A MULTI-LEVEL, CONTROL-THEORETIC FRAMEWORK FOR INTEGRATION OF PLANNING, SCHEDULING AND RESCHEDULING

Munawar, S. A. and Gudi, R. D.

Department of Chemical Engineering, Indian Institute of Technology, Bombay, Mumbai – 400 076, India.

Abstract: In this paper, an integrated multi-level, control-theoretic framework has been proposed for effectively handling integration of planning, scheduling and rescheduling. The overall problem is decomposed into three levels with different horizons; each level has an abstraction of the lower level, with the philosophy of decentralized decision making and the flexibility and amenability to rescheduling. The main feature of the proposed framework is the integration of reactive scheduling that is motivated by some of the process control principles like cascade control and the concepts of receding horizon. An illustrative case study of a simple refinery flow sheet involving continuous lube production in a hybrid flowshop is presented.

Keywords: Integration, Planning, Scheduling, Rescheduling /Reactive scheduling, Multilevel, Control-theory.

1. INTRODUCTION

In the recent past, development of methods for efficient integration of planning and scheduling has received momentous attention in the industrial sector and in the research community, largely because of the challenges and the high economic incentives involved. Most large-scale multi-enterprise facilities which are distributed over distinct geographic locations, try to integrate the planning and scheduling activities as best as possible, from the stand point of a central planning at the corporate office for production and distribution, thus setting targets for single-site plants. Often, they employ some improvised techniques for this integration and are generally discontented with the resulting inconsistencies in decision-making.

Over the last few years, though some progress has been made in this direction for development of better frameworks for such integrations, still there is a large scope for further improvements. Traditionally, the decisions in an enterprise flow in a top-down manner leaving less degree of freedom at lower levels for rescheduling, leading to frequent revision of targets set by the top levels. At each level the horizons of interest are widely different. The upper levels must also reflect accurate abstractions of the lower levels and should be revised as infrequently as possible compared to the lower levels. Embedding contingency measures for integration of rescheduling has been ignored in most of the works.

In the literature there are several works on planning and scheduling. Shah (1998) gives a detailed review and current status on single and multi-site planning and scheduling. Grossmann et al. (2001) revised the classification of planning and scheduling models arising in process operations and the recent developments in their solution techniques. They proposed a general disjunctive model for integration of planning and scheduling. Shobrys and White (2000) examined the incentives and barriers in the integration of planning, scheduling and control functions in the process industries. They recounted some of the success stories in this direction and analyzed the reasons for failure of other companies that could not achieve integration despite multiple initiatives. They identified two non-techno challenges of coping with the human and organizational behavior and made some recommendations to overcome these barriers to integration. Recently, Van den Heever and Grossmann (2003) presented a two-level decomposition model for integration of planning and reactive scheduling in hydrogen supply networks. They used a simplified pipeline model at the planning level and a detailed pipeline description for the scheduling level.

The area of process control is well matured and recently there have been increasing applications of control-oriented frameworks for supply chain management and integration of planning and scheduling. Perea et al. (2000) proposed a dynamic approach to supply chain management with ideas from process dynamics and control. Vargas-Villamil and Rivera (2000) proposed a model predictive control (MPC) formulation for scheduling of reentrant semiconductor manufacturing lines. Bose and Pekny (2000) proposed MPC for integration of planning and scheduling for a multi-period operation of consumer goods supply chain. In each period, they used a forecasted model to calculate target inventories (control variable) for future periods and a scheduling model to achieve these targets by scheduling tasks (manipulated variable). Perea-Lopez et al. (2003) proposed an MPC strategy for supply chain optimization with a rolling horizon approach for updating the changes to the supply chain.

In this paper, we consider a hierarchical decomposition for integration of planning and scheduling and embed some decision-making ability at each level to provide the necessary degrees of freedom to make the model amenable to rescheduling. The principles of cascade-control and a receding horizon approach are used for integrating reactive scheduling to handle unexpected machine breakdowns in a local fashion. In section 2, we present a general multi-level, control-theoretic formulation for multi-site facilities. Due to space constraints, as proof of concept we demonstrate the proposed methodology for a single-site lube production in a hybrid flowshop facility for which the problem definition and solution approach are presented in section 3. A case study with model details is illustrated in section 4 followed by conclusions in the last section.

2. A GENERAL MULTI-LEVEL, CONTROL-THEORETIC FRAMEWORK

Consider a general multi-site, multiproduct planning and scheduling problem with several plants located in different geographic locations. At each plant a generic hybrid flowshop configuration of various machines is assumed that can be easily simplified to any problem specific topology of series and/or parallel configuration of different stages. We additionally consider resource constraints on the feed side of each plant where the feed could be a continuous stream with finite storage space. Again as a problem specific situation these resource constraints may be easily simplified or dropped. The overall problem can be traditionally decomposed into two major levels, a primary level for strategic (or long-term) planning across multiple sites and a secondary level for tactical (or mid-term) planning and scheduling at each site. The primary level has demands over a longer horizon (say 1 year) and has abstractions of each plant in terms of the average production and inventory capacities. The global objective here is meeting commitments made to a large number of customers. Accordingly, based on the production and transportation costs, the model at this level (LP/MILP) sets production targets for each plant. The primary level is revised on a less frequent basis (say monthly/quarterly).

At the secondary level, for mid-term planning and scheduling the horizon of interest is smaller (1-3 months) catering to less number of customers, and the model here may be revised on a frequent (say weekly/monthly) basis. To meet the global objectives of the overall problem, as in cascade-control, we propose that any local disturbances (machine breakdowns etc.) at the plant level have to be attenuated locally before they affect the global performance. If a machine breakdown occurs in a given period then we consider a receding horizon window from that time instant onwards and meet the slacks with increased production rates (the degrees of freedom provided here) and if necessary we may intrude into the next time period without sending a feedback until the next scheduled revision of the upper level. In contrast, if we do the normal feedback control option then in the same time horizon we may have to trigger the primary level thus forcing frequent revisions of the commitments made to customers. Only when these disturbances cannot be handled locally, a feedback to the primary level is sent thus seeking a target revision.

The multi-site formulation presented above is for the sake of generality. In the present work we consider the secondary level for single-site planning and scheduling with detailed model formulations and results as described in the following sections. Nevertheless, the model at primary level can be easily integrated vertically with the following models for the secondary level.

3. PROBLEM DEFINITION AND SOLUTION APPROACH

Consider a single-site plant with a general *M*-stage hybrid flowshop facility as shown in Fig. 1.

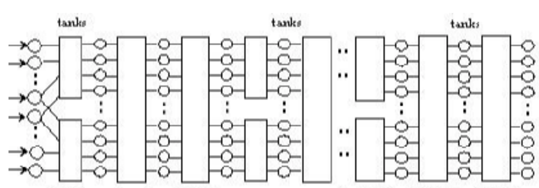

Fig. 1. General single-site hybrid flowshop facility

At each stage there could be either a single machine or multiple machines in parallel. The resource constraints are in terms of feed to the plant envelope being a continuous stream with finite storage space and hence the plant schedule is governed by both the feed side inventories and the demand side constraints. Earlier, we (Munawar et al., 2003a) proposed a generalized MINLP model for cyclic scheduling of this configuration with detailed inventory constraints. This model accounts for feed losses during grade changeovers (also known as slopping losses in continuous plants) through a modified time slot definition. It could handle special cases leading to empty slots (zero time duration) in the model resulting due to splitting of products across parallel lines. However, such model becomes intractable when we consider integration of planning and scheduling for a multi-period operation. Hence, at the secondary level for single-site plants, we propose a three level decomposition of the overall problem. Though, the case study in this paper is for cyclic scheduling of continuous multi-product plants, the proposed framework can however be appropriately extended to other cases of short-term scheduling.

At the single-site plant, the top level (level-1) for mid-term planning and scheduling has 1-3 month time horizon, the production targets for which are set from the primary level of the multi-site problem. At this level, we consider an abstract model with assumed slopping losses, and simplified inventory constraints in terms of upper bounds on processing times. In the next level (level-2) we consider detailed inventory constraints with the modified time slot definition to account for actual slopping losses. The horizon at level-2 is that of a single period and we do not consider detailed inventory management at this level, but consider an abstraction of the total inventory available for each product. The inventory management on an hourly/daily basis for individual product-to-tank assignments is done heuristically at the lower most level (level-3). For efficient usage of the available tank volumes an Inventory Slicing and Tank Reallocation (ISTR) algorithm is proposed at level-3. In the next section we present the model details for each level demonstrated for a simplified hybrid flowshop facility.

4. CASE STUDY AND MODEL DETAILS

Consider a single-site lube production in hybrid flowshop facility for producing 4 products on 3 stages as shown in Fig. 2, where stage 1 has two parallel machines line 1 and line 2 that relate to the same processing task (for e.g. extraction).

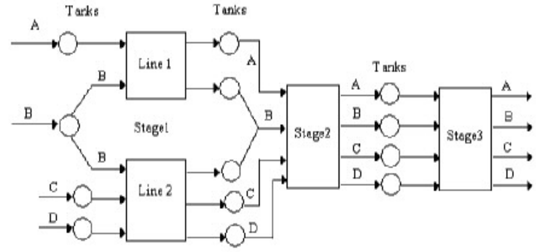

Fig. 2. Lube production in a hybrid flowshop facility

With the exception of component B, which can be processed in both the lines, all the products have to be processed one at a time in all the stages in the same sequence: line 1 or line 2 followed by stages 2 and 3.

The feed rates for base stocks of continuous streams A, B, C and D are assumed to be constant: 21, 28, 17 and 16 m^3/hr. There are some conversion losses in each stage and the processing rates and grade transitions are sequence and stage/line dependant, with the processing rates constrained between specified lower and upper bounds. The total inventory space available is finite and is about 3800 m^3 overall across all stages/lines, except for the product side where unlimited inventory space (UIS) is considered. Since the feed is a continuous stream received from upstream plants, the feed inventory tanks would almost always be busy. Nevertheless, if some inventory is unused then there is a potential for usage elsewhere in the plant from a reactive scheduling point of view. Hence, the objective at the operator level is to minimize the available inventory usage as best as possible.

4.1 Level-1 formulation:

Consider mid-term planning at level-1 for a multi-period operation over a small horizon, with demands specified for three periods each of 1000 hr, 900 hr and 800 hr duration respectively. As an abstraction of the inventories to be handled at level-2, upper bounds are specified on processing times for all products and are assumed to be 25 hr, 20 hr and 15 hr for stages 1, 2 and 3 respectively (based on past experience or some heuristics). The slopping losses are reflected by reducing the conversion rates of grades appropriately based on some heuristics. Here, we do not present the details of these heuristics for continuity purposes.

For a multi period operation, it is evident that an effective planning would require the processing tasks to be distributed uniformly, rather tightly in the early periods, so as to be optimal with respect to all the periods. For example, a high demand for a particular product in a given period would require its processing to be distributed over the entire time up to that period (including the previous periods), even though the demand for this product in the previous periods is low or zero. Even otherwise, from a reactive scheduling perspective we propose that the scheduler must push in tighter schedules in the early periods so that there is some leeway in the forthcoming periods to accommodate the unforeseen events. Hence, we allow over production in the first period of interest. As level-1 is anyway revised at the end of each period, the first period of interest rolls through, and finally we have the best possible production rates in each time period. The inventory costs for the overproduced quantities can always be weighed (trade-off) against the slacks in demands met otherwise if we do not produce in the early periods. We consider the objective function at level-1 to be maximization of production in the first period of interest subject to penalties for overall grade changeovers in all the time periods. Table 1(a) gives a comparison of the actual demands specified (set points) and the output from level-1 for the three periods considered.

Table 1. Summary of results for demand (m^3) predictions of level-1 and level-2

(a) Demands specified and projected by level-1 for all three periods

	set points			output		
	t=1	t=2	t=3	t=1	t=2	t=3
A	6237	7713	8190	7501.41	6751.27	6001.13
B	8414	7572	6731	8953.56	7032.44	6731
C	1470	2780	1975	3148.70	1101.30	0
D	1470	2780	1975	1817.14	3367.96	0

(b) Demands specified and projected by level-1 for last two periods

	set points		output	
	t=2	t=3	t=2	t=3
A	6751.27	7887.32	6751.27	6001.128
B	7032.44	6731	7032.44	6731
C	1101.30	3407.22	1101.30	0
D	3367.96	1039.9	3367.96	0

(c) Results of level-2 solved sequentially for each period

	t=1		t=2		t=3	
	set point	output	set point	output	set point	output
A	7501.41	7501.41	6751.27	6751.27	7887.32	6505.81
B	8953.56	8953.56	7032.44	7032.44	6731	6731
C	3148.70	1716.5	1101.30	1019.85	3488.67	1279.32
D	1817.14	1817.14	3367.96	1008.34	3399.52	2762.84

The model constraints at level-1 are similar to the model proposed by Munawar et al. (2003a) for level-2, except that we do not consider the detailed inventory constraints and the variables corresponding to slopping. Hence, we refer to Munawar et al. (2003a) for technical details.

For a demand limited scenario we may have to make provision for some idle time in the subsequent periods (except the first period because we are anyway allowing over production in the first period of interest) to avoid the case where machines are forced to run at their lower bound on processing rates to fill the gap in time period; instead an idle time may be preferred for the remaining time if the demands in some periods are low. However, we considered a capacity limited scenario and hence the total demands at the end of all time periods could not be met. As there is no incentive for production in the subsequent periods (except first period) the output predicted by level-1 shows some zero productions in Table 1(a). Here, the schedule is meant to be aggressive with respect to only the first period of interest and hence now we solve level-2 for the first time period (t=1) with the demands projected by level-1 as set points. Level-2 will predict the best possible demands in the presence of actual slopping losses and detailed inventory constraints. If all of the demands projected by level-1 cannot be met at level-2 (meaning heavy over production projected by level-1) then the slack (actually meaning less over production done at level-2) can be added on to the last period of the horizon. When level-1 is solved again for the remaining two periods (t=2 and 3), since we allow over production in the first period of interest (t=2), thus adding constraint of type ‘ $\geq$ ’on productions, it may lead to infeasibility if we add this slack to second time period. Hence, we always add the slacks to the last known period of interest. If it were feasible for this slack to be produced in t=2 itself, it is anyway feasible due to over production being allowed.

The output of level-1 for the last two periods is shown in Table 1(b). Now we again solve level-2 for t=2 with these demands as set points and check for feasibility. Finally for the last period (t=3) we need not solve level-1 as we can directly solve level-2. All the above runs are solved as an offline activity to find the best possible demands that can be met by level-2 in presence of real slopping losses and inventory constraints. Finally, the output of level-2 for each period is given in Table 1(c).

4.2 Level-2 formulation:

The model at level-2 involves detailed inventory constraints and actual slopping losses. The objective is maximizing profit, where the inventory and grade changeover costs are penalized subject to meeting maximum possible demands. In the continuous time domain representation the definition of a time slot is modified to account for the feed losses in slopping and additional slopping variables are defined. The model at this level is same as in Munawar et al., 2003a where we consider detailed inventory constraints in terms of the inventory breakpoints that define the total inventory profile for consumption and discharge of material.

The detailed product-to-tank assignments are not done here but are considered at level-3. At level-2 however, we consider abstractions of the available inventory volumes for each product. For the problem shown in Fig. 2 the inventory available is about 2500 m^3 for the feed tanks and 600 m^3 for tanks after line 1, 400 m^3 for tanks after line 2 and 300 m^3 for tanks between stage 2 and stage 3, totaling 3800 m^3. As we can see later in level-3 tanks are reused for storage of different grades over the time horizon and hence we consider an over estimation for the available inventory volumes based on some heuristics or past experience. Here, we use about 800 m^3 volume for each grade on the feed side, 600 m^3 for each grade for storage after line 1, 400 m^3 for each grade after line 2, and 300 m^3 for each grade after stage 2. The above volumes are used as upper bounds on the maximum inventory breakpoint for each grade at level-2. Though the feed tanks are almost always busy, some re-use of tanks is still observed at level-3, and hence we consider slightly higher upper bounds for inventories of the feed tanks as well. The problem at level-2 is solved sequentially as an offline activity for each period with the demand inputs taken from the first period of interest of level-1 as discussed earlier. The inventory profiles from level-2 in the first period are shown in Fig. 3.

Reactive scheduling formulation for machine breakdowns at level-2: Now we consider the model for reactive scheduling and analyze the interaction of level-1 and level-2 and simulate some scenarios of machine breakdowns leading to loss of available time for production in a given time period. We use a receding horizon time window as done by Munawar et al. (2003b) and find the amount of time required to be intruded into the next time period for compensating the slack in the current period. However, the proposed model in this paper is more generalized as it has provision for empty slots leading to zero time duration and simplified slopping representations at level-1.

In order to simulate the unforeseen breakdown of machines, a shut down time is introduced at the end of ten cycles (724.4 hrs) in the first time period. The current levels of inventory at the end of 724.4 hrs and the processing sequence are fixed as that of the nominal schedule. The objective function at level-2 for reactive scheduling, is posed as minimization of the amount of time required to be intruded into the next due date in order to meet the overall demand (slack in the current due date + demands in the next due date for the intruded period) at end of the receding horizon window. The shut down times are gradually increased from 1 hr to 110 hrs. For shut down times of 1, 10, 50, 75 and 99 hrs, the receding horizon approach predicted that we need to intrude 8.7, 87.8, 447.6, 673.9 and 891.5 hrs into the second period respectively. For shut down times beyond 100 hrs, even if all of the second period is included in the receding horizon the bulk demands could not be met, so, either the third period also has to be included in the receding horizon or the slack may be fed back to leve-1 for effective redistribution of demands.

4.3 Level-3 formulation:

We assume that we are given product sequencing and the total inventory profiles as input from level-2 and we need to figure out at level-3 if this volume can be met from the set of available tanks. We first focus on the triangular inventory breakpoints between stage 2 and stage 3 as shown in Fig. 3(d). We know that if these profiles do not overlap in the time frame, then we can use the same tanks repeatedly. For example the tanks that are used for storing either grade D or C can be used again for storing grades A and B. We exploit this feature in the proposed heuristic algorithm and make an efficient usage of the non-overlapping profiles. However, since the feed is a continuous stream received from upstream plants, the feed inventory tanks would almost always be busy as shown in Fig. 3(a), thus rendering less probability for reuse of these tanks elsewhere.

Note that the inverted triangular profiles for most of the grades in Fig. 3(a) continue to be in use as some feed is stored for use in next cycle, unlike the rest that are traditional triangular profiles which if freed can be used elsewhere. As already mentioned earlier, from the reactive scheduling point of view the objective for the operator at this level is to minimize the available inventory usage as best as possible.

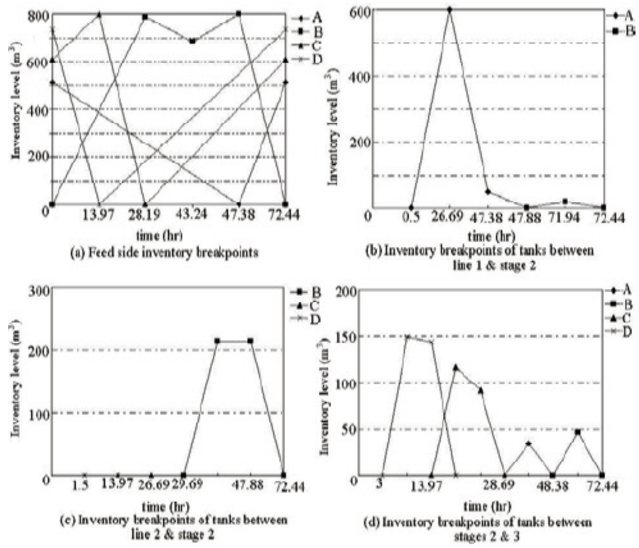

Fig. 3. Inventory profiles for level-2 in the first period of 1000 hr

Let us pose the problem here as one of finding the minimum number of tanks of each (say) 50 m^3 capacity required to store these grades. For each profile we first generate 'sub-profiles' by demarcating the 50 m^3 tank capacities as shown in Fig. 3(d) and find the corresponding timings on x-axis by linear interpolation. For example, for the profile of grade D, we get three sub-profiles of the shape stacked on each other as seen Fig. 3(d). For each of these profiles suppose if we use three tanks T1-T3 each of 50 m^3 size. Then we know the times at which each of these tanks would be occupied and free to use again. For example T3 would be empty at the end of 16.18 hr, T2 at the end of 18.68 hr and T1 at the end of 21.18 hr. Now we consider the pool of all such sub-profiles across all grades and look out for which of these sub-profiles do not overlap in the time frame and try to reuse these tanks. Using this heuristic for the inventory profiles between stage 2 and stag 3, we can find that we need a minimum of four tanks of each 50 m^3 size. Without reuse of tanks we would need eight such tanks.

When we consider many stages with numerous inventory profiles, it is difficult to visualize and apply this heuristic manually. With this motive, a simple heuristic algorithm termed as Inventory Slicing and Tank Re-allocation (ISTR) is proposed here (in appendix) which automates the generation of sliced profiles, checks for the non-overlapping zones (sub-profiles) and finds the minimum number of tanks of a given capacity (say 50 m^3 or100 m^3) required to manage these inventories efficiently at level-3. Using the ISTR algorithm for the total thirteen profiles across all lines/stages of Fig. 3., the minimum number of tanks of 50 m^3 capacity was found to be 50 (totaling 2500m^3) in the first period and 57 and 50 tanks respectively in the second and third periods. If the number of tanks suggested by ISTR at level-3 is less than the actual number of tanks available, then level-2 is revised with modified upper bounds on inventories.

Amenability for reactive scheduling at level-3: Consider the horizon at level-3 as one cycle, 72.44 hr (approx. 3 days) and the tank assignments as suggested by ISTR algorithm. From the view point of reactive scheduling, if there are any tank breakdowns in the last two days of the horizon, then they can be mostly taken care locally in a cascade control fashion, without sending any feedback to level-2, as most of the tanks are free in the last two third of the horizon. If there are some tank breakdowns on the first day of the horizon itself, then we consider a receding horizon model from that time instance onwards and if required intrude into the next cycle without sending feedback up to level-2 until its scheduled revision. Otherwise, the only way to reject such disturbance locally would be to provide some resiliency margin or back off from the best schedule, i.e. instead of deploying minimum number of tanks as suggested by ISTR, we may deploy additional tanks thus making the problem amenable to reactive scheduling. The analogy in process control applications is clear here in terms of a compromise between aggressive but non-resilient control and robust but relatively less aggressive control.

5. CONCLUSION

An integrated multi-level, control-theoretic framework has been proposed in this work for integration of planning, scheduling and rescheduling. The proposed methodology has been demonstrated for single-site lube production in a hybrid flowshop.

Appendix: *ISTR algorithm for triangular profiles:* Consider the triangular inventory breakpoints between stage 2 and stage 3. In this algorithm for each profile we first generate sub-profiles of the given tank capacity i.e. we generate time vs. volume data (say V_{ik}) for each grade by slicing the inventory profiles as discussed earlier. This completes inventory slicing. In the V_{ik} data, all the inventory breakpoints would then be multiples of the given tank capacity (50 m^3), and the volumes get repeated grade-wise, except the maximum breakpoint. We exploit this feature in our algorithm. We know that the first occurrence of these entries correspond to the time at which a tank needs to be deployed and the repeated occurrence corresponds to the time at which such tank would be freed. At each time instance (k), for each grade (i), for each breakpoint, V_{ik}, we start deploying new tanks until we encounter a maximum breakpoint (say V_i^{max}). For the entry corresponding to V_i^{max} we assign the same tank as was used for the previous V_{ik} of the same grade (indicating the same tank has been still in use). Now after crossing the maximum breakpoint, before assigning a new tank at each time instance, we additionally check for each grade if same V_{ik} entry already exists in the previous time instances. After finding the first such occurrence (in the backward search from the current time instance, k), we assign the same tank number as that was used earlier (indicating the same tank has been in use till now). And we mark the status of this tank number as freed and available from this time instance onwards. Now, for the subsequent time instances, before assigning new tanks we also check if some freed tanks are available and if so we use them and remove the corresponding entry from the available tank list. Only when the list of freed tanks is empty we deploy a new tank. For inverted triangular profiles the algorithm had to be modified to account for their typical characteristic features, the details of which are beyond the scope of this paper.

REFERENCES

Bose, S., and J.F. Pekny (2000). A model predictive framework for planning and scheduling problems: a case study of consumer goods supply chain. *Comp. Chem. Eng.*, **24**, 329-335.

Grossmann, I.E., S.A. Van den Heever and I. Harjunkoski (2001). Discrete optimization methods and their role in the integration of planning and scheduling. *AIChE Symp. Ser.*, **98** *(326)*, 150-168.

Munawar, S.A., M. Bhushan, R.D. Gudi and A.M. Belliappa (2003a). Cyclic scheduling of continuous multiproduct plants in a hybrid flowshop facility. *Ind. Eng. Chem. Res.*, **42**, 5861-5882.

Munawar, S.A., M. Bhushan, R.D. Gudi and A.M. Belliappa (2003b). A multi-level, control-theoretic approach to reactive scheduling in continuous plants. In: Grossmann, I.E. and C.M. McDonald (Eds.), *Proceedings of the conference in foundations of computer aided process operations, CACHE Publications*, 397-400.

Perea E., I.E. Grossmann, E. Ydstie and T. Tahmassebi (2000). Dynamic modeling and classical control theory for supply chain management. *Comp. Chem. Eng.*, **24**, 1143-1149.

Perea-Lopez E., B.E. Ydstie and I.E. Grossmann (2003). A model predictive control strategy for supply chain optimization. *Comp. Chem. Eng.*, **27**, 1201-1218.

Shah, N. (1998). Single and multisite planning and scheduling: current status and future challenges. *AIChE Symp. Ser.*, **94** *(320)*, 75.

Shobrys, D. E., and D. C. White (2000). Planning, scheduling and control systems: why can they not work together. *Comp. Chem. Eng.*, **24**, 163-173.

Van den Heever, S.A., and I.E. Grossmann (2003). A strategy for the integration of production planning and reactive scheduling in the optimization of a hydrogen supply network. *Comp. Chem. Eng.*, **27**, 1813-1839.

Vargas-Villamil, F.D. and D.E. Rivera (2000). Multilayer optimization and scheduling using model predictive control: application to reentrant semiconductor manufacturing lines. *Comp. Chem. Eng.*, **24**, 2009-2021.