

# Robust Management of Multi-product, Multi-echelon Demand Networks Using Model Predictive Control

Martin W. Braun<sup>†</sup>, Daniel E. Rivera<sup>†</sup>, W. Matthew Carlyle<sup>††</sup>, and Karl G. Kempf<sup>†††</sup>

<sup>†</sup>Department of Chemical and Materials Engineering  
Control Systems Engineering Laboratory,  
Arizona State University, Tempe, Arizona 85287-6006

<sup>††</sup>Department of Industrial Engineering,  
Arizona State University, Tempe, AZ, USA 85287-5906

<sup>†††</sup>Decision Technologies, Intel Corporation,  
5000 W. Chandler Blvd., Chandler, AZ, USA 85226

(480) 965-9476

email: daniel.rivera@asu.edu

<http://www.eas.asu.edu/~csel/presrecord01.pdf>

Prepared for presentation at the Annual AIChE Meeting, Reno, NV  
November 4–9, 2001

Session: [285] Integration Between Scheduling, Planning and Control

Copyright © Control Systems Engineering Laboratory, ASU.

November, 2001

UNPUBLISHED

AIChE shall not be responsible for statements or opinions contained in papers or printed in publications.

# Robust Management of Multi-product, Multi-echelon Demand Networks Using Model Predictive Control

M. W. Braun<sup>1\*</sup>, D. E. Rivera<sup>1†</sup>, W. M. Carlyle<sup>2</sup>, and K. G. Kempf<sup>3</sup>

<sup>1</sup>Department of Chemical and Materials Engineering  
Control Systems Engineering Laboratory,  
Arizona State University, Tempe, Arizona 85287-6006

<sup>2</sup>Department of Industrial Engineering,  
Arizona State University, Tempe, AZ, USA 85287-5906

<sup>3</sup>Decision Technologies, Intel Corporation,  
5000 W. Chandler Blvd., Chandler, AZ, USA 85226

November 5, 2001

## Abstract

Model Predictive Control (MPC) has long been recognized as an effective, practical solution for multivariate control in the process industries. With recent advances in computer and information technology, the opportunity exists to incorporate model predictive control into the supply chain management business function. In this paper, the interaction of control parameters in the decentralized controllers of a two-node supply chain simulation is investigated, providing insights into the behavior of the network under uncertainty. An MPC approach is shown to be flexible enough to handle a six-node demand network simulation proposed by Intel Corp. This simulation has been designed with topology, facility characteristics, and demand uncertainty similar to real semiconductor demand networks. This study also demonstrates the benefits of information sharing for networks with long process or transportation delays.

---

\*Currently with Texas Instruments Inc., 13570 N. Central Expressway, MS 3701, Dallas, TX, 75243

†To whom all correspondence should be addressed. phone: (480) 965-9476 fax: (480) 965-0037; e-mail: daniel.rivera@asu.edu

## 1 Introduction

Improperly managed supply chains or demand networks expose organizations to the risk of millions (if not billions), of dollars of excess cost and scrapped inventory in times of economic downturn, ultimately leading to the demise of entities in the chain (Chandra and Kumar, 2000). During ideal economic conditions, mismanagement prevents an organization from fully realizing maximum profit, providing superior customer service, and gaining market share. The most successful supply chain will gain market share by providing its customers with the right product, in the right amount, at the right time, for the right price, at the right place (Bodington and Shobrys, 1999; Kempf *et al.*, 2001).

The semiconductor sector is one of the most dynamic industries to study in this respect. Products in this industry experience short life-cycles in comparison with other industries. Demand is often cyclic and long production lead times are unavoidable due to the complexity of the product and the technology required to build the product. To combat uncertainty in supply and demand in the computer industry, the total inventory held throughout the chain may exceed one year's worth of demand (Lee *et al.*, 1997).

In this paper, an approach using Model Predictive Control (MPC) is shown to provide robustness in the face of inaccurate demand forecasts and production lead time estimates for a realistic demand network simulation of industrially relevant size and topology proposed by Intel Corp. This two-product, six-node network consists of assembly/test, warehouse, and retailer nodes as shown in Figure 1. An additional study to assess the impact of possible information sharing (feed-forward control) configurations on attainable safety stock levels demonstrates the flexibility of this approach. This study suggests a reduction of approximately 85% in safety stock levels can be achieved through information sharing, in combination with an appreciable improvement in network performance. Additional results and work developing meaningful MPC formulations for supply chain management can be found in Braun (2001).

Kapsiotis and Tzafestas (1992) were the first to use an MPC-type approach for inventory control. Their formulation adjusted replenishment to maintain a target inventory level in the face of disturbances (material removed from inventory due to demand). This work demonstrated robust setpoint tracking of a step change in target inventory level, subject to a stochastic demand, plant-model mismatch, and perishable inventory. This work offered a practical alternative to classical control approaches for inventory control, which had fallen out of favor with researchers (Axsäter, 1985).

Recently, the practical success of MPC in the process industries has inspired process control researchers to apply MPC in the area of supply chain management. Bose and Pekny (2000), provide analysis of the tradeoff between centralized and decentralized oriented designs and the factors that

impact customer service for networks handling fast moving consumer goods. Perea *et al.* (2000) have used MPC control on a discrete event simulation of a three-product, four-node, four-echelon supply chain to investigate improvements in system performance through centralized control of the nodes. Flores *et al.* (2000) have demonstrated the robustness of MPC strategies for lead time uncertainties for both deterministic and stochastic demand profiles.

This paper begins with a review of similar approaches to supply chain management and inventory control utilizing MPC. The paper continues with the formulation of the MPC approach for a two-node network to illustrate the technique, and demonstrate the effect of control parameter interactions on supply chain performance metrics. This work is an extension of the analysis of Flores *et al.* (2000). In Section 3, the Intel Corp. network is defined and modeled using a material balance approach. Subsequently, the MPC management configuration for the six node network is described. This approach is shown to handle realistic demand, and constrained shipping scenarios. The final result discussed illustrates the safety stock reduction possible through various information sharing configurations. Discussion of the financial impact of this research brings the paper to a close.

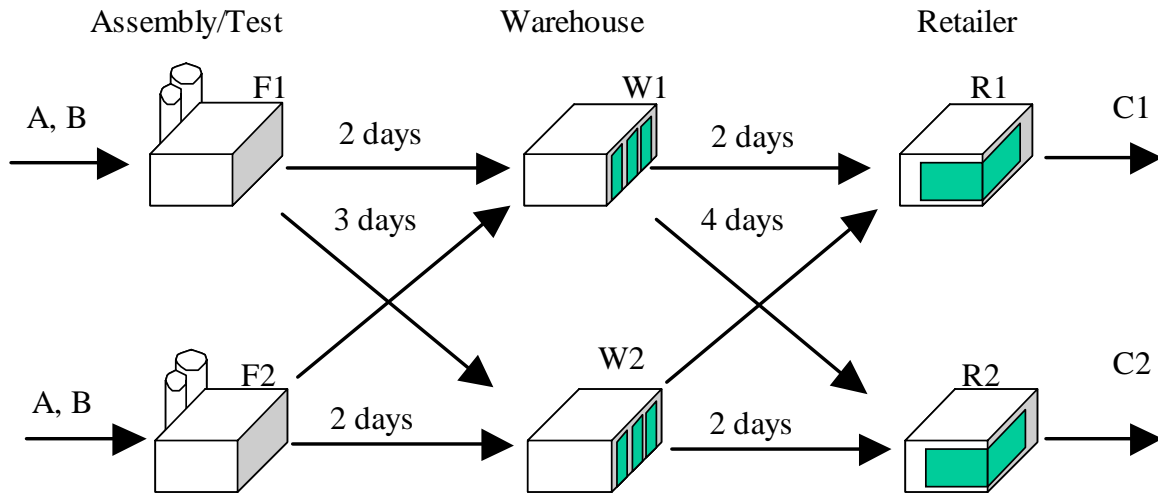


Figure 1: Six-node network material flow configuration.

## 2 Two-node Supply Chain Management

In this section, the two-node supply chain originally discussed in Flores *et al.* (2000) is re-examined to further understand the impact uncertainty has on the operable region of control parameters in terms of the metrics of inventory, backorders, and backorder lead times. Figure 2 shows the material flows for the two-node configuration. Both products  $A$  and  $B$  are shipped from the Factory to the Retailer, and from the Retailer to the customer. It is assumed that the Factory performs some value-added operation on  $A$  and  $B$ , which requires some processing time before the material shows up in the inventory level (i.e. either  $I_A$  or  $I_B$ ). It is assumed Products  $A$  and  $B$  are readily available to the Factory and thus the supply of these products is not relevant to the problem. Between the Factory and the Retailer, the materials may experience a transportation delay. Once at the Retailer, the materials are available for customer purchase as soon as they are visible in inventory.

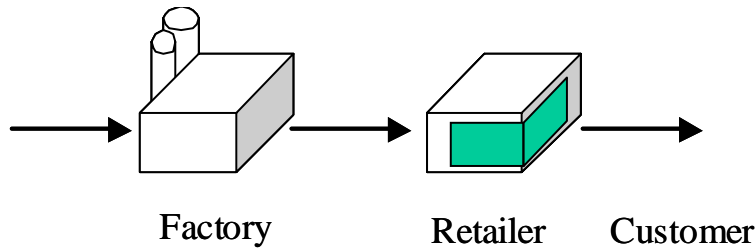


Figure 2: Two-node network material flow configuration.

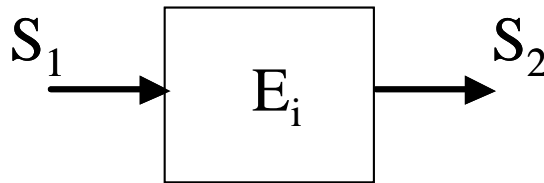


Figure 3: Generic entity material streams.

For the single-inflow, single-outflow modeling entity  $E_i$  shown in Figure 3, stream  $S_1$  enters from an upstream node or alternatively it may represent factory starts. Stream  $S_2$  leaves the entity and may follow a transportation link to arrive at a downstream node. Alternatively,  $S_2$  may represent the material taken from the node by the customer. The material balances can be written

$$I_A(k+1) = P_A(k) - S_{2A}(k) + I_A(k), \quad (1)$$

$$I_B(k+1) = P_B(k) - S_{2B}(k) + I_B(k), \quad (2)$$

Table 1: Variable mapping for two-node, MPC configuration.

Process Control Variable	Demand Network Information
setpoints $r$	inventory targets of species A & B
outputs $y$	inventories of species A & B minus cumulative outstanding backorders
estimated outputs $\hat{y}$	forecasted inventories of species A & B
measured disturbances $u_d$	demand or orders for species A & B being placed at the node
estimated future measured disturbances $\hat{u}_d$	forecasted demand or orders for species A & B being placed at the node
estimated inputs $\hat{u}$	forecasted orders for species A & B being placed at the upstream node
inputs $u$	orders for species A & B being placed at the upstream node

where

$$P_A(k) = S_{1A}(k - \Theta_p) \quad (3)$$

$$P_B(k) = S_{1B}(k - \Theta_p). \quad (4)$$

where  $\Theta_p$  represents the processing delay;  $P_A(k)$  and  $P_B(k)$  represent material which has completed processing at time  $k$  of species  $A$  and  $B$  respectively. These mass balances are subject to the constraints

$$I_A(k) + I_B(k) \leq C_I \quad (5)$$

$$S_{1A}(k) + S_{1B}(k) \leq C_1 \quad (6)$$

$$S_{2A}(k) + S_{2B}(k) \leq C_2, \quad (7)$$

which may or may not be imposed, depending on the experimental conditions. The first constraint represents an inventory capacity limit  $C_I$ . The second constraint represents an arrival capacity into the facility  $C_1$ . The third constraint represents a release capacity from the facility  $C_2$ .

The objective function of an MPC controller, can be written as

$$J = \sum_{\ell=1}^{N_p} Q_e(\ell) (\hat{y}(k + \ell|k) - r(k + \ell))^2 + \sum_{\ell=1}^m Q_{\Delta u}(\ell) (\Delta u(k + \ell - 1|k))^2 \quad (8)$$

$$+ \sum_{\ell=1}^{N_u} Q_u(\ell) (u(k + \ell - 1|k) - u_{target}(k + \ell - 1|k))^2$$

The three terms in the MPC cost function penalize predicted setpoint tracking error, excess movement of the manipulated variable, and deviation of the manipulated variable from a target value, respectively. The MPC optimization problem can be written

$$\min_{\Delta u(k|k)\dots\Delta u(k+N_u-1|k)} J \quad (9)$$

s. t.

$$u_{min} \leq u(k + \ell - 1|k) \leq u_{max}, \ell = 1, \dots, N_p \quad (10)$$

$$\Delta u_{min} \leq \Delta u(k + \ell - 1|k) \leq \Delta u_{max}, \ell = 1, \dots, N_p \quad (11)$$

$$\Delta u(k + j|k) = 0, j = N_u, \dots, N_p \quad (12)$$

The optimization problem is readily solved by standard quadratic programming (QP) algorithms. Only the first control element of the solution is implemented. At the next time step the optimization problem is solved again with updated information from the system. This is referred to as the receding-horizon property of MPC as illustrated in Figure 4. Note that the MPC controller explicitly uses a model relating the inputs, and measured disturbances to the outputs.

Each entity has its own MPC controller to manage inventory levels. Both controllers are developed with a slight modification of the interpretation of inventory. Namely, the net stock as defined by Silver *et al.* (1998) is used as the measured output of the system. Net stock is the measured inventory at time  $k$  less the cumulative outstanding backorders at time  $k$ . It is expected that this configuration will benefit the performance of network, since the backorders will now be observable by the MPC controllers. Table 1 specifies the exact mapping of the demand network definitions to process control variables. Figure 5 illustrates the information flows.

Current customer demand is fed directly to the retailer and the retailer can immediately fill that demand that day. For the deterministic case (concatenated step functions), it is assumed that this forecast is known, or it may be biased or randomized in a manner dictated by the simulation conditions. The demand (measured disturbance) and demand forecasts (estimated future measured disturbances) are fed to the first echelon MPC controller. Using the current net stock (outputs) information from the retailer, the first echelon MPC controller decides what orders (inputs) for products A and B should be placed with the factory, and what the order forecast (estimated inputs) will look like. This order forecast is shared with the second echelon MPC controller. The second echelon MPC controller uses the order forecast (now an estimated future measured disturbance) from the first echelon MPC controller, and the net stock information (outputs) from the factory to decide on production starts (inputs) for the day. Both MPC controllers contain models that determine the effect orders (measured disturbances) from downstream entities have on the future net stock levels (estimated outputs) in their node. This model also relates orders to the factory (orders for the first echelon MPC controller) and production starts (inputs for the second

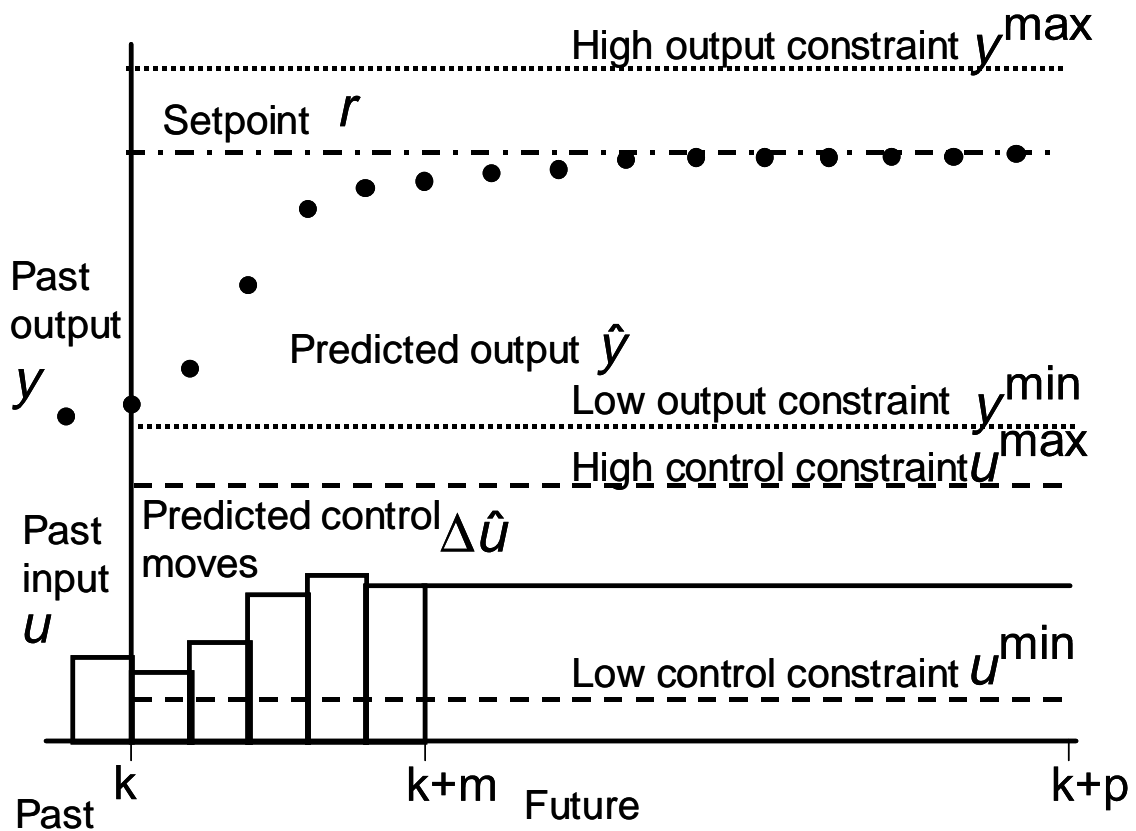


Figure 4: MPC receding horizon philosophy.

echelon controller) to the net stock levels (outputs). The net stock targets (setpoint trajectories) are a forward time shifted version of the estimated future measured disturbances (plus safety stock), for the first echelon controller. For the factory, the net stock targets are an exact replication (plus safety stock), of the estimated measured disturbances for the factory since there is no direct feed through (i.e. orders placed today are only on backorder if not filled tomorrow).

The MPC management strategy is now tested using a multiple step demand profile. The numeric values are deliberately kept small to determine the effect of round-off error on the algorithm. This round off error also contributes to the determination of optimal values for move suppression, since it also acts as a disturbance to the management system. To understand the interaction of move suppression values for the retailer and factory with no plant-model mismatch, the same simulation experiment was run for all combinations of move suppression between values of  $10^{-10}$ , 1, 5, and 10, and then by  $10$ 's up to 150. Table 2 contains the prediction and control horizon lengths used in both controllers. All six metrics of leadtime, backorder, and inventory (cumulative and maximum) are plotted as a function of Retailer and Factory move suppression in Figures 6 and 7. To provide

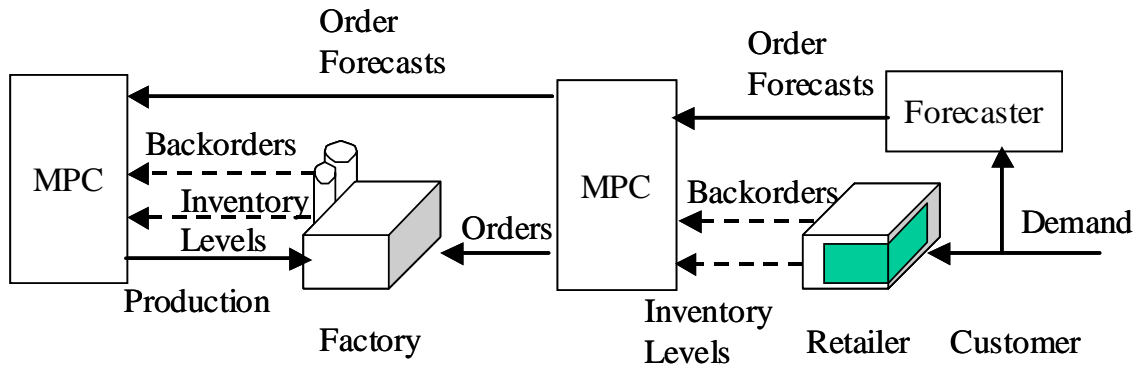


Figure 5: Two-node network MPC configuration information flow.

adequate scaling, some of the data has been omitted from the plots. The omitted results typically are orders of magnitude greater than those shown. The time-series results for the omitted cases typically exhibit highly oscillatory or unstable behavior. Therefore, operating with both move suppression values at  $10^{-10}$  provides very good performance as shown in Figures 8 and 9, but any other combination of values with  $10^{-10}$  provides very poor performance.

Table 2: Prediction and control horizons used in six node controllers.

Parameter	Retailer	Factory
$N_p$	15	13
$N_u$	13	9

To get a sense for the robustness of the MPC configuration, a one time unit transportation delay is introduced between the factory and the retailer. This delay provides plant-model mismatch, since the model in the first echelon MPC controller is not changed. This transportation delay is initialized with a shipment of 10 units. In addition, the production delay in the factory is increased by one time unit as well. The model in the second echelon MPC controller is not modified. The same demand profile is used and the same move suppression values are investigated. Using no move suppression (i. e.  $10^{-10}$ ) produces the results shown in Figures 10 and 11. By inspection of these figures, it is evident that “optimal” decisions calculated with imperfect models lead to a severe degradation in performance. By inspection of Figures 6 and 7, it is revealed that another minimum in the metrics can be found as the move suppression in the retailer and factory both approach<sup>1</sup> 150 . In contrast, the results from a move suppression value of 150 demonstrate remarkable robustness

<sup>1</sup>The space bounded by 100 and 200 for both move suppression values was investigated as well, but is not shown for brevity.

to plant-model mismatch. The cumulative backorders experienced in both nodes are an order of magnitude lower with the robust setting. When the plant models are perfect, the values of 150 for move suppression provide reasonable performance (Figures 14 and 15). Moreover, the inventory levels achieved without move suppression for the plant-model mismatch case would not be tolerated in real facilities. With plant-model mismatch, the advantages of move suppression become clear.

In Figures 16 and 17, the metrics as a function of move suppression in each of the controllers is presented for the case with plant-model mismatch. With plant-model mismatch the dependencies of the metrics on the move suppression have changed. The backorder and lead time metrics for the retailer now become dependent on both the retailer *and* the factory move suppression. In the factory, the inventory levels are also dependent on both values for move suppression. Backorders and lead time in the factory continue to have a significant dependence on the retailer's move suppression. The interaction plots in both the perfect information case and plant-model mismatch case illustrate the interdependence the two closed-loop systems have on each other. This is due in part to their cascade-like structure, but also because of the interdependence brought by the information and material flows of the network. It is evident that the area about 150 for move suppression for both controllers exhibits a minimum in the metrics.

In a financial sense, the move suppression represents a penalty on the change in order amount. This research suggests the utility of supply chain wide or demand network wide order change penalties that may otherwise be counter-intuitive from a business perspective. These penalties may be viewed as a barrier to the agility of the network or a restriction from attaining the optimal cost reduction/profit maximization for the network. Yet as this section has shown, in an uncertain world, the "optimal" decision does not necessarily produce the optimal result.

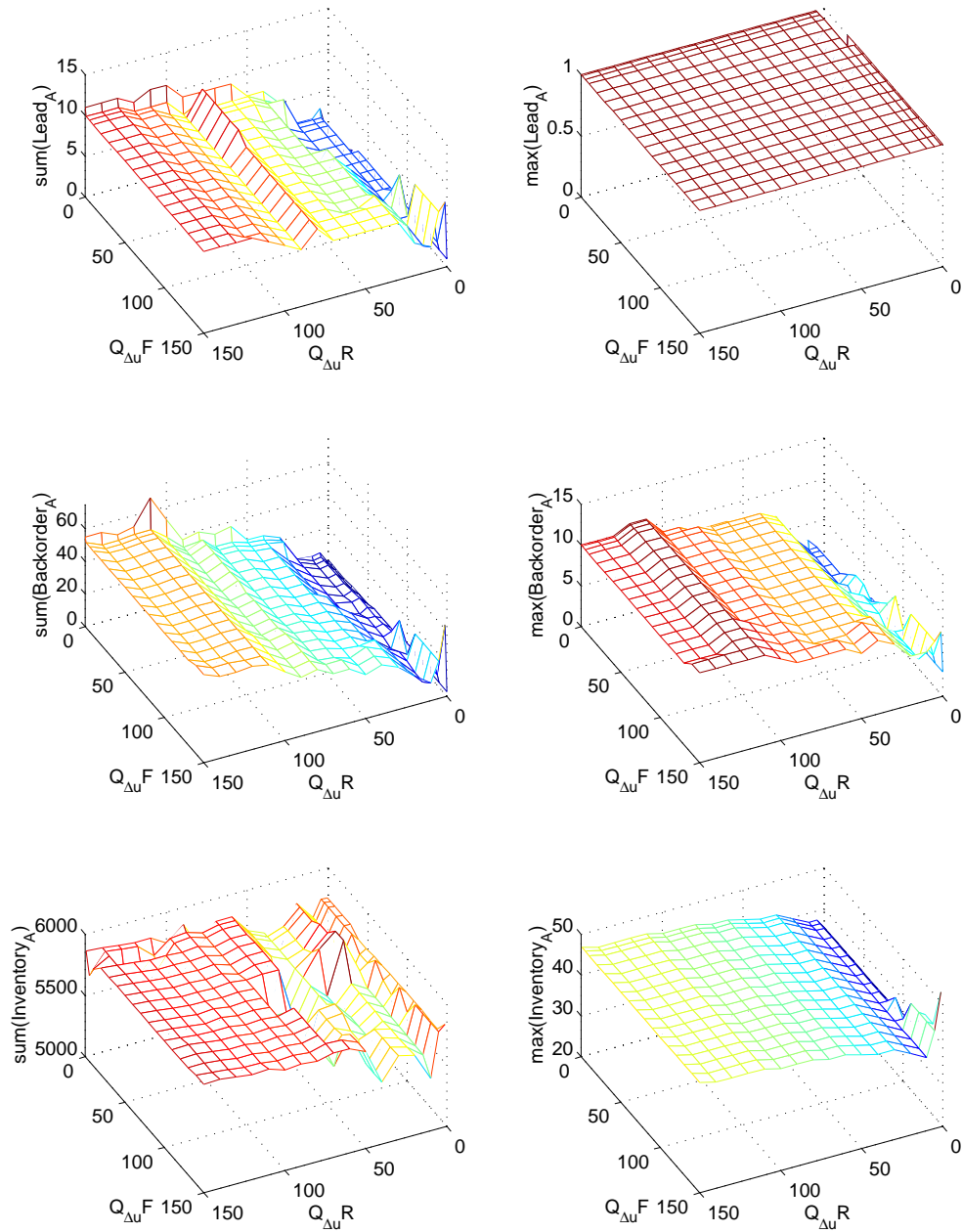


Figure 6: Retailer metrics for move suppression interactions, under no plant-model mismatch.

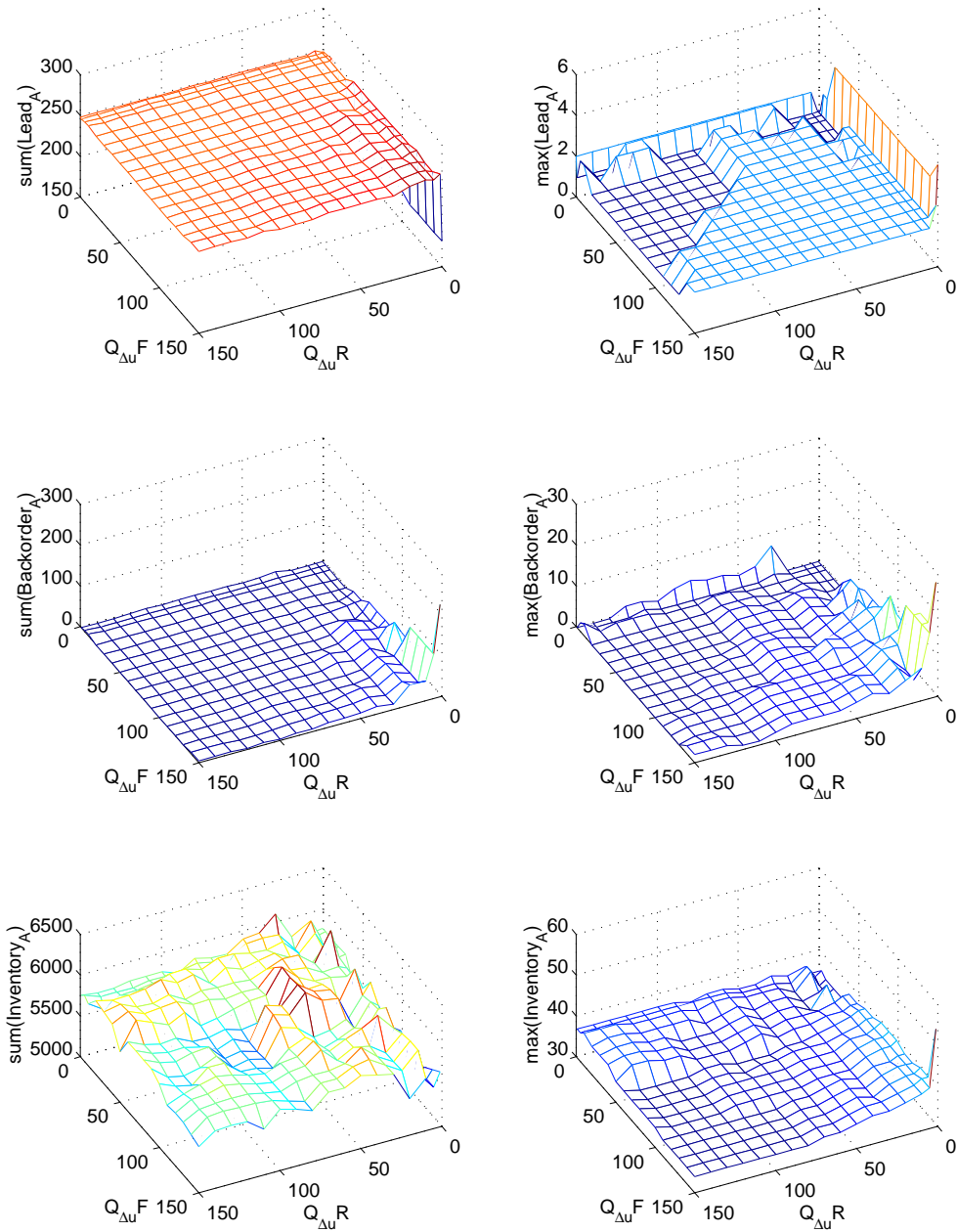


Figure 7: Factory metrics for move suppression interactions, under no plant-model mismatch.

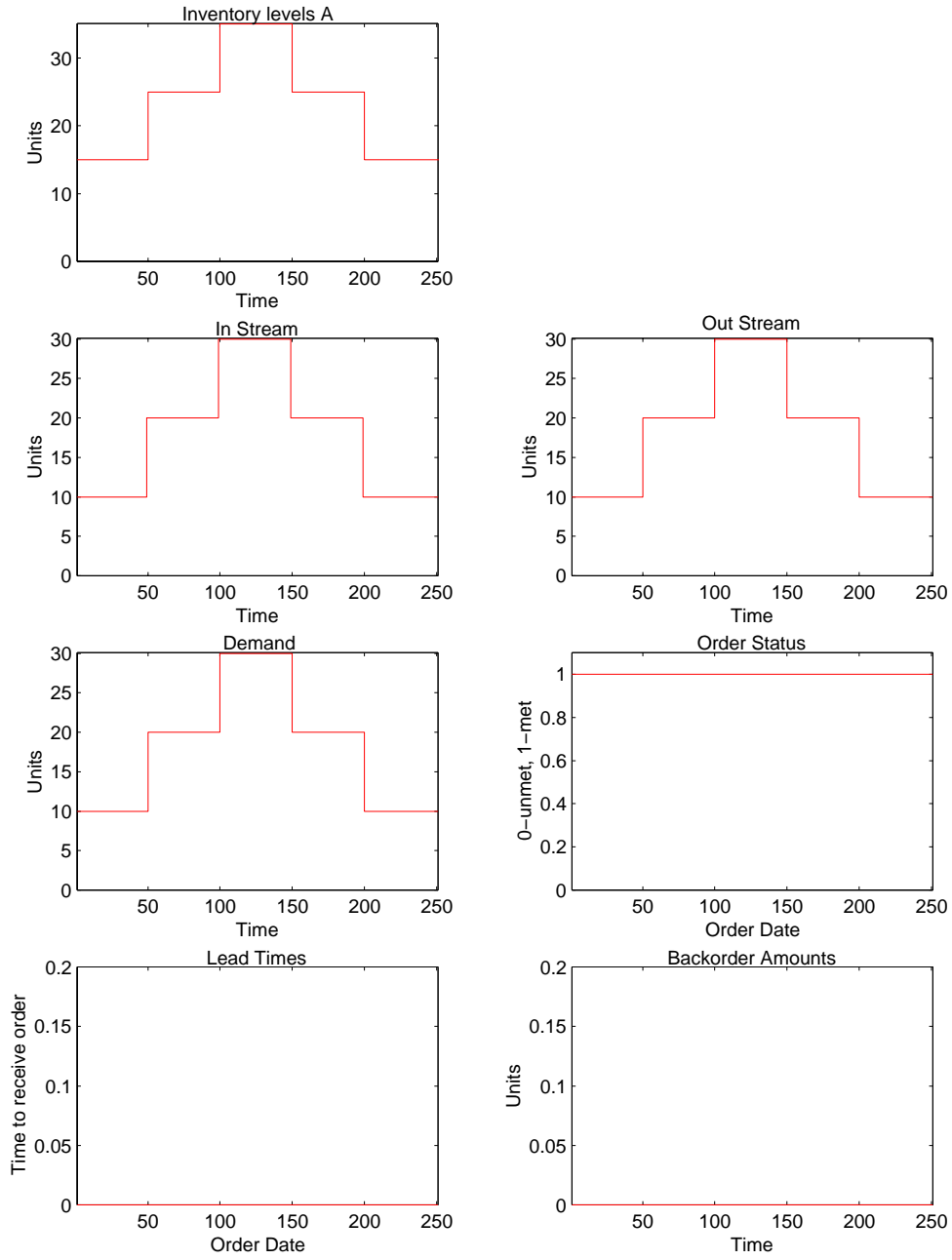


Figure 8: Retailer material balance and metric time series  $Q_{\Delta u}^R = Q_{\Delta u}^F = 10^{-10}$ , MPC configuration #1, no plant-model mismatch.

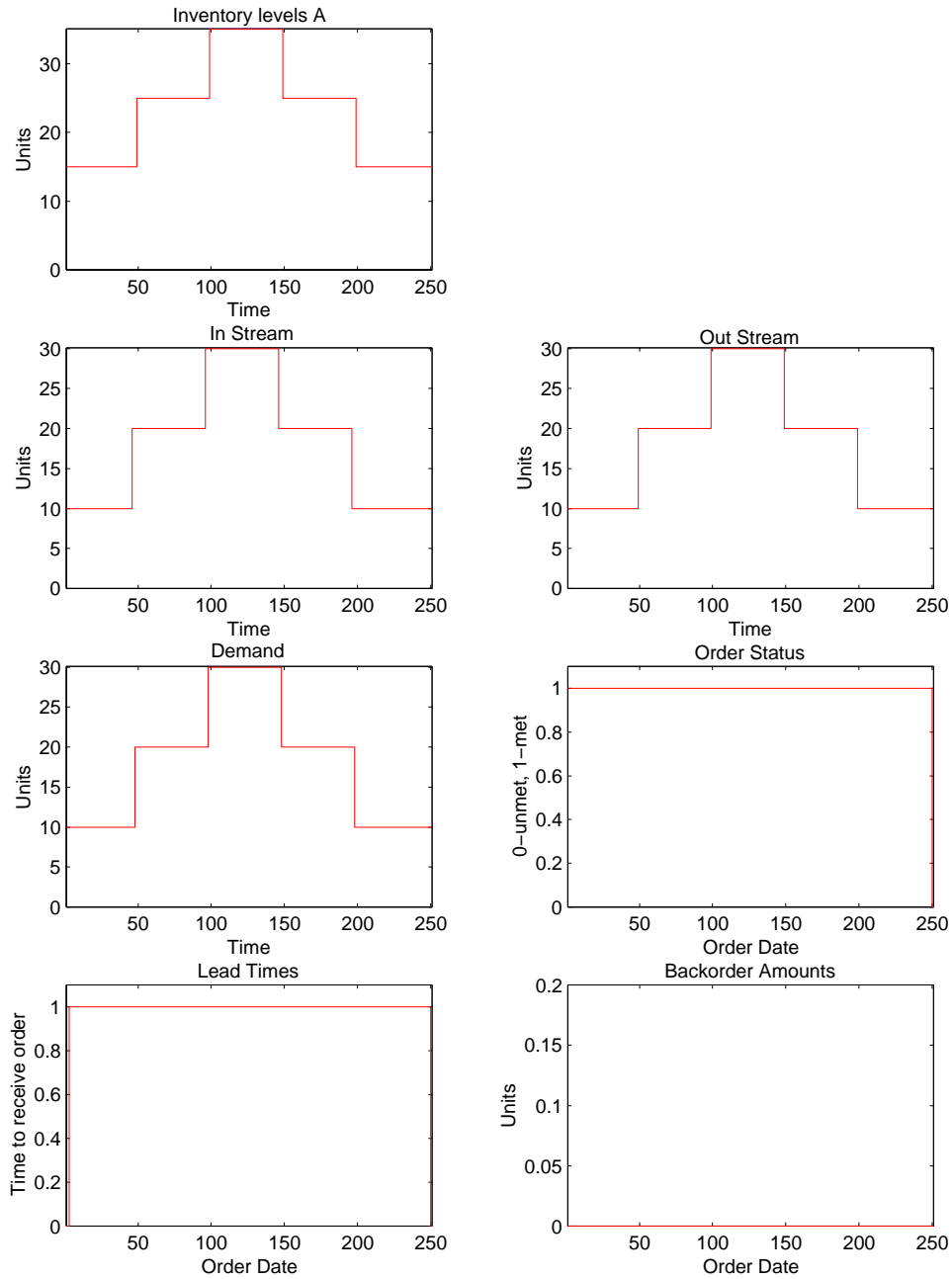


Figure 9: Factory material balance and metric time series  $Q_{\Delta u}^R = Q_{\Delta u}^F = 10^{-10}$ , MPC configuration #1, no plant-model mismatch.

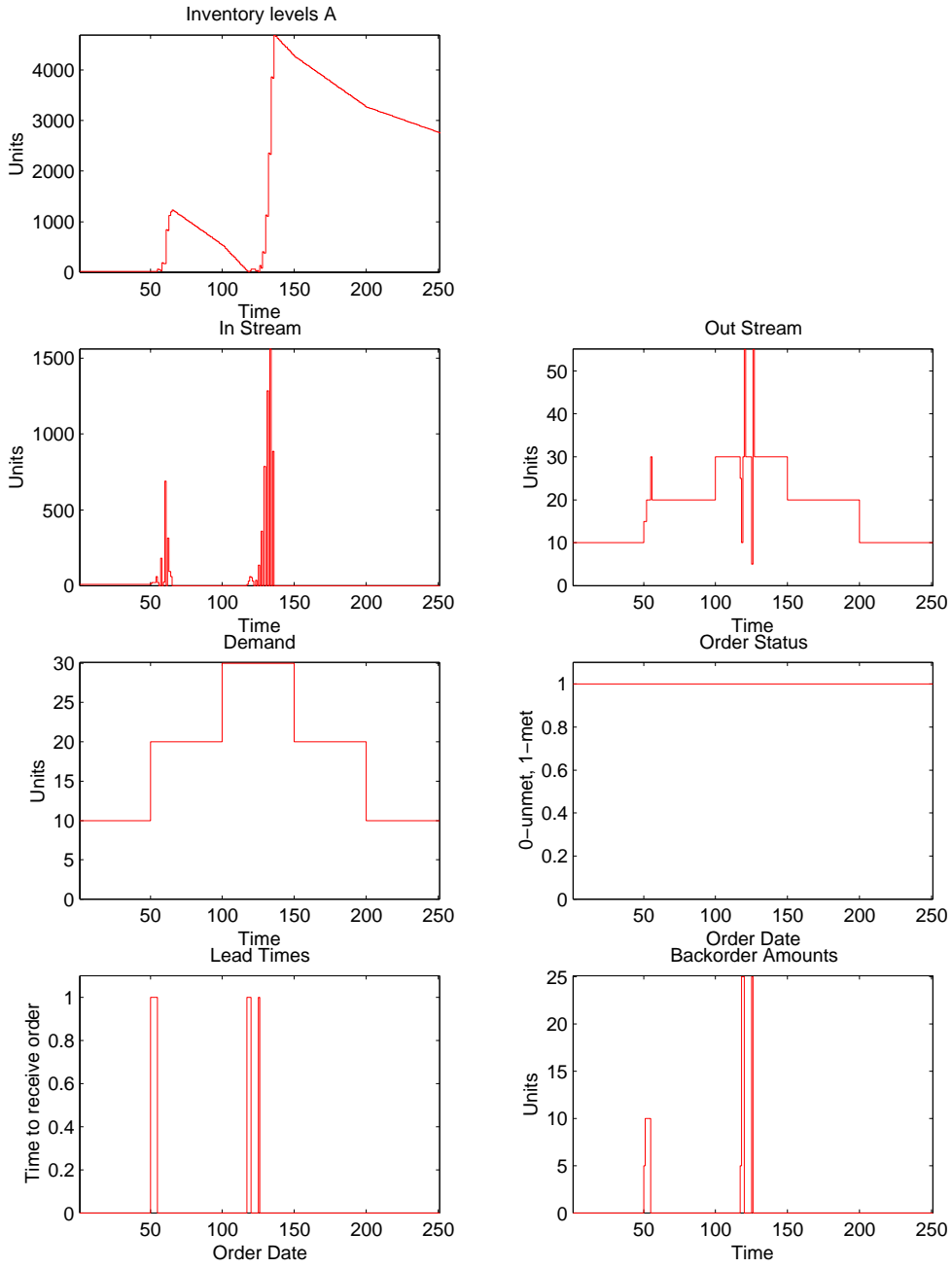


Figure 10: Retailer material balance and metric time series  $Q_{\Delta u}^R = Q_{\Delta u}^F = 10^{-10}$  under plant-model mismatch.

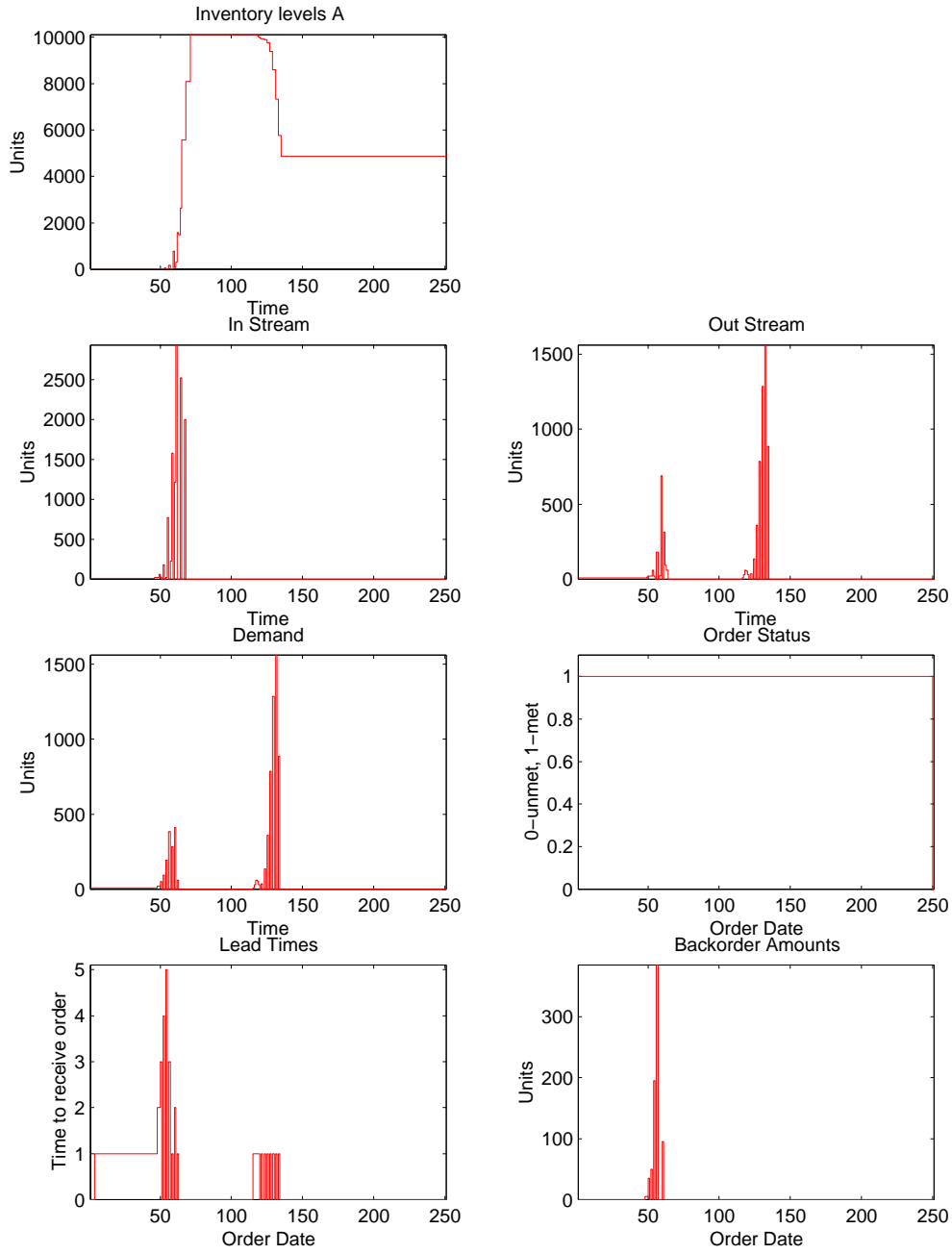


Figure 11: Factory material balance and metric time series  $Q_{\Delta u}^R = Q_{\Delta u}^F = 10^{-10}$  under plant-model mismatch.

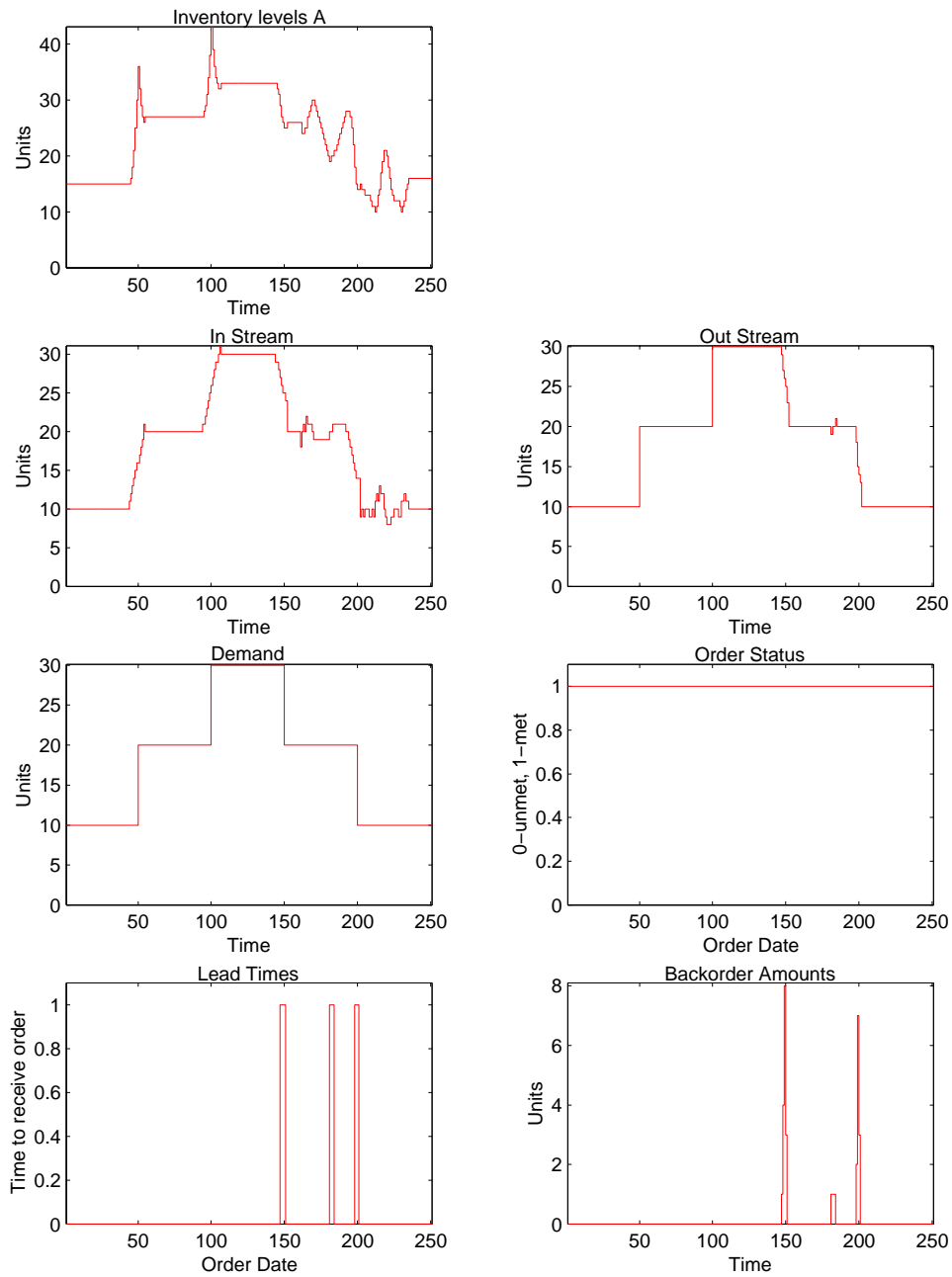


Figure 12: Retailer material balance and metric time series  $Q_{\Delta u}^R = Q_{\Delta u}^F = 150$ , under plant-model mismatch.

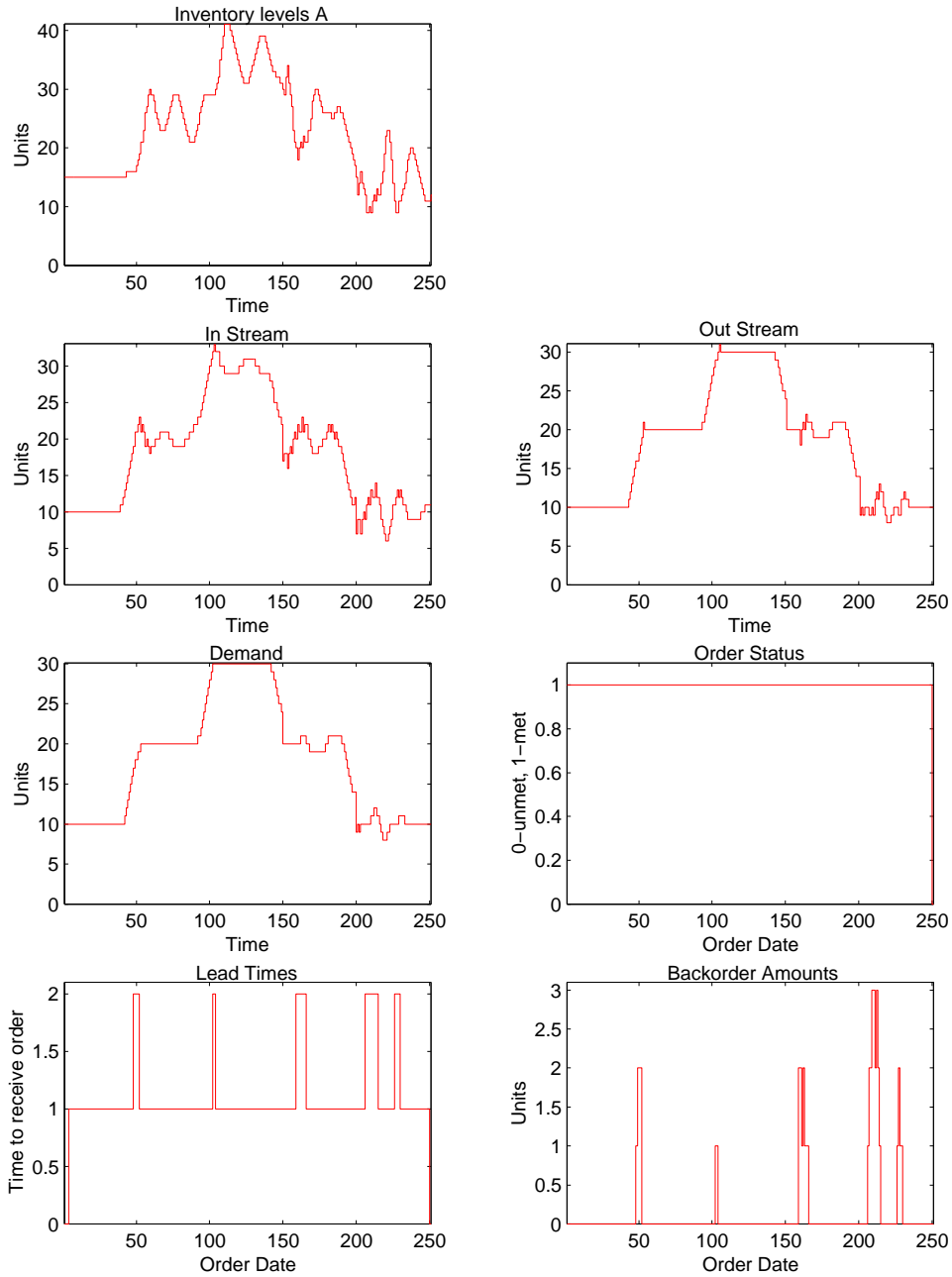


Figure 13: Factory material balance and metric time series  $Q_{\Delta u}^R = Q_{\Delta u}^F = 150$ , under plant-model mismatch.

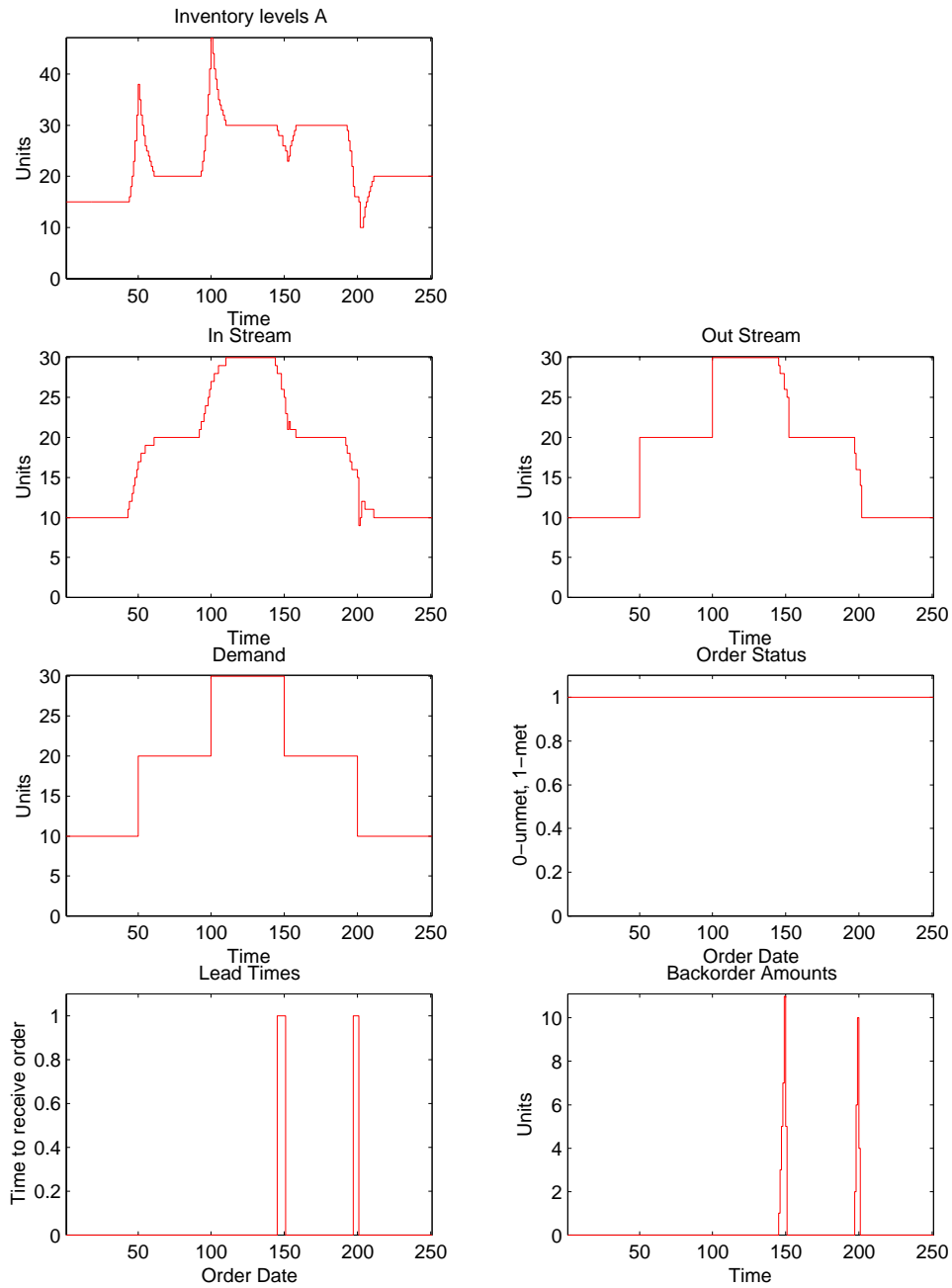


Figure 14: Retailer material balance and metric time series  $Q_{\Delta u}^R = Q_{\Delta u}^F = 150$ , under no plant-model mismatch.

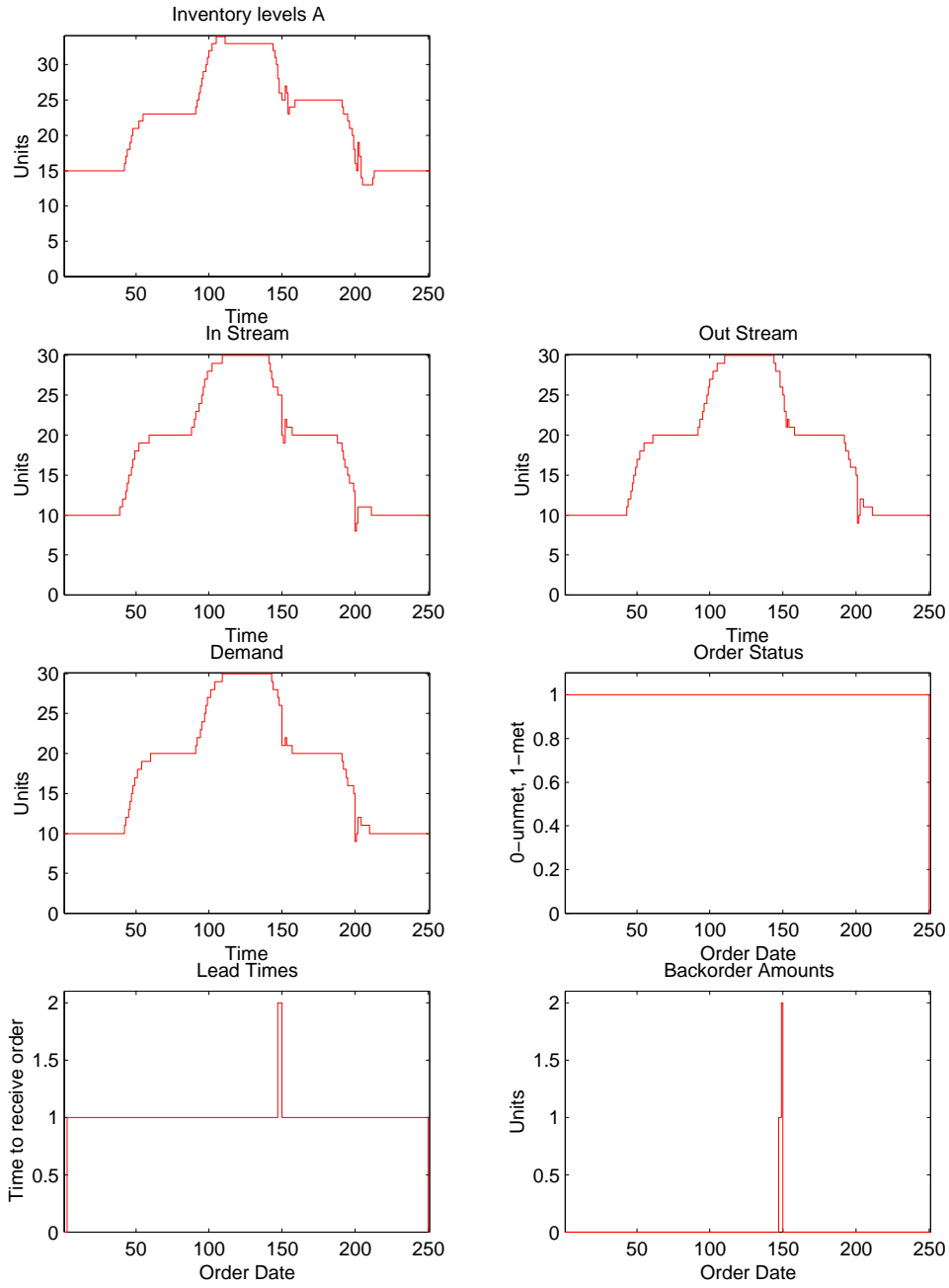


Figure 15: Factory material balance and metric time series  $Q_{\Delta u}^R = Q_{\Delta u}^F = 150$ , under no plant-model mismatch.

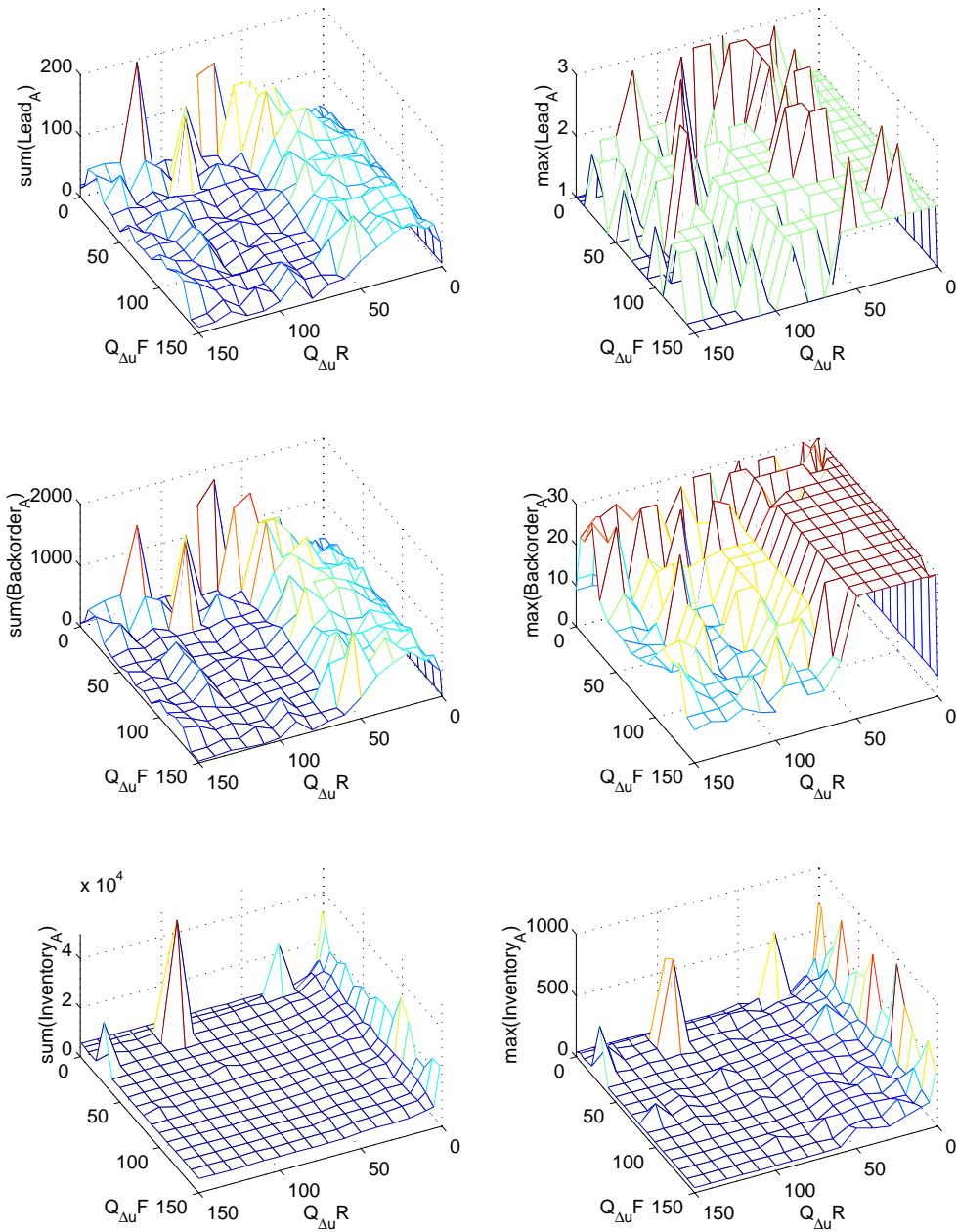


Figure 16: Retailer metrics for move suppression interactions, under plant-model mismatch.

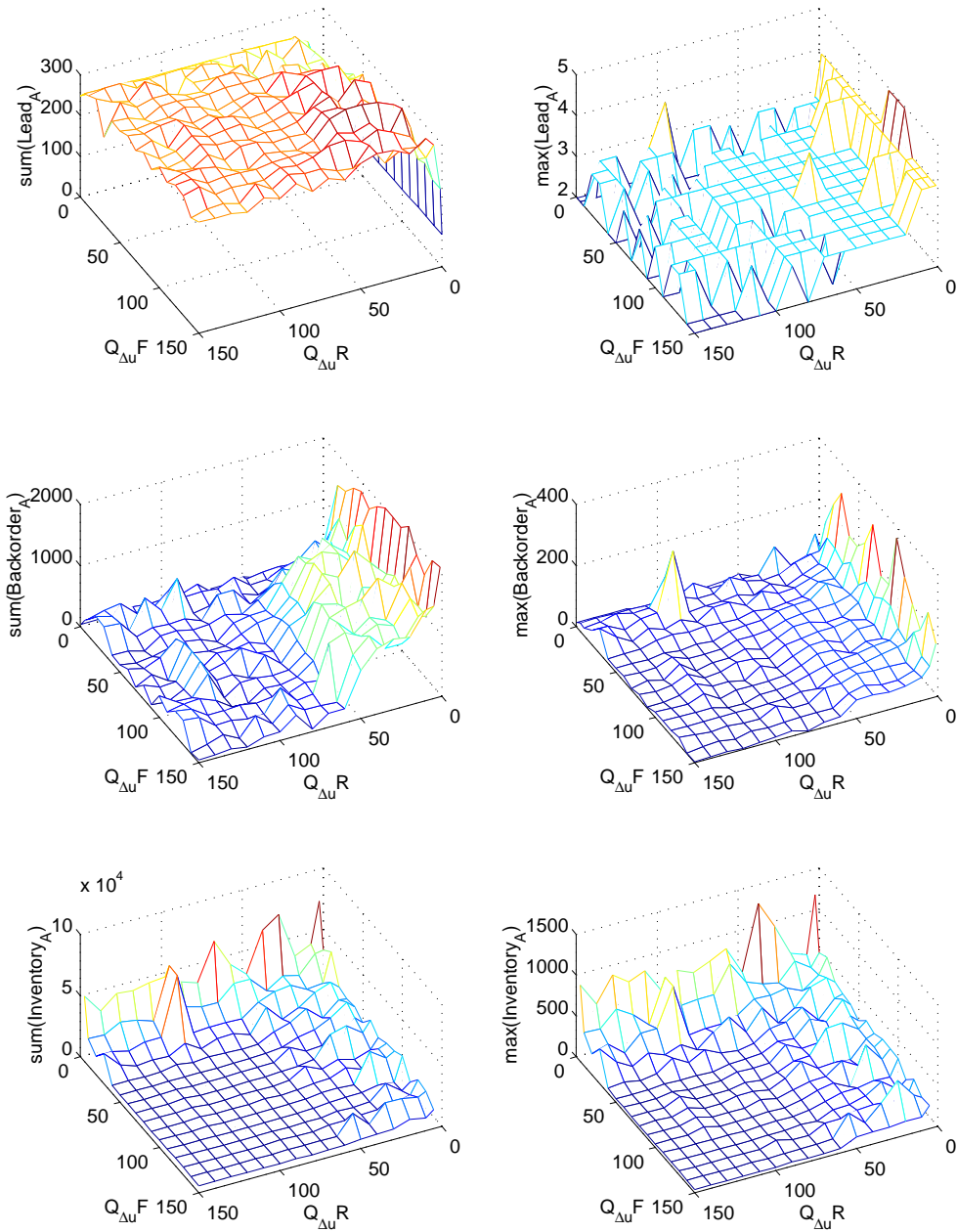


Figure 17: Factory metrics for move suppression interactions, under plant-model mismatch.

### 3 Six-node Network Management

In this section, the ability of an MPC-based approach to handle networks of industrially relevant size and topology is demonstrated. A six-node network suggested by researchers at Intel Corp. is used as a challenging testbed for robustness of the proposed approach. The need for robust solutions is motivated by the observations that inaccurate information is often provided between members of a supply chain, demand forecasts have limited accuracy, and the inherently long lead times in processing present a fundamental limit in the ability of a purely feedback only management policy. First, the six-node network is described, an MPC management policy is presented, and the following management scenarios/experiments are investigated:

- realistic demand pattern under demand forecast bias, and a plant-model mismatch for the assembly/test processing time
- constrained production, alternating demand pattern with plant-model mismatch for the assembly/test processing time
- investigation of the effect of safety stock levels on four different information sharing strategies

#### 3.1 Network Definition

Figure 1 illustrates the interconnections of the material flows for the six-node Intel Problem (Armbruster *et al.*, 2001). As noted in the diagram, transportation delays range from 2 to 4 days. The direct shipment routes each require 2 days. The cross-shipment routes require 3 days between Assembly/Test and Warehouse echelons. The cross-shipment routes require 4 days between Warehouse and Retailer echelons. In the Warehouse and Retailer nodes, material that enters the receiving dock does not show up in the inventory until the following day. The Assembly/Test nodes require an additional 10 days for processing.

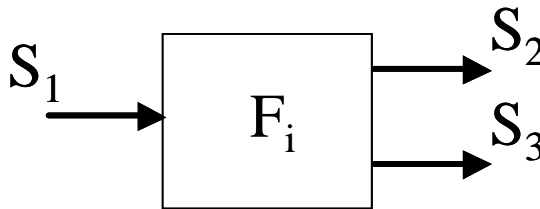


Figure 18: Single inflow, two outflow factory material streams.

For this demand network pattern the transportation delays can be modeled in the same way as was done with the 2-node supply chain. There are three new modeling entities to consider: Factories,

Geographic Warehouses, and Retailers. Each has a different configuration of input and output flows. For the single inflow, two outflow factory as shown in Figure 18, the material balances for each species can be written

$$I_A(k+1) = P_A(k) - S_{2A}(k) - S_{3A}(k) + I_A(k) \quad (13)$$

$$I_B(k+1) = P_B(k) - S_{2B}(k) - S_{3B}(k) + I_B(k) \quad (14)$$

with the constitutive relationships

$$P_A(k) = S_{1A}(k - \Theta_p) \quad (15)$$

$$P_B(k) = S_{1B}(k - \Theta_p) \quad (16)$$

These material balances are subject to the following constraints

$$I_A(k+1) + I_B(k+1) \leq C_I \quad (17)$$

$$S_{1A}(k) + S_{1B}(k) \leq C_1 \quad (18)$$

$$S_{2A}(k) + S_{3A}(k) + S_{2B}(k) + S_{3B}(k) \leq C_{23} \quad (19)$$

$$S_{2A}(k) + S_{2B}(k) \leq C_2 \quad (20)$$

$$S_{3A}(k) + S_{3B}(k) \leq C_3. \quad (21)$$

These constraints represent capacity limitations on inventory, shipping for stream #1, release rate from the Factory, shipping for stream #2, and shipping for stream #3, respectively.

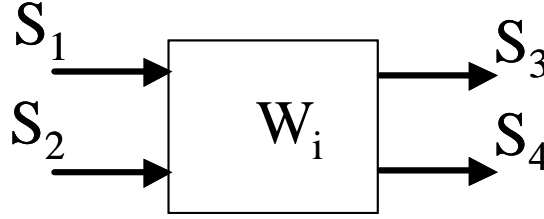


Figure 19: Two inflow, two outflow warehouse material streams.

The next entity to consider is the two inflow, two outflow warehouse as shown in Figure 19. The material balances of this system are a natural extension of the single inflow, two outflow material balance equations.

$$I_A(k+1) = S_{1A}(k) + S_{2A}(k) - S_{3A}(k) - S_{4A}(k) + I_A(k) \quad (22)$$

$$I_B(k+1) = S_{1B}(k) + S_{2B}(k) - S_{3B}(k) - S_{4B}(k) + I_B(k) \quad (23)$$

These material balances are subject to the following constraints

$$I_A(k+1) + I_B(k+1) \leq C_I \quad (24)$$

$$S_{1A}(k) + S_{1B}(k) \leq C_1 \quad (25)$$

$$S_{2A}(k) + S_{2B}(k) \leq C_2 \quad (26)$$

$$S_{1A}(k) + S_{2A}(k) + S_{1B}(k) + S_{2B}(k) \leq C_{12} \quad (27)$$

$$S_{3A}(k) + S_{3B}(k) \leq C_3 \quad (28)$$

$$S_{4A}(k) + S_{4B}(k) \leq C_4 \quad (29)$$

$$S_{3A}(k) + S_{4A}(k) + S_{3B}(k) + S_{4B}(k) \leq C_{34} \quad (30)$$

These constraints represent capacity limitations on inventory, shipping for stream #1, shipping for stream #2, total inflow rate, shipping for stream #3, shipping for stream #4, and release rate from the Warehouse, respectively.

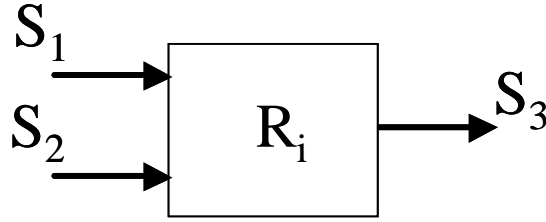


Figure 20: Two inflow, single outflow retailer material streams.

For the two inflow, single outflow Retailer of Figure 20 the material balances are written

$$I_A(k+1) = S_{1A}(k) + S_{2A}(k) - S_{3A}(k) + I_A(k) \quad (31)$$

$$I_B(k+1) = S_{1B}(k) + S_{2B}(k) - S_{3B}(k) + I_B(k). \quad (32)$$

These material balances are subject to the following constraints

$$I_A(k+1) + I_B(k+1) \leq C_I \quad (33)$$

$$S_{1A}(k) + S_{1B}(k) \leq C_1 \quad (34)$$

$$S_{2A}(k) + S_{2B}(k) \leq C_2 \quad (35)$$

$$S_{1A}(k) + S_{2A}(k) + S_{1B}(k) + S_{2B}(k) \leq C_{12} \quad (36)$$

$$S_{3A}(k) + S_{3B}(k) \leq C_3 \quad (37)$$

These constraints represent capacity limitations on inventory, shipping for stream #1, shipping for stream #2, total inflow rate, and shipping/release for stream #3 from the Retailer, respectively.

### 3.2 MPC For Management Of The Six Node Demand Network

Having investigated different facets of the demand network management problem with the two-node network, this insight can be applied to the six-node problem. The understanding of tuning parameter interaction can be used to readily tune the controllers employed in this problem. Figure 21 illustrates the information flow between the three MPC controllers and the six-nodes of the demand network. For this problem, one monolithic MPC controller would become computationally prohibitive. A fully decentralized structure (one controller for each node), would require intervention by an outer layer to manage the constraints due to the interconnections of the nodes, the release rate capacities, and the shipping capacities. Due to the nature of the constraints, a compromise can be made between the two strategies. Three controllers are used. Each controller handles the flow and composition decisions for each echelon, passing forecasted information to the upstream node and receiving forecasted information from the downstream node. By resolving the information flows in this manner the solution is scalable, yet still feasible for managing the constraints.

What makes this problem unique in comparison with the two node problem is the ability to cross-ship materials from the factory nodes to the warehouse nodes and from the warehouse nodes to the retailer nodes. This adds extra degrees of freedom to the first and second echelon controllers. Using the ideas from the MPC configuration of the two node problem, the six-node problem can be translated into process control variables according to Table 3. The 2<sup>nd</sup> echelon controller mapping is used as an example.

The way the extra degrees of freedom are handled is through the use of the  $u(k + \ell - 1|k) - u_{target}(k + \ell - 1|k)$  penalty term of the MPC cost function.  $u_{target}$  is set to 0 for all  $k$  and  $\ell$ . Orders placed for the direct shipment routes receive a different penalty for being other than zero, than the cross shipment routes. This allows the problem to have a unique solution and it allows the user to specify a preference of shipping routes, based on cost or preference. For most of this work, the cross-shipment routes will be penalized with a much higher weight than the direct shipment routes. Thus, the cross-shipment routes are used only when necessary to deal with shipping or release constraints. Material is thereby forced down the quicker direct-shipment routes, with the goal of improving agility of the overall system.

Constraints on the shipping and release from each node in an echelon are formulated in such a way that they become extra constraints to be satisfied by the inputs or decision variables of the downstream controller. In this manner, they can be dealt with as hard constraints. As an example,

Table 3: MPC variable mapping for six-node, 2<sup>nd</sup> echelon controller.

Process Control Variable	Demand Network Information
setpoints $r$	inventory targets of species A & B for Warehouses 1 & 2
outputs $y$	inventories minus cumulative backorders of species A & B in Warehouses 1 & 2
estimated outputs $\hat{y}$	estimated inventories of species A & B in Warehouses 1 & 2
measured disturbances $u_d$	actual shipments of A & B to Retailers 1 & 2
estimated future measured disturbances $\hat{u}_d$	estimated demand or orders for species A & B being placed by Echelon #1 MPC
estimated inputs $\hat{u}$	estimated orders for species A & B being placed at Factories #1 & #2 for all four routes from Echelon #3 to #2
inputs $u$	orders for species A & B being being placed at Factories #1 & #2 for all four routes from Echelon #3 to #2

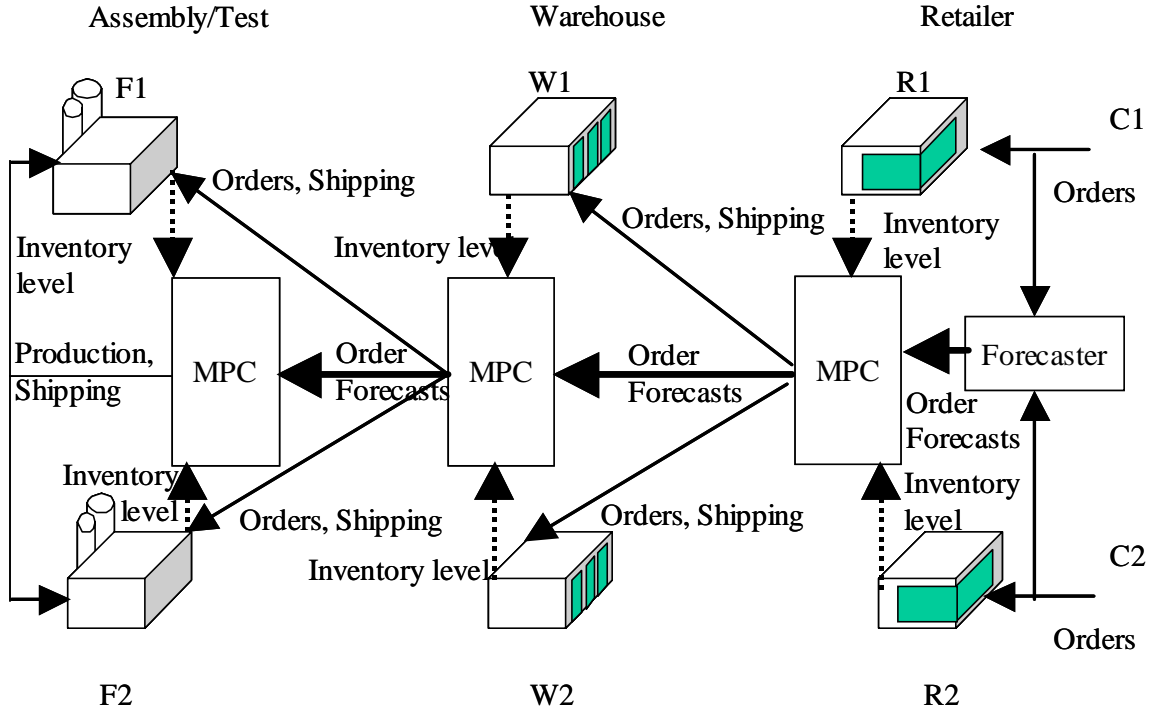


Figure 21: Information flow for control of the six-node Intel Demand Network.

the second echelon MPC controller would be required to satisfy

$$u_{2A}^{F1}(k) + u_{3A}^{F1}(k) + u_{2B}^{F1}(k) + u_{3B}^{F1}(k) \leq C_{23}^{F1} \quad (38)$$

$$u_{2A}^{F1}(k) + u_{2B}^{F1}(k) \leq C_2^{F1} \quad (39)$$

$$u_{3A}^{F1}(k) + u_{3B}^{F1}(k) \leq C_3^{F1} \quad (40)$$

$$u_{2A}^{F2}(k) + u_{3A}^{F2}(k) + u_{2B}^{F2}(k) + u_{3B}^{F2}(k) \leq C_{23}^{F2} \quad (41)$$

$$u_{2A}^{F2}(k) + u_{2B}^{F2}(k) \leq C_2^{F2} \quad (42)$$

$$u_{3A}^{F2}(k) + u_{3B}^{F2}(k) \leq C_3^{F2}, \quad (43)$$

where  $u_{ij}^{F\{k\}}$  represents the order placed to Factory  $k$ , for product  $j$  via shipment route  $i$ . Constraint  $C_{23}^{F\{k\}}$  represents the release capacity of Factory  $k$ .  $C_i^{F\{k\}}$  denotes the shipping capacity of route  $i$  from Factory  $k$ .

In the subsections that follow, the ability of this MPC framework to deal with active constraints, plant-model mismatch, and forecast bias will be illustrated. Subsequently, various information sharing configurations will be examined to 1.) demonstrate the flexibility of this framework to adapt in different ways to available information flows, 2.) support the conclusions regarding the benefits of information sharing in the qualitative supply chain research literature and 3.) verify the

assumption that the partially decentralized MPC topology does not amplify demand variability via order processing.

### 3.3 Performance Under Plant-Model Mismatch And Biased Forecast Error

The robustness of the six node management framework will now be demonstrated using realistic plant-model mismatch for the Assembly/Test Nodes, biased forecast error, and a realistic demand profile. These conditions were recommended for evaluation by Intel Corp. No constraints are active for this case.

Experience with the actual performance of Assembly/Test nodes and the estimated lead times by facility personnel, suggest that the facility personnel traditionally provide themselves a lead time buffer of one day. So for example, if in reality the process takes 9 days to complete, a 10 day lead time estimate will be quoted to others in the organization. Thus, a 9 day actual/10 day estimate plant-model mismatch is adopted for both Assembly/Test facilities.

The Sales and Marketing personnel have been generally known to determine forecasts which are biased in an optimistic manner. As an example, the sales forecast for the next time period might be 11,000 units, when in fact the actual sales will be 10,000 units. To mimic this type of forecast bias, all demand forecasts passed to the Retailer level MPC controller are biased by +1000 units.

Lastly, products have been observed to follow demand patterns which may be correlated at times, and uncorrelated at other times. To mimic this type of behavior, the demand patterns for Product A and Product B follow correlated, deterministic steps up until time 110. The remainder of the time, the demand patterns remain uncorrelated. This behavior can be observed in Figure 22.

The experiment is run with a value of 300 for all move suppression parameters  $Q_{\Delta u}$ , 0 for all penalties for values other than zero for direct shipment  $Q_u$ , and a penalty of 100 for all penalties for values other than zero for cross-shipments  $Q_u$ . Table 4 holds the prediction horizons  $N_p$  and control horizons  $N_u$  for all three MPC controllers. These horizon parameters are used for every six node management experiment in this document. Safety stock is set at 5,000 units per product, per node. All inventory control error weights  $Q_e$  are left at 1. All entities use a “pecking order” dispatch rule. The orders of Warehouse 1, Retailer 1, and Customer 1 take precedence over the orders of the corresponding counterparts, should shortages or constraints be encountered. Figures 23 and 24 demonstrate the performance of this approach with plots of inventories, demands, shipments, and factory starts.

At time 1, the Retailer MPC controller adjust orders and order forecasts to the upstream nodes to start bringing in more product, since the demand forecast is now 11,000 even though the actual amounts being demanded are 10,000. Because of the move suppression, the increases in order amounts are less than 1,000. The increase in orders is evident in the increase in direct and cross shipments from the Warehouse echelon. Soon the Retailer MPC echelon observes via the inventory

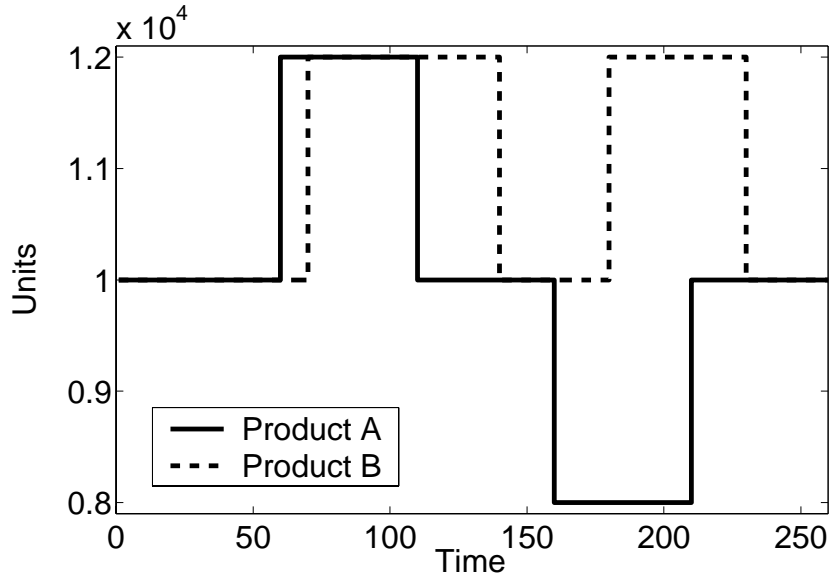


Figure 22: Qualitatively realistic demand profile for products A and B.

Table 4: Prediction and control horizons used in six node controllers.

Parameter	Echelon #1	Echelon #2	Echelon #3
$N_p$	49	41	33
$N_u$	42	34	12

levels that the actual amount supplied to customers is *not* increasing as suggested by the forecast and the Retailer MPC adjusts to account for the forecast error and reduce inventory.

In the first few time units, inventories in the Factory echelon and Warehouse echelon are drained below their target levels. The Factory echelon MPC controller then observes the effects of the plant-model mismatch, since changes in starts show up sooner than expected. The inventory levels of the Assembly/Test nodes fluctuates, but the fluctuation remains at reasonable levels. No backorders take place throughout the entire experiment. This is rather impressive, since the general rule of thumb practiced for this network requires a safety stock level equivalent to between two and four times the expected demand for the next time period (e. g. if tomorrow’s demand is expected to be 10,000 units, safety stock held today may range from 20,000 to 40,000 units).

Note that the cross-shipments in this example are used whenever there are rapid changes in the order/demand forecasts. The MPC controllers make use of the cross-shipments, since the costs associated with the move suppression weightings of the direct shipments become comparatively large

at these times. This may make sense not only from an optimization standpoint, but in a realistic setting it may also allow nodes to hedge against uncertainties or disturbances in transportation links or nodes connected by the direct shipment lanes.

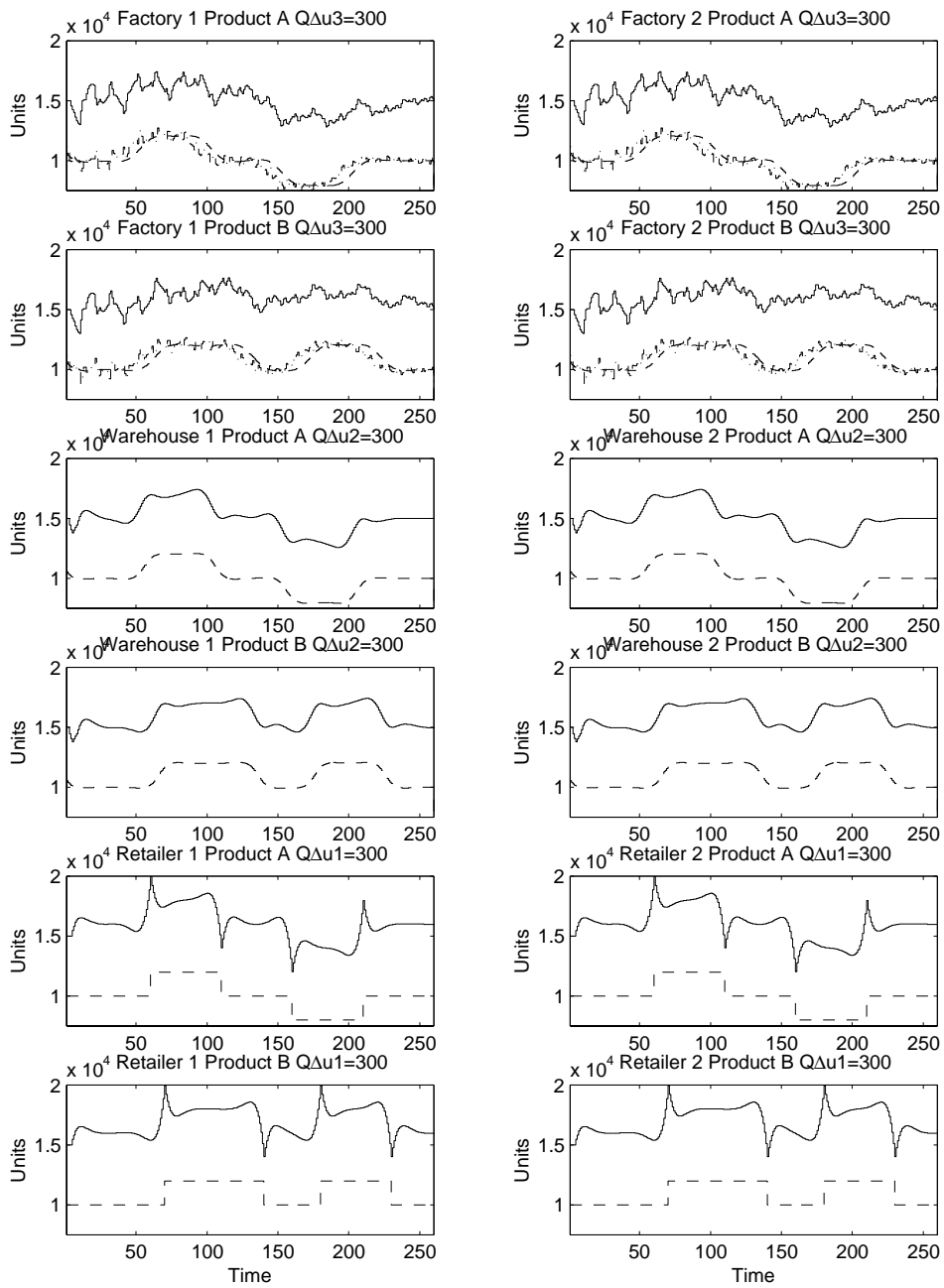


Figure 23: Plant-model mismatch & forecast error: Inventories (solid) and demand (dashed) by facility and product; Assembly/Test plots include starts (dash-dot).

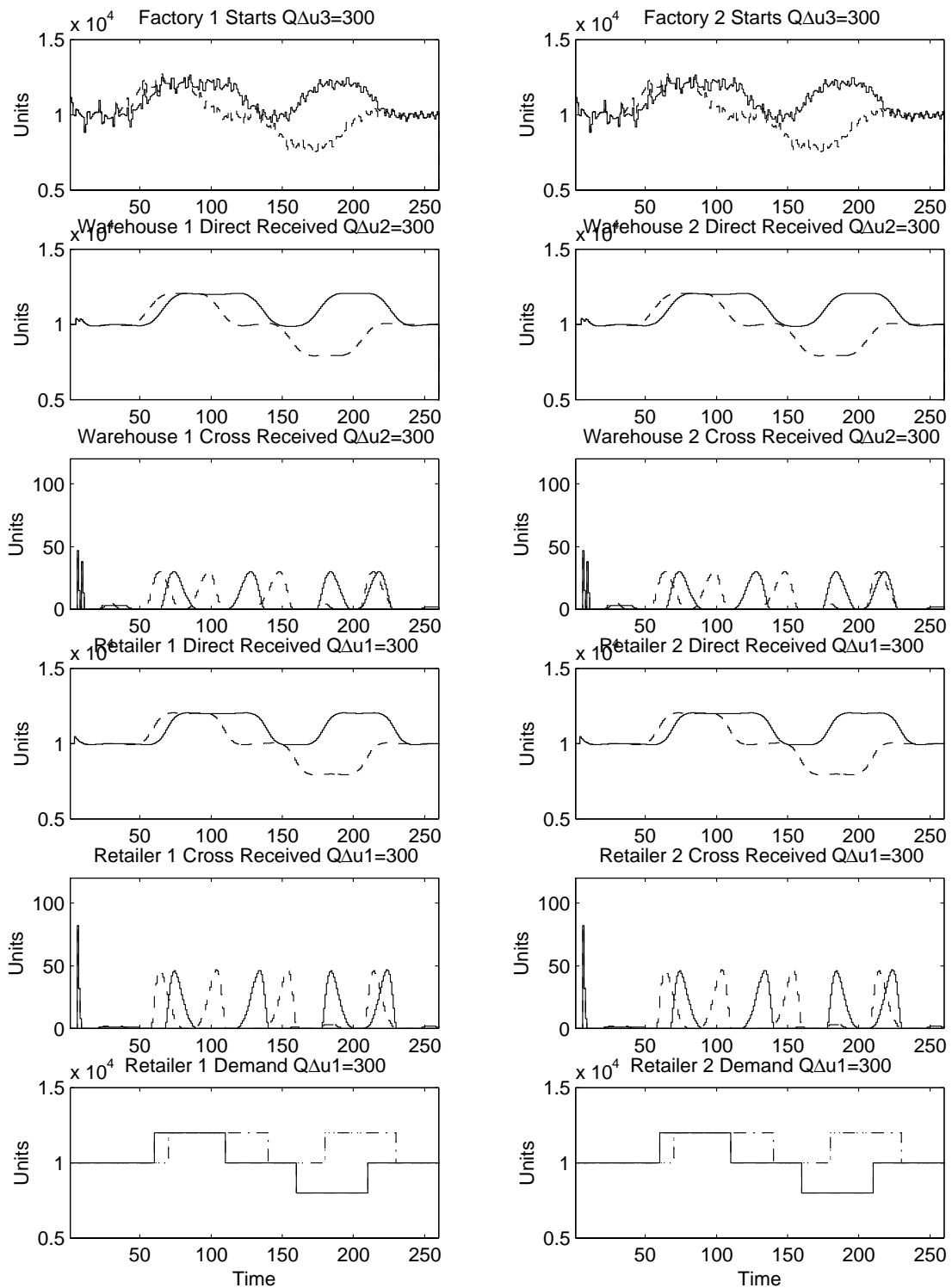


Figure 24: Plant-model mismatch & forecast error: Material flows Total (solid) and Product A (dashed); Demand Product A (dotted) and Total (dash-dot) and corresponding customer receipts Product A (dashed) and Total (solid).

### 3.4 Performance Under Constrained Factory And Warehouse Release

The ability to handle input and output constraints has long been noted as a major benefit of employing an MPC control approach in the process industries. In this example, the six node demand network is operated with constraints of 23,000 and 24,000 total units on the release rates of each of the Assembly/Test and Warehouse nodes, respectively. An alternating step demand pattern is used to determine whether the MPC framework can anticipate the large demand changes and bring additional material through to the Retailers via the cross-shipment routes. The pattern alternates between 20,000 and 26,000 units total for each retailer.

Figures 25 and 26 present the inventories, factory starts, shipments, and demand patterns for this example. The safety stock is set at 5000 units per product, per node. There is no error in the demand forecasts, however the “pessimistic” plant-model mismatch is used in the Assembly/Test nodes. All move suppression values are set at 300; values of 0.001 were used for all penalties for the input targets for direct shipment  $Q_u$ , and penalties of 0.1 were used for input target penalties for cross-shipments  $Q_u$ . All inventory control error weights  $Q_e$  are left at 1. All nodes make use of a “share the pain” release policy. With this release policy, all orders are shorted an equal percentage if inventory is short, or if constraints are active.

These results illustrate the ability of the MPC approach to make good use of the cross-shipment routes available between both echelons to circumnavigate the alternating bottlenecks in the system that result in the direct shipment paths. These alternating bottlenecks in the direct shipment lanes result because of the constraints as well as the alternating demand at the Retailers. Both Assembly/Test facilities run at capacity for the entire duration of the alternating demand. Moreover, no backorders are placed over the entire duration of the experiment.

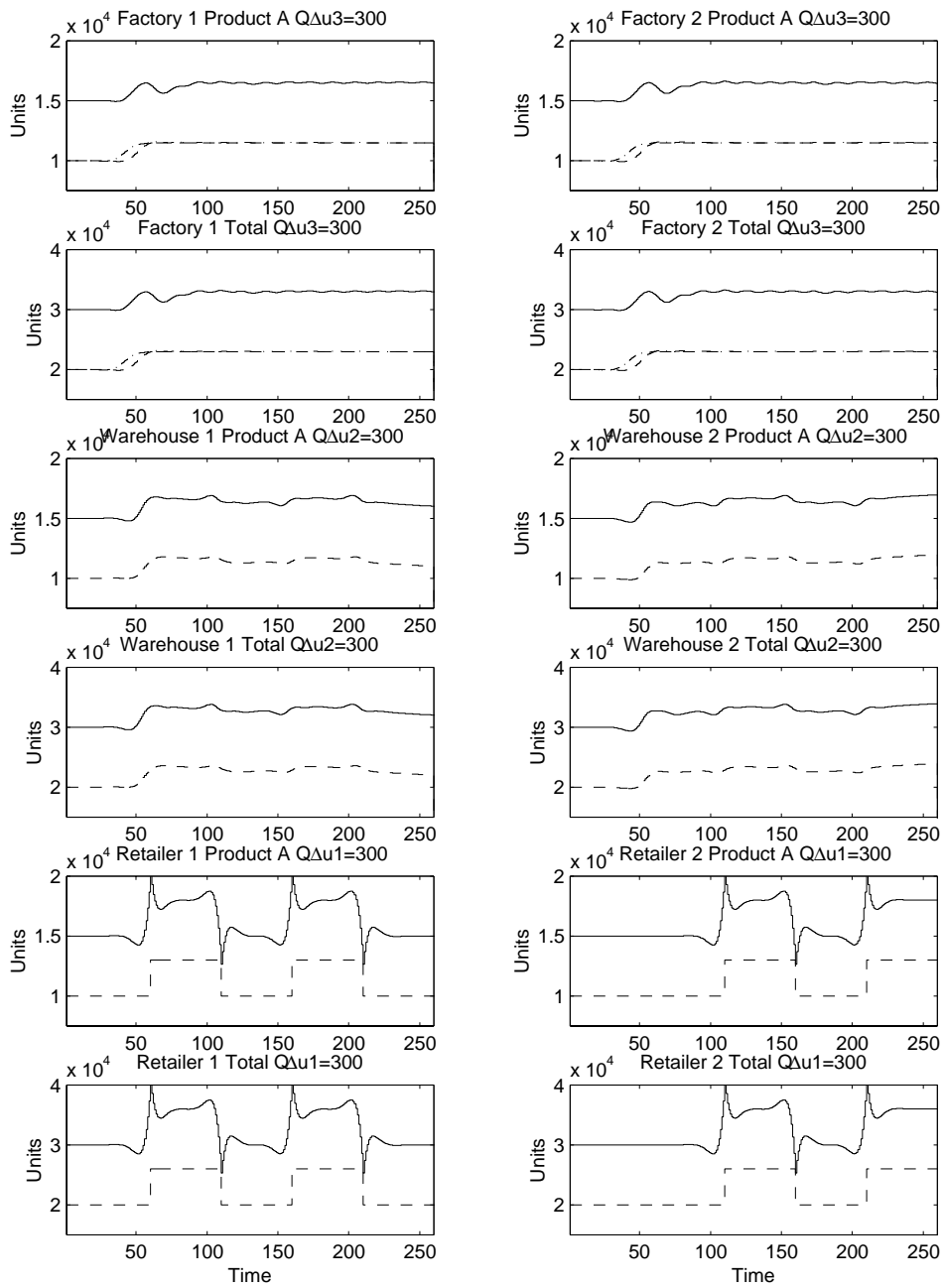


Figure 25: Constrained factories & warehouses: Inventories (solid) and demand (dashed) by facility and product; Assembly/Test plots include starts (dash-dot).

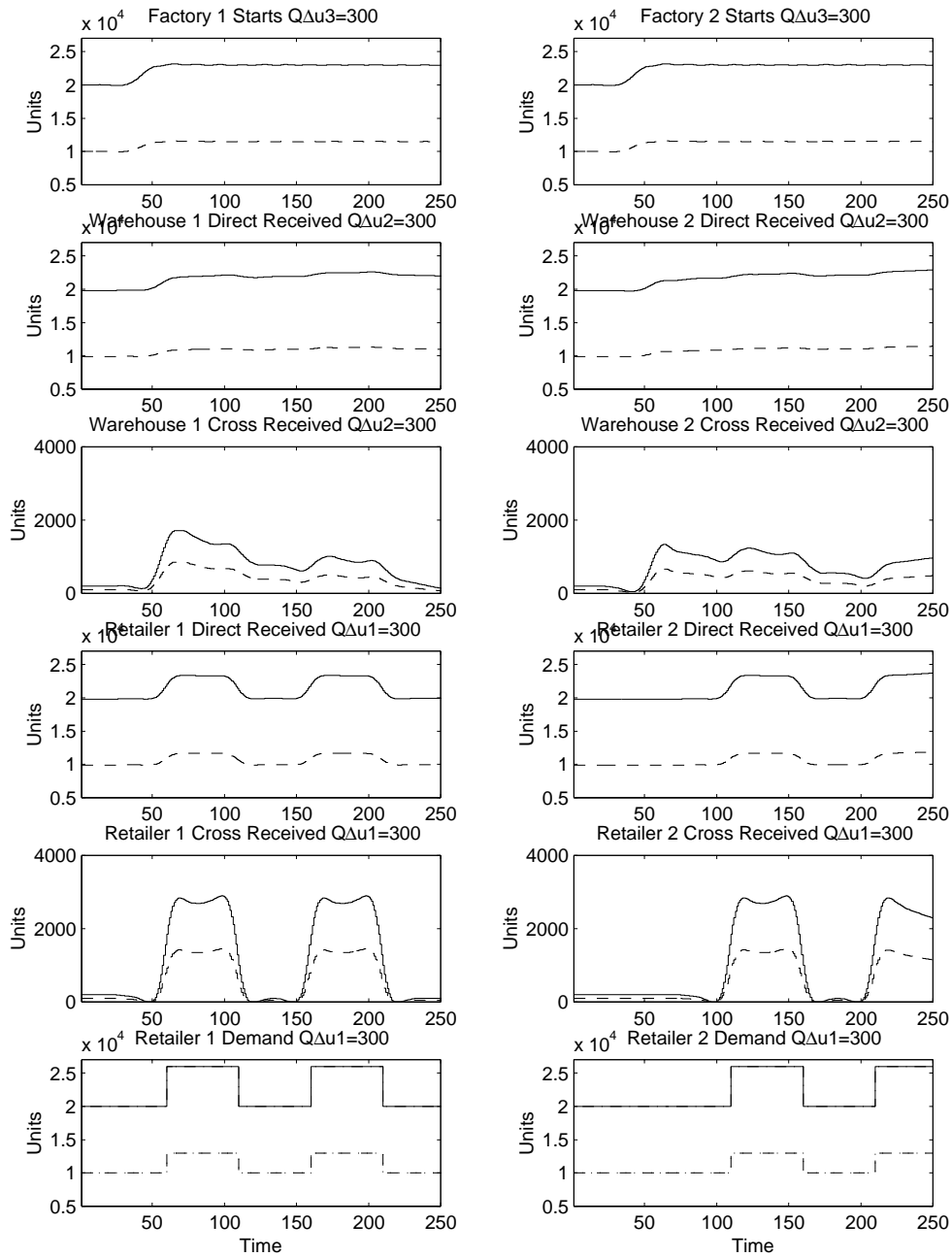


Figure 26: Constrained factories & warehouses: Material flows Total (solid) and Product A (dashed); Demand Product A (dotted) and Total (dash-dot) and corresponding customer receipts Product A (dashed) and Total (solid).

### 3.5 Information Sharing Strategy Analysis

In a recent study and survey by PriceWaterhouseCoopers (2000), it was made clear that information sharing is a critical issue for the improvement of supply chains in the semiconductor industry. Because of lead times on the order of months, it is easy to see how information sharing is critical for this industry. The PwC report suggests that a lack of visibility across the supply chain is perpetuating supply and demand mismatches. Based on results of the survey, PwC advocates sharing realtime forecasts, demand, and production information across the entire value web. The respondents of the survey advocate information sharing as well, since in the year 2000 only 30% of the respondents were sharing demand history and forecasts. It was reported that 70% of the respondents hoped to be sharing this information in 2001. Moreover, PwC recommend the implementation of inventory modeling and buffering programs to optimize inventory levels and dampen the supply and demand variability. The survey was completed by top executives in multi-national semiconductor companies. Over 50% of the the top 10 semiconductor companies were represented.

In this subsection, the six node demand network will be used to demonstrate the improvements brought about by information sharing of forecasts. This comparison will investigate different ways to share forecasts in the MPC framework. Four possible information sharing strategies will be investigated under plant-model mismatch and forecast bias. The first strategy (strategy #1) represents the minimum amount of information sharing possible. No forecasts are provided to upstream MPC controllers; only the current orders are provided. This level of information sharing is depicted<sup>2</sup> in Figure 27. The second information sharing strategy which is shown in Figure 28 is the same as that discussed in the beginning of this section. Each MPC controller shares its order forecast with the upstream node. The information flows for strategies #3 and #4 are shown in Figures 29 and 30. The third information sharing strategy shares the order forecast of the Retailer level MPC controller with both controllers upstream. This and strategy #4 are tested alongside strategy #2 to determine whether demand distortion is taking place as orders are transferred upstream. If demand distortion is taking place, strategies #3 and #4 would be expected to demonstrate improved performance over strategy #2. Strategy #4, is a natural continuation of the concept behind strategy #3. In this strategy, no filtering of orders takes place. All controllers in the network see the demand exactly as it is expected to happen from the Retailers point of view. Note that in all four strategies, the information sharing is appropriately time shifted to be accurate in the effect on echelon inventory models.

All information sharing strategies are tested under exactly the same simulation and controller parameter settings. The safety stock is initially set at 5000 units per product, per node. All move suppression values are set at 300; values of 0 were used for all penalties for the input targets for

---

<sup>2</sup>For clarity, other information flows such as orders to nodes, inventory levels, etc. are left out of the figure but are still implemented in the simulation.

direct shipment  $Q_u$ , and penalties of 100 were used for input target penalties for cross-shipments  $Q_u$ . All inventory control error weights  $Q_e$  are left at 1. All demand forecasts passed to the Retailer level MPC controller are biased by +1000 units. A 9 day actual/10 day estimate plant-model mismatch is used for both Assembly/Test facilities. All entities use a “pecking order” dispatch rule. The orders of Warehouse 1, Retailer 1, and Customer 1 take precedence over the orders of the corresponding counterparts.

Figures 27 through 38 present the demand, orders, starts, and inventory levels for all four information sharing strategies. By inspection of these figures it is observed that the sharing of order forecasts in combination with move suppression offers a huge benefit in reducing inventory levels, smoothing of order and inventory fluctuations, while at the same time providing better customer service. It should be noted that the combination of information sharing to provide feedforward information to the controller, and move suppression to temper the feedback action of the controller is required for the approach to demonstrate this level of performance. Without move suppression in the presence of plant-model mismatch or forecast error, poor performance is observed even with the sharing of order forecasts throughout the network.

In information sharing strategy #1 the MPC controllers are behaving in a purely feedback mode. Since lead times in transportation and production are significant, it takes time for the chain to respond to the change in demand quantity. When stockouts and backorders occur, the controllers respond by increasing the order amounts to try to compensate for the negative net stock levels which are now far below the inventory targets. For strategy #1, stockouts introduce the additional disturbances introduced by the pecking order release policies enforced in all six nodes. From a control perspective, the delays in the system fundamentally limit the achievable performance of a purely feedback strategy. Hence, the need for feedforward information as demonstrated with strategies #2 and #3. The presence of nonminimum phase elements (i.e. delays, inverse response) in the system dynamics require feedforward information use. This is well-established in the process control literature (Morari and Zafriou, 1989; Ogunnaike and Ray, 1994). Research in supply chain management has yet to fully recognize the value of the systems perspective, and the fundamental insights this perspective brings to the problem.

Both strategies #2 and #3 are able to anticipate the changes in demand and increase or decrease production and shipping levels appropriately. The variance in the performance of these two strategies comes from the rejection of the optimistic forecast bias by the controllers in combination with the pessimistic plant-model mismatch. At initial time, the retailer level controller notes the forecast bias and increases the order amounts to the warehouse level. The warehouse controller in turn increases the order amounts to the factory level. This causes plant-model mismatch to now be observable by the factory level controller. In a few time steps, the retailer level determines the

demand has not increased as suggested by the forecast. The controller begins integrating out the effects of the forecast bias and readjusts order levels accordingly.

From a qualitative standpoint, there is essentially no difference between the performance of information sharing strategies #2 and #3. This suggests that strategy #2 is not artificially amplifying the demand information as this information is passed upstream. This observation justifies the partially decentralized MPC management structure that has been adopted. These observations suggest that controllers in strategies #2 and #3 are well designed and tuned so that the resulting closed-loop speed of response of the controllers is nearly the same. The first echelon MPC controller filters out the high frequency contents of the step demand forecast and the second echelon MPC does very little additional processing of the forecast as it is passed on to the third echelon MPC controller. The control strategy therefore exhibits the properties of a good control system: good low frequency setpoint tracking, good low frequency disturbance rejection, and good high frequency noise attenuation. The performance of information sharing strategy #4 is not on par with information sharing strategy #2 and #3 partly because the first Echelon MPC controller serves as a filter for the discontinuity of the demand forecast. This discontinuity is the cause for the peaks in inventory level observed in Figure 37.

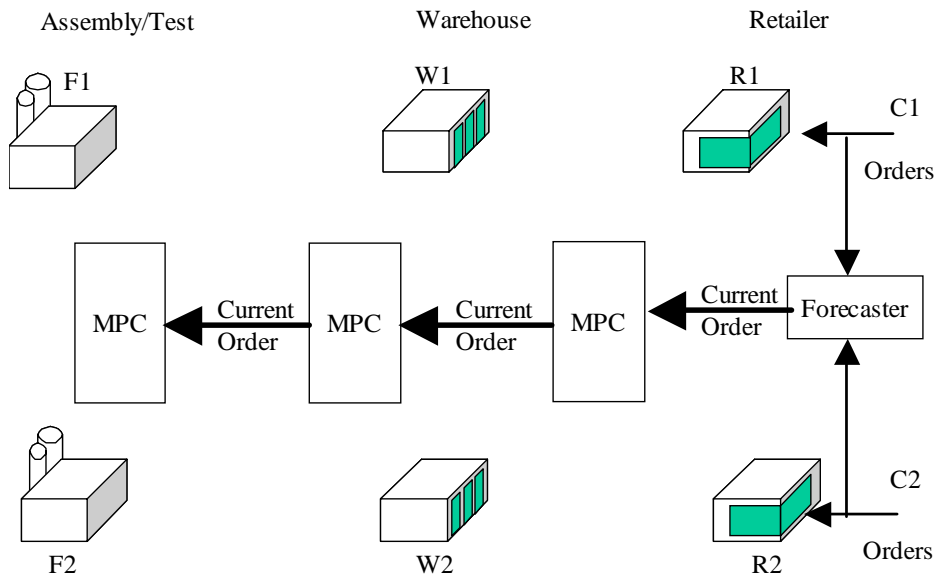


Figure 27: Information sharing strategy #1 for the six-node Intel Demand Network

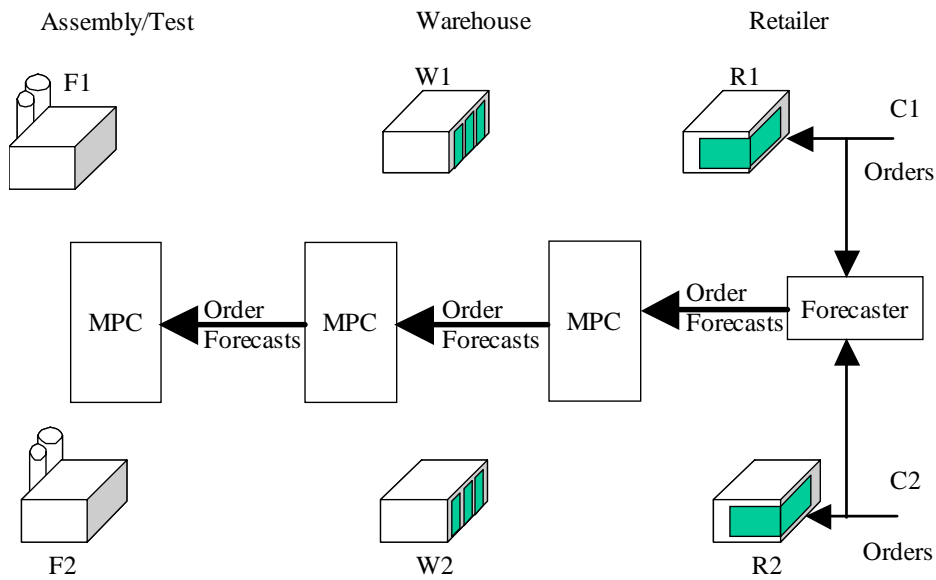


Figure 28: Information sharing strategy #2 for the six-node Intel Demand Network

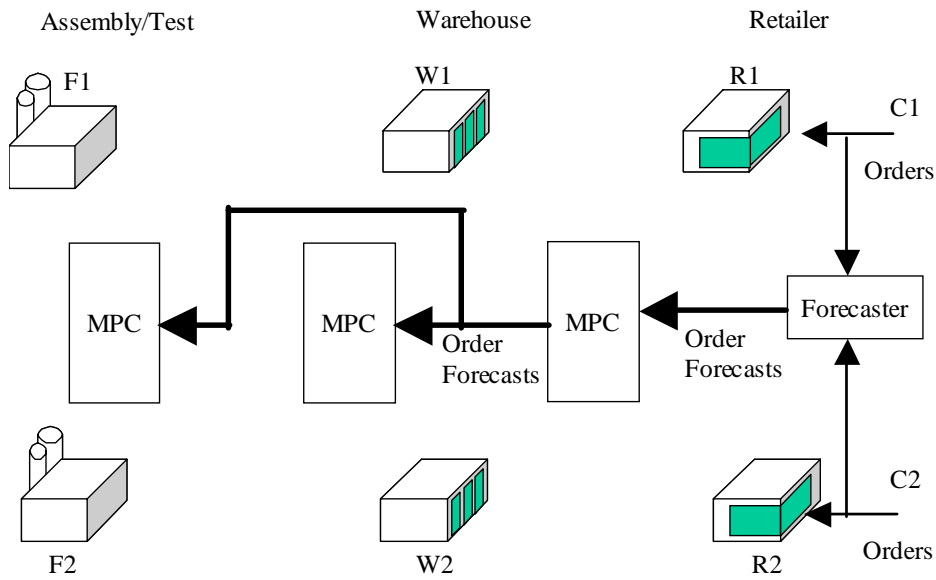


Figure 29: Information sharing strategy #3 for the six-node Intel Demand Network

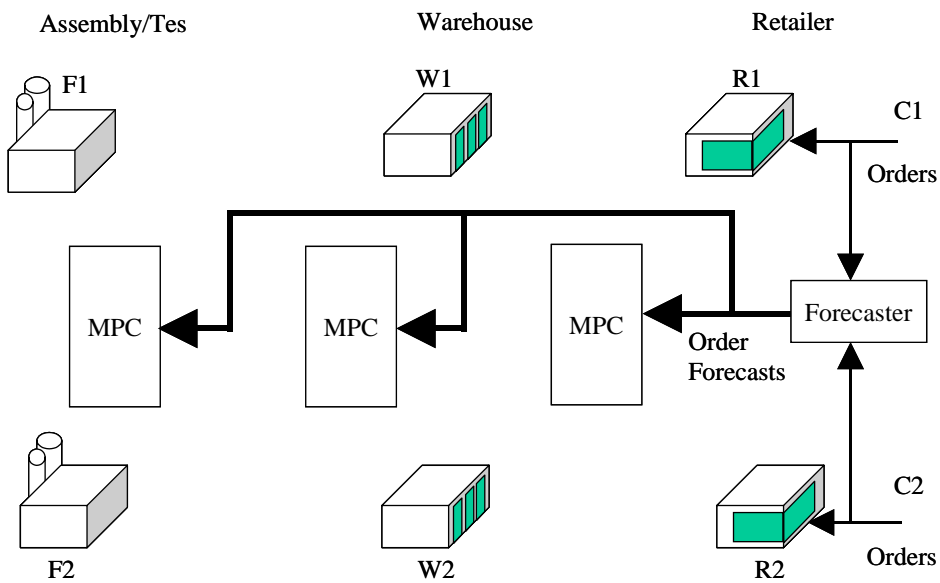


Figure 30: Information sharing strategy #4 for the six-node Intel Demand Network

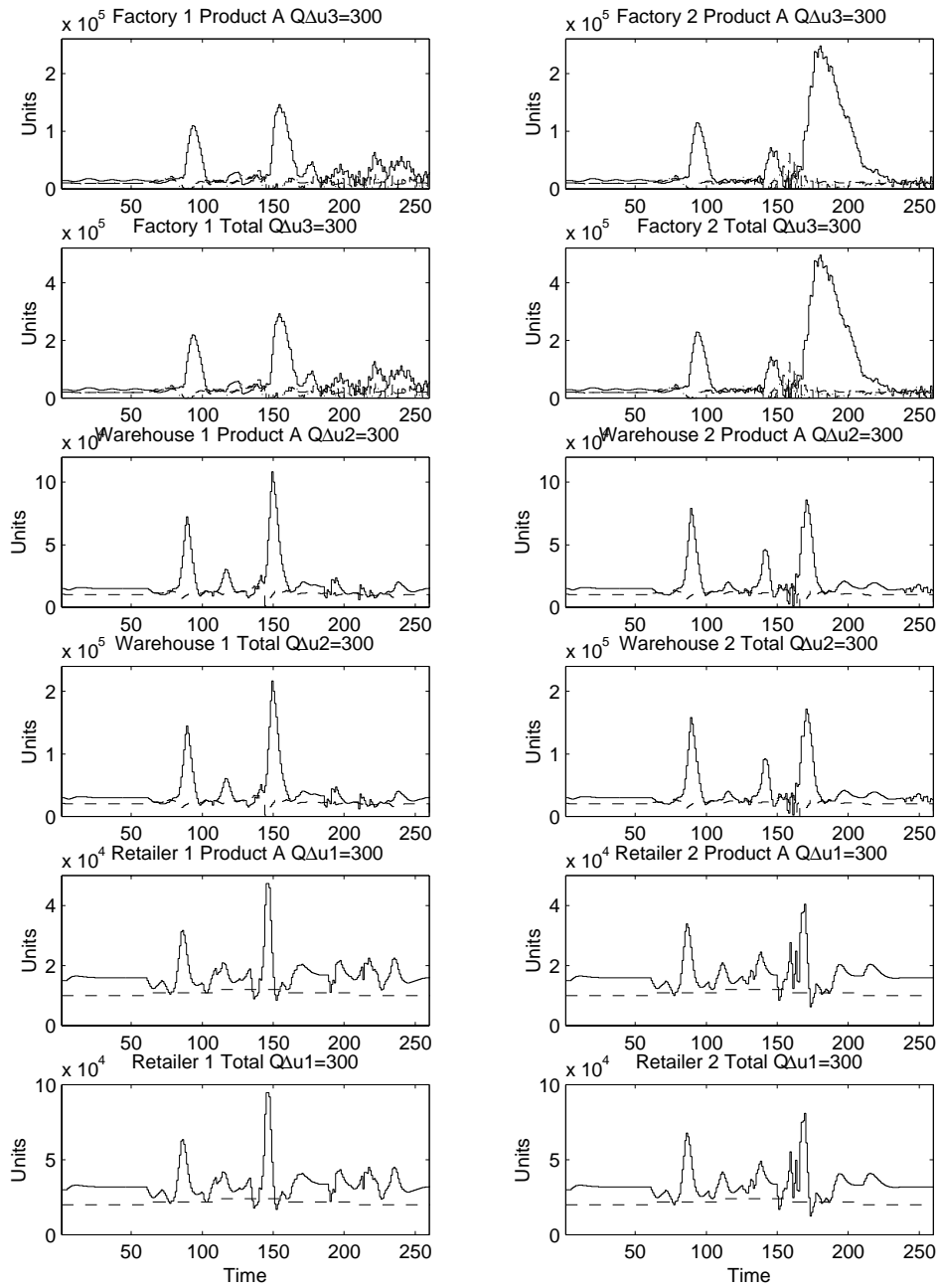


Figure 31: Information sharing strategy #1: Inventories (solid) and demand (dashed) by facility and product; Assembly/Test plots include starts (dash-dot).

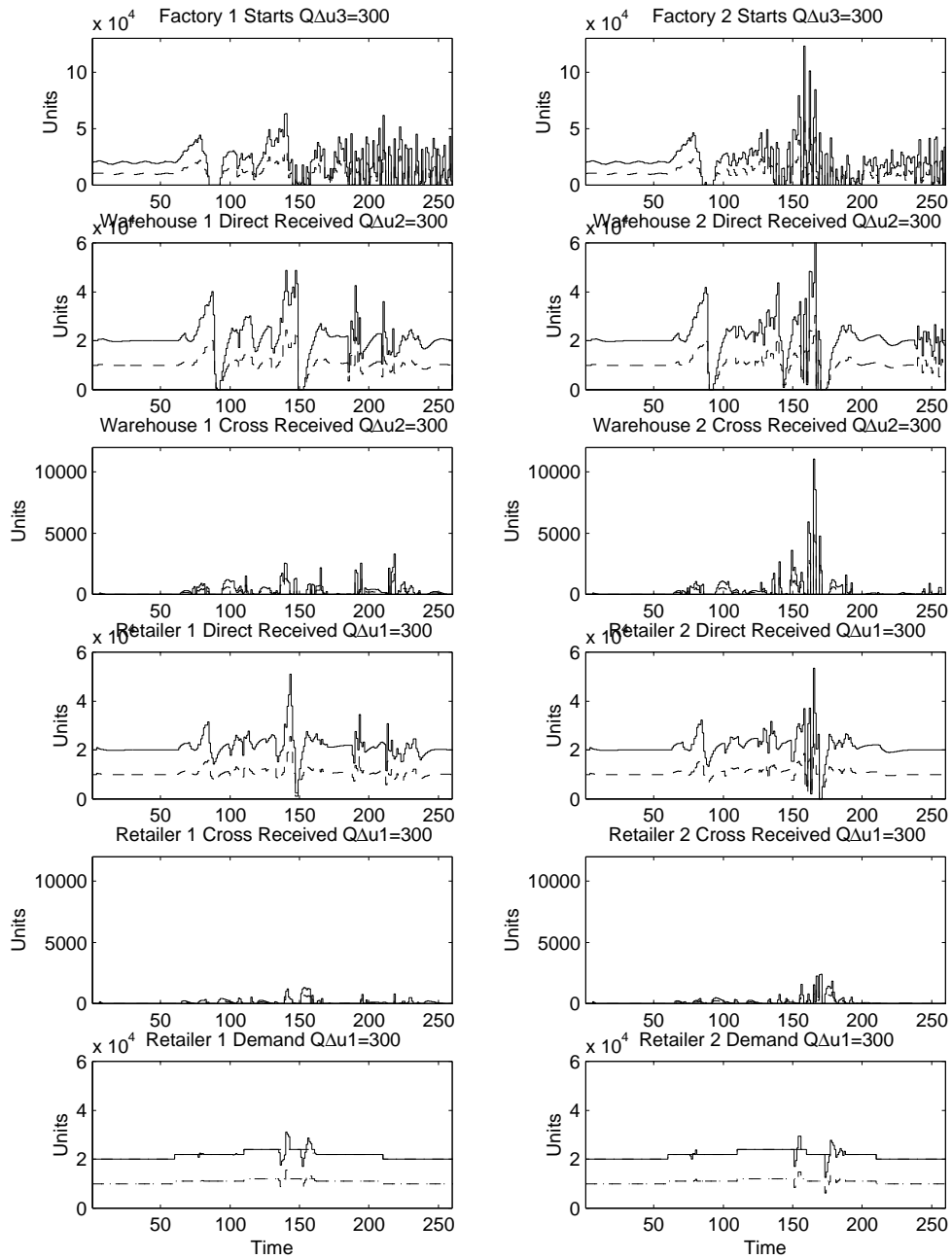


Figure 32: Information sharing strategy #1: Material flows Total (solid) and Product A (dashed); Demand Product A (dotted) and Total (dash-dot) and corresponding customer receipts Product A (dashed) and Total (solid).

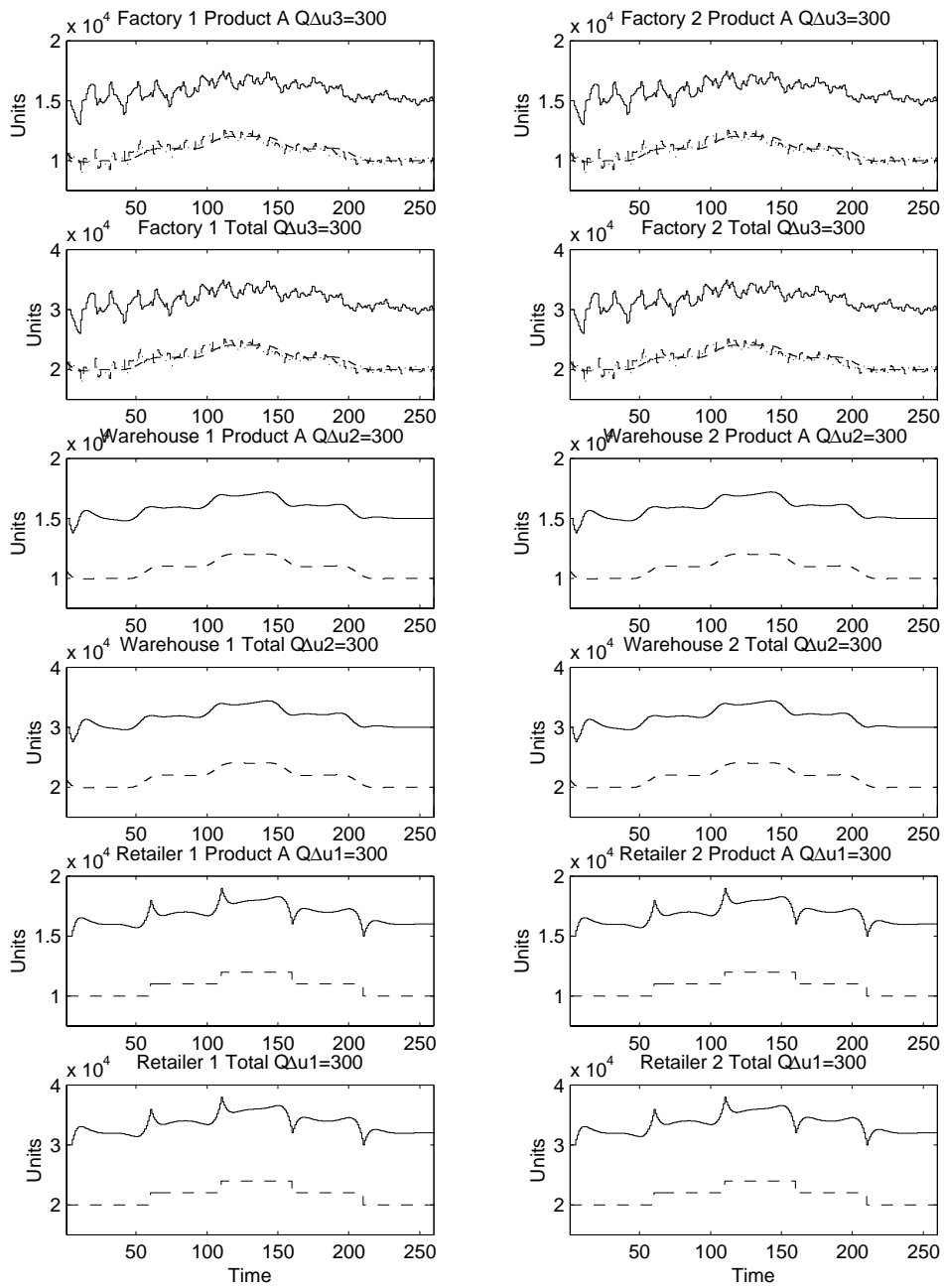


Figure 33: Information sharing strategy #2: Inventories (solid) and demand (dashed) by facility and product; Assembly/Test plots include starts (dash-dot).

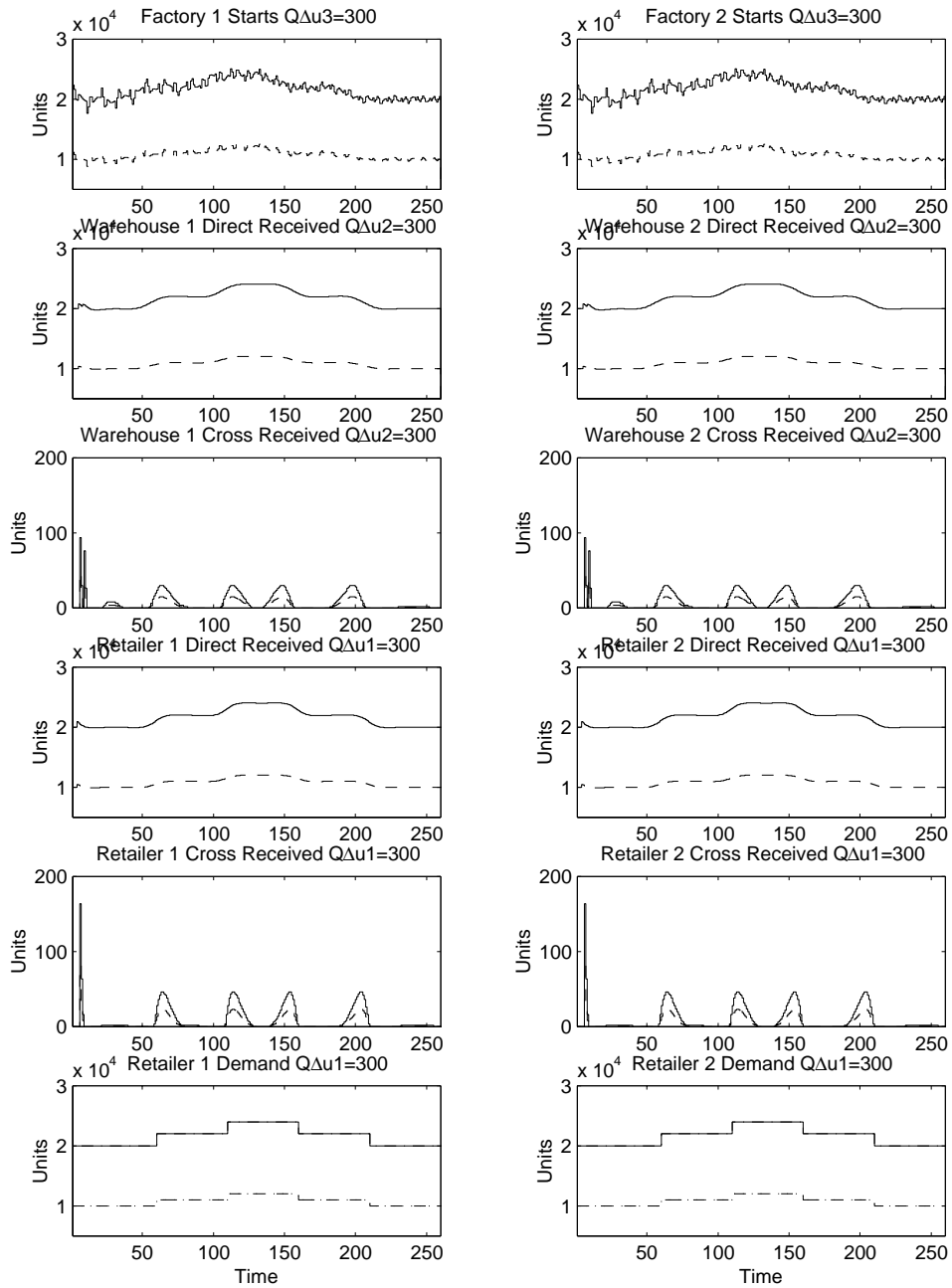


Figure 34: Information sharing strategy #2: Material flows Total (solid) and Product A (dashed); Demand Product A (dotted) and Total (dash-dot) and corresponding customer receipts Product A (dashed) and Total (solid).

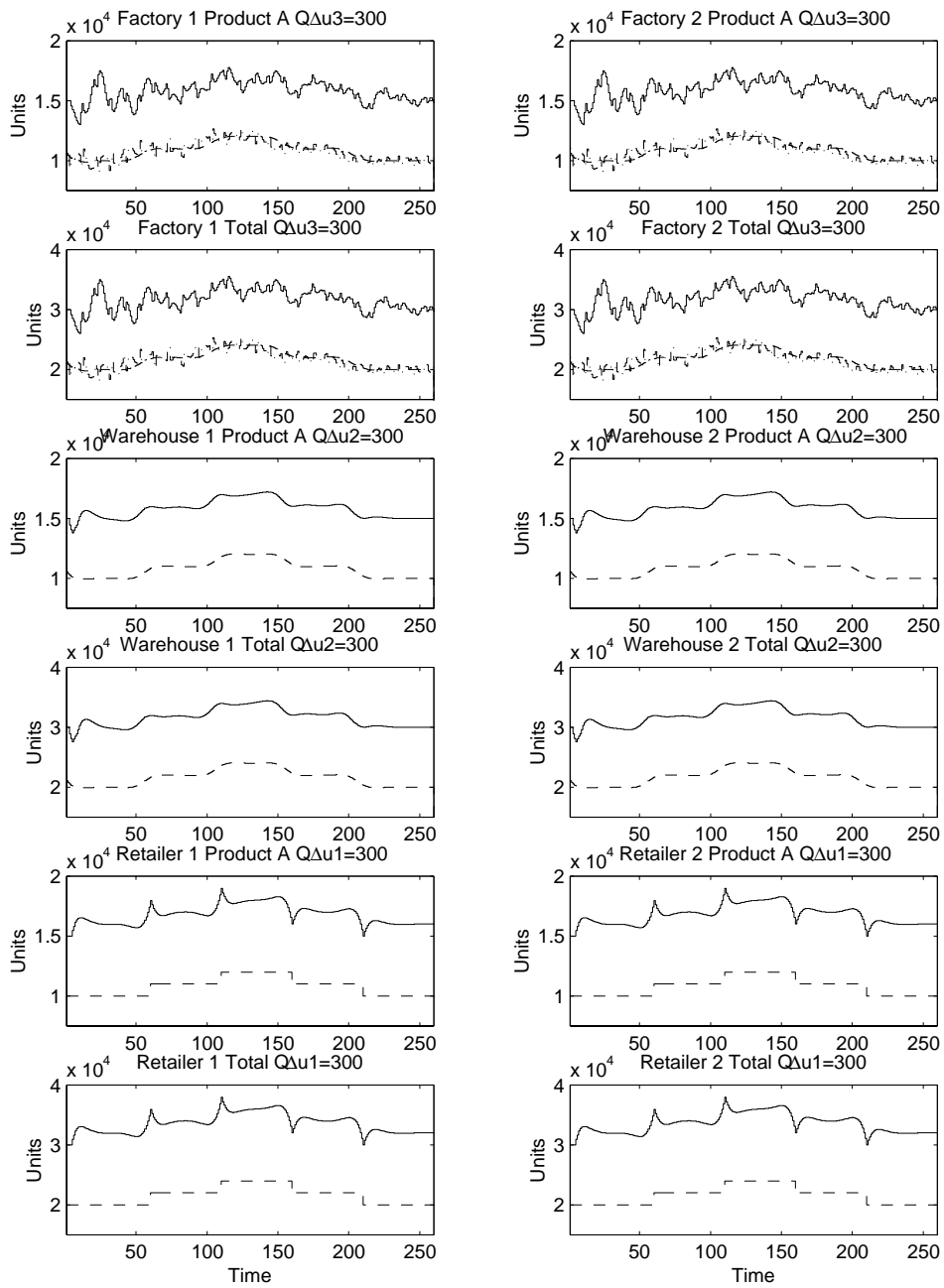


Figure 35: Information sharing strategy #3: Inventories (solid) and demand (dashed) by facility and product; Assembly/Test plots include starts (dash-dot).

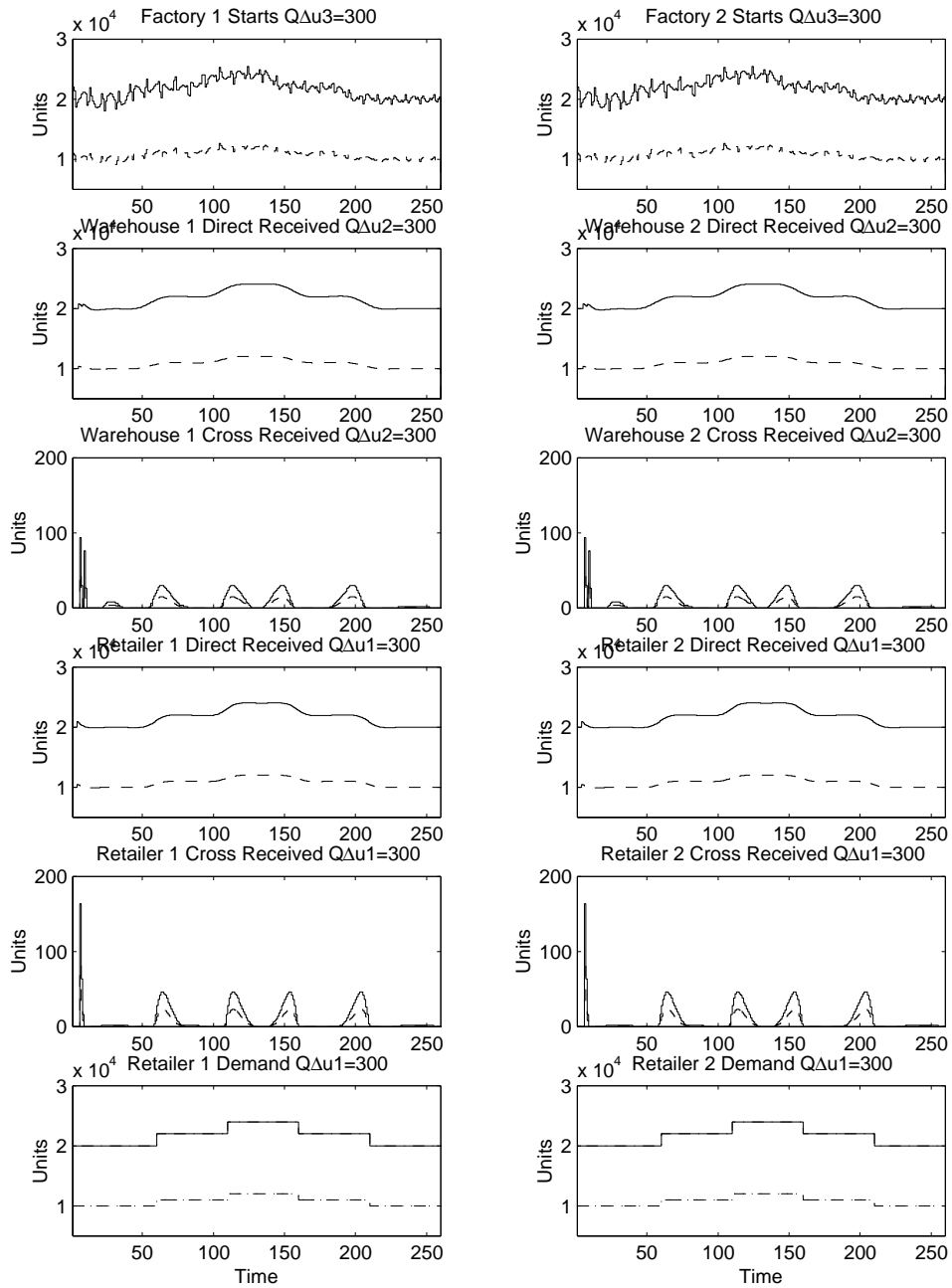


Figure 36: Information sharing strategy #3: Material flows Total (solid) and Product A (dashed); Demand Product A (dotted) and Total (dash-dot) and corresponding customer receipts Product A (dashed) and Total (solid).

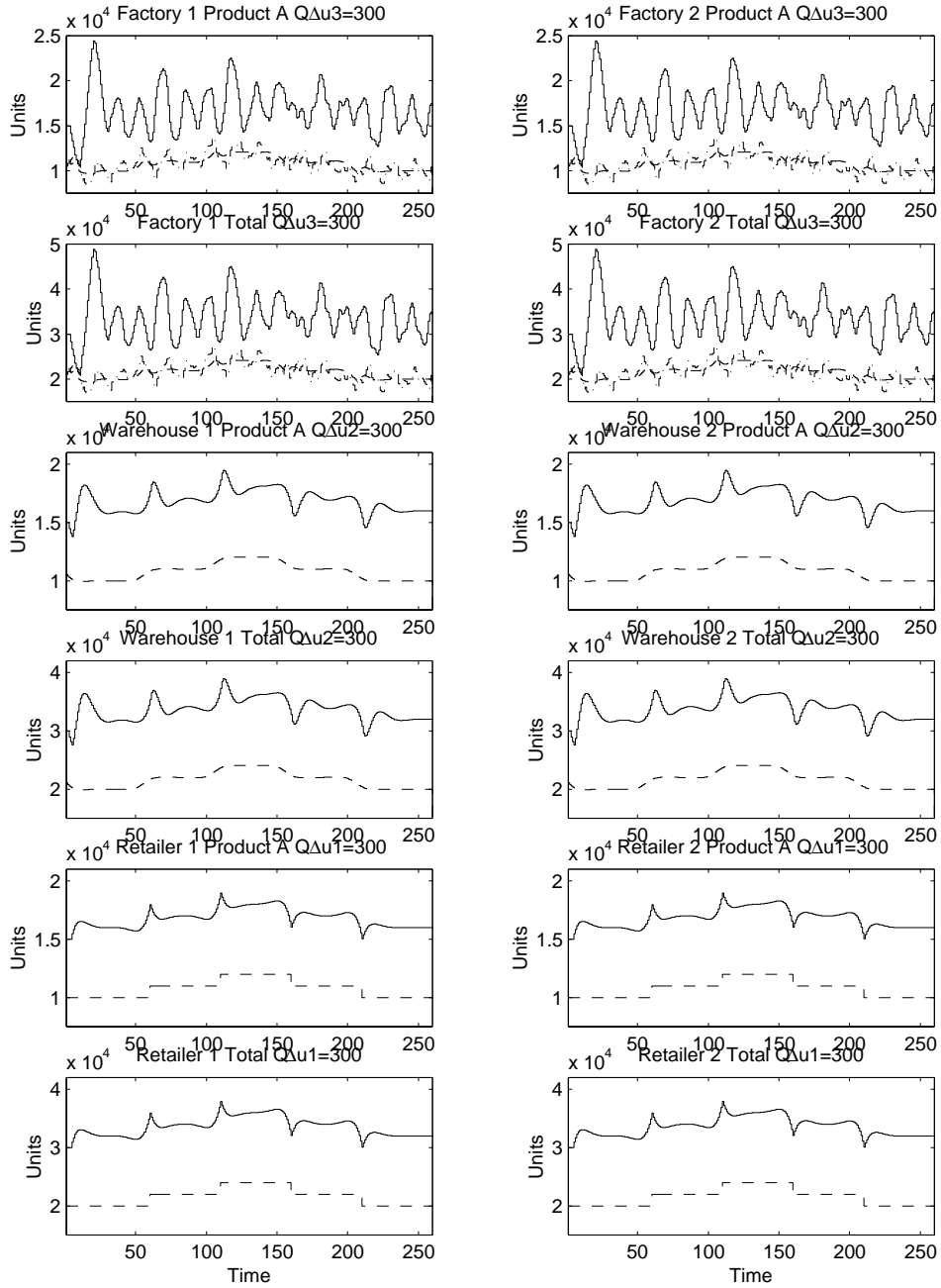


Figure 37: Information sharing strategy #4: Inventories (solid) and demand (dashed) by facility and product; Assembly/Test plots include starts (dash-dot).

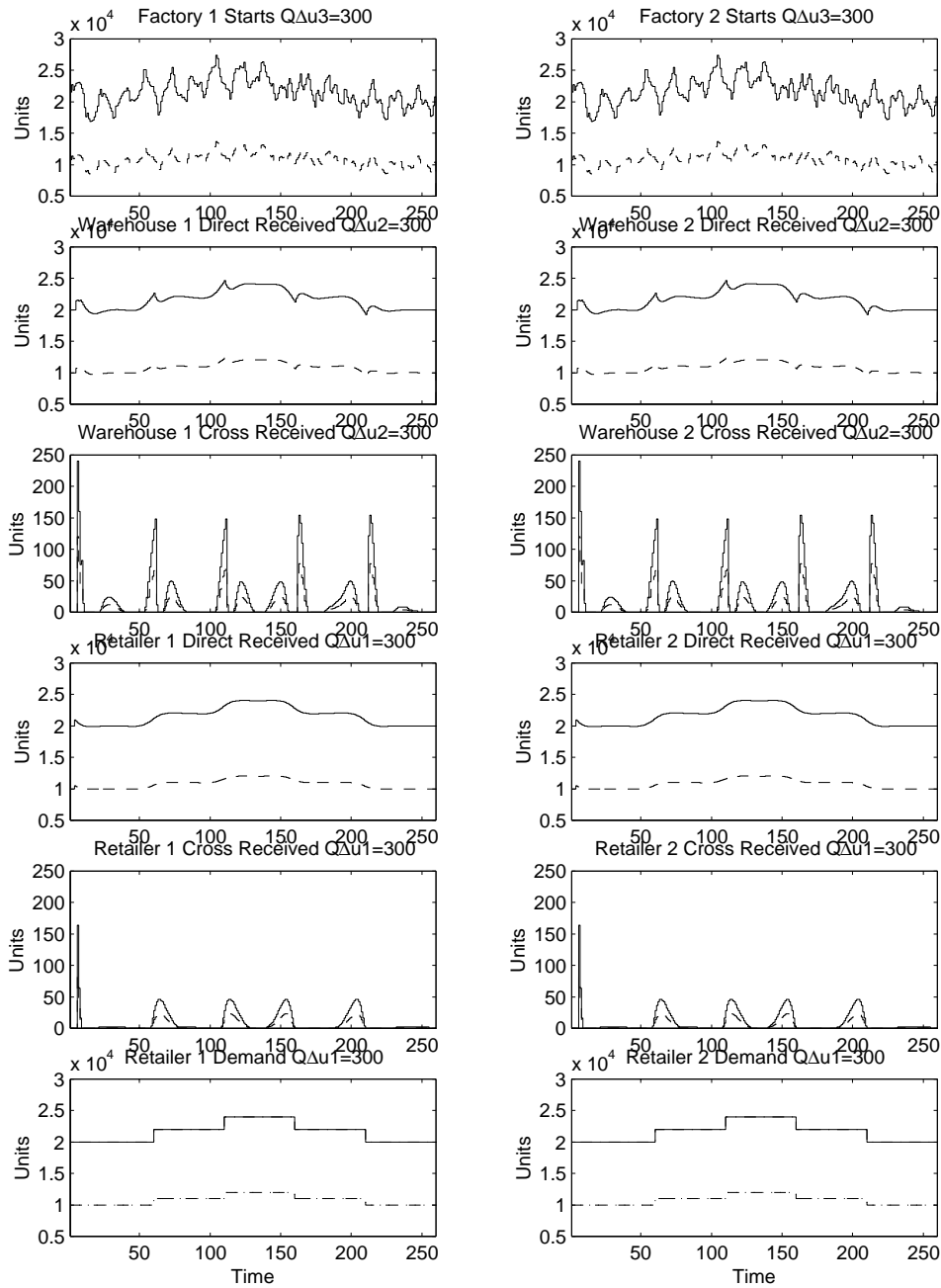


Figure 38: Information sharing strategy #4: Material flows Total (solid) and Product A (dashed); Demand Product A (dotted) and Total (dash-dot) and corresponding customer receipts Product A (dashed) and Total (solid).

The next set of simulations determines the lowest achievable safety stock levels possible with each information sharing configuration for the same simulation conditions as was just discussed. Safety stock levels are varied between 0 and 10,000 units per product, per node for all four strategies and the cumulative backorders and cumulative inventory levels are plotted for the Retailers, and for the sum of the whole network in Figures 39 and 40, respectively. These results demonstrate the value of feedforward information sharing in networks with significant leadtimes in production or transportation. Information sharing strategies #2 and #3 manage the system with no backorders, and minimal inventory levels with safety stock levels as low as 1500 units per product, per node. These results suggest a substantial reduction in inventory holdings, in combination with improved customer service levels for companies willing to adopt information sharing strategies. For this example, the network benefits by holding less inventory, and 85% less safety stock, while maintaining the same customer service level.

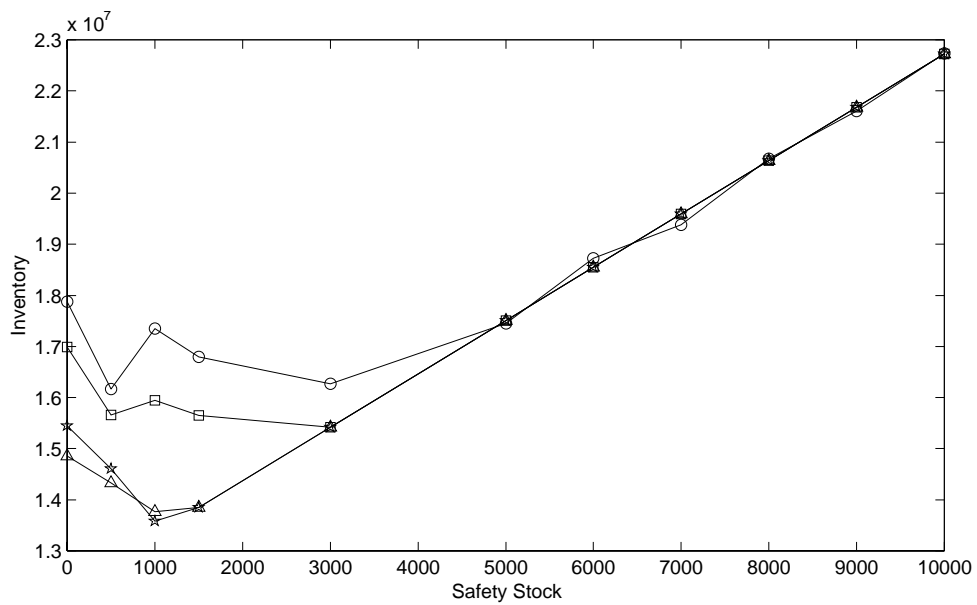
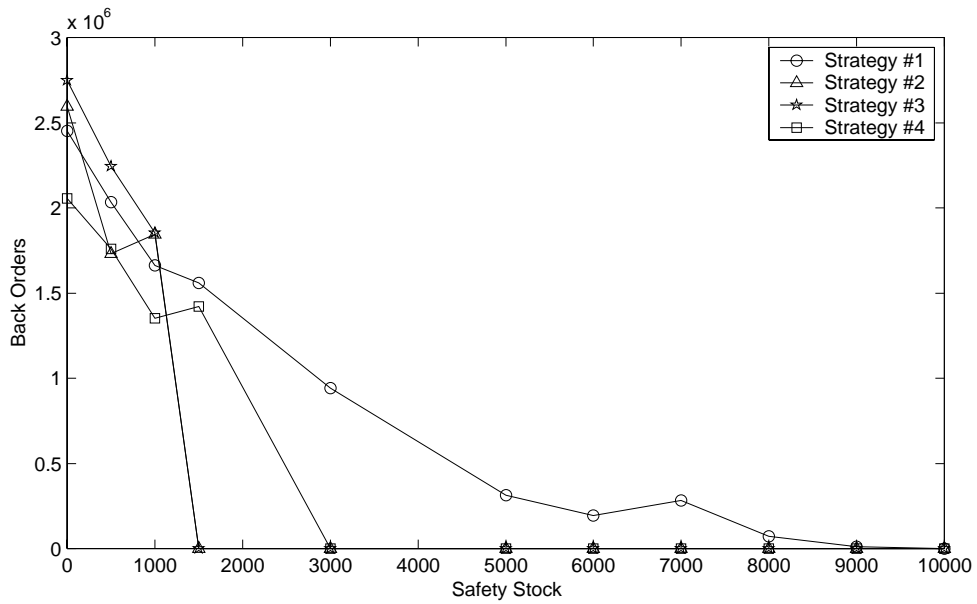


Figure 39: Safety stock - information sharing study: Retailer metrics

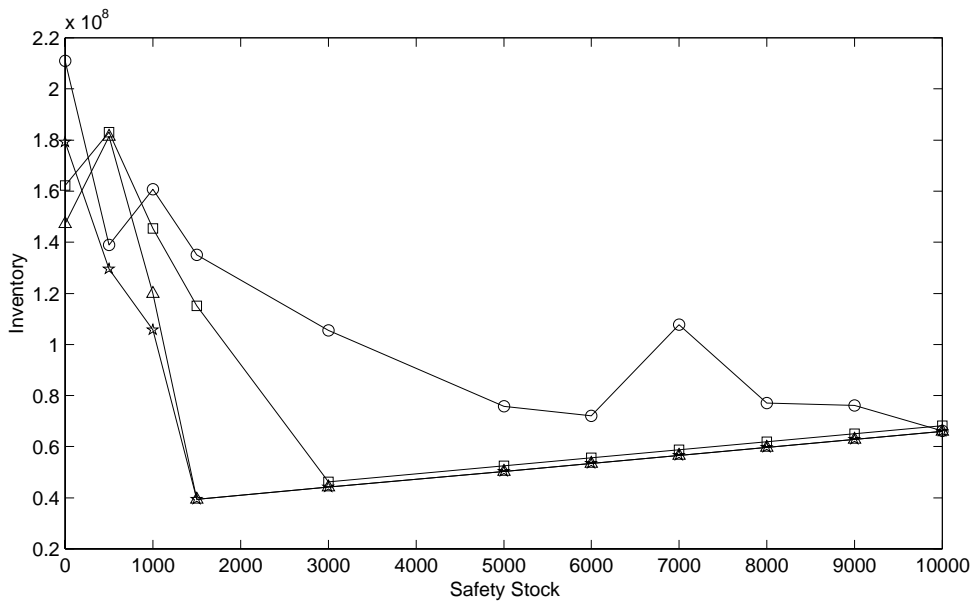
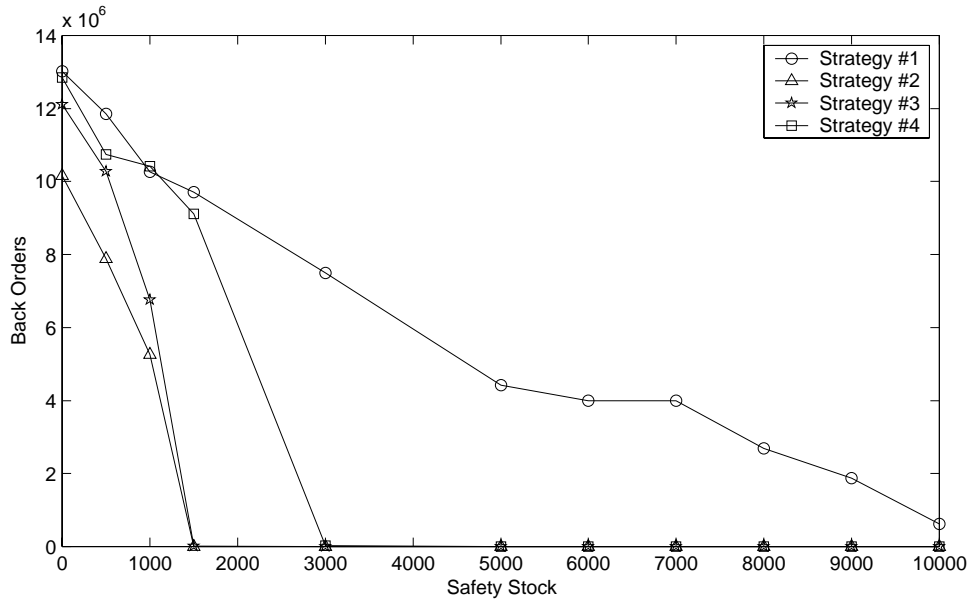


Figure 40: Safety stock - information sharing study: Retailer metrics

## 4 Conclusions & Future Research

As demonstrated throughout this paper, MPC offers the ability to robustify supply chain management systems to imperfect information regarding system performance, system bottlenecks, and inaccurate demand forecasts. The move suppression term in the MPC cost function can be interpreted as a financial penalty which could be set by the demand network management team to find minima in the cost curve which are not reachable with decisions made on a strictly cost optimal basis. This was demonstrated through experimentation with the two-node supply chain. An MPC approach was shown to be flexible enough to handle a system of industrial size and configuration - the six-node demand network proposed by Intel Corp. With this network, feedforward information sharing, well-known in the control community to be a necessity for improved performance of non-minimum phase systems, was deemed essential for cost reduction in networks with long production or transportation delays.

A natural extension of this work is to incorporate the *demand management* side of the organization. Sales and marketing functions serve to manipulate and predict demand forecasts which may also be used to further maximize profits or minimize costs for the organization. By incorporating advertising and marketing models to predict sales volume as a function of advertising dollars spent, the supply chain management function and sales and marketing function may be better integrated.

## References

- Armbruster, D., R. Chidambaram, G. Godding, K. G. Kempf and I. Katzorke (2001). Modeling and analysis of decision flows in complex supply networks. In: *Proceedings of IV SIMPOI/POMS 2001*. Guarujá, Brazil.
- Axsäter, S. (1985). Control theory concepts in production and inventory control. *International Journal of Systems Science* **16**(2), 161–169.
- Bodington, C. E. and D. E. Shobrys (1999). Optimize the supply chain. In: *Advanced Process Control and Information Systems for the Process Industries* (L. A. Kane, Ed.). pp. 236–240. Gulf Publishing Company. Houston, TX.
- Bose, S. and J. F. Pekny (2000). A model predictive framework for planning and scheduling problems: a case study of consumer goods supply chain. *Computers and Chemical Engineering* **24**, 329–335.
- Braun, M. W. (2001). Model-On-Demand Nonlinear Estimation And Model Predictive Control: Novel Methodologies For Process Control And Supply Chain Management. PhD dis-

- sertation. Arizona State University. Control Systems Engineering Laboratory, Tempe, AZ. <http://www.public.asu.edu/~mwbraun/dissertation.pdf>.
- Chandra, C. and S. Kumar (2000). Supply chain management in theory and practice: a passing fad or a fundamental change?. *Industrial Management & Data Systems* **100**(3), 100–113.
- Flores, M. E., D. E. Rivera and V. Smith-Daniels (2000). Managing supply chains using model predictive control. In: *AIChE 2000 Annual Meeting*. Los Angeles, CA. paper 262f.
- Kapsiotis, G. and S. Tzafestas (1992). Decision making for inventory/production planning using model based predictive control. In: *Parallel and Distributed Computing in Engineering Systems* (S. Tzafestas, P. Borne and L. Grandinetti, Eds.). pp. 551–556. Elsevier Science Publishers B. V.
- Kempf, K., K. Knutson, B. Armbruster, P. Babu and B. Duarte (2001). Fast accurate simulation of physical flows in demand networks. In: *Advanced Simulation Technologies Conference*. The Society for Computer Simulation International. Seattle, WA.
- Lee, H. L., V. Padmanabhan and S. Whang (1997). The bullwhip effect in supply chains. *Sloan Management Review* pp. 93–102.
- Morari, M. and E. Zafiriou (1989). *Robust Process Control*. Prentice Hall. Englewood Cliffs, NJ.
- Ogunnaike, B. A. and W. H. Ray (1994). *Process Dynamics, Modeling, and Control*. Oxford University Press. New York.
- Perea, E., I. Grossmann, E. Ydstie and T. Tahmassebi (2000). Application of Model Predictive Control for supply chain management. Paper 262e, *2000 AIChE Annual Meeting*, Los Angeles.
- PriceWaterhouseCoopers (2000). Silicon 2000: Millenium semiconductor survey on supply chain management. <http://www.ebusiness.pwcglobal.com/external/genreg.nsf/semi?OpenForm&Seq=1>.
- Silver, E. A., D. F. Pyke and R. Peterson (1998). *Inventory management and production planning and scheduling*. Wiley. New York.