Balancing Responsiveness and Economics in Process Supply Chain Design with Multi-Echelon Stochastic Inventory

Fengqi You and Ignacio E. Grossmann
Dept. of Chemical Engineering, Carnegie Mellon University, Pittsburgh, PA 15213

This article is concerned with the optimal design of multi-echelon process supply chains (PSCs) under economic and responsive criteria with considerations of inventory management and demand uncertainty. The multi-echelon inventory systems are modeled with the guaranteed service approach to handle the uncertain demands at each echelon. The maximum guaranteed service time of the last echelon of the PSC is proposed as a measure of a PSC's responsiveness. The problem is formulated as a bi-criterion mixed-integer nonlinear program (MINLP) with the objectives of minimizing the annualized cost (economic objective) and minimizing the maximum guaranteed service times of the markets (responsiveness objective). The model simultaneously predicts the optimal network structure, transportation amounts, and inventory levels under different specifications of the PSC responsiveness. An example on acetic acid supply chain is presented to illustrate the application of the proposed model and to comprehensively compare different measures of PSC responsiveness. © 2010 American Institute of Chemical Engineers *AIChE J,* 57: 178–192, 2011
Keywords: supply chain design, responsiveness, MINLP, inventory control, guaranteed service time, safety stock

Introduction

Because of the pressures from global competition, responsiveness is becoming a critical issue for the success of process supply chains (PSCs) since it allows chemical companies to achieve the best performance in the global marketplace.[1–5] Quick response enables supply chains to meet the customer demands with short lead times, and to synchronize the supply to meet the peaks and troughs of demand.[6] A major concern for chemical process companies has become how to effectively leverage the PSC design and operation to quickly satisfy the customer demands and achieve profitability.[7,8] This challenge requires addressing the optimal design

and development of "responsive" PSCs with effective inventory management to deal with the demand uncertainty.[9–11] You and Grossmann[5] have recently addressed this problem with a bi-criterion optimization framework by considering the optimal PSC design and operation under responsive and economic criteria. In their work, the economic criterion is measured by the net present value, whereas the criterion for responsiveness is measured by the lead time or expected lead time (ELT), which accounts for transportation times, residence times, cyclic schedules in multiproduct plants, and safety stocks in the distribution centers (DCs). Although the definition of the ELT integrates PSC responsiveness with safety stocks in DCs by using a probabilistic model for stockout, that model[4,5] was restricted to a single stage inventory model. The extension to multi-echelon inventory systems is nontrivial and, therefore, it is addressed in this article.

We consider the optimal design of multi-echelon PSCs and the associated inventory systems under demand
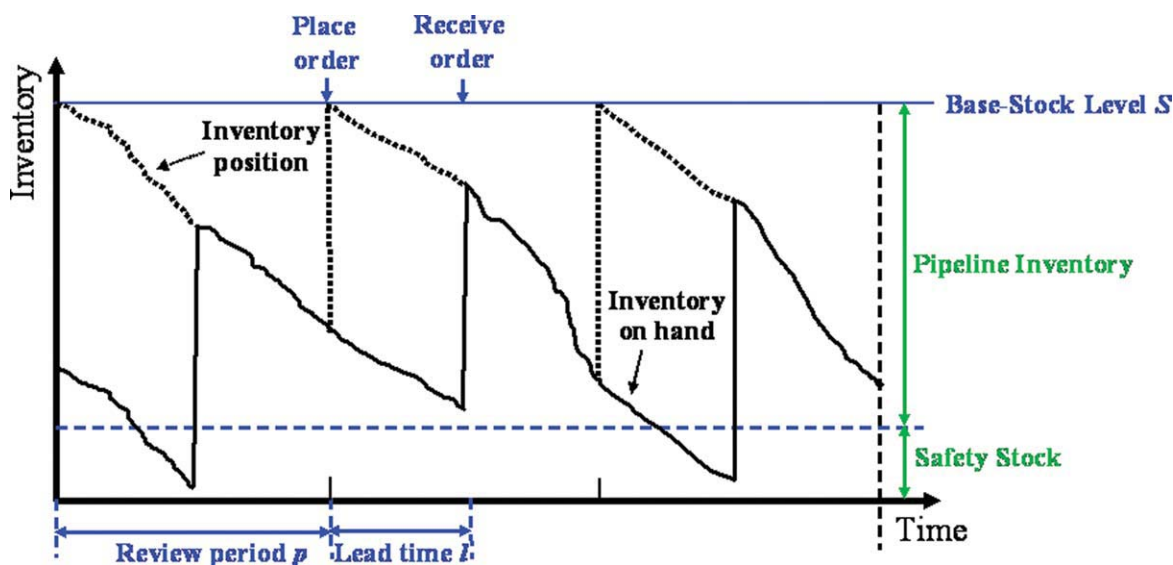
Figure 1. Inventory system controlled by base stock policy.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

uncertainty with considerations of economic performance and supply chain responsiveness. The guaranteed service approach[12–20] is used to model the multi-echelon stochastic inventory system in the PSC. Furthermore, the maximum of the guaranteed service times over all the markets (last echelon of the PSCs) is proposed as a quantitative measure of the responsiveness of the PSCs. For the case of a PSC with fixed network structure, this measure is compared with the ELT proposed by You and Grossmann.[5] We incorporate the proposed responsiveness measure into the joint multi-echelon supply chain design and inventory management model,[20] and formulate the problem as a bi-criterion mixed-integer nonlinear programming (MINLP) model with the objectives of minimizing the annualized cost (economic objective) and minimizing the maximum guaranteed service times of the markets (responsiveness objective). The model simultaneously determines the optimal network structure, transportation amounts, and inventory levels for different levels of responsiveness of the PSC. An example for a specialty chemical supply chain is presented to illustrate the application of the proposed model.

The rest of this article is organized as follows. In the next section, we propose our new measure for PSC responsiveness and compare it with the ELT for a PSC with a fixed design. A formal problem statement is given in the "Problem Statement" section. We present a multi-objective MINLP model for this problem in the "Bi-Criterion MINLP model" section. In the "Acetic Acid Supply Chain Example" section, we consider an illustrative example on acetic acid supply chain, and compare different measures of responsiveness for the PSC design problem. The last section concludes this article.

Guaranteed Service Time and PSC Responsiveness

Our proposed measure for PSC responsiveness is the maximum guaranteed service time of the last echelon of PSCs. This measure is integrated with multi-echelon inventory con-

trol and uncertain demands. We first discuss the guaranteed service approach and then introduce the concepts for PSC responsiveness.

Multi-echelon stochastic inventory model: Guaranteed service approach

In this section, we briefly review some inventory management models that are related to the problem addressed in this work. Detailed discussion on these models are given in our previous work,[20,21] as well as in Zipkin[22] and in Graves and Willem.[17,18]

For an inventory system controlled by base stock policy under demand uncertainty, the total inventory cost includes safety stock cost and pipeline inventory cost (Figure 1). The accepted practice in this field is to assume a normal distribution of the demand, although of course other distribution functions can be specified. If the demand rate at each unit of time is normally distributed with mean μ and standard deviation σ, the total demand over review period p and the replenishment lead time l is also normally distributed with mean $\mu(p + l)$ and standard deviation $\sigma\sqrt{p + l}$. It is convenient to measure safety stock in terms of the number of standard deviations of demand, denoted as safety stock factor, λ. Then, the optimal base stock level is given by:

$$S = \mu(p + l) + \lambda\sigma\sqrt{p + 1} \qquad (1)$$

We should note that if α is the Type I service level (the probability that the total inventory on hand is more than the demand), the safety stock factor λ corresponds to the α-quantile of the standard normal deviate of the service level, that is, $\Pr(x \leq \lambda) = \alpha$.

In Eq. 1, the safety stock factor λ and uncertain demand rate (mean μ and standard deviation σ) are usually given parameters, or else they can be easily inferred. Thus, as long as we can quantify lead time l, the optimal base stock level can be obtained. For single stage inventory system, lead

time, which may include material handling time and transportation time, is exogenous and generally can be treated as a constant. However, for a multi-echelon inventory system, lead time of a downstream node depends on its uncertain demand and upstream node's inventory level, and thus the lead time and internal service level are stochastic. The guaranteed service approach[13–19,23] can address this issue by modeling the entire system in an approximate way and allows a planner to make strategic and tactical decisions without the need of approximating portions of the system that are not captured by a simplified topological representation.

The main idea of the guaranteed service approach is that each node j in the multi-echelon inventory system quotes a guaranteed service time T_j, by which this node will satisfy all the demands from its downstream nodes. That is, the demand at time t must be ready to be shipped by time $t + T_j$. The guaranteed service times for internal customers are decision variables to be optimized, whereas the guaranteed service time for the nodes at the last echelon (facing external customers) is an exogenous input. Besides the guaranteed service time, we consider that each node j has a given deterministic order processing time, t_j, which is independent of the order size. The order processing time, which includes material handling time, transportation time, and review period, represents the time from all the inputs that are available until the outputs are ready to serve the demand. The net lead time of node $j(\text{NLT}_j)$ is the time span over which safety stock coverage against demand variations is necessary, and it is given by the guaranteed service time T_i of its direct predecessor node i plus its processing time t_j minus the guaranteed service time of node j.[13] Therefore, we can calculate the net lead time with the following formula:

$$\text{NLT}_j = T_i + t_j - T_j \qquad (2)$$

where node i is the direct predecessor of node j.

In the guaranteed service approach each node in the multi-echelon inventory system is assumed to operate under a periodic review base stock policy with a common review period. Furthermore, demand over any time interval is also assumed to be bounded with an associated safety stock factor λ_j.[20] This yields the base stock level at node j:

$$S_j = \mu_j \text{NLT}_j + \lambda_i \sigma_j \sqrt{\text{NLT}_j} \qquad (3)$$

This formula is similar but slightly different from the single stage inventory model (1) in terms of the expression for the lead time. Note that the review period has been taken into account as part of the processing time and considered in the net lead time.

With the guaranteed service approach, the total inventory cost consists of safety stock cost and pipeline inventory cost. The safety stock of node $j(\text{SS}_j)$ is given by the following formula as discussed above.

$$\text{SS}_j = \lambda_j \sigma_j \sqrt{\text{NLT}_j} \qquad (4)$$

The expected pipeline inventory is the sum of expected on hand and on-order inventories. On the basis of Little's law,[24] the expected pipeline inventory PI_j of node j equals to the mean demand over the processing time, and is given by:
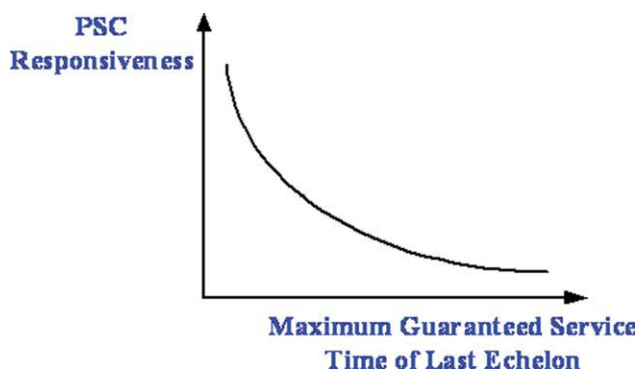

Figure 2. Conceptual relationship between PSC responsiveness and maximum guaranteed service time of the last echelon.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

$$\text{PI}_j = t_j \mu_j \qquad (5)$$

which is not affected by the guaranteed service time decisions.

Measures for PSC responsiveness

A major goal of this article is to develop a quantitative measure for PSC responsiveness in multi-echelon inventory systems under demand uncertainty. Responsiveness is defined as the ability of a PSC to respond rapidly to the changes of demand.[25,26] In our previous works,[4,5] we considered lead time, which is the time of a PSC network to respond to external demands, as the measure of PSC responsiveness. Specifically, we used first the worst case lead time,[4] corresponding to the response time when there are zero inventories as the measure of responsiveness under deterministic demand. For the case of uncertain demand, we used the ELT[5] as the measure of responsiveness. Although both measures capture the properties of PSC responsiveness and are integrated with safety stocks, they can not be readily extended to the case of the multi-echelon stochastic inventory of the PSCs.

In this article, we propose the maximum guaranteed service time quoted by the last echelon of a PSC (e.g., markets) to its external demand as a quantitative measure of PSC responsiveness. This service time of a market is the maximum time that all the demand of this market will be satisfied. If a PSC has more than one market, and each market k has a guaranteed service time R_k, the responsiveness of this PSC can be defined by considering the worst case, that is,

$$R = \max_k \{R_k\} \qquad (6)$$

where R is the maximum guaranteed service time of the last echelon of PSC. As shown in Figure 2, a PSC with long maximum guaranteed service time in the last echelon (R) implies that its responsiveness is low, and vice versa. We should also note that instead of using the maximum values in (6), that is, infinity norm, one could also use a weighted average value.

Compared to other measures for PSC responsiveness, such as lead time or ELT, this measure is more straightforward to
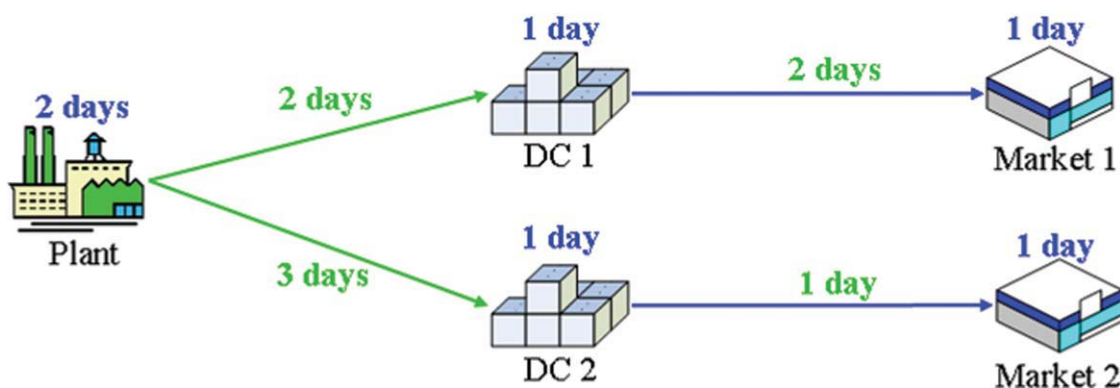
Figure 3. Network structure and processing times of the illustrative example.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

apply in PSCs, while still being able to capture the multi-echelon inventory structure of most of PSCs, and taking into account uncertain market demands. A comparison between the proposed measure for PSC responsiveness, maximum guaranteed service time of the markets (MGSTM), and the one used in our previous work,[5] ELT, is illustrated in the following example.

Illustrative example: Comparison of different measures of PSC responsiveness

In this example, we consider a PSC with fixed design including one plant, two DCs, and two markets as shown in Figure 3. The processing times of the plants, including residence time, material handling time, and review period is 2 days. The two DCs and the two markets all have the same processing time equal to 1 day. The transportation time from the plant to DC1 is 2 days, and 3 days from the plant to DC2. It takes 2 days to ship from DC1 to Market1, and 1 day to ship from DC2 to Market2. Each market has an uncertain demand following a normal distribution with the mean value μ and standard deviation σ (Figure 4). To compare the proposed responsiveness measure with those in our previous works,[4,5] we consider only single echelon inventory management in the PSC by assuming that only the DCs hold safety stock, that is, the net lead times of the plant and markets are 0 when using the guaranteed service approach.

Let us first address the responsiveness issue of this PSC using the guaranteed service approach. Because the net lead time of the plant is assumed to be 0, the guaranteed service time of the plant is equal to its processing time, 2 days. Let us denote the net lead times of DC1 and DC2 as NLT_1 and NLT_2. The order processing time from the plant to DC1 includes transportation from the plant to DC1 (2 days) plus the processing time of DC1 (1 day), so it should be 3 days. Thus, the guaranteed service time of DC1 equals to the guaranteed service time of the plant, 2 days, plus order processing time, 3 days, minus the net lead time of DC1, NLT_1. In this way, we can have the guaranteed service time of DC1 equal to $(5 - NLT_1)$ days, where $0 \leq NLT_1 \leq 5$. Similarly, we could have the guaranteed service time of DC2 equal to $(6 - NLT_2)$ days, where $0 \leq NLT_2 \leq 6$. It is also easy to derive that the order processing time of Market1 is 3 days and the one for Market2 is 2 days. Given the guaranteed

service times of these two DCs and the associated order processing times from the DCs to the markets, we can have the guaranteed service time of Market1 as $R_1 = (5 - NLT_1 + 3) = (8 - NLT_1)$ days and the guaranteed service time of Market2 as $R_2 = (6 - NLT_2 + 2) = (8 - NLT_2)$ days. On the basis of the definition of PSC responsiveness introduced in the previous section, we have the MGSTM of this PSC as:

$$R = \max\{R_1, R_2\} = \max\{8 - NLT_1, 8 - NLT_2\} \quad (7)$$

where $0 \leq NLT_1 \leq 5$ and $0 \leq NLT_2 \leq 6$. From this equation, we can see that when both DCs hold sufficient safety stock to ensure $NLT_1 = 5$ days and $NLT_2 = 6$ days, we have $R_{min} = \max\{3, 2\} = 3$ days. As the safety stock level in a DC decreases, the net lead time in this DC decreases quadratically in terms of safety stock based on Eq. 4. If the safety stock levels in both DCs decrease to 0, the net lead times of these two DCs are also 0, that is, $NLT_1 = NLT_2 = 0$. Thus, under zero inventories we have $R_{max} = 8$ days.

If we address this problem using the idea of worst case lead time as introduced in our previous work,[4] we first need
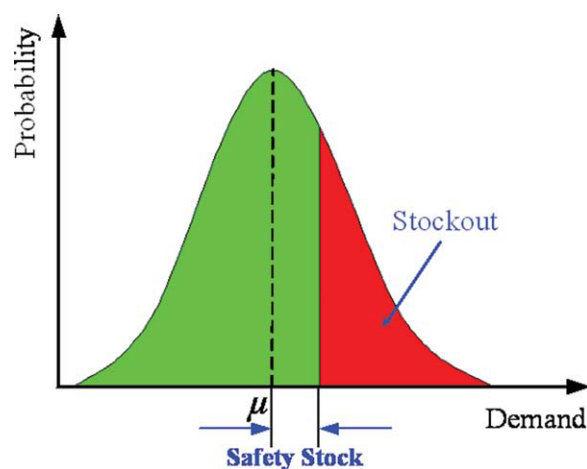

Figure 4. Probability distribution of the uncertain demands in the illustrative example.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

to decompose the PSC network into linear supply chains. There are two linear supply chains, from the plant to DC1 and then to Market1, and from the plant to DC2 and then to Market2. It is easy to figure out that all the time delays incurred in the first linear supply chain is 8 days, including the processing times of the plant, DC1, and Market1, as well as the transportation time between them. Similarly, the total time delay incurred in the second linear supply chain is 7 days. Thus, the worst lead time of the entire PSC is the longest lead time of all the linear supply chains included in the PSC network, that is, $LT = \max\{8, 7\} = 8$ days. We should note that this measure is based on the assumption of deterministic demand and zero inventory, thus 8 days correspond to the worst case with zero safety stocks in the DCs and the maximum guaranteed service time with $NLT_1 = NLT_2 = 0$ as shown in (7)

If we address this problem using the idea of ELT as introduced in our previous work,[5] the first step is also to decompose the PSC network into linear supply chains. There are two linear supply chains, from the plant to DC1 and then Market1, and from the plant to DC2 and then to Market2. As seen in Figure 3, in the first linear supply chain, the delivery lead time (LD_1) is 3 days, which includes the transportation time from DC1 to Market1 and the processing time of Market1, and the DC replenishment lead time (LP_1) is 5 days, which includes the processing times of the plant and DC1, as well as the transportation time from the plant to

DC1. Similarly, we can derive that the delivery lead time of the second linear supply chain (LD_2) is 2 days, and its production lead time (LP_2) is 6 days. On the basis of the definition,[5] the ELT of a linear supply chain is equal to its delivery lead time plus the stockout probability times its production lead time. Let us denote the stockout probability of DC1 as P_1 and the one of DC2 as P_2. As shown in Figure 4, the values of stockout probability P_1 and P_2 depend on the safety stock levels and demand uncertainty. The more safety stocks in a DC, the lower stockout probability is. For the normal demand distribution, we have the stockout probability given as follows:

$$P = 1 - \Phi\left(\frac{SS}{\sigma}\right) = \frac{1}{2}\left[1 - \text{erf}\left(\frac{SS}{\sqrt{2}\sigma}\right)\right],$$
$$\text{where } \Phi(x) = \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{x} e^{-(x^2/2)}dx, \quad 0 \leq SS \leq \lambda \cdot \sigma \quad (8)$$

where SS is the safety stock level in a DC and λ is the safety stock factor used in the guaranteed service approach. On the basis of this equation, we can see that the stockout probability can be very close to 0 when there is sufficient safety stock (if the safety stock factor is sufficiently large), and the stockout probability can be as high as 0.5 when there is no safety stock in the DC. Thus, we have the ELT (T) of the PSC equal to the maximum ELT of all the linear supply chains included in the PSC, given as follows:

$$T = \max\{(1 - P_1) \cdot LD_1 + P_i \cdot (LD_1 + LP_1), (1 - P_2) \cdot LD_2 + P_2 \cdot (LD_2 + LP_2)\}, 0 \leq P_1, P_2 \leq 0.5$$
$$= \max\{LD_1 + P_1 \cdot LP_1, LD_2 + P_2 \cdot LP_2\} = \max\{3 + 5 \cdot P_1, 2 + 6 \cdot P_2\} \quad (9)$$

From this equation, we can see that when both DCs hold sufficient safety stock to ensure $P_1 = 0$ and $P_2 = 0$, we have $T_{\min} = \max\{3, 2\} = 3$ days (note that the stockout probability would be sufficiently small and close to 0 if the safety stock factor is large enough). As the safety stock level in a DC decreases, the stockout probability in this DC increases based on Eq. 8. When both DCs do not hold any safety stocks, that is, $P_1 = P_2 = 0.5$, we have the maximum ELT of this PSC as $T_{\max} = \max\{3 + 5 \times 0.5, 2 + 6 \times 0.5\} = 5.5$ days.

From the above comparisons, we can draw the following conclusions. When there are sufficient safety stocks in the DCs, both MGSTM and ELT lead to the same minimum value, 3 days. When the safety stock level decreases, both MGSTM and ELT of a linear supply chain increases. When all the DCs hold zero safety stocks, the MGSTM has the maximum value, 8 days, which is the same as the worst case lead time, but the ELT has a maximum value of 5.5 days. The reason is that MGSTM accounts for the worst case in its definition, and thus it corresponds to the worst case lead time when all the DCs hold zero safety stocks. The ELT partially accounts for the worst case, but it makes use of the probability distribution of the uncertain demand and considers the "expected" value. As the stockout probability has a lower bound of 0.5, the maximum value of ELT for this PSC is less than MGSTM, although ELT also takes into account the worst case lead time through the use of stockout probability. We should note that both 8 days of MGSTM and 5.5 days of ELT correspond to the same inventory level,

and the difference between the values (8 days vs. 5.5 days) is due to the difference of measures, that is, "worst" case time in MGSTM vs. "average" time in ELT.

All these measures for PSC responsiveness share some similarities and have some differences. The major advantage of using the MGSTM as the measure is that it captures the interactions between different stages of a multi-echelon PSC and the corresponding inventory system, whereas the use of ELT is restricted to stochastic inventory in a single echelon. In the following sections, we first define the problem addressed in this article and then incorporate this measure into a joint PSC design and inventory management model as developed in our previous work.[20]

Problem Statement

To illustrate the application of this new measure for PSC responsiveness, we consider in this article the design and stochastic inventory management of a three-echelon PSC as in the example shown in Figure 5. The given potential PSC consists of a set of plants (or suppliers), a number of candidate DCs, and a set of markets. The markets are fixed and each of them has an uncorrelated normally distributed demand with known mean and variance. The candidate DCs are to be selected to install (and include in the PSC), and the investment costs for installing DCs are expressed by a cost function with fixed charges. The plants (or suppliers) are existing, but their presence in the optimal PSC network
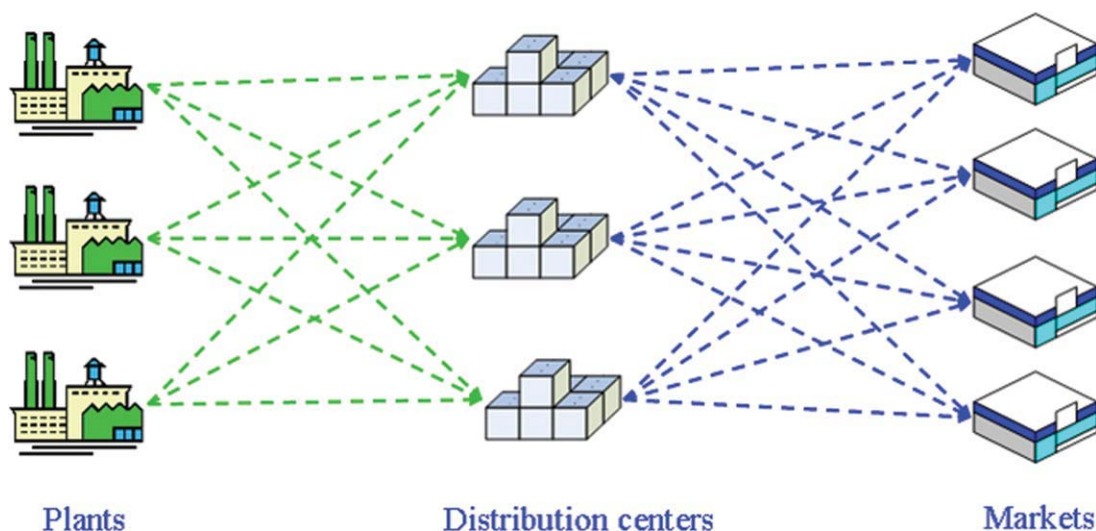
Figure 5. Acetic acid supply chain network superstructure.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

depend on their assignments with DCs, that is, a plant (or supplier) will not appear in the optimal PSC network if it is not selected to serve any DCs. Single sourcing restriction, which is common in industrial gases supply chains[27,28] and specialty chemicals supply chains,[29] is used for the assignment between plants and DCs and between DCs and markets. Linear transportation costs are considered for all the shipments. The service times of plants (or suppliers), and the deterministic processing times of DCs and markets are given. Inventories, including safety stocks and pipeline inventories, are hold at both the DCs and the markets, and the unit inventory costs are given. A common review period is used for inventory control throughout the PSC, and the safety stock factors for DCs and markets are also given.

The objective is to minimize the total installation costs of DCs, and the transportation, and inventory costs, and to maximize the responsiveness of the PSC, by deciding on how many distribution centers (DCs) to install, where to locate them, which plants to serve each DC, and which DCs to serve each market. Furthermore, the decisions also involve selecting the service time of each DC, and the level of safety stock to be maintained at each DC and market.

Bi-Criterion MINLP Model

The proposed PSC responsiveness measure can be readily incorporated into the joint multi-echelon supply chain design and inventory management model[20] to establish the tradeoff between economic performance and responsiveness of a PSC. The integrated model is a bi-criterion MINLP that deals with the supply chain network design for a given product, and considers its two-echelon inventory management and PSC responsiveness. The definition of sets, parameters, and variables of the model are given in the Appendix. The model formulation denoted as (P0) is given as follows:

$$\text{Min}: \quad \sum_{j\in J} f_j Y_j + \sum_{i\in I}\sum_{j\in J}\sum_{k\in K} A_{ijk} X_{ij} Z_{jk} + \sum_{j\in J}\sum_{k\in K} B_{jk} Z_{jk}$$
$$+ \sum_{j\in J} q1_j \sqrt{N_j \sum_{k\in K} \sigma_k^2 Z_{jk}} + \sum_{k\in K} q2_k \sqrt{L_k} \quad (10)$$

$$\text{Min}: \quad R \quad (11)$$

$$\text{s.t.} \quad \sum_{i\in I} X_{ij} = Y_j, \quad \forall j \quad (12)$$

$$\sum_{j\in J} Z_{jk} = 1, \quad \forall k \quad (13)$$

$$Z_{jk} \leq Y_j, \quad \forall j,k \quad (14)$$

$$N_j \geq \sum_{i\in I} \bar{S}_{ij} \cdot X_{ij} - S_j, \quad \forall j \quad (15)$$

$$L_k \geq \sum_{j\in J} \left(S_j + t2_{jk} \right) \cdot Z_{jk} - R_k, \quad \forall k \quad (16)$$

$$R \geq R_k, \quad \forall k \quad (17)$$

$$X_{ij}, Y_j, Z_{jk} \in \{0,1\}, \quad \forall i,j,k \quad (18)$$

$$S_j \geq 0, \quad N_j \geq 0, \quad \forall j \quad (19)$$

$$L_k \geq 0, \quad \forall k \quad (20)$$

where

$$\text{S}_{ij} = \text{SI}_i + t1_{ij}$$
$$A_{ijk} = \left(c1_{ij}\chi + \theta 1_j t1_{ij} \right) \cdot \mu_k$$
$$B_{jk} = \left(g_j\chi + c2_{jk}\chi + \theta 2_k t2_{jk} \right) \cdot \mu_k$$
$$q1_j = \lambda 1_j \cdot h1_j$$
$$q2_k = \lambda 2_k \cdot h2_k \cdot \sigma_k$$

(P0) is similar to the joint three-echelon supply chain design and inventory management model reformulated by You and Grossmann.[20] However, in this case, it corresponds

to a bi-criterion optimization problem where the new objective function (11) is added to minimize MGSTM, the new constraint (17) is introduced to define the measure of PSC responsiveness, and R_k, the maximum guaranteed service time of market k, is treated as variables with upper bound R. The economic objective function (10) is the annual cost of the PSC design that accounts for the fixed and variable costs of installing DCs, transportation costs from plants to DCs and from DCs to markets, pipeline inventory costs in DCs and markets, as well as safety stock costs in DCs and markets. Constraints (12–14) define the PSC network structure, and constraints (15) and (16) define the net lead times at the DCs and markets.

Similar to our previous work,[20] we can linearize[30] the bilinear terms (products of binary variables and continuous variables, or products of two binary variables) and reformulate the model as problem (P1):

$$\text{Min}: \sum_{j\in J} f_j Y_j + \sum_{i\in I}\sum_{j\in J}\sum_{k\in K} A_{ijk} XZ_{ijk} + \sum_{j\in J}\sum_{k\in K} B_{jk} Z_{jk}$$
$$+ \sum_{j\in J} q1_j \sqrt{NZV_j} + \sum_{k\in K} q2_k \sqrt{L_k} \quad (21)$$

$$\text{Min}: \quad R \quad (11)$$

$$\text{s.t.} \quad \text{Constraints } (12)-(15),\ (17)-(20)$$

$$L_k \geq \sum_{j\in J} SZ_{jk} + \sum_{j\in J} t_{jk} \cdot Z_{jk} - R_k, \forall k \quad (22)$$

$$XZ_{ijk} \leq X_{ij}, \quad \forall i,j,k \quad (23)$$

$$XZ_{ijk} \leq Z_{jk}, \quad \forall i,j,k \quad (24)$$

$$XZ_{ijk} \geq X_{ij} + Z_{jk} - 1, \quad \forall i,j,k \quad (25)$$

$$SZ_{jk} + SZ1_{jk} = S_j, \quad \forall j,k \quad (26)$$

$$SZ_{jk} \leq Z_{jk} \cdot S_j^U, \quad \forall j,k \quad (27)$$

$$SZ1_{jk} \leq (1 - Z_{jk}) \cdot S_j^U, \quad \forall j,k \quad (28)$$

$$NZ_{jk} + NZ1_{jk} = N_j, \quad \forall j,k \quad (29)$$

$$NZ_{jk} \leq Z_{jk} \cdot N_j^U, \quad \forall j,k \quad (30)$$

$$NZ1_{jk} \leq (1 - Z_{jk}) \cdot N_j^U, \quad \forall j,k \quad (31)$$

$$NZV_j = \sum_{k\in K} \sigma_k^2 \cdot NZ_{jk}, \quad \forall j \quad (32)$$

$$XZ_{ijk} \geq 0, \quad \forall i,j,k \quad (33)$$

$$NZV_j \geq 0, \quad \forall j \quad (34)$$

$$SZ_{jk} \geq 0,\ SZ1_{jk} \geq 0,\ NZ_{jk} \geq 0,\ NZ1_{jk} \geq 0, \quad \forall j,k \quad (35)$$

where constraints (23–32) are introduced for the exact linearization. The bounds of the variables are given in the constraints (36):

$$N_j^U = \max_{i\in I}\{SI_i + t_{ij}\} = \max_{i\in I}\{\bar{S}_{ij}\}, \quad \forall j \quad (36a)$$

$$S_j^U = N_j^U = \max_{i\in I}\{S_{ij}\}, \quad \forall j \quad (36b)$$

$$L_k^U = \max\left\{0, \max_{j\in J}\{S_j^U + t_{jk}\} - R_k\right\}, \quad \forall j \quad (36c)$$

$$SZ_{jk}^U = S_j^U, \quad SZ1_{jk}^U = S_j^U, \quad \forall j,k \quad (36d)$$

$$NZ_{jk}^U = N_j^U, \quad NZ1_{jk}^U = N_j^U, \quad \forall j,k \quad (36e)$$

$$NZV_j^U = \sum_{k\in K} \sigma_k^2 \cdot N_j^U, \quad \forall j \quad (36f)$$

Thus, to balance the economics and responsiveness, we have two objective functions to be minimized in (P1). One is the annual total PSC design cost, and the other one is the MGSTM. To obtain the Pareto-optimal curve for the bi-criterion optimization problem, one of the objectives is treated as an inequality with a fixed value for the bound which is considered as a parameter. There are two major approaches to solve the problem in terms of this parameter. One is to simply solve it for a specified number of points to obtain an approximation of the Pareto optimal curve (ε-constraint method).[31–34] The other one is to solve it as a parametric programing problem,[35,36] which yields the exact solution for the Pareto optimal curve. Although the latter provides a rigorous solution approach, the former is easier to implement for MINLP models. For this reason, we solve this multi-objective optimization problem with the ε-constraint method.

From the economic objective (10) and constraint (16) it is easy to see that by increasing R_k, the total annual cost of this PSC is reduced. Coupled with the responsiveness objective (11) and constraint (17), we can observe that (P0) is always feasible for any value of R_k. Note that the lower bound of the left hand side of constraint (16) is zero, whereas the lower bound of the right hand side of this constraint is $-R$, which is a nonpositive value. Thus, the problem has an infinite number of optimal solutions in the continuous space. To be more specific, if R_k^* is the optimal solution and we have the optimal value of variable R as $R^* = \max_{\forall k\in K}\{R_k^*\}$, by replacing the value of variables R_k, $\forall k \in K$ with R_*, we have $R^* = \max_{\forall k\in K}\{R_k^*\} = R_k^*$ is also a feasible and optimal solution of the problem. Therefore, we can always have an optimal solution with

$$R^* = \max_{\forall k\in K}\{R_k^*\} = R_k^* \quad \forall k \quad (37)$$

This property suggests that we can fix the guaranteed service times of all the markets (R_k) to a certain value when using the ε-constraint method. In other words, objective (11) and constraint (17) can be removed and variables R_k can be treated as parameters in the solution procedure. The lower bound of R and all the R_k can be easily determined to be 0. To obtain their upper bounds, we consider a modified objective function as follows:

$$\text{Min}: \sum_{j\in J} f_j Y_j + \sum_{i\in I}\sum_{j\in J}\sum_{k\in K} A_{ijk} XZ_{ijk} + \sum_{j\in J}\sum_{k\in K} B_{jk} Z_{jk}$$
$$+ \sum_{j\in J} q1_j \sqrt{NZV_j} + \sum_{k\in K} q2_k \sqrt{L_k} + \varepsilon \cdot R \quad (38)$$

where ε is a scaling parameter with sufficient small value, for instance, 0.001. By minimizing (38) subject to all the

Table 1. Parameters for Demand Uncertainty for Illustrative Example

	Mean Demand μ_i (ton/day)	Standard Deviation σ_i (ton/day)
Market1	250	150
Market2	180	75
Market3	150	80
Market4	160	45

Table 3. Order Processing Time ($t2_{jk}$) Between DC and Market (days), Including Material Handling Time, Transportation Time, and Review Period, for Illustrative Example

	Market1	Market2	Market3	Market4
DC1	2	2	3	3
DC2	4	4	1	1
DC3	4	4	3	3

constraints in (P1), we can determine the minimum upper bound of variable R. Therefore, in the solution procedure, we solve problem (P1) under different specifications of MGSTM between 0 and its upper bound to obtain an approximation of the Pareto optimal curve.

Acetic Acid Supply Chain Example

To illustrate the application of our model, we consider an example taken from our previous work[20] for an acetic acid supply chain with three plants, three potential DCs, and markets as shown in Figure 5.

Multi-echelon inventory system

In this instance, the annual fixed installation cost (f_j) is $200,000/year for all the DCs. The variable cost coefficient for installing a DC (g_j) at any candidate location is $0.05/ (ton year). The safety stock factors for DCs ($\lambda1_j$) and markets ($\lambda2_k$) are the same and equal to 1.96, which corresponds to 97.5% service level if the demands are normally distributed. We consider 365 days in a year (χ). The guaranteed service times of the three plants (SI_i) are 3, 3, and 4 days, respectively. The pipeline inventory holding cost is $0.5/(ton day) for all the DCs ($\theta1_{ij}/\chi$) and markets ($\theta2_{jk}/\chi$), and the safety stock holding cost is $1/(ton day) for all the DCs ($h1_j/ \chi$) and markets ($h2_k/\chi$). The data for demand uncertainty, processing times, and transportation costs are given in Tables 1–5. The models are coded in GAMS and solved with the global optimizer BARON 8.1.5 on an IBM T40 laptop with Intel 1.50 GHz CPU and 512 MB RAM (please refer to our previous work[20] for a comprehensive comparison of computational performance of other MINLP solvers for the model). The original model (P0) has 12 discrete variables, 22 continuous variables, and 27 constraints. The MINLP model (P1) includes 24 binary variables, 185 continuous variables, and 210 constraints.

We use the ε-constraint method to obtain the Pareto optimal curve and determine the tradeoff between the total annual cost of the PSC design and its responsiveness, which is measured by the maximum guaranteed service time of the markets. The first step of the ε-constraint method is to deter-

mine the optimal lower and upper bounds of MGSTM (R). Its lower bound can be easily determined as 0, which means that all the market demands are satisfied immediately. Its upper bound's minimum value can be obtained by minimizing (38) subject to all constraints in (P1). For this problem, we obtain 12 days as the optimal upper bound of MGSTM. After the lower and upper bounds of MGSTM are determined, we solve the problem with fixed values of R from 0 day to 12 days (e.g., 13 instances with increments of 1 day). Solving the original model (P0) with 0% optimality margin takes a total of 7101.51 CPU seconds for all the 13 instances. When we solve the reformulated model (P1) by minimizing (21), the 13 instances require only 86.89 CPU seconds and the optimal solutions are the same as what we obtained by solving model (P0).

The results are given in Figures 6 and 7, as well as in Tables 6 and 7. The line in Figure 6 is the Pareto optimal curve of this problem. As can be seen, the cost ranges from $1,721,685 to $2,519,886, whereas the guaranteed service time (R) ranges from 0 days to 12 days. We can see that the total cost decreases as the MGSTM increases. As the MGSTM is a measure of PSC responsiveness, we can conclude that the more responsive the PSC is, the more cost it requires. The columns in Figure 6 show the total safety stocks in the system under different specifications of MGSTM. As the MGSTM increases, the total safety stocks in this PSC decrease from 2187 tons to 0. However, the safety stock levels do not strictly decrease as MGSTM increases, but have a "valley" when MGSTM equal to 8 and 9 days. This is due to the change of supply chain network structure and inventory allocation decisions between DCs and markets.

Figure 7 shows the change of the optimal network structures under different specifications of MGSTM. We can see that the optimal network structures for most cases are the same as shown in Figure 7a. When MGSTM is equal to 8, 9, and 11 days, the optimal network structures are different from the "common" one, and they are given in Figures 7b–d, respectively. In Figures 7a, d, only one DC is installed and serves all the markets. This is due to the "risk pooling" effect, which tries to group markets to be served by one DC so as to reduce the total safety stocks. In Figures 7b, c there

Table 2. Order Processing Time ($t1_{ij}$) Between Plant and DC (days), Including Material Handling Time, Transportation Time, and Review Period, for Illustrative Example

	DC1	DC2	DC3
Plant1	4	4	2
Plant2	2	4	3
Plant3	3	4	4

Table 4. Unit Transportation Cost ($c1_{ij}$) from Plants to DCs ($/ton)

	DC1	DC2	DC3
Plant1	1.8	1.6	2.0
Plant2	2.4	2.2	1.3
Plant3	2.0	1.3	2.5

Table 5. Unit Transportation Cost ($c2_{jk}$) from DCs to Markets ($/ton)

	Market1	Market2	Market3	Market4
DC1	1.0	3.3	4.0	7.4
DC2	1.0	0.5	0.1	2.0
DC3	7.7	7.3	5.1	0.1

is one more DC installed that leads to higher DC installation costs, but presumably reduces the total transportation and inventory costs. This reveals the tradeoff between transportation cost, inventory cost, and facility location cost, and suggests that changing network structure may be more effective to improve the PSC responsiveness compared to holding safety stock. A similar conclusion is also presented in our previous work,[5] although in that case safety stock levels always decrease as the PSC responsiveness decreases due to the single echelon inventory model used in that model. When the MGSTM increases from 9 days to 10 days, the optimal number of DCs reduces from two to one, and the network structure "returns" to the one shown in Figure 7a. This is also presumably due to the tradeoff between the several cost items as discussed above. When the MGSTM increases to 11 and 12 days, the optimal safety stocks in the systems reduce to zero (as in Figure 6), and the two optimal network structures have the same distribution network, but different production plants. The reason is that Plant1 has a shorter guaranteed service time (3 days) than Plant3 (4 days), but higher unit transportation cost (as in Table 4)
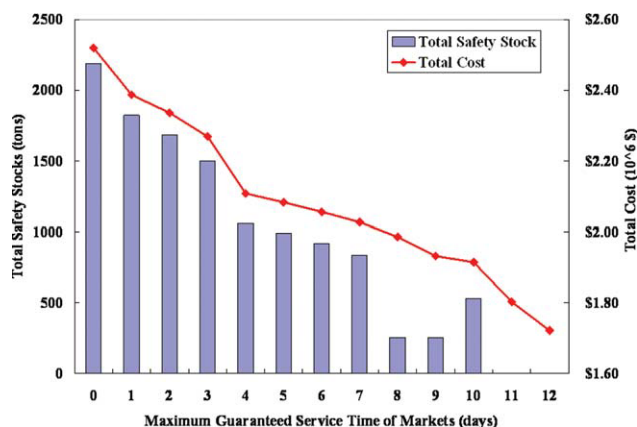

Figure 6. Pareto optimal curve and optimal safety stock levels for the acetic acid supply chain example.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

from the plant to DC 2. To reduce the overall transportation, inventory and supply chain design cost, it is optimal to include Plant3 instead of Plant1 in the supply chain network when MGSTM is 12 days. Note that these existing plants are considered to be acting as suppliers, which can be added or removed from the supply chain network at no additional instillation costs, although the transportation cost from plants to DCs are taken into account.
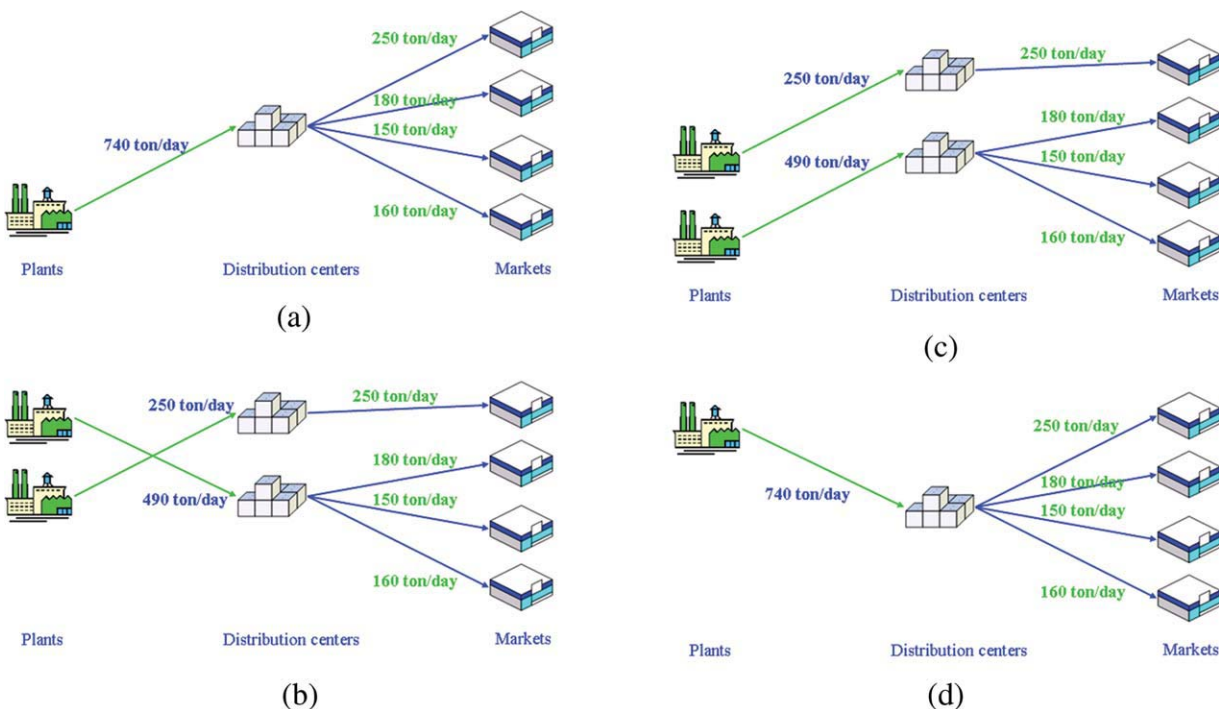

Figure 7. Optimal acetic acid supply chain network structure under different MGSTM (multi-echelon inventory case).

(a) Optimal network for MGSTM is 0–7 or 10, 12 days. (b) Optimal network for MGSTM is 8 days. (c) Optimal network for MGSTM is 9 days. (d) Optimal network for MGSTM is 11 days. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

Table 6. Optimal Net Lead Times (days) of DCs and Markets Under Different Specifications of Maximum Guarantee Service Time of Markets for the Illustrative Example with Multi-Echelon Inventory Systems

MGSTM	DC1 (days)	DC2 (days)	DC3 (days)	Market1 (days)	Market2 (days)	Market3 (days)	Market4 (days)
0 day	–	8	–	4	4	1	1
1 day	–	8	–	3	3	0	0
2 days	–	8	–	2	2	0	0
3 days	–	8	–	1	1	0	0
4 days	–	8	–	0	0	0	0
5 days	–	7	–	0	0	0	0
6 days	–	6	–	0	0	0	0
7 days	–	5	–	0	0	0	0
8 days	0	0	–	0	3	0	0
9 days	0	0	–	0	3	0	0
10 days	–	2	–	0	0	0	0
11 days	–	0	–	0	0	0	0
12 days	–	0	–	0	0	0	0

Tables 6 and 7 show the change of optimal net lead times and optimal safety stock levels in the DCs and markets under different specifications of the maximum guaranteed service time to the markets (MGSTM). It is interesting to note that the optimal net lead times of DCs and markets are at integer values, although they are not restricted to be integers in the optimization model. In the most "responsive" case, both DCs and markets hold maximum safety stock to guarantee the MGSTM equal to zero. As the MGSTM increases, the safety stock levels and net lead times in the markets decrease, whereas the safety stock levels and net lead times in the DCs remain unchanged until the safety stocks in markets deplete. This trend shows that holding safety stocks in the DCs (upstream) is more efficient to increase PSC responsiveness than holding safety stocks in the markets (downstream). When the MGSTM is between 5 days and 8 days, no market holds inventory and the safety stocks in the DCs decrease as the MGSTM increases. When the MGSTM increases to 8 and 9 days, the safety stocks are shifted from the DCs to the Markets. The change of inventory allocation decision is presumably due to the change of the PSC network structure as given in Figures 7b, c, and leads to the "valley" in the inventory profile as in Figure 6. When the MGSTM increases from 9 days to 10 days, we can see that the safety stocks "return" to the DCs and this change further reduces the total cost but almost double the total safety stocks. These interesting changes of inventory allocation and the associated network structure are due to

the fact that we consider the two-echelon inventory (DCs and markets).

Although this is a small example, the proposed approach for PSC responsiveness can easily be applied to large scale PSCs design problems by using the tailored algorithm presented in our previous work.[20] In the Appendix, we present the data and results of a large scale instance with 10 plants, 50 potential DCs, and 100 markets, by solving the MINLP model with the Lagrangean relaxation and decomposition algorithm as discussed in our previous work.[20]

Single-echelon inventory system

To illustrate the similarities and differences of the two measures of PSC responsiveness, MGSTM and ELT,[5] we consider a special case that only DCs of the acetic acid supply chain can hold safety stocks, that is, single echelon inventory.

To model and optimize the system with the guaranteed service approach, we need to make a few minor changes in model (P1). By fixing the values of the net lead times in the markets to zero in model (P1), that is, $L_k = 0$, $\forall k$, model (P1) reduces to the PSC design problem with single echelon inventory, as the markets are not allowed to hold safety stocks due to their zero net lead times. Because of the change of the inventory structure, the optimal lower and upper bounds of the MGSTM may change. Thus, to obtain the optimal lower bound we solve an optimization problem

Table 7. Optimal Safety Stock Levels (tons) of DCs and Markets Under Different Specifications of Maximum Guarantee Service Time of Markets for the Illustrative Example with Multi-Echelon Inventory Systems

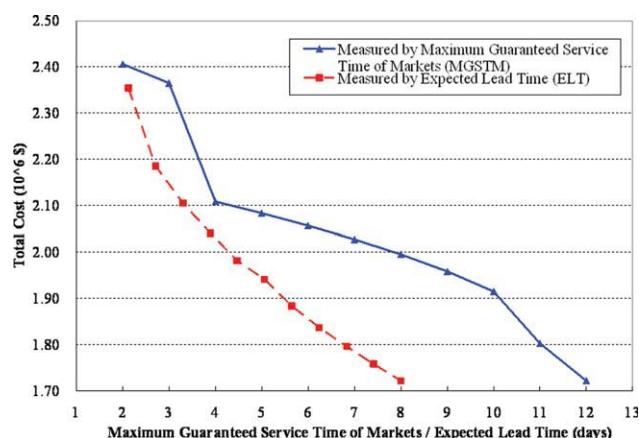| MGSTM	DC1 (tons)	DC2 (tons)	DC3 (tons)	Market1 (tons)	Market2 (tons)	Market3 (tons)	Market4 (tons)
0 day	–	1059.85	–	588.00	294.00	156.80	88.20
1 day	–	1059.85	–	509.22	254.61	0	0
2 days	–	1059.85	–	415.78	207.89	0	0
3 days	–	1059.85	–	294.00	147.00	0	0
4 days	–	1059.85	–	0	0	0	0
5 days	–	991.40	–	0	0	0	0
6 days	–	917.86	–	0	0	0	0
7 days	–	837.89	–	0	0	0	0
8 days	0	0	–	0	254.61	0	0
9 days	0	0	–	0	254.61	0	0
10 days	–	529.93	–	0	0	0	0
11 days	–	0	–	0	0	0	0
12 days	–	0	–	0	0	0	0

Figure 8. Pareto optimal curve for the acetic acid supply chain with single-echelon inventory for ELT and MGSTM.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
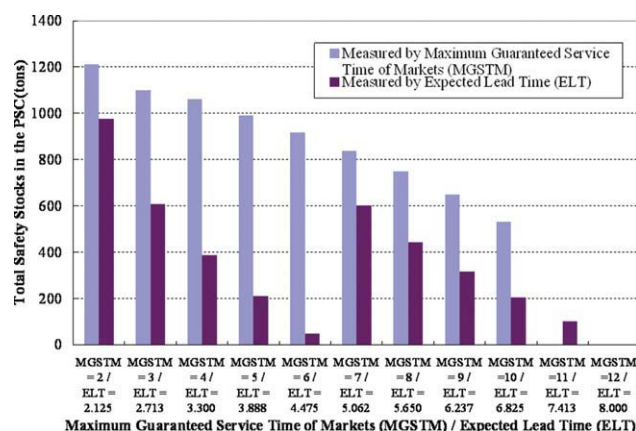

Figure 9. Total safety stock levels for the acetic acid supply chain with single-echelon inventory under the two measures for PSC responsiveness, ELT, and MGSTM.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

by minimizing (11) subject to all the constraints in model (P1), and to obtain the optimal upper bound we solve a problem by minimizing (38) subject to all the constraints in model (P1).

Using ELT as the responsiveness measure for PSC design problem leads to another MINLP model (Q1), the detailed formulation of which is given in the Appendix. The major difference between this model and the one introduced in our previous work[5] is that the standardized normal variables ϕ_j are restricted to be not greater than the safety stock factor $\lambda 1_j$, so as to be consistent to the data used in the guaranteed service time approach. Besides, the risk-pooling effect is also taken into account in this model.

We solve these problems using the same computer described in the previous section. All the instances are solved with the global optimizer LINDOGlobal in the GAMS modeling system, since BARON does not support the error function included in model (Q1).

For the guaranteed service time case, the resulting optimal lower bound of MGSTM is 2 days, and the optimal upper bound is 12 days. Then, we fix the value of R to integer values from 2 days to 12 days, and solve these 11 instances by minimizing (21) subject to all the constraints of (P1). The 11 instances require 228 CPU seconds. When using the ELT as the measure of responsiveness, the resulting optimal lower and upper bounds of the ELT are 2.125 days and 8 days, respectively. We then fix the value of T to 11 evenly distributed points ranging from 2.125 days to 8 days, and solve these 11 instances by minimizing (38) subject to all the constraints of (Q1). The 11 instances require a total of 248 CPU seconds. The resulting Pareto-optimal curves, optimal safety stock levels, and optimal PSC network structure under different specifications of MGSTM (R) and ELT (T) are given in Figures 8–11.

Figure 8 shows the Pareto optimal curves for this PSC with single-echelon inventory using two different measures for responsiveness. In the Pareto curve for guaranteed service approach, the total cost ranges from $ 1.72 MM to $ 2.41 MM, while the MGSTM (R) ranges from 2 days to 12 days. We can see that total cost decreases as MGSTM increase. The trade-off shows that the more responsive the

PSC is, the more cost is required in the PSC design. In particular, the total cost decreases significantly when MGSTM increases from 3 days to 4 days and from 10 days to 11 days, because the optimal PSC network structure changes (see Figure 11) at these points of the Pareto curve.

In the Pareto curve for measuring responsiveness with ELT, the total cost ranges from $ 1.72 MM to $ 2.35 MM, whereas the ELT ranges from 2.125 days to 8 days. We can see a similar trend as in the other curve that the total cost decreases as the ELT increases. By comparing these two curves, we can see the one using guaranteed service approach lies above the curve from the ELT. The reason is that guaranteed service approach is a worst case criterion to measure the response time, whereas ELT is a "weighted" measure that considers both the case of over-stock and the case of stockout. Thus, under the same total cost, the
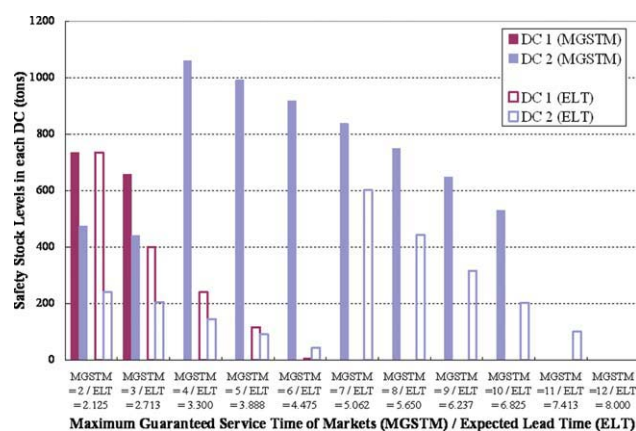

Figure 10. Safety stock levels at DCs for the acetic acid supply chain with single-echelon inventory under the two measures for PSC responsiveness, ELT, and MGSTM.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]
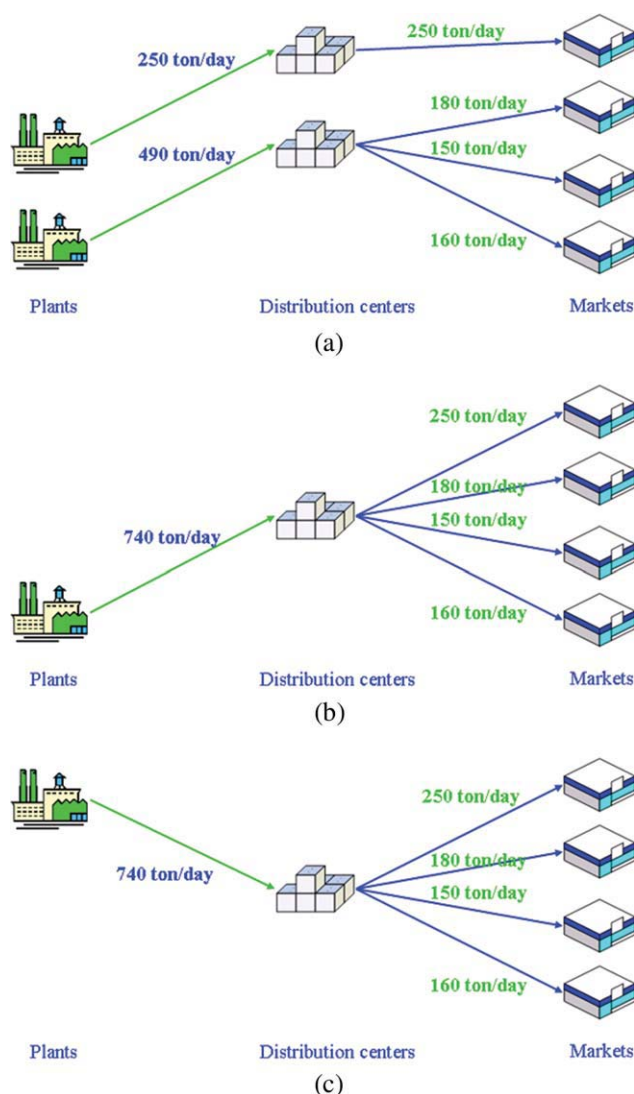
Figure 11. Optimal acetic acid supply chain network structure (with single-echelon inventory) under different specifications of MGSTM and ELT.

(a) Optimal network for MGSTM = 2, 3 days, ELT = 2.125, 2.713, 3.300, 3.888, 4.475 days. (b) Optimal network for MGSTM = 4–10, 12 days, ELT = 5.062, 5.650, 6.237, 6.825, 7.413, 8 days. (c) Optimal network structure for MGSTM = 11 days. [Color figure can be viewed in the online issue, which is available at wileyonlinelibrary. com.]

MGSTM usually has a larger value than the ELT. Note that the results are consistent with the observations discussed in Section "Illustrative Example: Comparison of Different Measures of PSC Responsiveness" where we consider a PSC with fixed design. For the extreme values of these two curves, we can see that both curves have the same minimum cost solution, which corresponds to the case of holding zero safety stock in all the DCs, although the corresponding MGSTM is 12 days whereas the ELT is 8 days. The lower bound of the MGSTM is 2 days, which corresponds to the maximum delivery lead time in the ELT approach. However, the lower bound of the ELT is not 2 days but 2.125 days,

because the stockout probability is always greater than 2.5% due to the "bounded" safety stock factor corresponding to 97.5% service level. It is interesting to see that in the most "responsive" case ($R = 2$ days and $T = 2.125$ days), these two curves have different total costs, although the optimal network structures are the same (Figure 11). The reason can be found from Figure 10. We can see that DC1 has the same optimal safety stocks in these two instances, whereas safety stocks in DC2 are higher in the case of using the guaranteed service. This is because both measures somewhat consider the "worst case." DC1 is included in the linear supply chain or serves those markets that have dominant effect of the maximum values of R and T. Thus, once DC1 holds the maximum safety stocks to allow MGSTM and ELT reach their minimum, DC2 gains certain flexibility to hold lower safety stock than its maximum level, while still keeping the MGSTM and ELT unchanged. The differences between these two measures allow DC2 to have different levels of flexibility to reduce its safety stocks and the associated total costs. Therefore, these two Pareto optimal curves have different maximum total costs.

Figure 9 shows the change of total safety stock level in the PSC under two measures of responsiveness. We can see that when the ELT is used as the responsiveness measure, the total safety stock level increases significantly when the ELT increases from 4.475 days to 5.062 days. The reason is that the optimal network structure changes with a reduction of the number of DCs from two to one as the ELT increases. This further reveals the trade-off between DC installation cost, transportation cost, and safety stock cost. The detailed safety stock levels in each DC using these two measures of responsiveness are given in Figure 10, whereas the optimal network structures under different specifications of responsiveness are given in Figure 11.

The comparison shows that these two measures of PSC responsiveness, MGSTM and ELT, are inherently consistent. The proposed new measure, MGSTM, represents the "worst case," but it has the advantage that it can be applied to PSCs with multi-echelon structure of the inventory system leading to more accurate results.

Conclusion

In this article, we have proposed a new measure for PSC responsiveness based on the concept of guaranteed service approach that allows modeling the multi-echelon stochastic inventory system of a PSC. This measure is incorporated into a joint PSC network design and inventory control model to tradeoff annualized PSC cost and responsiveness within a bi-criterion optimization approach. Some special model properties of the bi-criterion MINLP optimization model were exploited to speed up the solution for the Pareto-optimal solutions. An illustrative example on an acetic acid supply chain was presented to illustrate the application of the proposed approach and models. In addition, comprehensive comparisons between the proposed measure and the ELT, which is another measure for PSC responsiveness focus on the single-stage inventory system, were presented. The analytical results show that these two measures are consistent, although the proposed measure has the advantage that it can be applied to multi-echelon inventory systems.

Acknowledgments

Notation

Sets/Indices

I = set of plants (suppliers) indexed by i
J = set of candidate DC locations indexed by j
K = set of markets indexed by k

Parameters

$c1_{ij}$ = unit transportation cost from plant i to DC j
$c2_{jk}$ = unit transportation cost from DC j to market k
f_j = fixed cost of installing a DC at candidate location j (annually)
g_j = variable cost coefficient of installing candidate DC j (annually)
$h1_j$ = unit inventory holding cost at DC j (annually)
$h2_k$ = unit inventory holding cost at market k (annually)
R_k = maximum guaranteed service time of market k
SI_i = guaranteed service time of plant i
$t1_{ij}$ = processing time of DC j if it is served by plant i, including material handling time of DC j, transportation time from plant i to DC j, and inventory review period
$t2_{jk}$ = processing time of market k if it is served by DC j, including material handling time of DC j, transportation time from DC j to market k, and inventory review period
μ_k = mean demand at market k (daily)
σ_k^2 = variance of demand at market k (daily)
χ = days per year (to convert daily demand and variance values to annual costs)
$\theta1_{ij}$ = unit cost of pipeline inventory from plant i to DC j (annually)
$\theta2_{jk}$ = unit cost of pipeline inventory from DC j to market k (annually)
$\lambda1_j$ = safety stock factor of DC j
$\lambda2_k$ = safety stock factor of market k

Binary variables (0-1)

X_{ij} = 1 if DC j is served by plant i, and 0 otherwise
Y_j = 1 if we install a DC in candidate site j, and 0 otherwise
Z_{jk} = 1 if market k is served by DC j, and 0 otherwise

Continuous variables (0 to ∞)

L_k = net lead time of market k
N_j = net lead time of DC j
R_k = guaranteed service time of market k
R = maximum guaranteed service time of markets (measure of responsiveness)
S_j = guaranteed service time of DC j to its successive markets
T = ELT of the PSC network (measure of responsiveness)
SS_j = safety stock level at DC j
U_j = replenishment lead time of DC j
φ_j = standardized normal variable of DC j
P_j = stockout probability of DC j
XZ_{ijk} = auxiliary variable
NZV_j = auxiliary variable
SZ_{jk} = auxiliary variable
$SZ1_{jk}$ = auxiliary variable
NZ_{jk} = auxiliary variable
$NZ1_{jk}$ = auxiliary variable
UZ_{jk} = auxiliary variable
$UZ1_{jk}$ = auxiliary variable
PX_{ij} = auxiliary variable
$PX1_{ij}$ = auxiliary variable

Literature Cited

1. Varma VA, Reklaitis GV, Blau GE, Pekny JF. Enterprise-wide modeling and optimization: an overview of emerging research challenges and opportunities. *Comput Chem Eng.* 2007;31:692–711.
2. Fisher ML. What is the right supply chain for your product? *Harvard Bus Rev.* 1997;752:105–116.
3. Grossmann IE. Enterprise-wide optimization: a new frontier in process systems engineering. *AIChE J.* 2005;51:1846–1857.
4. You F, Grossmann IE. *Optimal design and operational planning of responsive process supply chains.* In: Papageorgiou L. G., Georgiadis M. C., editors. *Process System Engineering: Volume 3: Supply Chain Optimization.* Weinheim: Wiley-VCH, 2010:56:419–440.
5. You F, Grossmann IE. Design of responsive supply chains under demand uncertainty. *Comput Chem Eng.* 2008;3212:2839–3274.
6. Sabath R. Volatile demand calls for quick response. *Int J Phys Distrib Logist Manage.* 1998;289:698–704.
7. Rotstein G, Shah N, Sorensen E, Macchietto S, Weiss RA. Analysis and design of paint manufacturing processes. *Comput Chem Eng.* 1998;22:S279–S282.
8. Shah N. Process industry supply chains: advances and challenges. *Comput Chem Eng.* 2005;296:1225–1236.
9. Shah N. Pharmaceutical supply chains: key issues and strategies for optimization. *Comput Chem Eng.* 2004;286:929–941.
10. Tahmassebi T. Issues in the management of manufacturing complexity. *Comput Chem Eng.* 1999;23S:S907–S910.
11. Voudouris VT. Mathematical programing techniques to debottleneck the supply chain of fine chemical industries. *Comput Chem Eng.* 1996;20:S1269–S1274.
12. Minner S. Multiple-supplier inventory models in supply chain management: a review. *Int J Prod Econ.* 2003;81–82:265–279.
13. Minner S. Strategic safety stocks in reverse logistics supply chains. *Int J Prod Econ.* 2001;71:417–428.
14. Inderfurth K, Minner S. Safety stocks in multi-stage inventory systems under different service measures. *Eur J Oper Res.* 1998;106:57–73.
15. Inderfurth K. *Valuation of leadtime reduction in multi-stage production systems.* In: Fandel G, Gulledge T, Jones A, editors. *Operations Research in Production Planning and Inventory Control.* Springer: Berlin, Germany, 1993:413–427.
16. Inderfurth K. Safety stock optimization in multi-stage inventory systems. *Int J Prod Econ.* 1991;24:103–113.
17. Graves SC, Willems SP. *Supply chain design: safety stock placement and supply chain configuration.* In: de Kok AG, Graves SC, editors. *Handbooks in Operations Research and Management Science.* North-Holland, Amsterdam: Elsevier, 2003;11:95–132.
18. Graves SC, Willems SP. Optimizing strategic safety stock placement in supply chains. *Manuf Serv Oper Manage.* 2000;21:68–83.
19. Graves SC. Safety stocks in manufacturing systems. *J Manuf Oper Manage.* 1988;1:67–101.
20. You F, Grossmann IE. Integrated multi-echelon supply chain design with inventories under uncertainty: MINLP models, computational strategies. *AIChE J.* 2010;56:419–440.
21. You F, Grossmann IE. Mixed-integer nonlinear programing models and algorithms for large-scale supply chain design with stochastic inventory management. *Ind Eng Chem Res.* 2008;4720:7802.
22. Zipkin PH. *Foundations of Inventory Management.* Boston, MA: McGraw-Hill, 2000.
23. Graves SC, Willems SP. Optimizing the supply chain configuration for new products. *Manage Sci.* 2005;518:1165–1180.
24. Little JDC. A proof of the queueing formula $L = \lambda W$. *Oper Res.* 1961;9:383–387.
25. Christopher M, Towill D. An integrated model for the design of agile supply chains. *Int J Phys Distrib Logist Manage.* 2001;31:235–246.
26. Holweg M. The three dimensions of responsiveness. *Int J Oper Prod Manage.* 2005;25:603–622.
27. Savelsbergh M, Song J-H. Inventory routing with continuous moves. *Comput Oper Res.* 2007;346:1744–1763.
28. Campbell AM, Clarke LW, Savelsbergh MWP. *Inventory routing in practice.* In: Toth P, Vigo D, editors. *The Vehicle Routing Problem.* Philadelphia, PA: SIAM, 2001:309–330.
29. Hübner R. *Strategic Supply Chain Management in Process Industries: An Application to Specialty Chemicals Production Network Design.* Heidelberg: Springer-Verlag, 2007.
30. Glover F. Improved linear integer programming formulations of nonlinear integer problems. *Manage Sci.* 1975;224:455–460.
31. Mitra K, Gudi RD, Patwardhan SC, Sardar G. Midterm supply chain planning under uncertainty: a multiobjective chance constrained programming framework. *Ind Eng Chem Res.* 2008;4715:5501–5511.
32. Chen C-L, Lee W-C. Multi-objective optimization of multi-echelon supply chain networks with uncertain product demands and prices. *Comput Chem Eng.* 2004;28:1131–1144.

33. Cheng L, Subrahmanian EWAW. Multi-objective decisions on capacity planning and production: inventory control under uncertainty. *Ind Eng Chem Res.* 2004;439:2192–2208.
34. Raj TS, Lakshminarayanan S. Multiobjective optimization in multiechelon decentralized supply chains. *Ind Eng Chem Res.* 2008;4717:6661–6671.
35. Dimkou T, Papalexandri K. A parametric optimization approach for multiobjective engineering problems involving discrete decisions. *Comput Chem Eng.* 1998;22S:S951–S954.
36. Dua V, Pistikopoulos E. Parametric optimization in process systems engineering: theory and algorithms. *Proc Ind Natl Sci Acad.* 2003;693:429–444.
37. Eppen G. Effects of centralization on expected costs in a multi-echelon newsboy problem. *Manage Sci.* 1979;25:498–501.

Appendix: Large-Scale Example for PSC Design with Multi-Echelon Inventory Using MGSTM as Responsiveness Measure

We consider a large-scale acetic acid supply chain example with 10 plants, 50 potential DCs, and 100 markets. The MGSTM is used as the measure of supply chain responsiveness.

The input data of this example is given as follows. The safety stock factors for DCs ($\lambda 1_j$) and markets ($\lambda 2_k$) are the same and equal to 1.96, which corresponds to 97.5% service level in demand is normally distributed. We consider 365 days in a year (χ). The guaranteed service time of the last echelon customer demand zones (R_k) are set to 0. The annual fixed costs ($/year) to install the DCs (f_j) are generated uniformly on $U[150,000, 160,000]$ and the variable cost coefficient (g_j, $/ton year) are generated uniformly on $U[0.01, 0.1]$. The guaranteed service times of the plants (SI_i, days) are set as integers uniformly distributed on $U[7, 10]$. The order processing time ($t1_{ij}$, days) between plants and DCs are generated as integers uniformly distributed on $U[3, 7]$, and the order processing time ($t2_{jk}$, days) between DCs and customer demand zones are generated as integers uniformly distributed on $U[2, 5]$. The unit transportation cost from plants to DCs ($c1_{ij}$, $/ton) and from DCs to customer demand zones ($c2_{jk}$, $/ton) are set to $c1_{ij} = t1_{ij} \times U[0.05, 0.1]$ and $c2_{jk} = t2_{jk} \times U[0.05, 0.1]$. The expected demand ($\mu_i$, ton/day) is generated uniformly distributed on $U[75, 150]$ and its standard deviation (σ_i, ton/day) is generated uniformly distributed on $U[0, 50]$. The daily unit pipeline and safety stock inventory holding costs ($\theta 1_{ij}/\chi$, $\theta 2_{jk}/\chi$, $h1_j/\chi$, and $h2_k/\chi$) are generated uniformly distributed on $U[0.1, 1]$.

The resulting MINLP problem includes 5550 binary variables, 75,250 continuous variables, and 185,350 constraints. Solving this problem directly with a global optimizer is a nontrivial task. To address the computational challenge, we use the Lagrangean relaxation and decomposition algorithm as discussed in our previous work,[20] and integrate it with the ε-constraint method to obtain the Pareto optimal curve and determine the tradeoff between the total annual cost of the PSC design and its responsiveness. The first step of the ε-constraint method is to determine the optimal lower and upper bounds of MGSTM (R), which are 0 days and 18 days, respectively. Then, we solve the problem with fixed values of R from 0 day to 18 days (e.g., 19 instances with increments of 1 day). Solving the model (P1) takes a total of 479,563 CPU seconds for all the 19 instances.

The results are given in Figure A1. The red line in Figure A1 is the Pareto optimal curve of this problem. As can be
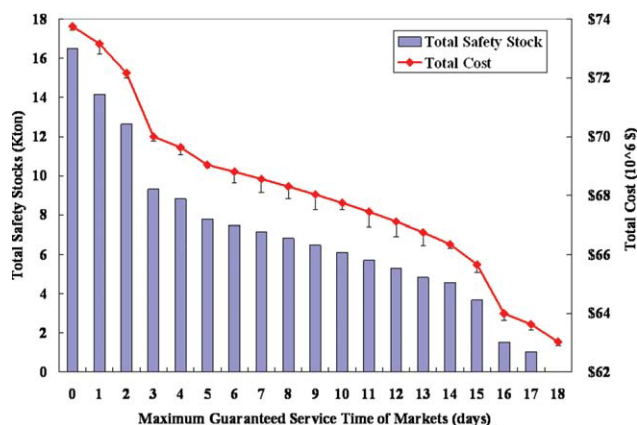

Figure A1. Pareto optimal curve and optimal safety stock levels for the large-scale acetic acid supply chain example with 10 plants, 50 DCs, and 100 markets in Appendix A.

[Color figure can be viewed in the online issue, which is available at wileyonlinelibrary.com.]

seen, the cost ranges from $73,752,953 to $63,028,935, whereas the guaranteed service time (R) ranges from 0 days to 18 days. We can see that the total cost decreases as the MGSTM increases. Thus, the more responsive the PSC is, the more cost it requires. Because of the integer variables and nonconvex terms, there are duality gaps by solving the problems with the Lagrangean relaxation and decomposition algorithm, and these gaps are reflected on the error bars in the Pareto optimal curve. We can see that these gaps are rather small and do not affect the trend of the Pareto optimal curve. The columns in Figure A1 show the total safety stocks in the system under different specifications of MGSTM. As the MGSTM increases, the total safety stocks in this PSC decrease from 6,302,893 tons to 0. It shows that the more inventories we have, the more responsive the PSC is.

Model Formulation for PSC Design with Single-Echelon Inventory Using ELT as Responsiveness Measure

Using ELT as the measure for PSC responsiveness for the design of the acetic acid supply chain leads to the following MINLP model (Q1):

$$\min: \quad \sum_{j \in J} f_j Y_j + \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} A_{ij} X_{ij} Z_{jk}$$
$$+ \sum_{j \in J} \sum_{k \in K} B_{jk} Z_{jk} + \sum_{j \in J} h1_j SS_j \quad (A1)$$

$$\text{Min}: \quad T \quad (A2)$$

s.t.

$$\sum_{i \in I} X_{ij} = Y_j, \quad \forall j \quad (12)$$

$$\sum_{j \in J} Z_{jk} = 1, \quad \forall k \quad (13)$$

$$Z_{jk} \leq Y_j, \quad \forall j, k \quad (14)$$

$$U_j = \sum_{i \in I} \left(SI_i + t1_{ij}\right) \cdot X_{ij}, \quad \forall j \tag{A3}$$

$$SS_j \geq \varphi_j \cdot \sqrt{\sum_{k \in K} \left(\sigma_k^2 \cdot U_j \cdot Z_{jk}\right)}, \quad \forall j \tag{A4}$$

$$P_j \geq 1 - \Phi\left(\varphi_j\right), \quad \forall j \tag{A5}$$

$$T \geq S_{ij} \cdot X_{ij} \cdot P_j + t2_{jk} \cdot Z_{jk}, \quad \forall i, j, k \tag{A6}$$

$$X_{ij}, Y_j, Z_{jk} \in \{0, 1\}, \quad \forall i, j, k \tag{18}$$

$$SS_j \geq 0, U_j \geq 0 \quad \forall j \tag{A7}$$

$$0 \leq P_j \leq 1, \quad 0 \leq \varphi_j \leq \lambda 1_j \quad \forall j \tag{A8}$$

where

$$A_{ijk} = \left(c1_{ij}\chi + \theta 1_j t1_{ij}\right) \cdot \mu_k$$

$$B_{jk} = \left(g_j\chi + c2_{jk}\chi + \theta 2_k t2_{jk}\right) \cdot \mu_k$$

$$\Phi(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} \exp\left(-\frac{x^2}{2}\right) dx$$

.

The objective function (A1) represents the total annual cost of the PSC design. The first term in (A1) is the annual DC installation cost, the second and the third terms are the total annual transportation and pipeline inventory cost, and the annual total safety stock cost in the DCs is given by the fourth term, where SS_j is the safety stock level in DC j. The objective function (A2) is to minimize the total ELT of the entire PSC (T). U_j is the replenishment lead time of DC j, and constraint (A3) defines its value. The variance of the demand over the replenishment lead time of DC j is given by $\sum_{k \in K} \left(\sigma_k^2 \cdot U_j \cdot Z_{jk}\right)$, which takes into account the risk-pooling effect.[37] Constraint (A4) shows that safety stock level (SS_j) at DC j is equal to the standard deviation of its demand over the lead time times the standardized normal variables ϕ_j, which is further used to define the stockout probability (P_j) in constraint (A5). In addition, constraint (46) restricts the standardized normal variables ϕ_j to be non-negative and not greater than the safety stock factor $\lambda 1_j$, so as to be consistent with the data used in the guaranteed service approach.

Note that we can similarly linearize[30] the bilinear terms (products of binary variables and continuous variables, $U_j \cdot Z_{jk}$ and $P_j \cdot X_{ij}$) with new variables UZ_{jk} and PX_{ij} using the following linear inequalities:

$$UZ_{jk} + UZ1_{jk} = U_j, \quad \forall j, k \tag{A9a}$$

$$UZ_{jk} \leq Z_{jk} \cdot U_j^U, \quad \forall j, k \tag{A9b}$$

$$UZ1_{jk} \leq \left(1 - Z_{jk}\right) \cdot U_j^U, \quad \forall j, k \tag{A9c}$$

$$UZ_{jk} \geq 0, \quad UZ1_{jk} \geq 0 \quad \forall j, k \tag{A9d}$$

$$PX_{ij} + PX1_{ij} = P_j, \quad \forall i, j \tag{A10a}$$

$$PX_{ij} \leq X_{ij}, \quad \forall i, j \tag{A10b}$$

$$PX1_{ij} \leq 1 - X_{ij}, \quad \forall i, j \tag{A10c}$$

$$PX_{ij} \geq 0, \quad PX1_{ij} \geq 0 \quad \forall i, j \tag{A10d}$$

.

To obtain the optimal lower bound of the ELT (T), we solve problem (Q1) by minimizing (A2). Similarly, to obtain the optimal upper bound, we consider a new objective function:

$$\text{Min}: \quad \sum_{j \in J} f_j Y_j + \sum_{i \in I} \sum_{j \in J} \sum_{k \in K} A_{ijk} X_{ij} Z_{jk} + \sum_{j \in J} \sum_{k \in K} B_{jk} Z_{jk}$$
$$+ \sum_{j \in J} h1_j SS_j + \varepsilon \cdot T \tag{A11}$$

where ε is a scaling parameter with sufficient small value, for instance, 0.001. By minimizing (A11) subject to all the constraints in (Q1), we can determine the minimum upper bound of the ELT (T).