i} \\ &= \frac{\partial E}{\partial y_m(t)} \cdot \frac{\partial y_m(t)}{\partial p_i(t)} \cdot \frac{\prod_{j=1}^{n} \mu_{ij}(x_j)}{\sum_{i=1}^{m}\prod_{j=1}^{n} \mu_{ij}(x_j)} \\ &= \frac{\partial E}{\partial y_m(t)} \cdot \frac{\partial y_m(t)}{\partial p_i(t)} \cdot \boldsymbol{x}_i(x)\end{aligned} \quad (8)$$

$\hat{\boldsymbol{p}} = [K\ T\ \boldsymbol{t}]$ is a 3 1 input vector of the FOPDT model (the output of vector of the fuzzy system)

$$\frac{\partial y_m}{\partial \boldsymbol{p}} = \left[\frac{\partial y_m(t)}{\partial K}, \frac{\partial y_m(t)}{\partial T}, \frac{\partial y_m(t)}{\partial \boldsymbol{t}}\right]^T \quad (9)$$

To find the partial derivatives of the output $y_m(t)$ of the model generator (FOPTD) w.r.t. gain (K), dominant time constant (T) and apparent dead-time (τ), respectively, please refer to Appendix A1.

$$\frac{\partial y_m(t)}{\partial T} = L^{-1}\left[\frac{-sKe^{-ts}}{(T\cdot s+1)^2}U(s)\right] \quad (10)$$

$$\frac{\partial y_m(t)}{\partial t} = L^{-1}\left[\frac{-sKe^{-ts}}{T\cdot s+1}U(s)\right] \quad (11)$$

$$\frac{\partial y_m(t)}{\partial K} = L^{-1}\left[\frac{e^{-\boldsymbol{t}s}}{T\cdot s+1}U(s)\right] \quad (12)$$

From equations (7) and (8), we can rewrite the update rule as follows:

$$c_i^1(t+1) = c_i^1(t) - \boldsymbol{h}_1 \cdot e(t) \cdot \frac{\partial y_m(t)}{\partial K}\boldsymbol{x}_i(\boldsymbol{x}) \quad (13)$$

$$c_i^2(t+1) = c_i^2(t) - \boldsymbol{h}_2 \cdot e(t) \cdot \frac{\partial y_m(t)}{\partial \boldsymbol{t}}\boldsymbol{x}_i(\boldsymbol{x}) \quad (14)$$

$$c_i^3(t+1) = c_i^3(t) - \boldsymbol{h}_3 \cdot e(t) \cdot \frac{\partial y_m(t)}{\partial T}\boldsymbol{x}_i(\boldsymbol{x}) \quad (15)$$

where *1,2* and *3* is the indication of the FOPTD model parameter gain, time constant and time delay and η_1 ,η_2 and η_3 are the learning rate of each fuzzy sub-system respectively. Figure 2 show the membership function of each sub-system .We have used the value of a equal to 1 in the following simulations. Therefore the fuzzy system consists of three fuzzy sub-systems as shown in Figure 2 and the output value can be obtained from equation 16.

$$\hat{p}_1(\boldsymbol{x}) = \boldsymbol{j}_1^T x(\boldsymbol{x}),\ \hat{p}_2(\boldsymbol{x}) = \boldsymbol{j}_2^T x(\boldsymbol{x}),\ \hat{p}_3(\boldsymbol{x}) = \boldsymbol{j}_3^T x(\boldsymbol{x}) \quad (16)$$

where $\boldsymbol{j}_3 = (c_1^3, \cdots\cdots, c_m^3)^T$, $\boldsymbol{j}_2 = (c_1^2, \cdots\cdots, c_m^2)^T$ and $\boldsymbol{j}_1 = (c_1^1, \cdots\cdots, c_m^1)^T$ are the regression vector of each sub-system as given in equation (5). The three fuzzy sub-systems have similar structures. In this paper, 2 fuzzy rules are used for each fuzzy sub-system.

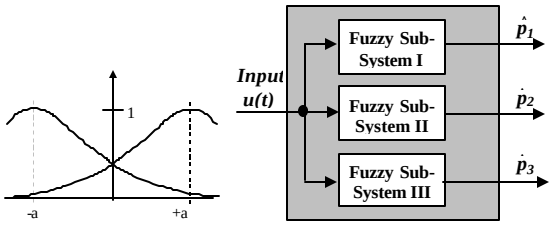


Fig 2 Fuzzy Subsystems

3. ON LINE PID TUNING METHOD USING FUZZY SYSTEM

In order to show the effectiveness of the proposed method, we combine the fuzzy algorithm with a standard PID controller to make an adaptive control algorithm. The control structure is shown in Figure 3. There are two parts in the control structure of the on line PID tuning method. The first part, which was described in the previous section, is the approximation of high order systems with FOPTD using fuzzy system, and the second part is the design of the PID controller. The parameters of the PID controller can be obtained from the corresponding parameters of the estimated FOPTD by fuzzy system. We have used the Ziegler-Nichols ultimate cycle tuning method (17) to compute the parameters of the PID controller:

$$K_p = 0.6\ K_u \qquad T_I = 0.5\ T_{u,,} \qquad T_d = 0.125T_u \quad (17)$$

Here, K_p, T_i, T_d, K_u and T_u are the proportional gain, integral time constant, derivative time constant, the ultimate gain and the ultimate period respectively. The ultimate gain and the ultimate period are calculated from the FOPDT model of the high order plant (Rad, 1997) It should be emphasized that other control algorithms could also be used. The PID controller is implemented in the following form:

$$u(t) = K_p[e_t(t) + \frac{1}{T_i s}\int e_t(t)dt - T_d\frac{dy_f}{dt} \quad (18)$$

$$e_t(t) = r(t) - y(t), \qquad Y_f(s) = \frac{Y(s)}{1 + 0.1T_d s}$$

where $u(t)$,$y(t)$,$r(t)$,$y(t)$and $y_f(t)$ are the controller output, process output, set-point and filtered derivative,

respectively. The implementation of the adaptive PID is as follows:

1. *Approximate the first-order with time delay model (FOPTD) parameters by fuzzy system.*
2. *Determine the ultimate gain (K_u) and ultimate period (T_u) by the FOPTD model.*
3. *Find the PID controller parameters K_p,T_i and T_d from equation (11) and calculate u(t.)*
4. *Find the FOPTD model output $y_m(t)$ from the FOPTD Model Generator.*
5. *Calculate error between the (FOPTD) model output and the process output.*
6. *Update $c_i(t)$ by using equation(13-15) (Gradient descent algorithm).*
7. *Update the error between the set-point and the process output. Go to step (1)*

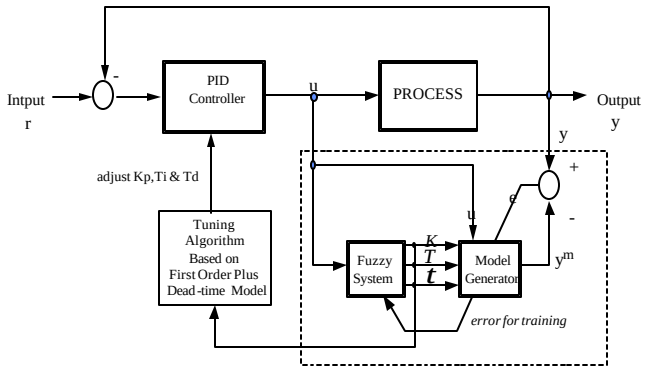


Fig. 3 On-line PID tuning using fuzzy system

4. SIMULATION RESULTS

To show the adaptive behaviour of the algorithm, let us consider three processes as:

Process I $$\frac{Y(s)}{U(s)} = \frac{1.5e^{-2.5s}}{(s+1)^2}$$

Process II $$\frac{Y(s)}{U(s)} = \frac{1-1.4s}{(s+1)^3}$$

Process III $$\frac{Y(s)}{U(s)} = \frac{1.5s^{-3.0s}}{(s+1)^4}$$

The first process is a second order with time delay system, the second process is a non-minimum phase system and the third process is a fourth order time delay system. First, adaptive control of Process I was simulated for t =120s after which the system was changed from Process I to Process II. For t =320s the system was switched from process II to process III. Furthermore, it should be noted that the gain in systems 1,2 and 3,2 are different (1.5 and 1.0). It is known that some adaptive controllers cannot cope with change in steady-state gain of the controlled system. However, as it is seen in Figure 5, the proposed method can successfully track the system change. Figure 4 shows the overall performance of the proposed algorithm. In the simulation, the set-point was selected to be a square wave with amplitude 0.6 and a period of 80s. A Gaussian noise with mean zero and variance of 0.001 was injected at the output of the system. We employed a fourth order Runge Kutta numerical integration algorithm for all time responses and the integration interval was selected to be 0.1s. The fuzzy system also used the same time interval for updating its parameters. The simulation proceeded as follows: the PID controller was initialized with $K_p = 1$, $T_i = 1000$, $T_d = 0.0$. The consequent values of fuzzy system were initialised with c^1=1.0, c^2=1.8, c^3=3.5. The learning rates were chosen as η_1=0.25, η_2=0.8 and η_3=0.8 respectively. Figure 4 shows the overall performance of the three controlled systems. In this figure, the set point and the output, the controller signal and the estimated parameters of gain, apparent time delay and the dominant time constant are shown in top, middle and bottom curves respectively. In all these system changes, the fuzzy system converged and the estimated parameters of the FOPDT also converged to their steady state values. The proposed method is shown to provide stable and robust control under various conditions. Tables 1, 2 and 3 show the parameters of FOPTD model approximated by several other methods such as Smith's (Smith, 1967), minimized error (Sundersan, 1978), and the corresponding ultimate gain and the ultimate period for processes I, II and III respectively. It should be noted that the parameters from all other methods except the proposed one were obtained off-line, from open loop excitation with unit step and were noise free. Furthermore, the values quoted for the proposed algorithm is based on the last measurement before each system change and not the average value.

Table 1
Process I FOPTD Model Parameters

	K	T	t	K_u	T_u
Smith Method[11]	1.5	1.65	3.00	1.06	8.38
Minimized-error[12]	1.5	1.46	3.11	0.98	8.44
Proposed method	1.5	1.33	3.19	0.936	8.48
Process I	-	-	-	1.036	8.438

Table 2
Process II FOPTD Model Parameters

	K	T	t	K_u	T_u
Smith Method[11]	1.0	1.89	2.43	1.93	7.22
Minimized-error[12]	1.0	1.67	2.55	1.74	7.35
Proposed method	1.0	1.39	2.45	1.64	6.91
Process II	-	-	-	1.54	6.83

Table 3
Process III FOPTD Model Parameters

	K	T	t	K_u	T_u
Smith Method[11]	1.5	2.49	4.86	1.03	13.39
Minimized-error[12]	1.5	2.057	5.1	0.923	13.49
Proposed method	1.5	2.66	4.75	1.07	13.33
Process III	-	-	-	0.987	13.48

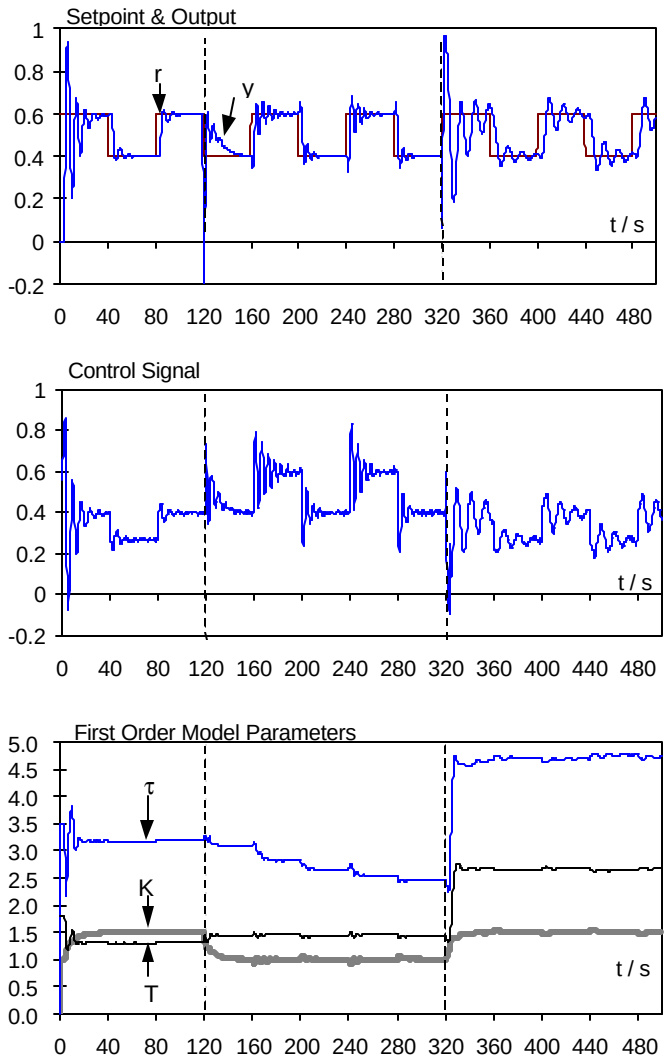


Fig 4. Simulation results of APID

5. CONCLUSIONS

In this paper, a new on-line FOPTD modelling method is proposed which is designed using fuzzy system theory. The proposed method is different from other fuzzy identification methods since it is integrated with a model generator to determine the parameters of FOPTD. The outputs of the fuzzy system are the three parameters of the FOPTD model. Combining with a PID controller, an on-line adaptive control using fuzzy system is designed and tested. The simplicity of the scheme for model-based control provides a new approach for implementing fuzzy applications for a variety of industrial control problems. Results presented clearly demonstrate the adaptive property of the proposed method.

6. ACKNOWLEDGEMENTS

The authors gratefully acknowledge the financial support of the Hong Kong Polytechnic University

REFERENCES

Ashwworth, M.J. (1982) Feedback design of systems with significant uncertainty. Lechworth, U.K. Research Studies Press.

Bhal, N. and Mecavoy, T.J. (1990) "Use of neural nets for dynamics modeling and control of chemical process systems." Comput. Chem. Eng. 14: 573-583.

Babuska, R. and Vebruggen, H.B. (1996). "An overview of fuzzy modelling for control." Control Eng. Practice. 4: 1593-1606.

Chen, C.L.; Chen P.C. and Chen, C.K. (1998) "Analysis and design of fuzzy control system." Fuzzy Sets Systs. 57:125-140.

Czogala, E. and Pedrycz, W. (1981) "On identification in fuzzy systems and its applications in control problems." Fuzzy Sets and Systems, 6: 73-83.

Gawthrop, P.J.; Nihtila, M.T. and Rad, A.B. (1989) "Recursive parameter estimation of continuous-time systems with unknown time delay." Control Theory and Advanced Technology (C-TAT). 5(3): 227-248.

Lu, Y.Z.; He, M. and Xu, C.W. (1997) "Fuzzy modeling and expert optimisation control for industrial process." IEEE Trans. Control Systems Technol. 5:2-12.

Narendra, K.S. and Parthasarathy, K. (1990) "Identificaion and control of dynamical systems using neural networks." IEEE Trans. On Neural Networks, 1: 4-27.

Rad, A.B., Lo, W.L., and Tsang, K.M. (1997) "Self-tuning PID controller using Netwon Raphson search method" IEEE Trans. On Industrial Electronics. 44 (5): pp. 717-725.

Smith, J.A.; Lopez, A.M.; and Murrill, P.W. (1967) "A comparison of controller tuning rules." Control Engineering.14, 72.

Smith, C. L. and Corripio, A.B. (1997) Principles and Practice of Automatic Process Control. John Wiley and Sons, Inc., Second Edition.

Sugeno, M. and Kang, G.T. (1986) "Structure identification of fuzzy model." Fuzzy sets syst. 28:329-346.

Sundaresan, K.R. and Krishnaswamy, P.R. (1978) "Estimation of Time Delay Time Constant Parameters in Time, Frequency, and Laplace Domains." The Canadian Journal of Chemical Engineering, 56, April: 257-262.

Takagi, T. and M. Sugeno. 1985. "Fuzzy identification of systems and its applications to modeling and control." IEEE trans. Syst. Man. Cybern. 15:116-132.

Tanaka, K. and Sugeno, M. (1992) "Stability analysis and design of fuzzy control systems." Fuzzy Sets Systs. 45:135-156.

Wang, H.O.; Tanaka, K. and Griffin, M. (1996) "An approach to fuzzy control of nonlinear systems: stability and design issues." IEEE Trans. Fuzzy Syst. 4(1):14-23.

Wang, L.X. (1997) A course in fuzzy systems and control. Englewood Cliffs, NJ :Prectice Hall.

Yu, C.C. (1999). Autotuning of PID Controllers. Springer, London.

Zadeh, L.A. (1965) "Fuzzy sets and systems" in: Proc. Sysmp. Systems Theory, Polytech. Inst. Brooklyn. 29-27.

Ziegler,J.G. and Nichols, N. B. (1942) Optimum Settings for Automatic controllers. Transactions on ASME, Nov: 759-768.

Appendix A1

$$\dot{\boldsymbol{X}}_m(t) = \boldsymbol{A}\boldsymbol{X}_m(t) + \boldsymbol{B}u(t-\boldsymbol{q}),$$
$$y_m(t) = \boldsymbol{C}\boldsymbol{X}_m(t) + \boldsymbol{D}u(t-\boldsymbol{q})$$
$$\boldsymbol{X}_m(t) = \boldsymbol{\Phi}(t-t_0)\boldsymbol{X}_m(t_0) + \int_{t_0}^{t}\boldsymbol{\Phi}(t-\boldsymbol{l})\boldsymbol{B}u(\boldsymbol{l}-\boldsymbol{q})d\boldsymbol{l}, \quad \text{(A1)}$$

$$\boldsymbol{\Phi}(t) = e^{\boldsymbol{A}t}$$
$$y_m(t) = \boldsymbol{C}\boldsymbol{\Phi}(t-t_0)\boldsymbol{X}_m(t_0) + \boldsymbol{C}\int_{t_0}^{t}\boldsymbol{\Phi}(t-\boldsymbol{l})\boldsymbol{B}u(\boldsymbol{l}-\boldsymbol{q})d\boldsymbol{l} + \boldsymbol{D}u(t-\boldsymbol{q})$$

Assuming that all initial states are zeros and $\boldsymbol{D=0}$, the output equation becomes:

$$y_m(t) = \int_{t_0}^{t} h(t-\boldsymbol{l},\boldsymbol{p})u(\boldsymbol{l}-\boldsymbol{q})d\boldsymbol{l}$$
$$e^{-sq}G_m(s) = e^{-sq}L[h(t,\boldsymbol{p})] = e^{-sq}\boldsymbol{C}(s\boldsymbol{I}-\boldsymbol{A})^{-1}\boldsymbol{B} \quad \text{(A2)}$$
$$= e^{-sq}\frac{b_0 s^n + b_1 s^{n-1} + \ldots + b_m}{s^n + a_1 s^{n-1} + \ldots + 1} = e^{-sq}\frac{B_m(s)}{A_m(s)}$$

where $\boldsymbol{q}$ is the model time delay and $h(t,\boldsymbol{p}) = L^{-1}[G_m(s)]$ is the impulse function of $G_m(s)$. The vector is defined as $\boldsymbol{p} = [a_1\ a_2 \ldots a_n\ b_0\ b_1 \ldots b_m]$. The partial derivatives of the model output with respect to the time delay and the model parameters are as follows.

$$\frac{\partial y_m}{\partial \boldsymbol{q}} = \frac{\partial}{\partial \boldsymbol{q}}\int_{t_0}^{t} h(t-\boldsymbol{l},\boldsymbol{p})u(\boldsymbol{l}-\boldsymbol{q})d\boldsymbol{l}$$
$$= \int_{t_0}^{t}\frac{\partial}{\partial \boldsymbol{q}}[h(t-\boldsymbol{l},\boldsymbol{p})u(\boldsymbol{l}-\boldsymbol{q})]d\boldsymbol{l} \quad \text{(A3)}$$
$$= \int_{t_0}^{t} h(t-\boldsymbol{l},\boldsymbol{p})\frac{\partial u(\boldsymbol{l}-\boldsymbol{q})}{\partial \boldsymbol{q}}d\boldsymbol{l}$$
$$= \int_{t_0}^{t}[-h_m(t-\boldsymbol{l},\boldsymbol{p})\dot{u}(\boldsymbol{l}-\boldsymbol{q})]d\boldsymbol{l} = -L^{-1}[se^{-sq}G_m(s)U(s)]$$

$$\frac{\partial y_m}{\partial a_i} = \frac{\partial}{\partial a_i}\int_{t_0}^{t} h_m(t-\boldsymbol{l},\boldsymbol{p})u(\boldsymbol{l}-q)d\boldsymbol{l}$$
$$= \int_{t_0}^{t}\frac{\partial}{\partial a_i}[h_m(t-\boldsymbol{l},\boldsymbol{p})u(\boldsymbol{l}-q)]d\boldsymbol{l} \quad \text{(A4)}$$
$$= \int_{t_0}^{t}\frac{\partial h(t-\boldsymbol{l},\boldsymbol{p})}{\partial a_i}u(\boldsymbol{l}-q)d\boldsymbol{l}$$

$$L[\frac{\partial h}{\partial a_i}] = \int_0^{\infty}\frac{\partial h}{\partial a_i}e^{-st}dt = \frac{\partial}{\partial a_i}\int_0^{\infty}h(t,\boldsymbol{p})e^{-st}dt$$
$$= \frac{\partial}{\partial a_i}L[h(t,\boldsymbol{p})] = \frac{\partial G_m(s)}{\partial a_i} \quad \text{(A5)}$$
$$\Rightarrow \frac{\partial h}{\partial a_i} = L^{-1}[\frac{\partial G_m(s)}{\partial a_i}]$$

Similarly $\frac{\partial h}{\partial b_i} = L^{-1}[\frac{\partial G_m(s)}{\partial b_i}]$

$$\frac{\partial y_m}{\partial a_i} = \int_{t_0}^{t}\frac{\partial h(t-\boldsymbol{l},\boldsymbol{p})}{\partial a_i}u(\boldsymbol{l}-q)d\boldsymbol{l}$$
$$= L^{-1}[e^{-sq}\frac{\partial G_m(s)}{\partial a_i}U(s)] \quad \text{(A6)}$$
$$= L^{-1}[-e^{-sq}\frac{s^{n-i}B_m(s)}{[A_m(s)]^2}U(s)]$$

$$\frac{\partial y_m}{\partial b_i} = \int_{t_0}^{t}\frac{\partial h(t-\boldsymbol{l},\boldsymbol{p})}{\partial b_i}u(\boldsymbol{l}-q)d\boldsymbol{l}$$
$$= L^{-1}[e^{-sq}\frac{\partial G_m(s)}{\partial b_i}U(s)] \quad \text{(A7)}$$
$$= L^{-1}[e^{-sq}\frac{s^{m-i}}{A_m(s)}U(s)]$$

$$\frac{\partial y_m}{\partial \boldsymbol{q}} = L^{-1}[-se^{-sq}\frac{B_m(s)}{A_m(s)}U(s)] \quad \text{(A8)}$$

For first order with time delay model $\frac{Ke^{-st}}{Ts+1}$

$$\frac{\partial y_m}{\partial K} = L^{-1}[\frac{e^{-st}}{Ts+1}U(s)] \quad \text{(A9)}$$
$$\frac{\partial y_m}{\partial T} = L^{-1}[-\frac{Kse^{-st}}{(Ts+1)^2}U(s)] \quad \text{(A10)}$$
$$\frac{\partial y_m}{\partial \boldsymbol{t}} = L^{-1}[-\frac{Kse^{-st}}{Ts+1}U(s)] \quad \text{(A11)}$$

For second order with time delay model $\frac{Ke^{-st}}{a_0 s^2 + a_1 s + 1}$

$$\frac{\partial y_m}{\partial K} = L^{-1}[\frac{e^{-st}}{a_0 s^2 + a_1 s + 1}U(s)] \quad \text{(A12)}$$
$$\frac{\partial y_m}{\partial a_0} = L^{-1}[-\frac{Ks^2 e^{-st}}{(a_0 s^2 + a_1 s + 1)^2}U(s)] \quad \text{(A13)}$$
$$\frac{\partial y_m}{\partial a_1} = L^{-1}[-\frac{Kse^{-st}}{(a_0 s^2 + a_1 s + 1)^2}U(s)] \quad \text{(A14)}$$
$$\frac{\partial y_m}{\partial \boldsymbol{t}} = L^{-1}[-\frac{Kse^{-st}}{a_0 s^2 + a_1 s + 1}U(s)] \quad \text{(A15)}$$

The control signal $u(t)$ is filtered by the filter function in eq. (A9-A11) to find the partial derivatives of $y_m(t)$ with respect to various model parameters.

A NOVEL SOFT SENSOR MODELING FOR GASOLINE ENDPOINT OF THE CRUDE UNIT

Xuemin TIAN, Guochu CHEN

Department of Automation
University of Petroleum
Dongying, P.R.China 257061

ABSTRACT: This paper presents a novel soft sensor model for identifying the gasoline endpoint of a crude unit. A hybrid model was developed, by combining in series a first principle model with a neural network. A nonlinear observer based on the prior knowledge of the process is designed to estimate the composition C of the upper unit. The neural network is used to predict the gasoline endpoint with C and other process parameters as its inputs. The error between real measurement and the network prediction is fedback for the network correction. Industrial applications of the proposed model indicate that the proposed model is accurate and adaptable. *Copyright©2003IFAC*

KEYWORDS: soft sensor modeling, gasoline endpoint, first principle models, neural network, hybrid model

1. INTRODUCTION

Soft sensing technique has been increasingly used as an attractive and effective method for process modeling (McAvoy, 1992) and for the replacement of expensive and inefficient analytical instrumentation (Santen A, et al, 1997). First principle modeling and empirical modeling are the common two methods that are used in the soft sensor modeling of a chemical process. First principle model (FPM) is based on the analysis of the mass, momentum, and energy balance as well as empirical state equations. However, only major characteristics and trends of the process can be well described by the FPM. In developing such a model, certain assumptions that may be strong have to be made. Also, disturbances that are common in a practical process are difficult to be modeled. These can often lead to poor model precision. Empirical models (EM) (T.Montin, 1998), on the other hand, are computationally efficient. Data-driven models, such as statistical analysis, neural network and fuzzy deduction, can model a nonlinear process accurately in the domain covered by the data, even if unmeasured disturbances are present. But, they often lack of good process interpretability for the dynamic behaviors of the system.

In this paper, we propose a soft sensor model, by combining a first principle model with a neural network, to overcome the drawbacks of these two approaches in their separate forms. Industrial application to the gasoline endpoint prediction shows that this model is more accurate and adaptable.

2. PROCESS DESCRIPTION

The crude distillation unit is a front operation for a refinery. This unit performs the initial distillation of crude oil into several fractions of different boiling ranges. The products of the crude distillation unit are either feedstock for other processing units or a part of product blending pool. The lighter fraction distillated from the upper section of the crude unit (Fig.1.) can be used as gasoline. End point is an important quality indicator for the gasoline produced from the unit. The operating economics of the crude unit

generally dictate that the units run as close as possible to the product specifications (end point). A soft sensor for the endpoint measurement needs to be constructed due to the lack of an effective online analytic instrument.

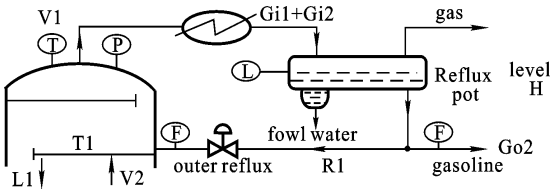


Fig.1. Schema of the upper section of a crude distillation unit

3. SOFT SENSOR MODELING SCHEME

A hybrid model is developed to identify the endpoint of gasoline for the crude unit, by combining, in series, a FPM model that predicts the upper section composition factor C and a multilayer neural network that predicts the gasoline endpoint by taking the composition C and a number process parameters as the network input. The error between the real value and the prediction is feedback to the network for correction. The general structure is illustrated as Fig.2. below.

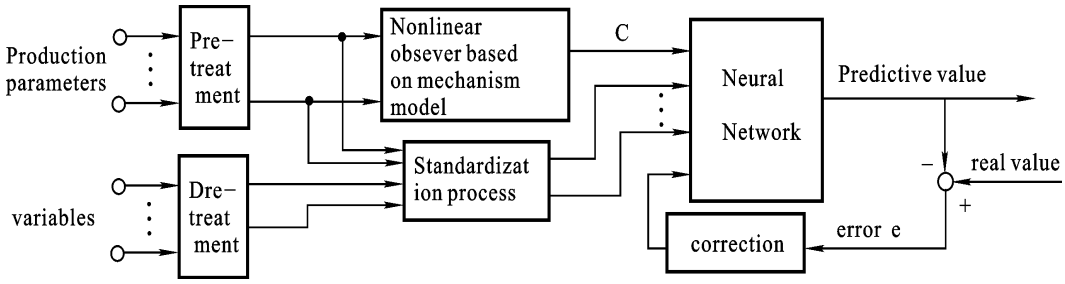


Fig.2. General structure of soft sensor model for gasoline endpoint

In the crude distillation operation, it is a common knowledge that the composition of the top inner section will affect the product quality. Disturbances can change the cut points of the fractions. There lacks an on-line measurement of the upper inner composition, a nonlinear observer is proposed to estimate it. We define that the composition of the top inner section as $C=L_1/V_2$. C increases, when the composition of the feed becomes heavier. The condensed liquid phase increases, as the inner reflux liquid phase L_1 increases, with an unchanged upper unit vapor phase V_2. For the similar reasons, C decreases with a lighter feed. So, C, as defined, can reflect the changes in the cut fraction composition of the upper unit. For this reason, C is referred as a composition factor.

4. ESTIMATIOR DESIGN FOR COMPOSITION FACTOR C

The composition estimator, developed for the crude distillation based on FPM, is derived from the first principles of the material and energy equations. The model equation for the process, based on the separation process principles, is:

$$A_0 \rho_0 \frac{dH}{dt} = G_{i2} - G_{02} - R_1 \tag{1}$$

The energy conservation of the upper unit results in:

$$\frac{dT_1}{dt} = \frac{R_1\left(H_1^R - H_1^L\right) + V_2\left(H_2^V - H_1^L\right) - V_1\left(H_1^V - H_1^L\right)}{n_1 A \rho_1 \left(a_{11} + 2a_{12}T_1\right)\left[0.00284\left(\frac{L_1}{\rho_1 l_w}\right)^{2/3} + h_w\right]} \tag{2}$$

By rearranging Eq. (1), the observation equation for V_1 is given as:

$$\begin{cases} Z_1(k+1) = (1 - g_1 d_1) Z_1(k) - g_1^2 d_1 H(k) + g_1 d_1 (G_{02}(k) + R_1(k)) \\ \hat{G}_{i2}(k) = Z_1(k) + g_1 H(k) \\ V_1(k) = G_{i1} + \hat{G}_{i2}(k) \end{cases} \tag{3}$$

By rearranging Eq. (2), the observation equation for V_2 is given as:

$$\begin{cases} Z_2(k+1) = (1 - g_2 d_2) Z_2(k) - g_2^2 d_2 T_1(k) - g_2 f_1 R_1(k) + g_2 f_2 V_1(k) \\ \hat{V}_2(k) = Z_2(k) + g_2 T_1(k) \end{cases} \tag{4}$$

Since the change of the retaining flow in the

tower tray is much faster than the change of the temperature, the dynamic behavior of the retaining flow can be neglected. By observing V_1 and V_2, we get the observation equations for the liquid flow L_1 and the composition factor C of the upper unit:

$$\begin{cases} L_1(k) = R_1(k) + V_2(k) - V_1(k) \\ C(k) = L_1(k) / V_2(k) \end{cases} \quad (5)$$

5. NEURAL NETWORK MODELS

Artificial Neural Network (ANN) is widely used for its ability in modeling complex nonlinear processes, with a minimal requirement of the process knowledge. BP network is one of the most widely used among numerous networks, so it is adopted here to construct the soft sensor model for the gasoline endpoint.

In the network, the output is the gasoline endpoint, the number of hidden layers is one and the number of nodes in the hidden layer is I, the number of nodes in the input layer is n, the input signal is $x_1^{kk} \sim x_n^{kk}$ (kk=1, 2, …, np, np is the total number of the samples), the weight between the input nodes and hidden nodes is W1(n, i), the weight between hidden nodes and output nodes is W2(i). They are determined by the net transfer function ϕ and the training data sets. The structure of this network is illustrated as in Figure.3.

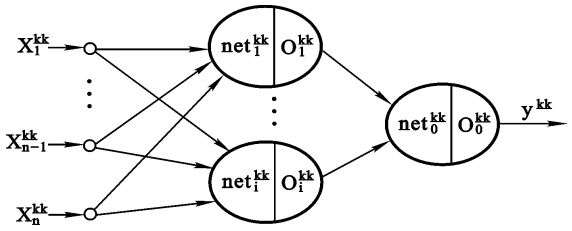


Fig.3. Network structure (NN1)

After setting the structure and the initial weights of the network, BP algorithm (Chen, 2000) can be used to train NN1, using the sample data of the inputs and outputs, to obtain the network weights.

In the production, there always exist uncertain disturbances and time-variance process factors that can influence the network prediction accuracy. A standard BP network is only suitable within a certain range of operation conditions. To improve its prediction, we need to make correction to the network model. Correction includes real-time dynamic correction and updating model periodically. Updating model means that, when sufficient new samples are accumulated, the system model is re-constructed based on these new training samples. Real-time dynamic correction refers to correct the model according to the difference between the real value and the model predictive value. In this paper, we would use the dynamic correction for the soft sensor modeling, by error-feedback correction.

The error correction structure is formed by adding an error feedback element to the structure of the BP network (NN1). With the basic structure of NN1 unchanged, an additional input node and its relevant weights are added. The added input node f_1^{kk}, i.e. the correction input, can be estimated by the error feedback. The new network (NN2) contains the new weights, and the weights of the previous network NN1.

6. INDUSTRIAL EXAMPLE

6.1 BPNN Model

By analyzing the crude unit process, we can determine the input variables that are relevant to the gasoline endpoint. They are the pressure of the crude unit x1, the top temperature of the crude unit x2, the temperature of reflux x3, the outer reflux flow x4, the distillation temperature of the first section x5, the flow of the first section x6, the flow of the first middle part x7, the pressure of input material x8, the temperature of input material x9, and the composition factor C of the upper section x10. The output is gasoline endpoint.

In order to show the effect of introducing composition factor C to the soft sensor model of the gasoline endpoint, a model is also constructed

for comparison by deleting the composition factor C and its related weights from the above-developed model. By comparing the prediction results of these two models, with C and without C, we can obviously see the effect of estimating C.

The trained network has the following structure. The number of hidden layer is 1, and in the hidden layer, the number of nodes is 8. NN1 represents the following nonlinear relationship:

$$y(k) = f(x1(k), x2(k), \cdots, x10(k)) \qquad (6)$$

Where, $f(\cdot)$ indicates the nonlinear function and k is sampling moment.

6.2 Results And Analysis

Comparison of learning results 170 pairs of samples are used, after pretreatment of error testing, filtering and normalization, to train three different soft sensor models as described above (see table 1)

Table 1 the contrast of the results of three models after learning

		average	Mini-Mum	maximum		|e|	|e|<1	1<|e|<2	|e|>2
Real value（℃）		185.8	178	193		total	170		
Com-putation value	without C	186.0	179.5	192.3		no.	141	21	8
	No Correction	185.9	178.8	192.99		no.	159	10	1
	Corrected	185.85	178.5	192.6		no.	169	1	0
Absolute average error	without C	1.15（℃）			deviation	without C	2.145（℃）		
	No Correction	0.881（℃）				No correction	1.223（℃）		
	Corrected	0.159（℃）				Corrected	0.064（℃）		

Comparison of validation results 50 pairs of samples are used to test the three trained models (see Table 2). The deviation of model prediction with C and error correction is less than 1℃. In Figures 4, and 5, solid lines stands for the real value of the gasoline endpoint, dotted-line stands for the model prediction. The abscissa stands for sample number and the vertical coordinate indicates gasoline endpoint (℃).

Table 2 the contrast of the results of industrial validation

		average	Mini-mum	Maxi-mum		|e|	|e|<1	1<|e|<2	|e|>2
Real value（℃）		185.7	178	190		total	50		
computation value	Without C	185.8	178.5	192.0		no.	35	9	6
	No correction	185.5	179.0	189.8		no.	39	7	4
	Corrected	185.7	178.6	190.0		no.	49	1	0
Absolute average error	Without C	1.61 (℃)			deviation	Without C	4.78 (℃)		
	No correction	1.241（℃）				No correction	2.583（℃）		
	Corrected	0.165（℃）				Corrected	0.077（℃）		

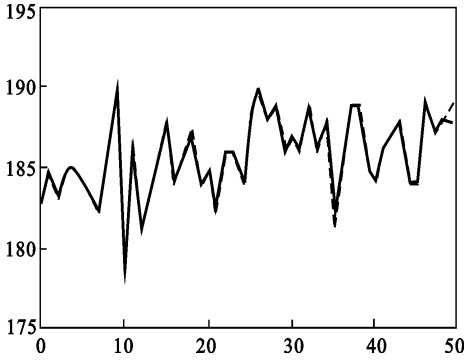


Fig.4. Industrial validation results with C and Error correction

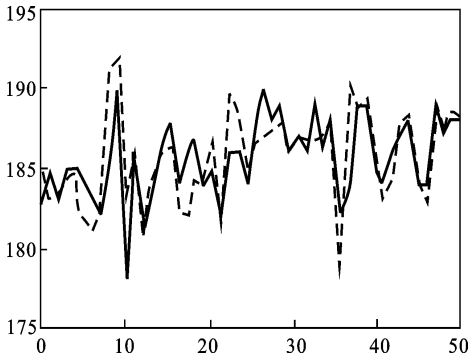


Fig.5. Industrial validation result without C

From Tables 1 and 2 and Figures 4 and 5, it is obvious that the average error and standard deviation of the soft sensor model with C is far smaller than those of the model without C. It is also shown that the modeling with correction has much better precision than the one without correction. The generalization performance of the model with error correction is also better than that of the model without the correction.

7 CONCLUSIONS

A modeling method has been proposed by combining a first principle model with a neural network to form a hybrid model. A soft sensor model with composition C and error correction has been presented for the endpoint prediction for a crude distillation with superior performance to the models without error correction or composition C estimation.

ACKNOWLEDGEMENT

Authors are grateful to China National Petroleum Corporation (CNPC) for its support, which made this study possible.

REFERENCES

A. Brambilla, F.Trivella (1996). *Estimate Product Quality with ANNs* [J] H. P., 1996, (9): 61-66

Guochu Chen (2000). *Simulation and Soft Sensor for Crude Unit*, Thesis of Master, University of Petroleum, 2000.

Jiang Qingyin (1986). et. al *An Approach to Estimate Unmeasurable Inputs*[J] J. Of East China Petroleum Institute 1986, 10(4): 105-113

McAvoy T J (1992). *Contemplative Stance for Chemical Process Control* [J] Automatica, 1992, 28 (2): 441-442

Santen A, et. al. (1997). *Statistical data analysis of a chemical plant* [J] Comput Chem Eng. 1997, 21 (supple): sl123-sl129

Thompson M L, and Kramer M A (1994). *Modeling Chemical Process Using Prior Knowledge and Neural Networks* [J] AIChE. J. 1994,40(8): 1328-1340

T.Mejdell, and S.Skogestad (1991). *Estimation of Distillation Composition from Multiple Temperature Measurements Using Partial-Least-Squares Regression* [J] Ind.Eng.Chem.Res. 1991,30:2543-2555

T.Mejdell, and S.Skogestad (1991). *Composition Estimator in a Pilot-Plant Distillation Column Using Multiple Temperatures*[J] Ind. Eng. Chem. Res. 1991, 30: 2555-2564

T.Montin (1998). *Online Flash Point Calculation Improves Crude Unit Control* [J] H. P. 1998. 77 (12): 81-84

Willis M J, et al. (1992). Artificial Neural Networks of Process Estimation and Control [J] Automatica, 1992, (6): 1181-1187

W. Zhong, and J. Yu (2000). *Improve Nonlinear Soft Sensing Modeling by Combining Multiple Models*[J] H．P．2000, (4): 108-112

ZHANG Naiyao, and YAN Pingfan (1998). *Neural Networks And Fuzzy Control* [M] Beijing: TsingHua University Press, 1998

STATE AND PARAMETER ESTIMATION THROUGH DYNAMIC BAYESIAN FORECASTING

Zhen Lu, Elaine Martin and Julian Morris

Centre for Process Analytics and Control Technology
School of Chemical Engineering and Advanced Materials,
University of Newcastle, Newcastle upon Tyne, NE1 7RU, England, UK

Abstract: Successful application of model based control depends on having good estimates for the system dynamic states and parameters. A multivariate dynamic linear model is developed for the estimation of the states from limited measurements in a non-linear system comprising model uncertainties. Since the noise statistics are rarely available *a priori*, the noise covariance matrix is treated as a tuning parameter and determined through repeated simulations. For non-linear, time varying processes, the assumption of a constant process noise covariance matrix does not realise accurate estimates. In this paper Monte Carlo simulations are used to obtain the time-varying noise covariance matrix. The methodology is demonstrated on a benchmark polymerisation process.

Keywords: State estimation; Parameter estimation; Dynamic linear model; Monte Carlo simulation; Extended Kalman filter

1. INTRODUCTION

In batch polymerisation processes, the operating objectives require the satisfaction of complex property requirements for the final polymer whilst reducing production costs. Most mechanical and rheological properties of polymer products are directly, or indirectly, linked to the molecular structural properties of the polymer chains that are not usually measured on-line. Average polymer molecular weight properties (e.g. number and weight average molecular weight), and particle size distribution which can be inferred from on-line measurements, are often selected as the major controlled variables that need to be maintained within well-determined limits so that the desired product quality criteria can be achieved.

Recursive stochastic state estimation techniques, such as the Extended Kalman filter have been traditionally used for state and parameter estimation especially for polymerization processes. The main bottleneck in the application of recursive stochastic state estimation techniques to real world situations is that the process noise statistics are rarely available *a priori*. In most applications, they serve as tuning parameters and are determined through a trial-and error procedure using repeated simulations.

Few techniques for determining the process noise covariance matrix have been developed for chemical engineering applications. Zhou and Luecke (1995) used maximum-likelihood estimation with linear regression to obtain the diagonal elements of the covariance matrices for linear systems. For non-linear systems, the use of innovation processes for estimating the noise statistics was proposed by Myers and Tapley (1976). However they assumed the covariance matrix to be constant. For batch processes with time-varying process dynamics that operate over a range of process conditions, this is not the case. The specification of a constant process noise covariance matrix may not be sufficient to provide sufficiently accurate estimation. Using a fixed value of noise statistic can lead to poor estimation or potentially result in filter divergence.

Valappil and Georgakis (2000) introduced two approaches to systematically estimate the process noise covariance matrix. The first method was based on Taylor series expansion whereas the second method used Monte Carlo simulations to calculate the time-varying values of the process noise

covariance matrix on-line. Both methods require information about the plant-model mismatch in the form of a parameter covariance matrix. The process noise covariance matrix is obtained from the parameter covariance matrix. If the user is not certain about the process-model mismatch, or the model uncertainty can not be represented by the parameter covariance matrix, it is difficult to apply these methods.

.

In this paper, a new approach is proposed where a multivariate dynamic model is constructed for the estimation of the states. Monte Carlo simulations are then used to calculate the time-varying process noise covariance matrix on-line from the prediction errors.

2. BAYESIAN DYNAMIC MODEL

2.1 Multivariate Dynamic Linear Model

The dynamic linear model (DLM) is a Bayesian forecasting tool based on a state space model that allows a variety of adaptive linear and generalised linear models to be fitted iteratively to univariate or multivariate time series data. A DLM incorporates information from any relevant source, including subjective expert views, leading to amended and updated model structures. The general multivariate dynamic linear model developed by West and Harrison (1997), is given by the following system of equations:

Observation equation:

$$\mathbf{Y}_t = \mathbf{F}_t \mathbf{x}_t + \mathbf{v}_t \qquad \mathbf{v}_t \sim N[0, \mathbf{V}_t] \tag{1}$$

System equation:

$$\mathbf{x}_t = \mathbf{G}_t \mathbf{x}_{t-1} + \mathbf{w}_t \qquad \mathbf{w}_t \sim N[0, \mathbf{W}_t] \tag{2}$$

Initial information:

$$(\mathbf{x}_0 \mid D_0) \sim N[\mathbf{m}_0, \mathbf{C}_0] \tag{3}$$

where $\mathbf{Y}_t$ is the observed vector of the series at time point t, $\mathbf{x}_t$ is the state vector, $\mathbf{v}_t$ is the observational error, $\mathbf{w}_t$ is the vector of process noise that is assumed to be independent and normally distributed and $\mathbf{D}_0$ is the initial prior information at $t = 0$. At any future time point, t, the available information set is:

$$D_t = \{\mathbf{Y}_t, D_{t-1}\} \tag{4}$$

At time t-1, for some mean $\mathbf{m}_{t-1}$ and variance matrix $\mathbf{C}_{t-1}$, the posterior is given by:

$$(\mathbf{x}_{t-1} \mid D_{t-1}) \sim N[\mathbf{m}_{t-1}, \mathbf{C}_{t-1}] \tag{5}$$

and the prior for the state vector at time t can be derived from the system equation:

$$(\mathbf{x}_t \mid D_{t-1}) \sim N[\mathbf{a}_t, \mathbf{R}_t] \tag{6}$$

where

$$\mathbf{a}_t = \mathbf{G}_t \mathbf{m}_{t-1} \quad \text{and} \quad \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t \tag{7}$$

According to the observation equation, the one step ahead forecast can be given by:

$$(\mathbf{Y}_t \mid D_{t-1}) \sim N[\mathbf{f}_t, \mathbf{Q}_t] \tag{8}$$

where

$$\mathbf{f}_t = \mathbf{F}_t' \mathbf{a}_t \quad \text{and} \quad \mathbf{Q}_t = \mathbf{F}_t' \mathbf{R}_t \mathbf{F}_t + \mathbf{V}_t \tag{9}$$

The feedback of information obtained at time t from vector $\mathbf{x}_t$ is achieved through the application of linear Bayes methods. As stated previously, the model at time (t-1) requires only the mean vector and covariance matrix of the posterior for $(\mathbf{x}_{t-1} \mid D_{t-1})$. Thus at time t, the corresponding moments:

$$(\mathbf{x}_t \mid D_t) \sim N[\mathbf{m}_t, \mathbf{C}_t] \tag{10}$$

are required in order to progress to time point, (t+1), and subsequent observations. The information obtained at time point, t, is used to update the prior moments to give:

$$\mathbf{m}_t = \mathbf{a}_t + \mathbf{A}_t \mathbf{e}_t \text{ and } \mathbf{C}_t = \mathbf{R}_t - \mathbf{A}_t \mathbf{Q}_t \mathbf{A}_t' \tag{11}$$

where

$$\mathbf{A}_t = \mathbf{R}_t \mathbf{F}_t' \mathbf{Q}_t^{-1} \quad \text{and} \quad \mathbf{e}_t = \mathbf{Y}_t - \mathbf{f}_t \tag{12}$$

2.2 Multivariate Non-linear Dynamic Model

For a non-linear model, the process can be expressed as:

$$\begin{aligned} \mathbf{Y}_t &= f_t(\mathbf{x}_t) + \mathbf{v}_t \quad \mathbf{v}_t \sim N[0, \mathbf{V}_t] \\ \mathbf{x}_t &= g(\mathbf{x}_{t-1}) + \mathbf{w}_t \quad \mathbf{w}_t \sim N[0, \mathbf{W}_t] \end{aligned} \tag{13}$$

Before the usual DLM updating procedure is applied, the model requires to be linearized. The most straightforward and easily interpreted linearization technique is the Taylor series approximation. Applying Taylor series expansion to the updating function about the mean $\mathbf{m}_{t-1}$:

$$\begin{aligned} \mathbf{g}(\mathbf{x}_t) = \mathbf{g}(\mathbf{m}_{t-1}) + \mathbf{G}_t(\mathbf{x}_{t-1} - \mathbf{m}_{t-1}) \\ + \text{quadratic and higher order terms} \end{aligned} \tag{14}$$

where $\mathbf{G}_t$ is the matrix derivative of the updating function evaluated at the estimate $\mathbf{m}_{t-1}$:

$$\mathbf{G}_t = \left[\frac{\partial g_t(\mathbf{x}_{t-1})}{\partial \mathbf{x}_{t-1}} \right]_{\mathbf{x}_{t-1}=\mathbf{m}_{t-1}} \tag{15}$$

$$\mathbf{a}_t = \mathbf{g}_t(\mathbf{m}_{t-1}) \text{ and } \mathbf{R}_t = \mathbf{G}_t \mathbf{C}_{t-1} \mathbf{G}_t' + \mathbf{W}_t \tag{16}$$

Proceeding to the observation equation, a similar approach applies. The non-linear regression function is linearized about the expected value $\mathbf{a}_t$ for $\mathbf{x}_t$, giving:

$$\mathbf{f}_t(\mathbf{x}_t) = \mathbf{f}_t(\mathbf{a}_t) + \mathbf{F}_t'(\mathbf{x}_t - \mathbf{a}_t) + \text{quadratic and higher order terms} \tag{16}$$

where $\mathbf{F}_t$ is the matrix derivative of $\mathbf{f}_t(.)$ evaluated at the prior mean $\mathbf{a}_t$:

$$\mathbf{F}_t = \left[\frac{\partial f_t(\mathbf{x}_t)}{\partial \mathbf{x}_t} \right]_{\mathbf{x}_t=\mathbf{a}_t} \tag{17}$$

The standard updating equations continue to apply.

Most process models constructed from limited experimental observations involve significant uncertainties. For batch and semi-batch processes, this is especially true. Model accuracy is obtained by tuning the covariance matrix ***Q*** of the process noise using a repeated simulation procedure.

If a covariance matrix whose entries are of small magnitude is selected, greater confidence will be expressed in the model and less on-line measurement information will be required to update the states. However, this may result in degraded estimates and possibly estimator divergence. Too much state compensation can cause the state estimates to be noisy and unreliable.

The process noise **w(t)** is mainly due to uncertainties in the model and can be either parametric or structural. Monte Carlo simulations are utilized to estimate a time-varying covariance matrix ***Q*** on-line.

3. ESTIMATION OF THE COVARIANCE MATRIX

The concept of applying Monte Carlo simulation to estimate the process noise covariance matrix is to capture the effect of uncertainties in the model through the statistics of **w(t)**. The key issue is how to derive the information from the measurements and model evolution to represent the process noise. In the algorithm described above, the prediction error vector $\mathbf{e}_t$ gathers the information about the errors that caused by process disturbance and model uncertainties. The updating information from prior to the posterior for the states is obtained by multiplying the prediction errors $\mathbf{e}_t$ with a gain $\mathbf{A}_t$. The idea in this paper is that, using the information drawn from the prediction errors $\mathbf{e}_t$ and the updating information to estimate the process noise covariance matrix.

At a desired time instance *t*-1, a set of samples $\{\mathbf{x}_{t-1}^k\}$ is randomly selected from the posterior distribution function, equation (5), of the state vector while the observation $\mathbf{Y}_{t-1}$ is given. For the k^{th} Monte Carlo simulation, a non-linear model is used to generate random samples of $\mathbf{x}_{t|t-1}$

$$\mathbf{x}_{t|t-1}^k = \mathbf{g}_t(\mathbf{x}_{t-1}^k) \tag{18}$$

Because $\mathbf{x}_{t-1}^k$ is randomly sampled directly from a probability distribution function, some samples may be located in the tails of the distribution. For non-linear systems, the presence of such samples will seriously affect the performance of the estimate. One solution is to reject these samples. The simplest way is to set a rejection bound. Those samples that have a probability larger than the rejection bound will be accepted, otherwise they are discarded and new samples generated. Once the new measurement $\mathbf{Y}_t$ is obtained, the prediction errors are calculated

$$\mathbf{e}_t^k = \mathbf{Y}_t - \mathbf{f}_t(\mathbf{x}_{t|t-1}^k) \tag{19}$$

and the information used to update the prior moments is calculated as follow:

$$\mathbf{?}_t^k = \mathbf{A}_t \mathbf{e}_t^k \tag{20}$$

The process noise is also obtained from the samples of the updating information:

$$\mathbf{w}_t^k = \mathbf{?}_t^k - \overline{\mathbf{?}}_t \tag{21}$$

where $\overline{\mathbf{?}}_t$ denotes the mean of the updating information. The process noise is normally distributed with zero mean. The process noise covariance matrix $\mathbf{Q}_t$ can be calculated from $\mathbf{w}_t^k$, and is a non-diagonal and time-varying matrix. Because the process measurements are available at discrete time instances, the preceding calculation of $\mathbf{w}_t^k$ and $\mathbf{Q}_t$ is performed for discrete time intervals.

For the model development described in the previous section, the process noise is assumed to be a white, Gaussian random process. Thus the approximation of normally distributed process noise needs to be tested using the values of the process noise data set $\{\mathbf{w}_t^k\}$ that were obtained from the Monte Carlo simulations.

For this, normal probability plots were used and the distribution was observed to be approximately normal. For the simulation case presented, 500 Monte Carlo simulations of the different state values were used, resulting in 500 evaluations of the process noise for each state.

4. RESULTS

4.1 MMA polymerization

The process studied is the free radical polymerisation reactor of methyl-methacrylate (MMA) (Mourikas *et al*, 2001). A mathematical model describes the dynamic behaviour of an experimental pilot scale system (Fig 1). Heating and cooling of the reaction mixture is achieved by controlling the flows of hot and cold water stream, through the reactor jacket. The polymerisation temperature is controlled by a cascade control system consisting of a primary PID and two secondary PI controllers. The polymerisation is highly exothermic and exhibits a strong acceleration in polymerisation rate due to gel-effects. Batch duration is 120 minutes.

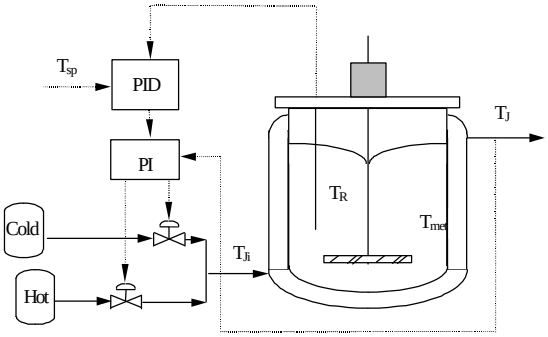


Fig. 1. Plant polymerisation reactor

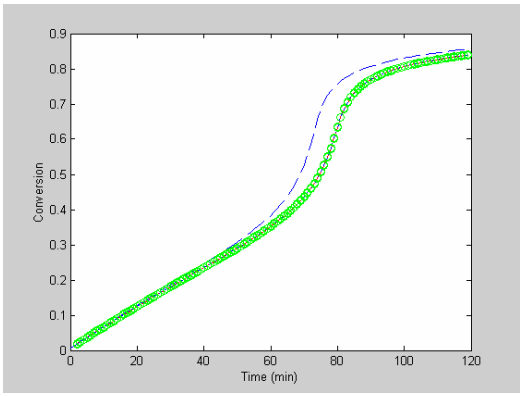
The MMA system consists of 11 states, which are monomer conversion; three moments of dead polymer that are used to calculate the molecular weight distribution; reactor and metal wall temperatures; and four jacket zone temperatures. In polymerization, frequent measurements of the reactor and the jacket inlet and outlet temperatures are usually available along with possibly jacket flow. Monomer conversion measurements can also be obtained from an on-line densitometer. The measurements in a real process environment will be corrupted with measurement noise. In the following simulation, Gaussian white noise is added to the measurements.

In this particular study, the process model mismatch is introduced in the form of time variation in a kinetic parameter. In practice important kinetic parameters such as k_p, the propagation rate constant, cannot be determined accurately and may vary during the polymerisation. In this study, the propagation rate constant is represented by $k_p = k_{p0} g_p g^s_{p,corr}$, where k_{p0} is an intrinsic chemical rate constant, g_p is a diffusion controlled function which includes a number of parameters that are often unknown. The stochastic correction term $g^s_{p,corr}$ is used to account for the imprecise knowledge of g_p. In the model, a random walk is assumed for the behaviour of the stochastic state. In the process, the actual value of $g^s_{p,corr}$ is assumed to decrease linearly from an initial value of 1.0 to 0.76.

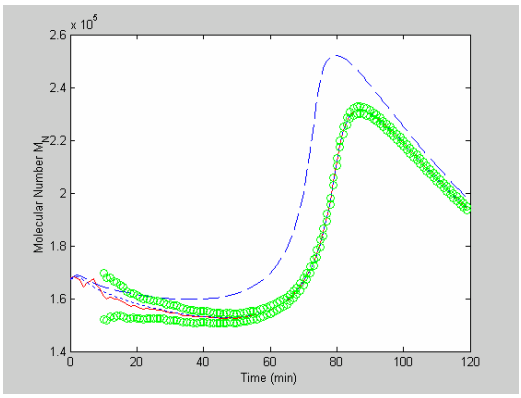
4.2 Discussion

The results of the estimation studies for the MMA polymerization reactor are shown in Fig. 2. It can be seen that the uncorrected model (dashed line), in which the stochastic correction term of the propagation rate constant is fixed at 1.0, differs significantly from the actual plant (dotted line), in which the propagation rate constant is time-varying. The estimates (solid line) closely match the actual process. These, results are compared with the estimation results from an EKF with fixed process noise covariance matrix, Fig. 3. The estimates of number average and weight average molecular weights in Fig. 2. match the actual process while the estimates in Fig. 3. show a discrepancy between the actual process and the estimates.

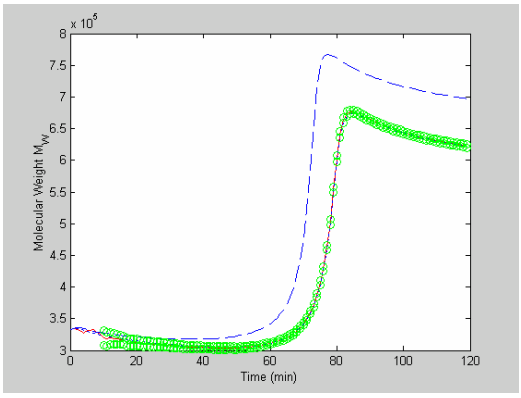
Comparing Fig. 2(e) and Fig. 3(e), the estimator involving Monte Carlo simulation tracks the decrease in the rate constant more closely and faster. This results in better state estimation performance. The 95% confidence bound (circle) of the estimate for the Bayesian approach is narrower than for the EKF, indicating that the estimates are more accurate and reliable. Since only a limited number of observation data can be used to update the estimate, the confidence bounds at the beginning are wide and hence the estimates are less reliable. As the observation data increases, the limits decrease in magnitude. Thus the structure of the confidence bounds for parameter estimation by the Bayesian approach are more reasonable than those of the EKF.



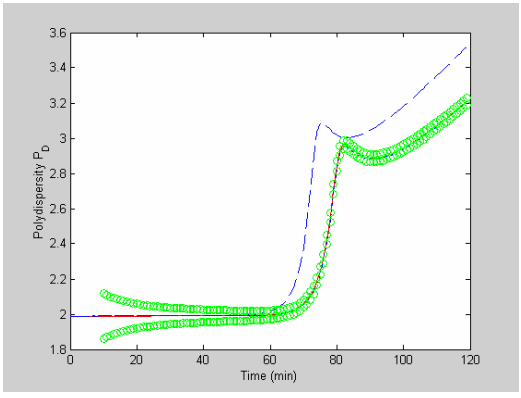
(a) Actual versus estimated (conversion)



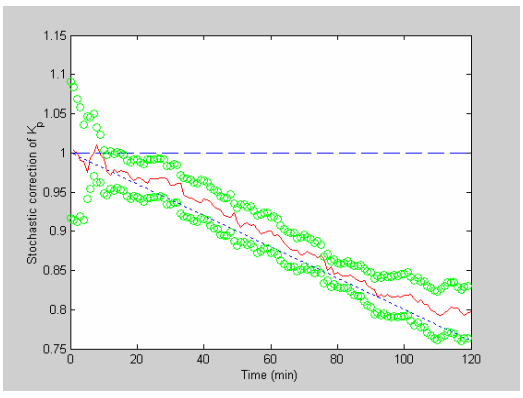
(b) Actual versus estimated (M_N)



(c) Actual versus estimated (M_W)



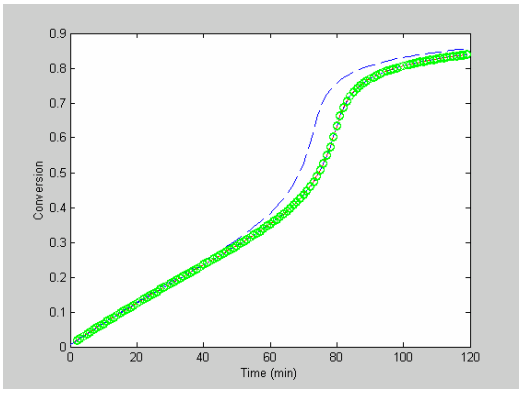
(d) Actual versus estimated (polydispersity)



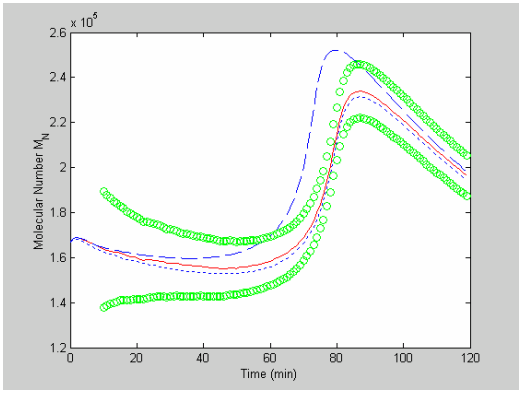
(e) Actual versus estimated (rate constant K_P)

Fig.2. Multivariate DLM state and parameter estimation.

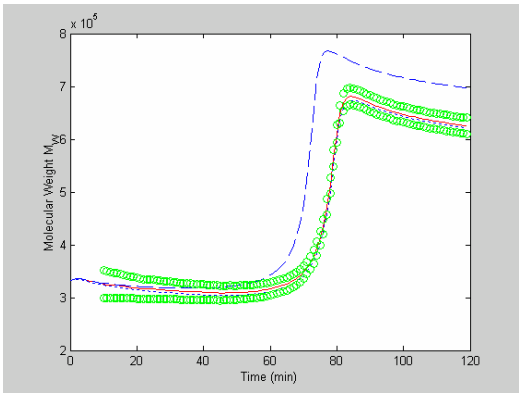
Key: Dotted line - actual plant; solid line - estimates; dashed line - process model with mismatch; circle - 95% confidence bounds



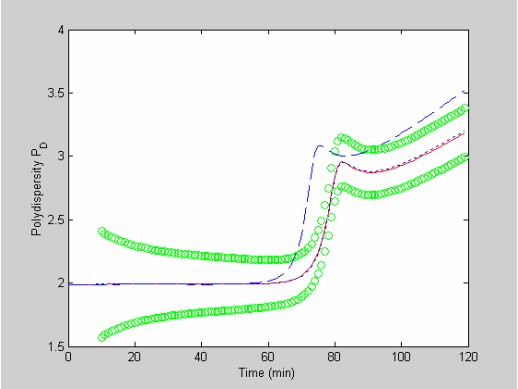
(a) Actual versus estimated (conversion)



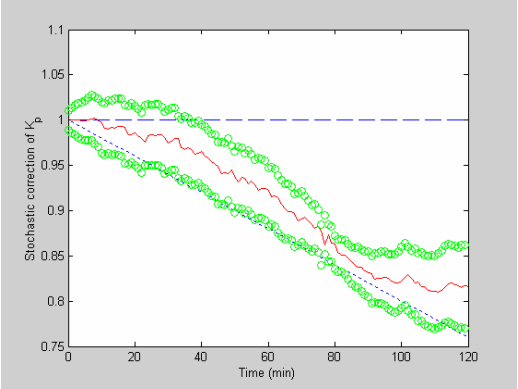
(b) Actual versus estimated (M_N)



(c) Actual versus estimated (M_W)



(d) Actual versus estimated (polydispersity)



(e) Actual versus estimated (rate constant K_P)

Fig.3. EKF state and parameter estimation with fixed process noise covariance matrix.

Key: Dotted line - actual plant; solid line - estimates; dashed line - process model with mismatch; circle - 95% confidence bounds

5. CONCLUSIONS

The feasibility of extending multivariate DLM to estimate the process states, monomer conversion and the molecular weights in MMA batch polymerizations has been demonstrated. The methodology provides a new approach to state estimation for possible application in on-line model-based optimising control. The estimator uses on-line measurements of key process variables including monomer conversion and reactor and jacket temperatures, to provide reliable estimates of the state variables. Monte Carlo simulation is used for the calculation of the process noise covariance matrix.

The results show that the approach presented can improve the performance of the state and parameter estimation. It also makes the design and the application of dynamic Bayesian forecasting more robust, since the methodology proposed eliminates the need for the tuning of the process noise covariance matrix. The non-diagonal and time-varying covariance matrix is obtained on-line in contrast to a diagonal and constant covariance matrix which is not able to adapt to non-linear systems with model uncertain. The algorithm can also be enhanced by Bayesian parameter estimation to provide a significantly enhanced overall state and parameter estimation methodology (Lu *et al*, 2001).

6. AKNOWLEDGMENTS

Zhen Lu acknowledges CPACT and the University of Newcastle for the financial support of his PhD and Prof. Kiparissides, CPERI, Thessaloniki, for the MMA reactor simulation.

7. REFERENCES

Crowley, T.J. and K.Y. Choi (1997). Discrete optimal control of molecular weight distribution in a batch free radical polymerization process. *Ind. Eng. Chem.*, **36**, pp. 3676-3684.

Crowley, T.J. and K.Y. Choi (1998). Experimental studies on optimal molecular weight distribution control in batch free radical polymerization process. *Chem. Eng. Sci.*, **53**, pp. 2769-2790.

Lu, Z., E.B. Martin and A.J. Morris (2002). Bayesian parameter estimation in batch polymerisation. *ESCAPE-12, European Symposium on Computer Aided Process Engineering-12*, 517-522.

Myers, K.E. and B.D. Tapley (1976). Adaptive sequential estimation with unknown noise statistics. *IEEE Trans. Automat. Contr.*, AC**21**, pp. 520-523.

Ruppen, D., D. Bonvin and D.W.T. Rippin (1997). Implementation of adaptive optimal operation for a semi-batch reaction system. *Comput. Chem. Eng.*, **22**, pp. 185-199.

Thomas, I. M., and C., Kiparissides (1984). Computation of the near-optimal policies for a batch polymerization reactor. *Can. J. Chem. Eng.*, **62**, pp. 284-291.

Valappil, J. and C. Georgakis (2000). Systematic estimation of state noise statistics for extended Kalman filters. *AIChE Journal.*, **46(2)**, pp. 292-308.

West, M., and P.J. Harrison (1997). *Bayesian forecasting and dynamic models.* Springer-Verlag, New York.

Zhou, J. and R. Luecke (1995). Estimation of the covariance of the process noise and measurement noise for a linear discrete dynamical system. *Comput. Chen. Eng.*, **19**, pp. 187-195.

HEAT TRANSFER IN A CABLE PENETRATION FIRE STOP SYSTEM

Seong-Pil Kwon, Jaekyu Cho*, Sang-Oak Song*, Wonguk Kim and En Sup Yoon

Institute of Chemical Processes, Seoul National University, Korea (ROK)
** School of Chemical Engineering, Seoul National University, Korea (ROK)*

Abstract: In this work the dynamic heat transfer occurring in a cable penetration fire stop system built in the firewall of nuclear power plants is three-dimensionally investigated to develop a test-simulator that can be used to verify effectiveness of the sealants. The dynamic heat transfer can be described by a partial differential equation (PDE) and its initial and boundary conditions. For the shake of simplicity PDE is divided into two parts; one corresponding to the heat transfer in the axial direction and the other corresponding to the heat transfer on the vertical layers. Two numerical methods, SOR (Sequential Over-Relaxation) and FEM (Finite Element Method), are implemented to solve these equations respectively. The axial line is discretized, and SOR is applied. Similarly, all the layers are separated into finite elements, where the time and spatial functions are assumed to be of orthogonal collocation state at each element. The heat fluxes on the layers are calculated by FEM. It is shown that the penetration cable influences the temperature distribution of the fire stop system very significantly. The simulation results are shown in the three-dimensional graphics for the understanding of the transient temperature distribution in the fire stop system.

Keywords: Dynamic heat transfer, Finite element method, Sequential over-relaxation, Partial differential equation, Cable penetrated fire stop system.

1. INTRODUCTION

For a few decades a great number of nuclear power plants have been constructed and are operating worldwide to supply the industrial and household electricity. Now, 14 nuclear power plants are operating night and day in the republic of Korea, and 6 nuclear power plants are newly under construction. According to the long-term policy, about 20 additional nuclear power plants will be constructed in South Korea until 2020. Due to the so rapid increase of the number of nuclear power plants, a great deal of social concern has been concentrated on the accidents risk of the nuclear power plant in recent times. The accidents risk is well known previously, and fires among the accidents are especially dangerous. Fires can critically affect the control systems of the nuclear power plant because of their quick spread. The fire penetration seal systems that prevent the passage of the fire, gas and heat between compartments so as to reduce the damage and to save lives are deeply associated with the products assembled in the field or pre-manufactured. Silicone and Latex sealant fire stop systems are usually employed in sealing around metal pipes, joints and gaps.

All fire stop systems are tested under the same ASTM standard to ensure repeatability and suitability for the specific application. The Tennessee Valley Authority Browns Ferry nuclear power plant accident that was on March 22 in 1975 resulted in the change of ASTM E-119 to ASTM E-814 or UL-1479, as a standard for testing the performance of fire penetration seal systems. According to the new testing method, previous fire stop systems were revaluated and safety was improved. The fire stop systems have a major responsibility in defense-in-depth.

In this work a simplified fire stop system is considered, shown in Fig. I. The cable penetration fire stop systems built in a nuclear power plant is from 3,000 to 10,000 in number. Because of the great number of the fire stop systems constructed under the old standard of ASTM E-119, safety of all the systems did not verify with the new test method of ASTM E-814 up to now. Corresponding to ASTM E-814, not only the F-rating test but also the T-rating test should be carried out to verify the fire stop system. For that purpose the complementary use of a test-simulator is suitable. Especially, the unsteady-state heat conduction in the fire stop system should be investigated in order to develop the test-simulator that the T-rating test of the fire stop system can be carried out with.

The dynamic heat transfer phenomenon occurring in the fire penetration seal system is formulated in a parabolic partial differential equation subjected to a set of boundary conditions. First, the PDE model is divided into two parts; one corresponding to the heat transfer in the axial direction and the other corresponding to the heat transfer on the vertical faces. The first partial differential equation is converted to a series of ordinary differential equations at finite discrete axial points for applying the numerical method of sequential over-relaxation (SOR) to the problem. After that, we can solve the ordinary differential equations by using an integrator, such as an ODE (ordinary differential equation) solver. In such manner the axial heat flux can be calculated at least at the finite discrete points. For the shake of simplicity a few assumptions are given in this work. There is no heat transfer between the fire stop system and the firewall. The surface of fire site of the fire stop system is always at the temperature of the standard curve of ASTM-119, and the penetration cable is also at the same temperature of the surface. These assumptions are summarized as the boundary condition equations. According to the standard method of ASTM E-814, the fire stop system is exposed to a standard temperature-time fire, and to a subsequent application of cable streams for testing the cable penetration fire stop system. Ratings are established on the basis of the period of resistance to the fire exposure, prior to the first development of through openings, flaming on the unexposed surface, limiting thermal transmission criterion, and acceptable performance under application of the cable stream. This test method specifies that pressure in the furnace chamber with respect to the unexposed surface shall be that pressure which will be applicable to evaluate the fire stop system with respect to its field installation. This pressure shall be determined by a specific code requirement, by the special pressures in the building, in which the fire stop system is to be installed, or by the test sponsor requesting a special environment to evaluate the fire stop specimen (ASTM, 1993). The fire test is performed with the standard temperature-time curve, as shown in Fig. I.

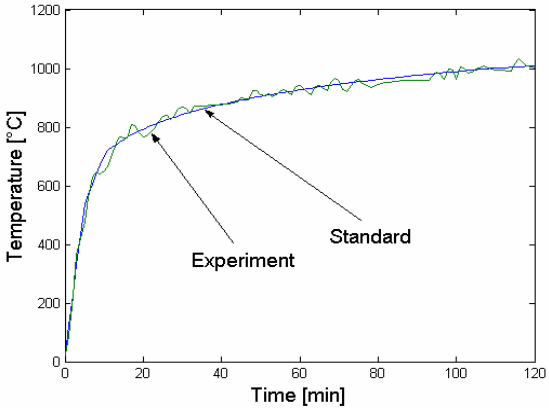


Fig. I: The standard temperature-time curve of ASTM E-119 and the experimental time-temperature curve.

2. MATHEMATICAL MODEL

It is assumed that heat transfer is constant against the change of temperature and pressure, and there is no additional heat generation in the cable penetration fire stop system. In this case the unsteady-state heat transfer in the fire stop system can be described by the parabolic partial differential equation, as follows:

$$\frac{\partial T}{\partial t} = \alpha\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} + \frac{\partial^2 T}{\partial z^2}\right) \qquad (1)$$

The thermal diffusivity α represents the physical property of a sealing material of the fire stop system. The cable penetration fire stop system can be simplified as a cubic, in which several cables are through-passed, at Cartesian coordinate, as shown in Fig. II.

For the shake of simplicity we assumed that the initial temperature of the cable penetration fire stop system and its surface are constant at the temperature of T_0, and the temperature of the inner faces and the whole cables are constant at the temperature of T_h.

$$T(x, y, z, 0) = T_0 \qquad (2)$$

$$T(x, y, 0) = T_h \qquad (3)$$

for $0 \le x \le X$, $0 \le y \le Y$.

$$T(x, y) = T_h \qquad (4)$$

for $(x - x_i)^2 + (y - y_i)^2 \le r_i^2$, $i = 1, 2, \ldots$

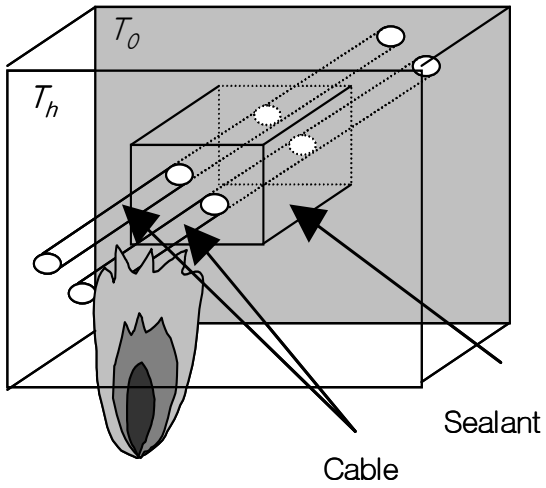


Fig. II: Simplified cable penetration fire stop system.

In addition it is assumed that there is no thermal exchange between the fire stop system and the firewall through the four interfaces – the bottom, up, left and right sides surrounding the sealant cubic, i.e. they are under the adiabatic condition. As a result, it can be described by following equations.

$$-k\frac{\partial T}{\partial x}\Big|_{x=0} = -k\frac{\partial T}{\partial x}\Big|_{x=X} = -k\frac{\partial T}{\partial y}\Big|_{y=0} = -k\frac{\partial T}{\partial y}\Big|_{y=Y} = 0 \qquad (5)$$

The last assumption is that the opposite surface (z=Z) is at a constant temperature of T_Z, and the heat flux, which spreads from the solid surface, is proportional to the temperature difference between the solid surface and the bulk of air, as follows.

$$-k\frac{\partial T}{\partial z}\Big|_{z=Z} = h(T - T_0) \qquad (6)$$

The unsteady-state heat transfer phenomena in the fire stop system can be modelled by a parabolic partial differential equation *Eq*. (1) subjected to an initial condition *Eq*. (2) and a series of boundary conditions *Eqs*. (3), (4), (5) and (6).

3. NUMERICAL CALCULATION

To solve the complex three-dimensional initial value partial differential equation, we used two numerical methods SOR (Sequential Over-Relaxation) and FEM (Finite Element Method) in turn. In this algorithm we calculated the z-axis components derived from the initial value partial differential equation at Cartesian coordinate, and several two-dimensional rectangular systems can be calculated with the results of the z-axis components iteratively. The heat transferred in the z-axis can be calculated by the numerical method of SOR, and the heat transfer on the x-y-layers can be estimated with the numerical method of FEM. Originally, the SOR method is developed as a very sophisticated hand computation technique for solving large sets of simultaneous linear equations iteratively. The overall approach is not well suited to digital computer use because of the extensive logic required, but the original concepts are embodied in the simple but powerful computer-oriented method. Basically, SOR works by using an initial guess of the solution and then progressively improving guesses until an acceptable level of accuracy is reached (Southwell, 1940; Monte, 2002).

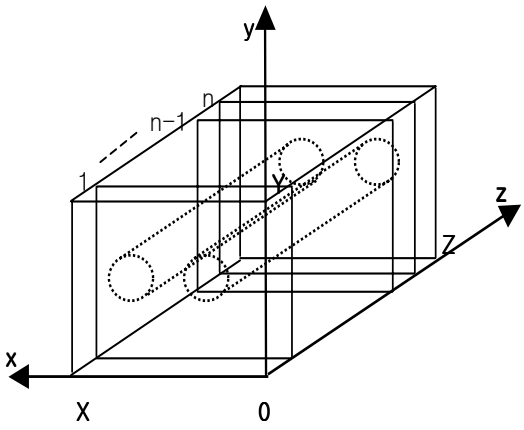


Fig. III: Schematic figure of the fire stop system.

Consider n+1 parallel layers in the fire stop system, as shown in Fig. III. First of all, One-dimensional (z-axis) heat conduction in the fire stop system is described as follows:

$$\frac{\partial T}{\partial t} = \alpha\left(\frac{\partial^2 T}{\partial z^2}\right) \qquad (7)$$

By using the numerical method of SOR the partial differential equation Eq.(7) is discretized to *n* ordinary differential equations.

$$\begin{aligned}
\frac{\partial T_1}{dt} &= \alpha\left(\frac{T_2 - 2T_1 + T(0,t)}{\Delta z^2}\right) \\
\frac{\partial T_2}{dt} &= \alpha\left(\frac{T_3 - 2T_2 + T_1}{\Delta z^2}\right) \\
&\vdots \\
\frac{\partial T_n}{dt} &= \alpha\left(\frac{T_z - 2T_n + T_{n-1}}{\Delta z^2}\right)
\end{aligned} \qquad (8)$$

For the first calculation the initial temperature is constant at the temperature of T_0 for every finite element, but the next step temperature of finite element T_P's can be calculated by FEM (Finite Element Method). Often in the finite element approach, the partial differential equation describing the desired

quantity (such as displacement) in the continuum is not dealt directly. Instead, the continuum is divided into a number of finite elements, which assumed to be joined at a discrete number of points along their boundaries. A functional form is then chosen to represent the variation of the desired quantity over each element in terms of the values of this quantity at the discrete boundary points of the element (Becker, *et al*., 1981; Feirweather, 1978). By using the physical properties of the continuum and the appropriate physical laws, a set of simultaneous equations in the unknown quantities at the element boundary points can be obtained. The temperature of the front surface is constant at the hottest temperature T_h.

$$\begin{aligned} T(z,0) &= T_0 \\ T(z,t) &= T_P \\ T(0,t) &= T_h \end{aligned} \qquad (9)$$

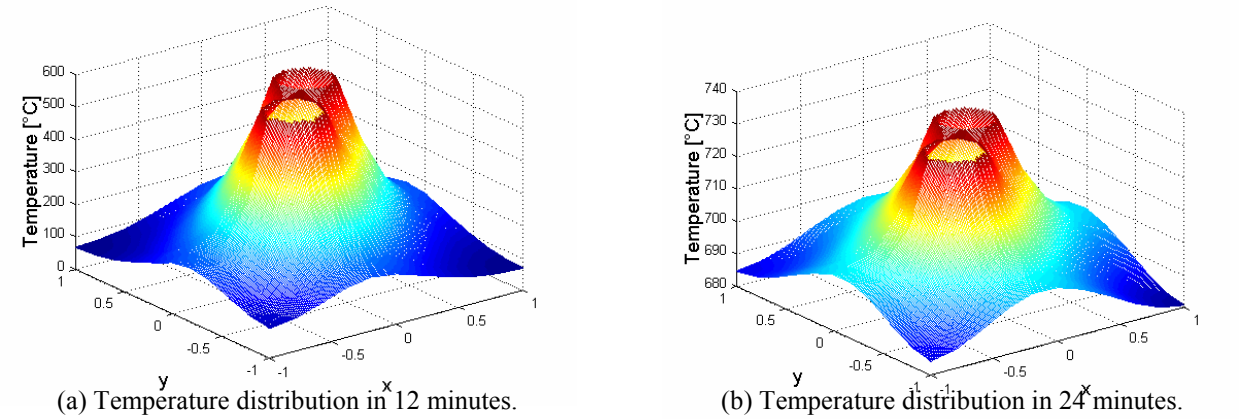
On the assumption that the temperature of the back surface is constant at the initial temperature of T_0, the initial temperature, the heat flux that passes through a layer is proportional to the temperature difference between the layer ($z = z_i$) and the next layer ($z = z_{i+1}$). Therefore, the equation *Eq*. (6) can be approximated to the following equation by using the forward difference method.

$$T_z = \frac{h \cdot T_0 + \dfrac{k}{2 \cdot \Delta z}(4T_n - T_{n-1})}{h + \dfrac{3k}{2 \cdot \Delta z}} \qquad (10)$$

Moreover, the horizontal heat transfer of the back surface can be estimated and can be applied to p finite elements by the partial differential equation *Eq*. (11).

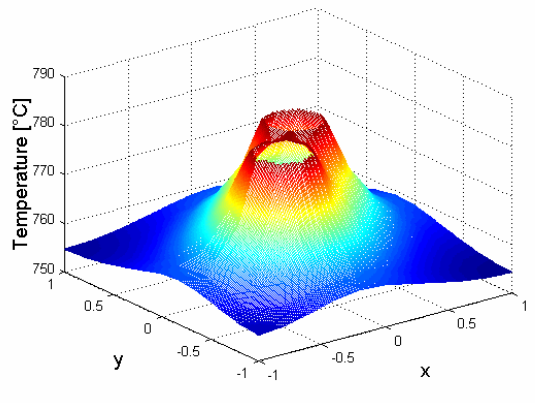
$$\frac{\partial T}{\partial t} = \alpha \left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2} \right) \qquad (11)$$

where the specific functions $\{\Phi_j(x) \mid j = 1,\cdots, N_P\}$ and $\{T_j(t) \mid j = 1,\cdots, N_P\}$ are piecewise continuously differentiable (Golebiowski and Kwieckowsk, 2002; Alazmi and Vafai, 2002). The initial condition of each finite element can be obtained from the solution of *Eq*. (8) repeatedly. The whole cable temperature is always assumed to be constant at the temperature of T_h. The temperature $T(x, t)$ as a function of time t and space x can be expressed by the multiplication of the temperature function $T_j(t)$ and the element function $\phi_j(x)$ at the state of orthogonal collation as follows:

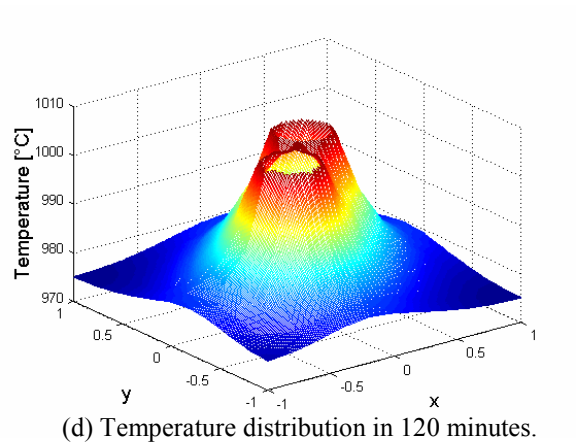


(a) Temperature distribution in 12 minutes.

(b) Temperature distribution in 24 minutes.



(c) Temperature distribution in 36 minutes.



(d) Temperature distribution in 120 minutes.

Fig. IV: Dynamic changes of the temperature distribution on the outer surface of the fire stop system.

$$T(x,t) = \sum_{i=1}^{N_P} T_i(t)\,\phi_i(x) \qquad (12)$$

The partial differential equation *Eq*. (11) subjected to the initial condition and the boundary conditions can be expressed as following ordinary differential equations.

$$M\frac{dT}{dt} + KT = L, \qquad (13)$$

where $M = \sum_i \left\{ \int_\Omega (\phi_j \cdot \phi_i) dx \right\}$,

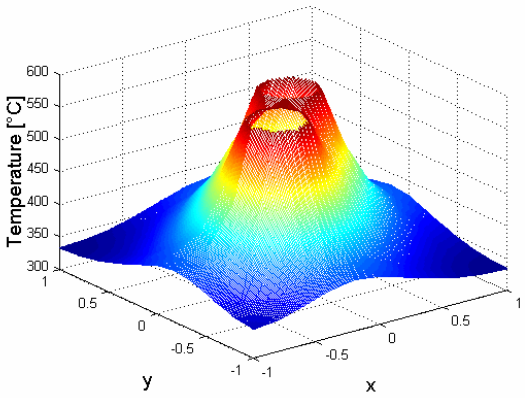
$K = \sum_i \left\{ \int_\Omega (\nabla \phi_j \cdot \alpha \nabla \phi_i) dx + \int_{\partial\Omega} (\phi_j \cdot \phi_i) ds \right\}$,

$L = \sum_i \left\{ \int_{\partial\Omega} (l \cdot \phi_i) ds \right\}$.

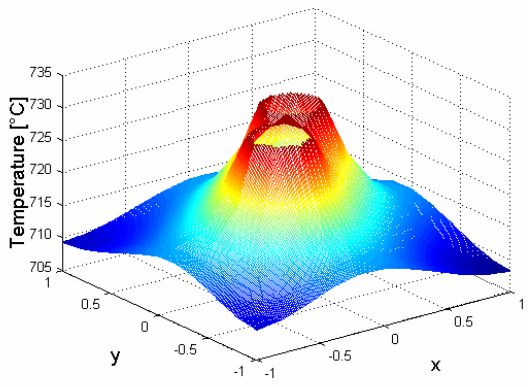
By solving these ordinary differential equations, we can obtain the nodal values as an approximate solution. The computations are performed on the computer Pentium IV-2.0 GHz. The program packet MATLAB is used for realizing the recommended algorithm. The initial temperature of T_0 is fixed at the temperature of 20 ℃, and the fire side wall temperature of T_h follows the ASTM E-119 standard temperature-time curve, changing from the initial temperature of 20 ℃ to the final temperature of 1,200 ℃ for two hours. As a simple geometry, the fire stop system with two penetrated cables is simulated in this work. The temperature values of 2323 elements are estimated simultaneously. The result of the estimation is used as the initial temperature values for calculating the temperature values at the five discrete axial points on each node of the elements. The calculation is carried out by SOR. The result is used again for the initial temperature values of the FEM calculation. In this manner the temperature distribution in the fire stop system is computed repeatedly.

4. RESULTS AND DISCUSSION

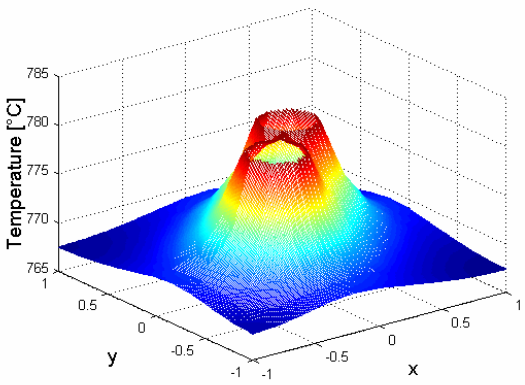
Simulation results are shown in Fig. IV and in Fig. V. The temperature distributions on the surface outer wall of the cable penetration fire stop system, which was built between compartments of the nuclear power plant, were calculated and the simulation results were three-dimensionally presented so as to show the dynamical heat conduction in the fire stop system. As shown in Fig. IV (a), the temperature around the cables reaches about 600 °C in 12 minutes. The temperature values



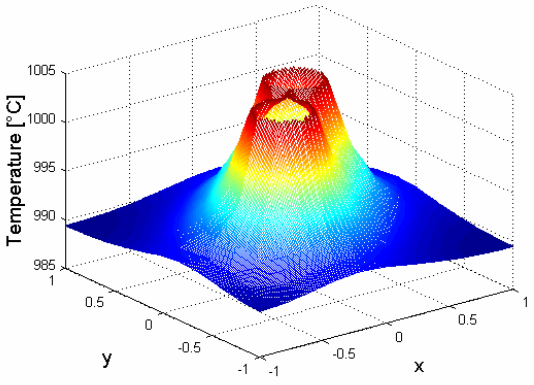
(a) Temperature distribution in 12 minutes.



(b) Temperature distribution in 24 minutes.



(c) Temperature distribution in 36 minutes.



(d) Temperature distribution in 120 minutes.

Fig. V. Dynamical changes of the temperature distribution on the inner face placed in the fire stop system.

near the firewall are under 100 °C. The temperature of the system rises very quickly in keeping with the standard temperature-time curve, as shown in Fig. I. According to the temperature distribution shown in Fig. V (b), the fire heat is transferred along to the cables quickly at the start, while the heat conduction occurs on the layers slowly. On the other hand, it is over the temperature of 700 °C around the cables in 24 minutes, and the temperature near the firewall is over the temperature of 680 °C. That means, the heat transfer on the layer is already progressed very much as well as the heat transfer along the penetrated cables. While the temperature rises in 36 minutes, the temperature difference becomes smaller with time. Furthermore, the temperature distribution in two hours is ranged from the temperature of 675 °C to the temperature of 1,000 °C, as shown in Fig. IV (d).

Similarly, the temperature distribution of the inner site located at a quarter distance of the wall thickness is shown in Fig. V. It consists of four figures. These figures could be compared with each other. By comparing the figures we could get more information for understanding the temperature distribution in the cable penetration fire stop system. The temperature distribution shown in Fig. V (a) is ranged from the temperature of 325 °C to the temperature of 600 °C in 12 minutes. In this case the temperature difference is much less than the case shown in Fig. IV (a). That means that the heat conduction is not working in a steady state, but is working dynamically. Fig. V (b) shows the results as follows. In 24 minutes the temperature variation around the penetrated cables is almost the same to the standard curve, and the difference between the maximum and minimum temperatures is about 20 °C. As shown in Fig. V (c), the temperature difference is decreased with time. The axial temperature difference near the firewall is more than 5 °C, as indicated in Fig. IV (d) and in Fig. V (d). Consequently, we can find the fact that the heat transfer through the cable stream is very significant and the heat conduction is not in a steady state.

5. CONCLSION

This work was aimed to know how the dynamics of heat conduction came about in the cable penetration fire stop system between compartments of nuclear power plants. Furthermore, the interest has focused on the thermal development around the penetration cables. The cable penetration fire stop system was modelled, simulated and analysed. The simulation results were illustrated in three-dimensional graphics. Through the simulations it was shown clearly that the temperature distribution was influenced very much by the number, the position and the temperature of the penetrated cables. Another significant contribution of this work is the development of an efficient numerical algorithm that consists of SOR and PEM for solving special partial differential equations. This numerical algorithm could be applied to the dynamic heat conduction problem successfully. At last, it was found that the dynamic heat transfer through the cable stream was one of the most dominant factors, and the feature of heat conduction could be understood as an unsteady-state and dynamic process.

ACKNOWLEDGEMENT

We acknowledge the financial aid for this research provided by the Brain Korea 21 Project supported by the Ministry of Education. In addition we would like to thank the Institute of Chemical Processes of Seoul National University.

REFERENCES

ASTM (1993), *ASTM Standards in Building Codes: Specifications, Test Methods, Practices, Classifications, Terminology*, American Society for Testing and Materials, 30th ed., Philadelphia, Pa

Southwell, R.V. (1940), *Relaxation Methods in Engineering Science*, Oxford Press, London

Monte, F. (2002), An Analytic Approach to Unsteady Heat Conduction Processes in One-dimensional Composite Media, *Int. J. Heat Mass Transfer*, 45, 1333-1343

Becker, E.B., Graham, F.C. and Oden, J.T. (1981), *Finite Elements*, Prentice-Hall

Feirweather, G. (1978), *Finite Element Galerkin Methods for Differential Equation*, Marcel Dekker, New York

Golebiowski, J. and Kwieckowsk, S. (2002), Dynamics of Three-dimensional Temperature Field in Electrical System of Floor Heating, *Int. J. Heat Mass Transfer*, **45**, 2611-2622

Alazmi, B.K. and Vafai, S. (2002), Constant Wall Heat Flux Boundary Conditions in Porous Media under Local Thermal Non-equilibrium Conditions, *Int. J. Heat Mass Transfer*, **45**, 3071-3087

MODELING OF METABOLIC SYSTEMS USING GLOBAL OPTIMIZATION METHODS

[†]Eberhard O. Voit and [‡]Edward P. Gatzke

[†]*Department of Biometry and Eppidemiology*
Medical University of South Carolina
Charelston, SC 29425

[‡]*Department of Chemical Engineering*
University of South Carlona
Columbia, SC 29208
Corresponding author: gatzke@sc.edu

Abstract: This paper considers the ***metabolic engineering*** problem of dynamic modeling in complex biological systems. New areas under consideration include ***dynamic system modeling*** of metabolic systems using a Generalized Mass Action (GMA) representation. The modeling problem will be presented as a nonconvex global optimization problem to be solved using deterministic optimization techniques. Advanced control and estimation methods can be devised based on the input-output model of the nonlinear dynamic system. A five-state fermentation pathway is considered using global optimization techniques for modeling and a discrete-time GMA formulation.

Keywords: Biomedical systems, nonlinear dynamic modeling, global optimization, sampled-data systems

1. INTRODUCTION

The hallmark of biological systems is their organizational complexity, which is manifested in large numbers of components and multitudes of intricate nonlinear interactions. For instance, in a biochemical system, various metabolites are consumed or created through enzyme-catalyzed reactions. These reactions are often modulated by regulatory components that are produced and consumed by these same reactions in the same pathway or may be constituents of entirely different pathways. When such modulations are present, intuitive analyses by cause-and-effect reasoning are no longer sufficient for system analysis, and systematic mathematical approaches are needed to gain useful insight. These numerical approaches are commonly based on systems of ordinary differential equations.

In metabolic engineering, the analysis of biochemical systems is often directed toward manipulation and optimization. For instance, one goal may be the improvement of metabolic yield in a microorganism. Two structured approaches currently dominate the field. One is a linear analysis of the flux distribution within the system, where the key concept is the well-known stoichiometric matrix. The other approach is a convenient nonlinear representation of the individual reactions. Over the past three decades, several groups around the world have developed and furthered a

mathematical framework specifically dealing with this latter approach.

The basis of this framework, known in the field as Biochemical Systems Theory (BST) is the representation of reaction rates with products of power-law functions that include those and only those metabolites and modulators that directly affect a given rate. See Savageau (1969) and Voit (2000). As an example, if enzyme E catalyzes a bimolecular reaction between A and B, and if this reaction is inhibited by end product P, the power-law term for the rate v in BST may be written as:

$$v = \alpha A^{\gamma_A} B^{\gamma_B} E^{\gamma_E} P^{\gamma_P} \tag{1}$$

where α is the rate constant of the reaction, the concentrations of the biochemical species are A, B, E, and P, and γ_A, γ_B, γ_E, γ_P are apparent kinetic orders.

Under some assumptions which have been discussed extensively in the literature, the nonlinear BST models (in the so-called S-system form) can be effectively optimized. See Voit (1992). However, in the alternative representation of a Generalized Mass Action (GMA) system, which is more intuitive to most biochemists, such optimization falls into the realm of NonLinear Programming (NLP) problems, which are notoriously difficult to handle. Preliminary work by Torres and Voit (2002) indicates that the special power-law structure of GMA systems might be amenable to streamlined, efficient methods of optimization. The development and refinement of such methods is a long-term goal. Achieving this goal would have great reward because GMA systems are the simplest systems that contain both the stoichiometric approach and the S-system approach as immediate special cases. Furthermore, GMA systems contain mixtures of linear and S-systems and have been shown by Savageau and Voit (1987) to provide mathematically equivalent representations for essentially all smooth, nonlinear phenomena. If all GMA rates that determine the dynamics of variable X_i $(i = 1..n)$ are symbolically coded as $\phi_i(X_1, X_2, ...X_n, ...X_m)$, the dynamic response of a GMA system can be modeled as follows:

$$\begin{aligned} \frac{dX_1}{dt} &= \phi_1(X_1, X_2, ...X_n, ...X_m) \\ &\vdots \\ \frac{dX_n}{dt} &= \phi_n(X_1, X_2, ...X_n, ...X_m) \end{aligned} \tag{2}$$

Note that variables X_1 through X_n are time dependent, while variables X_{n+1} to X_m may be independent of time for a given experiment.

One method to modify the rate of change of dependent variables is to over-express a gene. This changes the activity of an enzyme, which is usually modeled as an independent variable in a GMA model. Additionally, other independent variables, such as the substrate concentration or some inhibitor or cofactor, could be manipulated to different degrees, thereby evoking a dynamic response in the system.

In the presented case study, which is adapted from the work of Galazzo and Bailey (1990) and Curto et al. (1995), the external glucose concentration will be manipulated for the system, forcing changes in the dependent variables as glucose is absorbed into the cell at different rates. The metabolic pathway under consideration for this work is shown in Figure 1. Solid arrows represent reactions and dotted arrows show modulations. State variables in the model are: X_1 = cytosolic glucose; X_2 = glucose-6-phosphate; X_3 = fructose-1,6-diphosphate; X_4 phosphoenol pyruvate; X_5 = ATP. Independent variables with constant values are: X_6 = effective hexose transport; X_7 = hexokinase/glucokinase; X_8 = phosphofructokinase; X_9 = glyceraldehyde dehydrogenase; X_{10} = pyruvate kinase; X_{11} = glycogen and trehalose production; X_{12} = glycerol production; X_{13} = ATPase; X_{14} = NADH/NAD+ ratio.

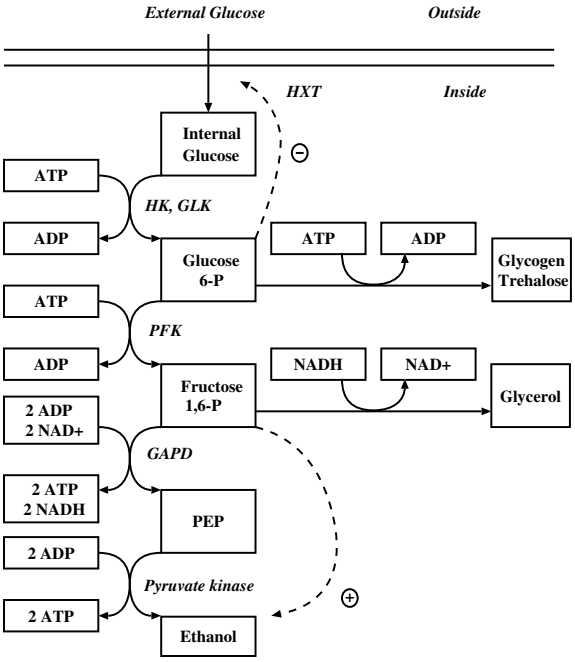


Fig. 1. Simplified model of anaerobic fermentation of glucose to ethanol, glycerol, and polysaccharides in Saccharomyces cerevisiae.

In general, a metabolic network may become quite complex for system involving many species. There may be many uses of such a model. Typically, a fully parameterized model is used for simulation, prediction, or optimizations. However, one might also have available measured concentrations at different points in time and attempt to deduce the structure of the pathway from these "metabolic profiles". See Voit and Almeida (2003). In general, this is a

formidable task, but in the case of a well-structured model such as a S-system or GMA model within BST, the task reduces to the simpler, yet challenging determination of parameter values that best describe the system and the measured profiles. The parameter estimation problem may be formulated as a nonconvex optimization problem to be solved using global optimization techniques. Due to the complexity of metabolic systems, a single set of model parameters not be readily apparent. Deterministic numerical methods may be used to approach this type of problem and determine the best model given the data.

2. MODELING FORMULATION

Given a system at steady state, a small perturbation in a metabolite concentration or in external conditions may cause significant transient response in metabolite concentrations, which provide insight into the structure of the metabolic network and the existence and magnitudes of the fluxes. In the GMA formulation, each flux representation requires determination of values for the rate constant α and the kinetic orders γ_i. The following global optimization scheme, based on system discretization and dynamic programming, can be used to determine the optimal values for these parameters. Here, $\hat{X}_i(k)$ is the metabolic concentration of species i in the model at time k. P is the number of measurement time points.

$$\begin{aligned}
\min_{\alpha,\gamma} \quad & \sum_{k=1}^{P} |e_i(k)| \ \forall i = 1..n \qquad (3)\\
s.t. \ \ & \hat{X}_1(k+1) = \hat{X}_1(k) + p_1(\hat{X}_1(k), \hat{X}_2(k), ...\hat{X}_m(k)) \\
& \forall k = 1..P \\
& \vdots \\
& \hat{X}_n(k+1) = \hat{X}_n(k) + p_n(\hat{X}_1(k), \hat{X}_2(k), ...\hat{X}_m(k)) \\
& \forall k = 1..P \\
& \hat{X}_i(k) - X_i(k) = e_i(k) \\
& \forall i = 1..n, k = 1..P \\
& \hat{X}_i(k) > 0 \\
& \forall i = 1..n, k = 1..P
\end{aligned}$$

This formulation minimizes the total sum of absolute errors for the nonlinear discrete-time system. Note that the concentrations are constrained to take only positive values. Also note that the nonlinear functions p_i in the formulation take the form of a sum of GMA reaction terms, as represented in the general GMA model reaction rate in Equation 1. The optimization problem can be seen as a nonconvex optimization problem in the general form:

$$\begin{aligned}
\min_{x} \quad & f(x) \qquad (4)\\
s.t. \quad & g_1(x) = 0 \\
& \vdots \\
& g_m(x) = 0 \\
& LB \le x \le UB
\end{aligned}$$

Here, the functions $f(x)$ and $g_i(x)$ may be nonconvex. In the original formulation of Equation 3 the objective function is a convex function but the model constraint equations are nonconvex nonlinear equality constraints. A nonconvex optimization problem in this form can be solved using standard branch-and-bound techniques. Deterministic branch-and-bound methods similar to those used in Mixed-Integer Linear Programming algorithms can be used to solve nonconvex NLP problems as in Soland (1971), Adjiman et al. (1998), and Tawarmalani and Sahinidis (2000). These methods rely on derivation of a convex relaxed lower bounding problem as described in McCormick (1976), Adjiman et al. (1996), Tawarmalani and Sahinidis (2000), and Gatzke et al. (2002). Recent range-reduction techniques have been shown to play a vital role in rendering more problems tractable as seen in Ryoo and Sahinidis (1995). As in all combinatorial optimization problems, reducing the problem dimension and solution space can lead to large improvements in solution efficiency. In this problem, the actual formulation may contain many variables (variables for α values , γ_i values, and model concentration values $\hat{X}_i(k)$). However, for estimation purposes, the actual solution space is significantly reduces, because branching only applies to α and γ_i.

Given a general nonconvex problem with continuous variables, $x \in \mathbb{R}^n$, any local solution using existing NLP methods will possibly provide an upper bound. The upper bounding problem can be expressed as described in Problem 4, where f and / or g may be nonconvex. After reformulation and introduction of new variables $w \in \mathbb{R}^o$, $z \in \mathbb{R}^{n+o}$. See McCormick (1976), Smith (1996), Tawarmalani and Sahinidis (2000), and Gatzke et al. (2002). An equivalent nonconvex problem can be expressed as:

$$\begin{aligned}
\min_{z} \ & c_1^T z \qquad (5)\\
& A_1 z \le b_1 \\
& h(z) = 0 \\
& z^L \le \ \ z \ \ \le z^U
\end{aligned}$$

Here, the nonconvex constraints $h(z)$ are simple nonlinear expressions involving two or three variables where one variable is explicitly defined using a single nonlinear operation, *e.g.* $z_1 = z_2 z_3$ or $z_4 = e^{z_5}$. This reformulation is required so that the simple nonconvex expressions can be relaxed using known convex envelopes, and outer approximation of the nonlinear

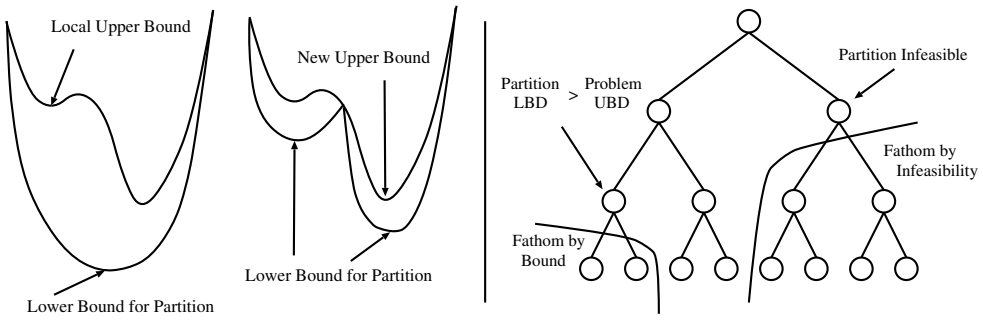


Fig. 2. Left: A single branch-and-bound step for a nonconvex function of a continuous variable. Right: Demonstration of implicit enumeration search for a branch-and-bound tree.

convex expressions described in Tawarmalani and Sahinidis (2000) and Gatzke et al. (2002) leads to convex lower bounding problem for a partition expressed as a Linear Programming (LP) problem:

$$\begin{aligned} \min_z \; & c_2^T z \\ & A_2 z \le b_2 \\ z^L \le \; & z \; \le z^U \end{aligned} \tag{6}$$

In this problem, c_2, A_2, b_2, z^L, and z^U depend on the current variable bounds for the variables in the original problem: x^L and x^U. This lower bounding LP problem can be solved using any valid LP technique.

The deterministic NLP solution proceeds according to the branch-and-bound algorithm illustrated in Figure 2. The original region is partitioned, and lower bounds are determined for each new partition. A partition is discarded if its lower bound exceeds the current upper bound for the problem, or if the lower bounding problem is infeasible. Once a feasible solution to Problem 4 is found, it serves as an upper bound for the global solution. The algorithm attempts to verify that the solution is the true global solution by systematically fathoming the remaining solution space. Range reduction methods can also be used to reduce portions of the solution space, possibly speeding convergence, Ryoo and Sahinidis (1995).

Range reduction techniques play a pivotal role in efficient solutions of nonconvex NLP problems. Reduction methods attempt to shrink the variable space without removing a region that may possibly contain the global solution of the problem. Interval analysis methods of Moore (1979) can be used to analyze the constraints in Problem 5 in order to modify the upper and lower bounds on z, reducing the possible solution space. Solution of Problem 6 provides a lower bound on the solution for a given partition. The Lagrange multipliers at the solution of the convex lower bounding problem can also be used to reduce the solution space, Ryoo and Sahinidis (1995). These bounds-tightening procedures may be repeated for a single partition, producing new variable bounds and a new lower bound for the partition while avoiding branching a partition. This may in some cases avoid the combinatorial growth of active subproblems.

In the formulation described by Problem 3, the only nonconvexity arises from the power-law rate terms of each GMA reaction. Each of these terms can be reformulated by introducing a logarithmic transformation as follows:

$$\begin{aligned} v &= \alpha A^{\gamma_A} B^{\gamma_B} E^{\gamma_E} P^{\gamma_P} \\ \frac{1}{\alpha}\ln(v) &= \gamma_A \ln(A) + \gamma_B \ln(B) + \gamma_E \ln(E) + \gamma_P \ln(P) \end{aligned} \tag{7}$$

Each logarithmic term is then replaced by a new variable, w_i as described by Torres and Voit (2002). After introducing these new variables, almost all constraints in the original formulation are linear. The only nonconvex relationships are the simple constraints of the form $w_i = \gamma_i \ln(X_i)$. As illustrated in Figure 3, X_i^L and X_i^U are known for a given partition and a secant can be used as a convex lower bound for this nonlinear function, while multiple linear first-order approximations may serve as linear upper bounding constrains for the nonlinear function.

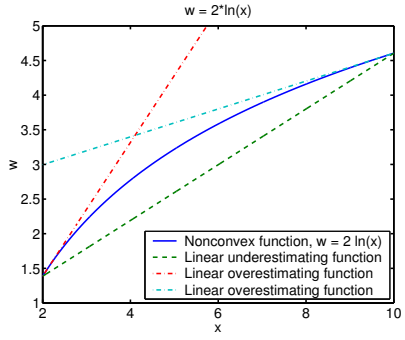


Fig. 3. Convex relaxation using linear constratins.

3. EXAMPLE SYSTEM

For this system, we consider the fermentation pathway in *Saccharomyces cerevisiae* described in Curto et al. (1995). This is a relatively simple metabolic pathway system with five time-dependent states and, thus, five differential equations. The metabolic pathway map is given in Figure 1. Each reaction is modeled separately in the GMA formulation. For illustration, this GMA model is used as the allegedly "true" model for the generation of data and testing the optimization algorithm. The model equations are given in Figure 4. While this is a nonlinear continuous time dynamic model, it can also be represented with a discrete-time nonlinear model. For this illustration, we use a discrete time sampling rate of 0.001. Slower sampling rates for the discrete time model may result in unstable or inaccurate dynamic systems. This rapid sampling time is an obvious limitation, which is to be considered in future work.

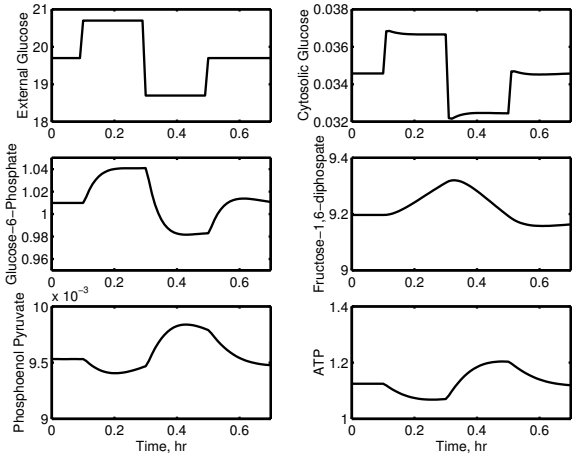


Fig. 5. Dynamic response for metabolite concentrations given step changes in external glucose concentration.

A typical response for the continuous time model, given the initial conditions and changes in the external glucose is shown in Figure 5. Data from a five-hour simulation were used to develop initial parameter estimates for the system. The model system was parameterized, resulting in 22 total parameters to be considered. Using a multistart unconstrained nonlinear optimization algorithm, these parameters were found using Matlab / Simulink, MathWorks (2000), to evaluate the objective function for the system for a given set of parameter values. The error terms for each species were scaled by the expected maximum deviation from the normal steady state operation. The scaling values used were:

$$\begin{bmatrix} 0.0025 & 0.05 & 0.3 & 0.0005 & 0.1 \end{bmatrix}$$

The resulting parameter values serve as an upper bound on the global solution. A comparison of the dynamic response of the continuous time process and the resulting discrete time model is given in Figure 6. The objective function (sum absolute error) at the resulting solution was 9.8479. It is the goal of the global optimization procedure to guarantee that the upper bound value is the global solution.

A branch-and-reduce algorithm was developed to determine optimal parameter values for this system. The lower bounding problems are posed as linear programming relaxations of the convex lower bounding problem for each partition. Each LP problem is solved using OSL from IBM, I. B. M. (1997). The branch-and-reduce procedure includes range reduction techniques that can be used to reduce the total number of nodes visited. The problem formulation was developed using a general purpose Maple script, Maplesoft (2000), that automatically generates code that can be automatically translated using DAEPACK. See Tolsma and Barton (2000) and Gatzke et al. (2002). DAEPACK generates convexification subroutines and gradient information for a given problem.

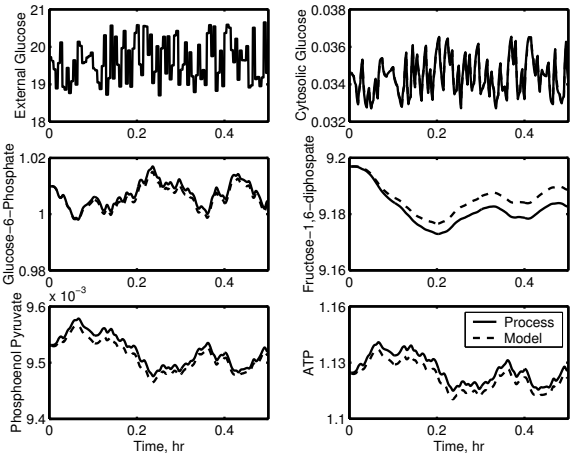


Fig. 6. Comparison of process and dynamic model response for changes in external glucose concentration.

For a given modeling problem with horizon length P with n dependent variables, $n(P-1)$ equations can be written. These equations serve as the nonlinear equality constraints for the prediction of the model concentrations over the data set of interest. The model equations include variables expressing explicitly the concentration of the n species at the $P-1$ points in the horizon of interest. The model equations also include variables representing the GMA reaction rate parameters for each mode flux (α and γ values). The number of new variables and constraints in the

$$
\begin{aligned}
\dot{X}_1 &= 0.8122X_2^{-0.2344}X_6 - 2.8632X_1^{0.7464}X_5^{0.0243}X_7 \\
\dot{X}_2 &= 2.8632X_1^{0.7464}X_5^{0.0243}X_7 - 0.5232X_2^{0.7318}X_5^{-0.3941}X_8 - 0.0009X_2^{8.6107}X_{11} \\
\dot{X}_3 &= 0.5232X_2^{0.7318}X_5^{-0.3941}X_8 - 0.011X_3^{0.6159}X_5^{0.1308}X_9X_{14}^{-0.6088} - 0.04725X_3^{0.05}X_4^{0.533}X_5^{-0.0822}X_{12} \\
\dot{X}_4 &= 0.022X_3^{0.6159}X_5^{0.1308}X_9X_{14}^{-0.6088} - 0.0945X_3^{0.05}X_4^{0.533}X_5^{-0.0822}X_{10} \\
\dot{X}_5 &= 0.022X_3^{0.6159}X_5^{0.1308}X_9X_{14}^{-0.6088} + 0.0945X_3^{0.05}X_4^{0.533}X_5^{-0.0822}X_{10} - 2.8632X_1^{0.7464}X_5^{0.0243}X_7 \\
&\quad -0.0009X_2^{8.6107}X_{11} - 0.5232X_2^{0.7318}X_5^{-0.3941}X_8 - X_5X_{13}
\end{aligned}
$$

Fig. 4. GMA model equations for continuous time system

resulting convex reformulation will depend on the complexity of each term in the GMA formulation and the total number of terms. Linear equations are written to explicitly express the error for each species at each point in time.

For a horizon of 6 samples, the resulting optimization formulation required 68 variables and 41 constraints. The lowerbounding problem resulted in 292 total variables and 688 total constraints. Given initial bounds on the parameter values, the branch-and-reduce algorithm was able to reduce the initial partition size, tightening the lower bound on the partition. The algorithm only considers parameter value (α and γ) variables for branching. These values were constrained to $\pm 5\%$ of their original values. The algorithm was able to guarantee within $\varepsilon = 0.1$ that the initially found upper bound was the global solution.

4. CONCLUSIONS

A nonconvex optimization formulation has been presented for determination of metabolic pathway parameters using a GMA representation of the metabolic system. This formulation can then be solved using branch-and-bound methods to global optimality. The branch-and-bound method proceeds in a deterministic manner, providing lower and upper bounds on the quality of the solution as the solution proceeds. The proposed problem can be reduced significantly by only considering a subset of variables for branching. This reduction in problem dimensionality can significantly improve the convergence aspects of the algorithm.

Acknowledgements

The Authors would like to acknowledge financial support from the South Carolina Collaborative Grants Program.

References

C. S. Adjiman, I. P. Androulakis, C. D. Maranas, and C. A. Floudas. A Global Optimization Method, α BB, for Process Design. *Comput. Chem. Eng.*, 20:S419–S424, 1996.

C. S. Adjiman, S. Dalliwig, C. A. Floudas, and A. Neumaier. A Global Optimization Method, αBB, for General Twice-Differentiable Constrained NLPs - I Theoretical Advances. *Comput. Chem. Eng.*, 22(9):1137–1158, 1998.

R. Curto, A. Sorribas, and M. Cascante. Comparative characterization of the fermentation pathway of Saccharomyces cerevisiae using biochemical systems theory and metabolic control analysis. Model definition and nomenclature. *Math. Biosc.*, 130:25–50, 1995.

J. L. Galazzo and J. E. Bailey. Fermentation pathway kinetics and metabolic flux control in suspended and immobilized Saccharomyces cerevisiae. *Enzyme Microb. Technol*, 12:162–172, 1990.

E. P. Gatzke, J. E. Tolsma, and P. I. Barton. Construction of Convex Function Relaxations Using Automated Code Generation Techniques. *Optimization and Engineering*, 3(3): 305–326, 2002.

I. B. M. . IBM Optimization Solutions and Library Linear Programming Solutions. Technical report, I. B. M., 1997.

Maplesoft. *Maple Reference Guide*. Springer Verlag, 2000.

The MathWorks. *Matlab 6.1*. Prentice Hall, 2000.

G. P. McCormick. Computability of Global Solutions to Factorable Nonconvex Programs: Part I - Convex Underestimating Problems. *Mathematical Programming*, 10:147–175, 1976.

R. E. Moore. *Methods and Applications of Interval Analysis*. SIAM, Philadelphia, 1979.

H. S. Ryoo and N. V. Sahinidis. Global Optimization of Nonconvex NLPS and MINLPs with Application to Process Design. *Comput. Chem. Eng.*, 19(5):551–566, 1995.

M. A. Savageau. Biochemical Systems Analysis, I. Some Mathematical Properties of the Rate Law for the Component Enzymatic Reactions. *J. Theor. Biol.*, 25(365-369), 1969.

M. A. Savageau and E. O. Voit. Recasting Nonlinear Differential Equations as S-Systems: A Canonical nonlinear Form. *Math. Biosci.*, 87:83–115, 1987.

E. M. B. Smith. *On the Optimal Design of Continuous Processes*. PhD thesis, Imperial College, London, 1996.

R. M. Soland. An Algorithm for Separable Nonconvex Programming Problems II. *Management Science*, 17:759–773, 1971.

M. Tawarmalani and N. V. Sahinidis. Global Optimization of Mixed Integer Nonlinear Programs: A Theoretical and Computational Study. Technical report, University of Illinois, 2000.

J. Tolsma and P. I. Barton. DAEPACK: An Open Modeling Environment for Legacy Models. *Ind. Eng. Chem. Res.*, 39(6): 1826–1839, 2000.

N. V. Torres and E. O. Voit. *Pathway Analysis and Optimization in Metabolic Engineering*. Cambridge University Press, Cambridge, UK, 2002.

E. O. Voit. Optimization in integrated biochemical systems. *Biotechn. Bioengin.*, 40:572–582, 1992.

E. O. Voit. *Computational Analysis of Biochemical Systems*. Cambridge University Press, New York, 2000.

E. O. Voit and J. Almeida. Dynamic Profiling and Cononical Modeling: Powerful Partners in Metabolic Pathway Identification. In R. Goodacre and G. G. Harrigan, editors, *Metabolite Profiling: Its Role in Biomarker Discovery and Gene Function Analysis*. Kluwer Academic Publishing, Dordrech, The Netherlands, 2003.

A STATE-SHARED MODELING APPROACH TO TRANSITION CONTROL

Zhenhua Tian and Karlene A. Hoo [1]

Department of Chemical Engineering
Texas Tech University, Lubbock, TX

Abstract: A rigorous theoretical derivation of a state-shared model structure for multiple-input multiple-output (MIMO) systems is proposed. When a nonlinear system transitions in a large operating space, this state-shared modeling approach can be used to approximate the nonlinear system, such that effective model-based controllers can be applied. A MIMO nonlinear reactor system illustrates the proposed approach.

Keywords: Adaptive identifier, reduced-order model, multiple models, parameter adaptation, nonlinear reactor

1. INTRODUCTION

In the chemical industry, it has become quite common for plants to produce more than one grade of product, the choice being dictated by market forces. This is particularly true for polymer industries, where product and grade transitions occur frequently. This necessitates the use of a controller that can successfully regulate the plant not only at the operating points, but also during the transition.

Useful measures to evaluate a transition control strategy are no violation of constraints, speed of response, satisfactory performance, and closed-loop stability (Narendra *et al.*, 1995).

Nonlinear behavior is not an uncommon characteristic during the transition. This feature serves to not only heighten the control problem but also to require a nonlinear dynamic model of the process for control studies. However, due to a lack of available high fidelity nonlinear models, because nonlinear control methods are not well-understood, and closed-loop stability arguments are difficult to prove, nonlinear strategies are very difficult to implement (Tian and Hoo, 2002*b*). Additionally, the maintenance costs of nonlinear control algorithms are usually substantially higher than those of linear control algorithms for the same process (Eker and Nikolaou, 2002).

A popular option is to use multiple linear models that together represent the nonlinear system. Thus, many transition control strategies are based on linear models with fixed parameters so that linear control theory can be applied. For large operating spaces, the issue of how many fixed model/controller pairs are needed remains unanswered. A variety of the multiple model strategies employ a conditional probability to determine how to combine a set of fixed models to generate a new model that better represents the plant outputs (Aufderheide and Bequette, 2001). Others adapt the existing models and controller parameters using a model reference adaptive control (MRAC) framework (Narendra *et al.*, 1995; Gundala *et al.*, 2000).

It is well known that different control strategies can be attempted once the estimation model is chosen. Whether a more robust controller can be used to assure stability and improved perfor-

[1] author to whom all inquires should be addressed, khoo@coe.ttu.edu

mance, is an intriguing question. Sun and Hoo (1999a, 1999b) presented a robust dynamic transition control structure for time-delayed systems using a set of fixed models and controllers. Tian and Hoo (2000b) used H_∞ controllers based on fixed and adaptive models.

In this work, a state-shared model framework represents multiple fixed and adaptive models. The state-shared model consists of a non-minimal realization and a non-minimal identifier. The former is developed at the known operating points, while the latter is used at any other point. The parameters in the measurement equation are adapted. Thus, all models, adaptive and fixed, can be cast into such a structure, in which all the models share the same states but the parameters in the measurement equation represent different operating points.

The rationale for using adaptive and fixed models is to ensure that there is at least one model with parameters sufficiently close to those of the unknown plant to provide accurate controller response. The fixed models together with stable switching provide speed, while the intelligent adaptive models and tuning provide accuracy.

From a system identification point of view, the state-shared model has attractive properties such as the uniqueness of the identified parameterization and convergence of the adaptable parameters.

The coefficient matrices in the state-shared model are selected to be controllable by the designer. The existence and uniqueness criteria for this type of parameterized model are based on a uniquely identifiable parameterization known as a matrix fraction description (MFD) (Kailath, 1980). The convergence of the parameter adaptation is also proven. For an m-input m-output linear MIMO system, the total number of identified parameters is derived to be $N_\theta = n(m^2 + 1)$, where n is the upper bound of the McMillan degree of all nominal linear models.

The organization of the paper is as follows. Section two begins with a brief review of some relevant results and properties of linear MIMO systems. Section three develops the state-shared model and presents the existence and uniqueness criteria. The convergence of the identifier is also analyzed. Section four demonstrates the construction of the state-shared model on a MIMO nonlinear system studied by (Gundala *et al.*, 2000) that undergoes a production rate transition. Lastly, section five summarizes the findings.

2. PRELIMINARIES

An m-input p-output linear time-invariant (LTI) system has a transfer function $H(s) \in R^{p\times m}(s)$, the set of $(p \times m)$ matrices with polynomial elements (Antsaklis and Michel, 1997; Kailath, 1980).

Definition 1. A rational transfer function matrix $H(s)$ is said to be *proper* if

$$\lim_{s\to\infty} H(s) < \infty$$

and *strictly proper* if

$$\lim_{s\to\infty} H(s) = 0$$

Theorem 1. (Antsaklis and Michel, 1997) H(s) is realizable as the transfer function matrix of a linear time-invariant system given by

$$\dot{x} = Ax + Bu$$
$$y = Cx + Vu$$

if and only if $H(s)$ is a proper rational matrix. If V is the null matrix, then $H(s)$ is a strictly proper rational matrix.

Rewrite $H(s)$ as

$$H(s) = \frac{N(s)}{d(s)}$$
$$d(s) = s^n + d_1 s^{n-1} + \cdots + d_n$$

where $d(s)$ is the least common multiple of the denominators of $H(s)$ and deg $d(s) = n$. Thus,

$$H(s) = D^{-1}(s)N(s), \quad D(s) = d(s)I_p \quad (1)$$

and define the *degree* of the denominator matrix as

$$\deg D(s) \equiv \deg \det(D(s)) = np$$

The pair $\{N(s),\ D(s)\}$ is called a *left* matrix fraction description (*left* MFD) (Kailath, 1980).

Proposition 1. Given any left MFD of $H(s) = D^{-1}(s)N(s)$, a state-space realization of order

$$\deg \det(D(s)) \equiv \text{deg left MFD}$$

can always be found.

Kailath (1980) provides a procedure to obtain the realization from the left MFD.

Lemma 1. If H(s) is a strictly proper (proper) transfer function and the left MFD is given by Eq (1), then every row of N(s) has degree strictly $<$ ($\leq$) that of the corresponding row of D(s).

The proof is similar to that of Kailath (1980).

3. MIMO STATE-SHARED MODEL

The aim is to achieve a non-minimal realization/non-minimal identifier, such that while all the models share the same state space representation, each model has unique input/output parameter set.

3.1 A non-minimal realization

Theorem 2. Any controllable and observable m-input m-output LTI system given by

$$Y_p(s) = H(s)U(s) = D^{-1}(s)N(s)U(s) \quad (2)$$

with $D(s) = d(s)I_m$ and $d(s)$, a polynomial of degree n, is input-output equivalent to the LTI system described by the differential equations

$$\begin{aligned} \dot{\omega}_1 &= F\omega_1 + GU \qquad \omega_1 \in R^{nm\times 1} \\ \dot{\omega}_2 &= F\omega_2 + GY_p \qquad \omega_2 \in R^{nm\times 1} \\ Y_p &= \Theta^T\omega \end{aligned} \quad (3)$$

by suitable choice of the parameter vector $\Theta \in R^{2nm\times m}$, and Θ is uniquely determined by the given $H(s)$. The pair (F, G) should be controllable with $F \in R^{nm\times nm}$, an asymptotically stable matrix, and $G \in R^{nm\times m}$. The (F, G) can be arbitrarily chosen under the given restrictions.

Proof
Consider Figure 1. The transfer functions from $U(s)$ to v_1 and Y_p to v_2 are given by

$$\begin{aligned} \Theta_1^T(sI-F)^{-1}G \qquad \Theta_1 \in R^{nm\times m} \\ \Theta_2^T(sI-F)^{-1}G \qquad \Theta_2 \in R^{nm\times m} \end{aligned}$$

respectively. Define, $\Phi_I \equiv (sI-F)^{-1}G$. Then, the transfer function from the input U to the output Y_p can be expressed as

$$\begin{aligned} H(s) &= \left(I - \Theta_2^T\Phi_I\right)^{-1}\left(\Theta_1^T\Phi_I\right) \equiv D^{-1}(s)N(s) \\ &= (D(s)/\Omega_I(s))^{-1}(N(s)/\Omega_I(s)) \end{aligned} \quad (4)$$

where, $\Omega_I(s) \equiv \det(sI-F) = s^n + a_n s^{n-1} + \cdots + a_2 s + a_1$. From Eq (4), the following equivalent relations can be obtained

$$I - \Theta_2^T\Phi_I = \frac{D(s)}{\Omega_I(s)} \qquad \Theta_1^T\Phi_I = \frac{N(s)}{\Omega_I(s)} \quad (5)$$

Eq (5) includes $n(m^2+1)$ linear equations with the same number of unknown variables. It is not difficult to show that the coefficient matrix of this system of linear equations is nonsingular, which means that Θ_1 and Θ_2 can be determined uniquely. It then follows that any linear time-invariant plant can be parameterized as shown in Eq (3). *QED*

Remark 1: Theorem 2 implies the existence of parameter vector Θ such that the transfer function of the state-space model, given in Eq (3), is equivalent to $H(s)$, given in Eq (2).

Remark 2: The value of Θ depends on the pair (F, G) and the coefficients of $H(s)$. Since (F, G) are under the influence of the designer, Θ is uniquely determined.

Remark 3: $D(s)$ is block diagonal with the same elements in each block. As a result, there are only n elements to be identified to determine Θ_2; the others are zeros. There are nm^2 elements in Θ_1, the total number of parameters to be identified is $n(m^2+1)$.

It is assumed that the space of operating conditions is large such that there are many possible operating states. At known operating states, a suitable linear model can be developed or identified and such a model will have fixed parameters in its input/output form.

3.2 A non-minimal identifier

At any unknown operating point, it is not difficult to construct a non-minimal identifier (not interpolation). Assume the non-minimal identifier has the same form as the non-minimal realization. Because the operating point is not known *a priori*, let the parameters in the measurement equation be adapted to obtain an accurate representation. The state-space equation of the non-minimal identifier will be driven by the measured signals, U and Y_p.

The non-minimal identifier is given by (3)

$$\begin{aligned} \dot{\omega}_1 &= F\omega_1 + GU \qquad \omega_1 \in R^{nm\times 1} \\ \dot{\omega}_2 &= F\omega_2 + GY_p \qquad \omega_2 \in R^{nm\times 1} \\ \hat{Y} &= \hat{\Theta}^T\omega \end{aligned} \quad (6)$$

3.2.1. Parameter adaptation The adaptation of the parameters must be done in a stable fashion. In this work, a normalized least-squares for an m-input m-output system is proposed. For details see (Tian and Hoo, 2002*a*).

Let the model-plant mismatch be given by $\tilde{Y} \equiv \hat{Y} - Y_p$. It can be shown that

$$\lim_{t\to\infty}\tilde{Y}(t) = 0, \qquad \lim_{t\to\infty}\hat{\Theta}(t) = \Theta$$

It is understood that the state-shared model is both the non-minimal realization of the plant and the adaptive non-minimal identifier. From the procedure described in the previous section, it can be concluded that the parameters of the state-shared model are uniquely determined by the transfer functions of the input/output models. Note that any such realization necessarily fulfills the requirement that the output of model j, Y_j be an asymptotically correct estimate of output of the plant Y_p if the process model transfer function were $H(s)$, i.e. $Y_j \to Y_p$.

3.3 Model reduction

For control implementable solutions, it is desirable to use low-order controllers whenever possible. One means of reducing the order of the

controller is to generate a reduced-order approximation of the plant before designing the controller (Mahadevan and Hoo, 2000; Zheng and Hoo, 2002).

Reduced-order approximation of plant dynamics is not an uncommon engineering practice. In general, mathematical models are reduced-order approximations of the true system generated by ignoring *minor* effects during modeling. Clearly, the number of parameters to be adapted affects the rate of convergence and the computational burden.

From the previous section, it is known that the order of state-shared model for a MIMO system is $2nm$. The input/output number, m, cannot be changed. To reduce the order of the state-shared model, the order of $d(s)$ (n) must be reduced. In this work, a balanced truncation approach is used (Burl, 1999).

4. EXAMPLE: NONLINEAR REACTOR

The chemical reactor consists of a continuous-stirred tank reactor in which a single, isothermal, irreversible reaction given as $A(g) + C(g) \rightarrow D(l)$, occurs in the vapor phase (Ricker, 1993). Components A and C are non-condensible gases and component D, the product, is a non-volatile liquid.

The molar balance of each component in the system is given by,

$$\begin{aligned}
\frac{dN_A}{dt} &= y_{A1}F_1 + F_2 - \frac{N_A}{N_3}F_3 - r_D \\
\frac{dN_B}{dt} &= y_{B1}F_1 - \frac{N_B}{N_3}F_3 \\
\frac{dN_C}{dt} &= y_{C1}F_1 - \frac{N_C}{N_3}F_3 - r_D \\
\frac{dV_L}{dt} &= (r_D - F_4)/\rho_L
\end{aligned} \tag{7}$$

$$\begin{aligned}
N_k &= \sum_{i=A}^{C} N_{ik} \qquad k = 1, \ldots, 4 \\
p_i &= \frac{N_i}{\sum_{j=A}^{C} N_{jk}} P \qquad i = A, B, C \\
r_D &= k_0 p_A^{v_1} p_C^{v_2} \\
F_4 &= (\chi_4 + u_4) c_{v4} \sqrt{P - P_r} \\
u_4 &= K_c (V_L^* - \frac{V_L}{V_{L,max}} 100\%)
\end{aligned}$$

where r_D is the reaction rate (kmol/h) that depends on the partial pressures of components A and C, N_i is the number of moles of component i, and y_{ij} is the mole fraction of component i in stream j. There are two feed streams (F_1, F_2) a purge (F_3), and a product stream (F_4) with units of kmol/h. The ideal gas law is assumed to be valid and the liquid density (ρ_L) is constant. Measured outputs include the reactor pressure (P), the liquid volume (V_L), and the mol% of unreacted A in F_3 (y_{A3}).

For economic, safety, and operational considerations, F_4, P, and y_{A3} should be controlled. The three manipulated variables are F_1, F_2 and F_3. The product rate is adjusted by a proportional feedback controller in response to variations in the liquid inventory. The control signal from a liquid inventory controller is u_4. There are safety and production constraints. The reactor pressure must be maintained below the shutdown limit of 3000 kPa, and F_1 and F_2 cannot be larger than their maximum values of 330.46 kmol/h and 22.46 kmol/h, respectively. More details can be found in (Tian and Hoo, 2002*a*).

It is desired to transition the reactor between two production rates, OPI: 100 and OPII: 130 kmol h^{-1}. An analysis of the model shows that both states are feasible, stable operating states.

4.1 Modeling

Linearization at the known operating states yields a linear model in the form,

$$\begin{aligned}
&\dot{x} = Ax + Bu \\
&y = Cx \\
&x = x_p - \bar{x} \qquad y = y_p - \bar{y}_p
\end{aligned} \tag{8}$$

where $\bar{x}$ and $\bar{y_p}$ are the steady state values of x_p and y_p, respectively; $A = \nabla_{x_p} f$, $B = \nabla_{u_p} f$, and C is found from the measurement equation. An eigenvalue analysis shows that the matrix A, at both OPI and OPII, is asymptotically stable. Additionally, the linear systems at both operating states are stable and output controllable implying that the nonlinear system at these operating states are at least locally stable. The system is also observable.

At each operating point, a 2-state, 3-input 3-output reduced-order model corresponding to the 4-state, 3-input 3-output full-order linear model is obtained by the model reduction method of balance truncation.

Assume that at any point in the operating space, the nonlinear system can be approximated by a reduced-order linear model of order 2 (the upper bound of the McMillan degree), but each model may be different at each operating state. The aim of the state-shared model structure is to represent all the linear models by one state-shared framework. Their measurement equations represent their differences. From the theory presented in §3, the state-shared model will have order $nm = 6$, with $n(m^2 + 1) = 20$ adaptable parameters. Let the pair (F, G) be given by,

$$F_j = \begin{bmatrix} 0 & 1 \\ -a_1 & -a_2 \end{bmatrix} \quad G_j = \begin{bmatrix} 0 \\ 1 \end{bmatrix} \quad j = 1, 2, 3$$

with $a_1 = a_2 = 1$ such that the pair (F, G) are controllable. It then follows that,

$$F = \begin{bmatrix} F_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & F_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & F_3 \end{bmatrix}_{6\times 6} \quad G = \begin{bmatrix} G_1 & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & G_2 & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & G_3 \end{bmatrix}_{6\times 3}$$

At any operating point, given $N(s)$ and $D(s)$, the measurement equation parameters, (Θ_1, Θ_2), can be calculated.

4.2 Model validation

In a neighborhood of OPI, random input disturbance signals (zero mean and standard deviations of 15% of their nominal values at OPI) is introduced to both the nonlinear system and linear models. To quantify the differences between the model responses, define the Average Relative Error (ARE) of the j^{th} measurement after k sample points by,

$$\text{ARE}(j) = k^{-1} \sum_{i=1}^{k} \left| \frac{y_p(i,j) - \bar{y}_p(j) - y(i,j)}{y_p(i,j)} \right| \quad (9)$$

Here, $y_p(i,j)$ is the j^{th} output of the nonlinear system, $\bar{y}_p(j)$ is the j^{th} nominal value of the nonlinear model, and $y(i,j)$ is the j^{th} model response.

Table 1 lists the AREs among the different models. The largest errors are associated with F_4 but they are $\leq 1\%$. Similar results were obtained at OPII.

4.3 Transition control

The system is forced to transition from 100 kmol/h to 130 kmol/h, while satisfying all other constraints. First order reference trajectories (dotted lines in the figures) are selected for the three outputs. A model predictive controller is used with the state-shared model and measurement equations to achieve the transition. A controller horizon of 4 and a prediction horizon of 10 are selected. To represent a preference among the controlled variables, output weights of 3, 1, and 10 for P, y_{A3}, and F_4, respectively are selected. Equal weighting of the rate of change in the inputs is used. The system has constraints on the outputs and inputs. No special attempts were used to determine optimal values for these parameters.

Figure 3 shows the closed-loop responses of F_4 and P. The production rate achieves its set point in about 10 hours. There is no violation of the pressure constraint. The production rate can be made to reach its the set point within 5 hours without violation of any constraint, but the controller action is more aggressive.

Figure 4 shows the the closed-loop responses of F_4 and P when the system has unmeasured disturbances as it transitions. Here, unmeasured output disturbances, with a signal noise ratio of 10:1, are introduced into the system. A first order filter is used to filter out any noise. There are no constraint violations, and although F_4 does not track closely the reference trajectory, the set point change is achieved within 8 hours.

In practice, the composition measurement can not always be obtained in a timely fashion. Assume y_{A3} must be inferred from the other measures signals. Closed-loop performance based on estimates of y_{A3} are shown in Figure 5. The transition is achieved within 8 hours. There is no violation of the pressure constraint (not shown).

5. SUMMARY

In this work, a method to construct a state-shared model for MIMO systems is developed and its properties analyzed. This approach can represent the plant in such a fashion that all the unknown parameters of the plant appear as the elements of a single matrix in the measurement equation of the state-shared model. A solution of the parameter vector is obtained when the transfer function is known. Existence and uniqueness for the parameterization are proven. The parameter vector can be adapted by least-squares, such that the adaptive model can be obtained with the same state-space representation.

A nonlinear chemical reactor system that transitions from one production rate to another was used to demonstrate the concept of the state-shared model in a model predictive control framework. Satisfactory closed-loop performance was achieved even in the face of unmeasured output disturbances and when composition was estimated. Future work will be to apply the state-shared model framework to a plant wide problem such as the Tennessee Eastman challenge problem.

Acknowledgments The authors would like to acknowledge the financial support of NSF grant CTS-9703252.

REFERENCES

Antsaklis, P. J. and A. N. Michel (1997). *Linear System*. McGraw-Hill. New York, NY.

Aufderheide, B. and B. W. Bequette (2001). A variable tuned multiple model predictive controller based on minimal process knowledge. Technical report. Proc. Amer. Contr. Conf.. Arlington, VA.

Burl, J. B. (1999). *Linear Optimal Control, H_2 and H_∞ Methods*. Addison Wesley. Menlo Park, CA.

Eker, S. A. and M. Nikolaou (2002). Linear control of nonlinear systems: Interplay between nonlinearity and feedback. *AIChE* **48**(9), 1957–1980.

Gundala, R., K. A. Hoo and M. J. Piovoso (2000). Multiple model adaptive control design for a multiple-input multiple-output chemical reactor. *Ind. & Eng. Chem. Res.* **39**, 1554–1564.

Kailath, T. (1980). *Linear Systems*. Prentice-Hall. Englewood Cliffs, NJ.

Mahadevan, N. and K. A. Hoo (2000). Wavelet-based model reduction of distributed parameter systems. *Chemical Engineering Science* **55**, 4271–4290.

Narendra, K. S., J. Balakrishnan and M. K. Ciliz (1995). Adaptation and learning using multiple models, switching, and tuning. *IEEE Contr. Syst. Mag.* pp. 37–51.

Ricker, N. L. (1993). Model predictive control of a continuous nonlinear, two-phase reactor. *J. Proc. Cont.* **3**(2), 109–132.

Tian, Z. and K. A. Hoo (2002*a*). A state-shared model approach for multiple-input multiple-output systems. *submitted to IEEE Control Systems Technology*.

Tian, Z. and K. A. Hoo (2002*b*). Transition control using multiple adaptive models and an H-infinity controller design. *In Proc. 2002 Amer. Control. Conf., Anchorage, AK* pp. 2621–2626.

Zheng, D. and K. A. Hoo (2002). Low-order model identification of distributed parameter systems by a combination of singular value decomposition and the karhunen-loève expansion. *Compt. & Chem. Engng.*

Table 1. AREs (%).

	P	y_{A3}	F_4
Nonlinear-Full linear	0.02	0.01	0.07
Nonlinear-Reduced linear	0.02	0.29	0.93
Nonlinear-State Shared	0.25	0.32	0.93
Full linear-Reduced linear	0.01	0.30	0.83
State Shared-Reduced linear	0.01	0.04	0.18

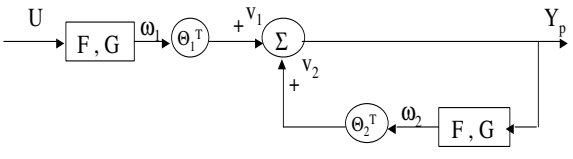


Fig. 1. Non-minimal realization.

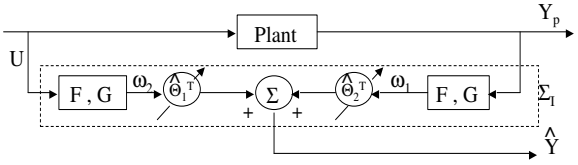


Fig. 2. Non-minimal identifier.

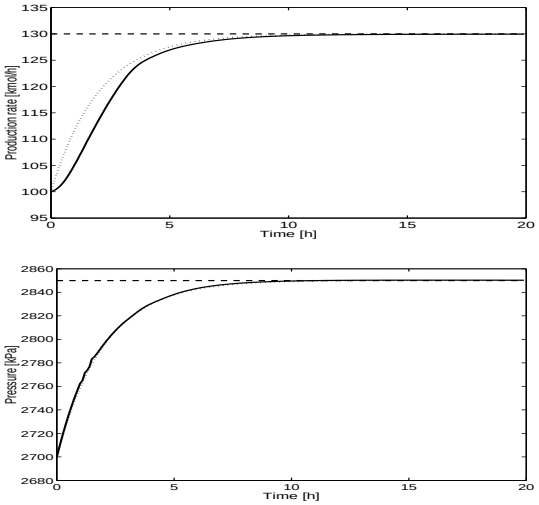


Fig. 3. Transition: ideal conditions.

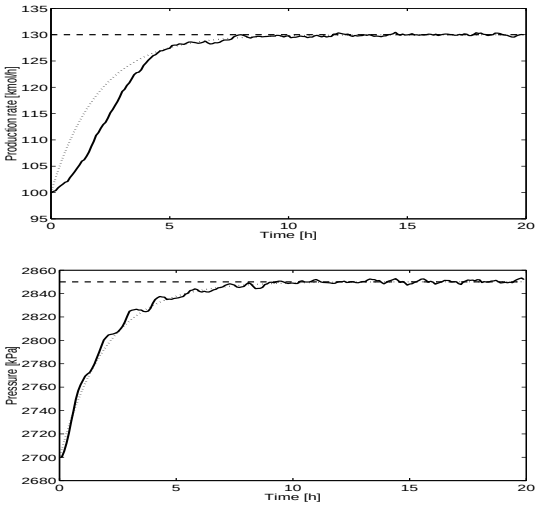


Fig. 4. Transition: unmeasured output disturbances.

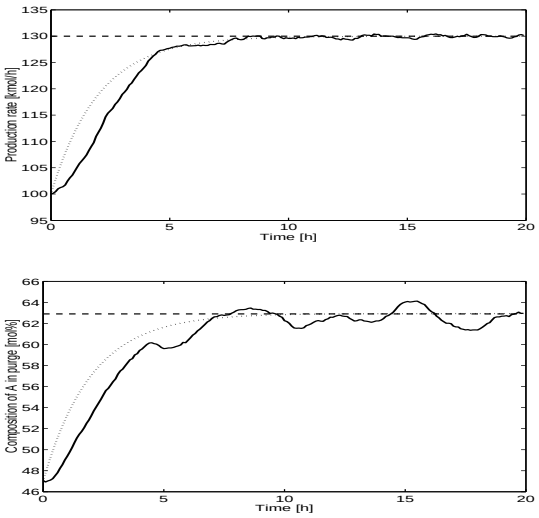


Fig. 5. Transition: estimate composition in the purge.