Stochastic Modeling for Inventory and Production Planning in the Paper Industry

K. Karen Yin
Dept. of Bio-based Products, University of Minnesota, St. Paul, MN 55108

G. George Yin
Dept. of Mathematics, Wayne State University, Detroit, MI 48202

Hu Liu
3M Commerce Services, 3M Center, St. Paul, MN 55144

DOI 10.1002/aic.10251
Published online in Wiley InterScience (www.interscience.wiley.com).

Problem formulations and solution procedures of production planning and inventory management for manufacturing systems under uncertainties is discussed. Markov decision processes and controlled Markovian dynamic systems are used in the models. Considering an inventory problem in discrete time and formulating it by a finite-state Markov chain lead to a Markov decision process model. Using the policy-improvement algorithm yields the optimal inventory policy. In controlled dynamic system modeling, the random demand and capacity processes involved in planning are described by two finite-state continuous-time Markov chains. Such an approach enables us to embed the randomness in the differential equations of the system. The optimal production rates that minimize an expected cost are obtained by numerically solving the corresponding Hamilton–Jacobi–Bellman (HJB) equations. To overcome the so-called curse of dimensionality, frequently encountered in computation, we resort to a hierarchical approach. Illustrative examples using data collected from a large paper manufacturer are provided. © 2004 American Institute of Chemical Engineers *AIChE J,* 50: 2877–2890, 2004
Keywords: inventory management; production planning; Markov chain; optimal policy; hierarchical approach

Introduction

In the pulp and paper industry, the search for good inventory control policies and production plans started several decades ago. Many models have been developed and used (see Leiviskä, 1999 and references therein). Nevertheless, similar to the situation in many other industries, inventory management and production planning remain to be challenging problems to paper manufacturers.

Correspondence concerning this article should be addressed to K. K. Yin at kyin@umn.edu.

Production planning and inventory management are crucial components of any supply chain. Efficient planning and inventory policies are necessary for the successful operation of any modern-day enterprise. It is well known that manufacturing systems are subject to random events such as raw material variation, demand fluctuation, and equipment failures. The dynamic and random nature of the demands makes their forecasting very difficult or sometimes impossible. Despite the existence of the various models, very often managers cannot find a single one that is suitable to their needs. As a result, decisions on inventory have to be based on a combination of experience, mathematical models, and even on the gut feeling of a few individuals, whereas production is planned following an everyday practice without concerning the optimality. It is desirable to shift such experience-based decision making to an information-based decision making model. This will require a systematic use of historical data and a theoretically sound mathematical model that is applicable to the real situation. This work is intended to contribute in this direction.

A large manufacturer/corporation usually has a certain number of plants/mills at different geographical locations. Having a collection of machines/equipment, each plant has its own inventory with a fixed carrying cost. Characterized by its production rates for a given product type, each machine can produce multiple products. At any given time, the enterprise has a set of orders in hand. Each order specifies a customer, a product, an order quantity, and a due date. Orders are fulfilled on one or more machines of specified plant(s). To minimize the setup costs and disruptions arising from transitions, it is customary to batch orders of a common product category together to form production runs, each of which consists of a specified subset of the current orders. In the process industries, scheduling defines a sequence of production runs for all machines, their starting and ending times as well as the volumes to be produced. The strategic planning has a long-term horizon, whereas short-term planning focuses on making decisions in a time span of only a few days to a few months.

Inventory management entails managing product stocks, in-process inventories of intermediate products as well as inventories of raw material, equipment and tools, spare parts, supplies used in production, and general maintenance supplies. In a broader sense, it constitutes all kinds required to run a business including factors such as storage, personnel, cash, and transportation facilities. A manufacturing company needs an inventory policy for each of its products to govern when and how much it should be replenished. Good inventory management offers the potential not only to cut costs but also to generate new revenues and higher profits. On the contrary, undersupply causes stockout and leads to loss of sales, whereas oversupply hinders free cash flow and may cause forced markdowns. Improper inventory policies will not only result in diminished earnings but can also generate a sufficiently negative impact that it renders a company nonprofitable.

Scheduling, planning, and inventory management have attracted growing attention in many industries. Numerous articles in the area of design, operation, and optimization of both batch and continuous plants have been published (see, for example, Applequist et al., 1997; Bassett et al., 1997; Gupta et al., 2000; Yan et al., 1994; Yin and Zhang, 1997; Yin et al., 1995, 2002 and references therein). There are many criteria by which a given plan may be evaluated. Issues such as manufacturing efficiency, on-time delivery as well as other customer-service requirements, inventory limits, and distribution costs need to be considered. These competing objectives should be prioritized according to the company's near-term economic consideration and its long-term strategic plan. The main objective of inventory management is to increase profitability. A frequently used criterion for choosing the optimal policy is to minimize the total costs, which is equivalent to maximizing the net income in many cases. In an inventory system, the main costs that affect profit may include such factors as manufacturing cost, holding cost, shortage cost, salvage cost, and discount rates. To seek inventory policies that enable a company to maintain a high level of customer service at a minimum cost requires evaluating and balancing these competing factors. The average inventory level during a period has a direct impact on the carrying costs. Given that inventory stock represents a considerable investment, it is desirable to maintain it at the lowest possible level, provided that the customer service can be guaranteed.

Inventory and production planning require making timely decisions for a manufacturing system, in which many different types of events may occur at the same time. The complexity of the system under consideration; the dynamic nature of the process; the close and complicated relationship among resource and production planning, scheduling, and product inventory management; the large number of competing alternatives; and the huge amount of information needed make the task very challenging. Additionally, limitations and constraints on the production procedure, capacity, and storage space, as well as the multiple options of production, add more difficulties in the decision-making process. In particular, various uncertainties such as the randomness that originates from raw material variation, equipment failures, and demand fluctuation, and such unpredictable events as potential technological advances and global market uncertainty render decision making very difficult. Breakdowns of papermaking machinery, frequently encountered in paper mills, are typical examples of such random events that affect the production capacity. To better understand and more effectively deal with uncertainties from various sources require sound mathematical models. The models must be able to capture the salient aspects of the system and to characterize the unique feature of each major event and, at the same time, it must permit the use of efficient algorithms that can scale-up effectively to handle large-scale problems.

This article reports on discrete-time Markov decision processes and continuous-time dynamic systems involving Markov chains, both to model and to solve the inventory management and production planning in the paper industry. The rationale is as follows. In production and inventory planning, we deal with stochastic processes, each of which may be considered as an ensemble of random variables, defined on a common probability space and evolving over time. The observed data are statistical time series, which are single realizations of the underlying process. Considering that, in practice, stock replenishments take place every week or every few weeks, we treat inventory management as a discrete-time problem. As for the production planning, similar to many other processes in the chemical and process industries, pulping and papermaking are continuous. Therefore we model such dynamic systems using continuous-time dynamic systems. To deal with the inherent jumps in the systems of interest, we resort to Markovian models for solution.

Markov chains have been used in the design, optimization, and control of queuing systems, manufacturing processes, reliability studies, and communication networks, where the underlying systems are formulated as a stochastic control problem driven by Markovian noise. Markovian formulations have proved useful in solving numerous real-world problems under uncertainties, such as determining the inventory levels for retailers, maintenance scheduling for manufacturers, and scheduling and planning in production management. This communication is concerned with two stochastic optimization methods and their uses in inventory management and production planning. Using Markov decision process models, we model the product stock on hand by a finite-state, discrete-time Markov chain and develop optimal inventory policies for the finished products. Modeling the random demand and capacity processes by two finite-state, continuous-time Markov chains, we seek the optimal production rate by minimizing an expected cost of the system in production scheduling. The rationale, problem formulation, and solution procedures are discussed and compared. Examples are presented for illustration.

The rest of the article is organized as follows. A brief review of the Markov decision process model, its application to the inventory decision making, and the use of a policy-improvement procedure to obtain the optimal policy are presented in Section 2 along with an illustrative example. The problem formulation and numerical method for production planning is detailed. The hierarchical approach, its corresponding procedure, and the near-optimal solution is discussed. An application example using real industrial data is also provided for illustration, followed by further discussion and a summary.

Markov Decision Processes and Inventory Planning

Inventory management requires making decisions on whether and how much to replenish at a given time. Since the first inventory model (Wilson, 1934) appeared in the literature more than 70 years ago, various models have been developed and used, such as the economic order quantity model and its modifications, multiple reorder system models, and a replenishment model and many of its variations (see Buffa, 1980; Hopp and Spearman, 1996; Zipkin, 2000). To deal with demand fluctuation, forecasting product demand, using mathematical models derived from historical data, has become a common practice in industries. This can lead to reasonably good prediction, provided that the demand variation is relatively small or that certain trends, such as seasonality, are easily identifiable. The basic assumption of forecasting is that markets and demands are predictable within a certain range of accuracy. Given the rapid and often unpredictable changes in today's global economy, numerous uncertainties involved in the dynamic process, as well as the complicated relationships among the elements and links of any given supply chain, nevertheless in many cases this assumption is not true, which renders the forecasting unreliable or totally mistaken. As a result, production and inventory decisions can no longer be made solely based on forecasting alone. Other tools are needed to adequately address the variability issue. Similar to many other dynamic processes in the real world, inventory systems have uncertainty associated with them. In the meantime, they exhibit a certain degree of regularity. It is desirable to incorporate both variability and regularity into mathematical models and treat them quantitatively from a probability perspective. In this work, we resort to Markov decision process (MDP) models for solution.

Introduced in the 1950s (see Puterman, 1994 for references and a historical note), discrete-time Markov decision processes, which use stochastic dynamic programming, are models for sequential decision making when outcomes are uncertain. In the past decades, MDPs have become an important research area in both theory and applications. MDP models have been used to solve real-world problems in ecology, economics, communication, electrical engineering, and computer science among other fields. Making decisions on inventory control, queuing, stock options, resource allocations, and machine maintenance are several of the many application examples. The key components in this sequential decision model include a set of states of the system; a set of decisions and their corresponding actions; and rewards/costs and transition probabilities, both dependent on the state and the action.

We are interested in inventory policies capable of handling situations in which the demand during a period is random, and where stock replenishments take place periodically (such as every week, every 2 weeks, or once every month, etc.). Assume that the total aggregate demands for a specific product during any given period n is a random variable d_n. Given the randomness and regularity in the inventory process, we may describe it with discrete-time, finite-state Markov chains. We need to specify the key components of the sequential decision model, which entails designating the state space of the Markov chain, the possible decisions, their corresponding actions, policies, and the rewards/costs structure. It is conceivable that the evolution of the system is affected by the random demands as well as the replenishment activities that are governed by the inventory policy. To quantify this requires the chain's transition probabilities.

Discrete-time Markov chains and transition probabilities

Recall that a stochastic process is a Markov chain if it possesses the Markovian properties and its state space is finite or countable. The Markov property states that, given the current state, the probability of the chain's future behavior is not altered by any additional knowledge of its past behavior. Let $\alpha(t)$ be a Markov chain with a finite-state space $\mathcal{M} = \{0, \ldots, m\}$. For discrete time cases with $t \in T = \{1, 2, \cdots\}$, the Markov property states that $P\{\alpha(t+1) = j \,|\, \alpha(1) = i_1, \ldots, \alpha(t) = i\} = P\{\alpha(t+1) = j \,|\, \alpha(t) = i\}$, whereas for continuous time cases with $t \geq 0$, the Markov property states that for each $i \in \mathcal{M}$, the conditional probability $P\{\alpha(t) = i \,|\, \alpha(r) : r \leq s\} = P\{\alpha(t) = i \,|\, \alpha(s)\}$.

Each element $P_{ij}^{n,n+1}$ $(i, j = 0, 1, \ldots, m)$ in the one-step transition matrix denotes the probability of the chain's transition from state i to state j at the period of n to $n+1$. Note that for time-homogeneous Markov chains, the one-step transition probabilities are independent of n, and thus the superscripts are often omitted for brevity. For an irreducible and aperiodic Markov chain, it can be shown that the limit of the chain's n-step transition probability, $P_{ij}^{(n)}$, exists and is independent of the starting state i. Denote this limiting probability distribution by π_j, that is

$$\pi_j = \lim_{n\to\infty} P_{ij}^{(n)} \qquad \text{for } i, j = 0, 1, \ldots, m \tag{1}$$

Also known as the stationary distribution, π_j satisfies

$$\pi_j \geq 0 \qquad \pi_j = \sum_{i=0}^{m} \pi_i P_{ij} \qquad \text{and} \qquad \sum_{j=0}^{m} \pi_j = 1 \tag{2}$$

This stationary distribution is needed to obtain the optimal inventory policy for our inventory planning problem. To completely define a Markov process requires specifying its initial state (or the initial probability distribution, in general) and its transition probability matrix.

Let d_1, d_2, $\cdots$ represent the demands (in kg) for a particular product during the first week, the second week, and so forth. Assume that d_n are independent, identically distributed random variables whose future values are unknown. Let $\tilde{X}_n$ denote the stock of certain product on hand at the end of the nth week. The states of the stochastic process, $\{\tilde{X}_n\}$, consist of all possible values of its stock size. The stock levels at two consecutive periods are related by the current demand d_n and the inventory policy chosen, such as under the well-known (s, S) policy, which requires replenishing to S if the stock level is lower than s; otherwise do not replenish, so that the dynamics of the stock satisfy

$$\tilde{X}_{n+1} = \begin{cases} \tilde{X}_n - d_{n+1} & \text{if } s < \tilde{X}_n \leq S \\ S - d_{n+1} & \text{if } \tilde{X}_n \leq s \end{cases} \tag{3}$$

Because the successive demands d_1, d_2, $\cdots$ are independent random variables, the amounts in stock, $\tilde{X}_0$, $\tilde{X}_1$, $\cdots$ constitute a Markov chain whose transition probability matrix can be calculated according to relation 3. The weight in stock at time n, $\tilde{X}_n$, may take values in a subset of $\mathcal{R}$ (that is, it is a continuous random variable). To simplify the solution procedure, we discretize the state space, which results in a discrete random variable, X_n. Such a procedure enables us to model the inventory system by a discrete-time, finite-state Markov chain. The random variable X_n indicates the stock level at the end of the nth period and takes values in a finite-state space, denoted by $\mathcal{M} = \{0, \ldots, m\}$. Additional information and detailed description of discrete-time, finite-state Markov chains can be found in many publications on stochastic processes (see, for example, Ross, 2000).

Decisions, actions, and policies

The state of the system, characterized by the stock levels X_n, is governed by both the random demand and the replenishment action taken. The stock on hand at the end of each period is recorded. Subsequently, a decision is made and an action is taken. At the end of each period the stock level may undergo $m + 1$ possible events: It may jump from the current level, i, to a higher one, j ($i < j \leq m$); it may fall into a lower level k ($0 \leq k < i$); or it may stay in the same state. For a given inventory system, there are a number of possible decisions, each having its corresponding action. We label all possible decisions by $k = 0, 1, \ldots, K$. The question that needs to be answered is which decision should be chosen at any given time and state. In other words, an inventory policy is needed.

A policy is a rule that prescribes decisions to be made for each state of the system during the entire time period of interest. There are a number of possible policies for each problem. Characterized by the values $\{\delta_0(R), \delta_1(R), \ldots, \delta_m(R)\}$, any policy R specifies decisions $\delta_i(R) = k$ ($k = 0, 1, \ldots, K$) for all states i ($i = 0, 1, \ldots, m$) at every time instant. We consider stationary policy only, which requires that the decision $\delta_i(R)$, be made whenever the system is in state i, regardless of time. Affected by this policy as well as the random demand, the system will move to a new state j according to the corresponding probabilities $P_{ij}(k)$.

Markov decision processes

The theory of discrete-time Markov decision processes studies sequential optimization of discrete-time stochastic systems, whose transition mechanisms are action or decision dependent. Each control policy has both immediate and future impacts on the dynamics of the underlying stochastic process: it determines values of the objective functions of the process. The goal is to choose an optimal control policy for the system of interest.

For an inventory system, let the stock on hand at the end of the nth period, X_n, be the state of the system; let Δ_n be the decision/action chosen. Under any given policy R, the pair $Y_n = (X_n, \Delta_n)$ forms a two-dimensional Markov chain whose transition probabilities are obtained by an application of the law of total probability

$$P[X_{n+1} = j, \Delta_{n+1} = k_1 | X_n = i, \Delta_n = k] = P_{ij}(k)p(k_1|j) \tag{4}$$

where $P_{ij}(k)$ is the conditional probability of the chain moving to state j at time $n + 1$, provided that the current state is $X_n = i$ and a decision $\Delta_n = k$ is taken, and $p(k_1 | j)$ is the probability of a specific decision $\Delta_{n+1} = k_1$ being chosen at a particular state $X_{n+1} = j$. For a given stationary feedback policy, the decision $\delta_i(R) = k$ is prescribed for every state $i = 0, 1, \ldots, m$, and thus $p(k | i) = 1$. Consequently, when the system is in state i and an action based on the decision $\delta_i(R) = k$ is excised, the probability of its moving to state j at the next time period is given by

$$P[X_{n+1} = j | X_n = i, \delta_i(R) = k] = P_{ij}(k) \tag{5}$$

Starting from X_0, the realization of the underlying stochastic process is X_0, X_1, $\cdots$ and the decisions made are Δ_0, Δ_1, $\cdots$. Note that $\Delta_n = \delta_{X_n}(R) \in \{0, 1, 2, \ldots, K\}$ if the feedback policy is used. It can be seen that the inventory problem, consisting of a sequence of observed states and decisions, fits to the general framework of a finite-state, discrete-time Markov decision process.

Among the many candidate policies, we seek the "optimal" one in the sense that it will minimize the (long-run) expected average cost per unit time. It should be noted that one of the considerations in practice is that the policy should be relatively simple and easily implementable.

Suppose a cost C_{X_n,Δ_n} is incurred when the process is in state X_n and a decision Δ_n is made. Assume that the Markov chain or its corresponding transition matrix $P(k) = [P_{ij}(k)]$ is irreducible. The long-run expected average cost per unit time can be written as

$$\lim_{n\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} E[C_{X_n,\Delta_n}] = \sum_{i=0}^{m} \sum_{k=0}^{K} \pi_{ik} C_{ik} \tag{6}$$

where π_{ik} is the stationary (limiting) probability distribution associated with the transition probabilities in Eq. 5. Our objective is to find a policy that minimizes the long-run expected average cost (Eq. 6).

The policy-improvement algorithm

A policy R can also be written in a matrix form, as follows

$$D = \begin{pmatrix} D_{00} & D_{01} & D_{02} & \cdots & D_{0K} \\ D_{10} & D_{11} & D_{12} & \cdots & D_{1K} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ D_{m0} & D_{m1} & D_{m2} & \cdots & D_{mK} \end{pmatrix} \tag{7}$$

The first subscript i of any element D_{ik} in the matrix D represents the state and the second subscript k stands for the decision. Observe that D_{ik} take values of 0 and 1 only. $D_{ik} = 1$ means $\delta_i(R) = k$, which calls for decision k and its corresponding action when the system is in state i, whereas $D_{jk} = 0$ means $\delta_j(R) = 0$, that is, no action will be taken when the system is in state j.

Because for any given policy, the decision to be made and the action to be taken in any state i has been specified, Eq. 6 can be further simplified by replacing π_{ik} with π_i, yielding

$$\lim_{n\to\infty} \frac{1}{N} \sum_{n=0}^{N-1} E[C_{X_n,\Delta_n}] = \sum_{i=0}^{m} \sum_{k=0}^{K} \pi_i C_{ik} \tag{8}$$

The problem can be solved by the *policy-improvement algorithm*. Let $g(R)$ represent the long-run expected average cost per unit time following any given policy R, that is

$$g(R) = \sum_{i=0}^{m} \pi_i C_{ik} \tag{9}$$

Denote by $v_i^n(R)$ the total expected cost of a system starting in state i and evolving in a period of length n. By definition, it satisfies the following recursive formula

$$v_i^n(R) = C_{ik} + \sum_{j=0}^{m} P_{ij}(k) v_j^{n-1}(R) \tag{10}$$

which means that $v_i^n(R)$ consists of two parts, the cost incurred in the first time period, C_{ik}, and the total expected costs thereafter. Note that C_{ik} is an expected cost and

$$C_{ik} = \sum_{j=0}^{m} q_{ij}(k) P_{ij}(k) \tag{11}$$

where $q_{ik}(k)$ and $P_{ik}(k)$ are the expected cost and the probability when the system moves from state i to state j after decision k is made, respectively. It can be shown, as in Hillier and Lieberman (1990), as $n \to \infty$, $v_i^n(R) \to v_i(R)$, and as a result

$$g(R) = C_{ik} - v_i(R) + \sum_{j=0}^{m} P_{ij}(k) v_j(R) \qquad \text{for } i, j = 0, 1, \ldots m \tag{12}$$

For a system of $m + 1$ states, Eq. 12 consists of $m + 1$ simultaneous equations but $m + 2$ unknowns, $g(R)$ and $v_i(R)$ ($i = 0, 1, \ldots, m$). To obtain a unique solution, it is customary to specify $v_m(R) = 0$. Solving the system in Eq. 12 yields the long-run average expected cost per unit time $g(R)$ if the policy R is used. An optimal policy is one that results in the lowest cost $g(R^*)$. Using the policy-improvement algorithm, an iteration procedure consisting of two steps, allows us to obtain the optimal policy. Given that there are only a finite number of possible stationary policies when the state space is finite, we will be able to reach the optimal policy in a finite number of iterations (Ross, 1983). For more detailed accounts for the MDP models, the policy improvement algorithm, and their applications in determining inventory policies, the readers are referred to Hillier and Lieberman (1990) and Puterman (1994).

An application example

To implement the method outlined above, all possible decisions and the corresponding actions must first be designated. The Markov decision process model of the underlying system is established by defining its state space and the transition probabilities, specifying the cost structure and evaluating its individual component. Then using the policy-improvement algorithm will yield the optimal policy. An illustrative example of making inventory decisions from real sales data is provided in this section. The data were collected from a large paper manufacturer, where the inventory of each product is checked weekly to determine whether and how much this item should be produced. An order is placed upon determination. It usually takes 3 weeks to fill an order. Therefore we consider the lead time to be a constant of 3 weeks. Other quantities necessary in making inventory decisions are the size of the buffer stock, the service level, and the replenishment amount. Their determination and other related information can be found in Zipkin (2000).

Our database includes customer demands for several hundred products during a period of 18 months. To protect private commercial information, the original data have been rescaled with the main features retained. A careful examination showed that demand for some of the products had trends and/or seasonality, whereas many of them exhibited random, unpredictable, and even erratic changes: the ratio of the standard deviation to the mean was as high as 9. Such random behavior renders forecasting impossible.

In practice, there is usually a minimum order/production amount u. The choice of u will affect the state space of the Markov chain and hence the final policy. In this work, we choose u to be the mean value of the 3-week demands, and assume the amounts of replenishment to be u or its multiples. Let $s \geq 0$ and $S > s$ be a low and the highest possible levels of the inventory, respectively. Let

$$X_n = \left\lfloor \frac{\tilde{X}_n - s}{u} \right\rfloor \qquad \text{and} \qquad m = \left\lfloor \frac{S - s}{u} \right\rfloor \tag{13}$$

where $\lfloor x \rfloor$ denotes the integer part of x. The discrete random variable X_n indicates the level of the stocks, which takes values in $M = \{0, \ldots, m\}$. Such discretization allows us to model this inventory system by an $(m + 1)$-state Markov chain.

Table 1. States, Decisions, and Actions (P-6)

State	Inventory Amount ($\tilde{X}_i \times 10^{-4}$ kg)	Decision	Action ($\Delta_i \times 10^{-4}$ kg)
0	$0 < \tilde{X}_i \leq 5.21$	0	Do not replenish
1	$5.21 < \tilde{X}_i \leq 7.31$	1	Replenish 2.1
2	$7.31 < \tilde{X}_i \leq 9.41$	2	Replenish 4.2
3	$9.41 < \tilde{X}_i$	3	Replenish 6.3

The first two columns in Table 1 show the states of the chain of a product named P-6, for which $M = \{0, \ldots, 3\}$ and $u =$ 21,000 kg. The upper bound of state 0, 52,100 kg, is the level of the buffer stock. We prescribed four decisions, coded 0, 1, 2, and 3, each corresponding to an action of replenishing the amount of (21,000 kg × the decision code): do not replenish, replenish 21,000, 42,000, or 63,000 kg.

As mentioned in the previous sections, under a given policy the pair of random variables $Y_n = (X_n, \Delta_n)$ forms a two-dimensional Markov chain. To evaluate a policy requires $P_{ij}(k)$. For a system having four states and four decisions/actions, there are four (4 × 4) transition probability matrices to be evaluated. We denote them by $P(k)$, and $P(k) = [P_{ij}(k)]$ ($i = 0, 1, \ldots, 3$, $j = 0, 1, \ldots, 3$, $k = 0, 1, \ldots, 3$). The elements in the matrix $P(k)$ are the transition probabilities $P_{ij}(k)$ under decision k. Equation 14 illustrates the determination of the transition probabilities $P_{ij}(0)$ for a four-state system if the policy prescribes decision 0 for all states, and thus there is no replenishment at any time. The other three matrices $P(1)$, $P(2)$, and $P(3)$ can be similarly obtained. Equation 15 presents the four transition probability matrices of product P-6 under the four decisions tabulated in Table 1.

$$P(0) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ P\left\{d > \frac{u}{2}\right\} & P\left\{d \leq \frac{u}{2}\right\} & 0 & 0 \\ P\left\{d > \frac{3u}{2}\right\} & P\left\{\frac{u}{2} < d \leq \frac{3u}{2}\right\} & P\left\{d \leq \frac{u}{2}\right\} & 0 \\ P\left\{d > \frac{5u}{2}\right\} & P\left\{\frac{3u}{2} < d \leq \frac{5u}{2}\right\} & P\left\{\frac{u}{2} < d \leq \frac{3u}{2}\right\} & P\left\{d \leq \frac{u}{2}\right\} \end{pmatrix} \quad (14)$$

$$P(0) = \begin{pmatrix} 1 & 0 & 0 & 0 \\ 0.59 & 0.41 & 0 & 0 \\ 0.22 & 0.37 & 0.41 & 0 \\ 0.11 & 0.11 & 0.37 & 0.41 \end{pmatrix}$$

$$P(1) = \begin{pmatrix} 0.59 & 0.41 & 0 & 0 \\ 0.22 & 0.37 & 0.41 & 0 \\ 0.11 & 0.11 & 0.37 & 0.41 \\ 0 & 0.11 & 0.11 & 0.78 \end{pmatrix}$$

$$P(2) = \begin{pmatrix} 0.22 & 0.37 & 0.41 & 0 \\ 0.11 & 0.11 & 0.37 & 0.41 \\ 0 & 0.11 & 0.11 & 0.78 \\ 0 & 0 & 0.11 & 0.89 \end{pmatrix}$$

$$P(3) = \begin{pmatrix} 0.11 & 0.11 & 0.37 & 0.41 \\ 0 & 0.11 & 0.11 & 0.78 \\ 0 & 0 & 0.11 & 0.89 \\ 0 & 0 & 0 & 1 \end{pmatrix} \quad (15)$$

In an inventory system, the main costs that affect profit may include manufacturing cost, holding cost, shortage cost, salvage cost, and discount rates. In this work, we consider two components, an average manufacturing cost of $\$C_m$/kg and an average shortage cost of $\$C_s$/kg. The manufacturing costs include all those incurred during the production. The shortage costs are attributed to loss of sales resulting from insufficient inventories. Let d denote the demand. Let Q_k be the ordering amount under decision k; and $\bar{\tilde{x}}_i$ be the average inventory amount when the system is in state i. Then the expected cost

$$C_{ik} = C_m Q_k + C_s \sum_{j=0}^{m} \max\{(d - \bar{\tilde{X}}_i), 0\} P_{ij}(k) \qquad \text{for } i, k = 0, \ldots, 3 \quad (16)$$

where $P_{ij}(k)$ is given in Eq. 15. To obtain the optimal policy, we use the policy-improvement algorithm. The procedure begins by choosing an arbitrary policy R_1. For the given policy R_1, the transition probabilities $P_{ij}(k)$ as well as the expected costs C_{ik_R1} in Eq. 11 can be computed. Subsequently, the values of $g(R_1)$, $v_0(R_1)$, $v_1(R_1)$, $\ldots$, $v_{m-1}(R_1)$ can be obtained from Eq. 12. In the second step, the current values of $v_i(R_1)$ are used to find an improved policy R_2. Specifically, for each state i, choose such decision $\delta_i(R_2)$ that makes the right-hand side of Eq. 12 a minimum. The best decisions for all states $(0, 1, \ldots, m)$ constitute the second, or the improved policy R_2. Repeat this iteration procedure until the two successive R values are the same.

By using the MDP model and the policy-improvement algorithm, we obtain the optimal policy as shown in Eq. 17, in which an entry $D_{ik} = 1$ means that the policy calls for the kth decision if the system is in state i. For example, $D_{11} = 1$ means that decision 1 and its corresponding action (to replenish 21,000 kg) is to be exercised if the system is in state 1 (stock at hand is between 52,100 and 73,100 kg); $D_{20} = D_{30} = 1$ requires choosing decision 0, that is, do not replenish, if the stock at hand is more than 73,100 kg

$$D_{\text{P-6}} = \begin{pmatrix} D_{00} & D_{01} & D_{02} & D_{03} \\ D_{10} & D_{11} & D_{12} & D_{13} \\ D_{20} & D_{21} & D_{22} & D_{23} \\ D_{30} & D_{31} & D_{32} & D_{33} \end{pmatrix} = \begin{pmatrix} 0 & 0 & 1 & 0 \\ 0 & 1 & 0 & 0 \\ 1 & 0 & 0 & 0 \\ 1 & 0 & 0 & 0 \end{pmatrix} \quad (17)$$

Although the concepts and development of the MDP model are seemingly more involved than other commonly used inventory models, once a policy becomes available, however, its implementation is straightforward. It can usually provide us with better results, as illustrated in Figure 1, which compares the MDP policy with the conventional (s, S) inventory policy based on the real demand data for product P-6, of which the weekly mean was 7000 kg. Its actual weekly demand ranges from the lowest value of zero to the highest of 46,000 kg. Such high variability makes the prediction of future demand very difficult. Using conventional policies often leads to stockout or requires a higher inventory level. Because the randomness has been included in the MDP model, better results can be expected. It is indeed the case as shown in our comparison.

Figure 1 shows that no stockout occurs under either policy.

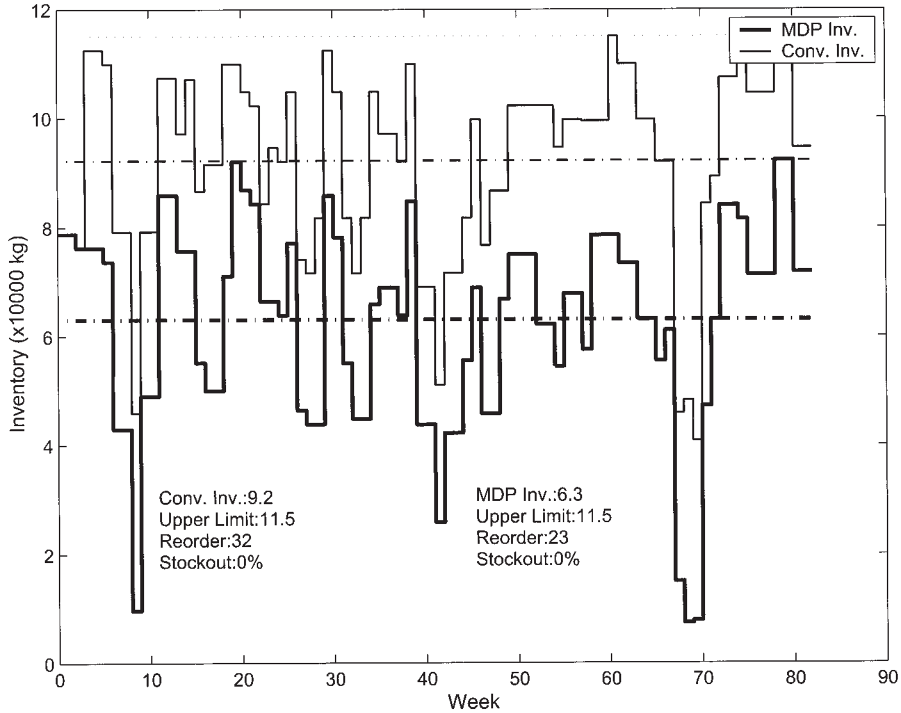

Figure 1. A comparison of the MDP policy and the conventional replenishment policy. Product: P-6; service level: 97.6%.

The same upper bound $M = 115{,}000$ kg was applied to both methods in their policy development. However, the MDP policy results in an average inventory level of 63,000 kg that is much lower than the level of 92,200 kg required by the conventional replenishment policy. The service level in the figure is a prescribed number. A service level of $\alpha = 97.6\%$ means that the allowable probability of stockout is $1 - 97.6\% = 2.4\%$. The calculation shows that using MDP method, a total of 23 reorders have been placed during the 83-week period, which is also lower than the 32 times required in the conventional method.

The same procedure was applied to several other products. Similar conclusions were obtained. Of the two different methods, the MDP system consistently yielded better policies than the traditional replenishment ones. In general, the MDP system results in lower average inventory level and requires fewer reorders to be placed. These advantages are more pronounced for products with highly variable demands, for which most of the other methods do not perform well. Table 2 compares the performances of the two policies for six products in determining their inventories. It can be seen that the MDP policy leads to lower inventory and fewer reorders when a comparable or a small amount of stockout is allowed. It should be noted that inventory stock represents a considerable investment; therefore it is desirable to lower the average inventory level as much as possible. As a trade-off, manufacturers often prescribe a service level less than 100%, that is, the allowable probability of stockout is greater than zero.

Markovian Production Planning Models

Many manufacturing systems are subject to such random and/or discrete events as raw material variation, demand fluctuation, and equipment failures, which may lead to jump discontinuity in their evolution. Production planning and scheduling require making decisions and are usually formulated as optimization problems. The functional relationship between the state variables and the inputs and disturbances can be described by dynamic systems. For instance, the surplus amount of the production system can be taken as the state variable that is determined by both demand rate and production rate. The latter, in turn, is governed by the production capacity. The uncertainties originating from random demand and random capacity processes are embedded in the system equation. Using finite-state, continuous-time Markov chains to model these

Table 2. Comparison of MDP and (*s*, *S*) Policy

	MDP Policy			(*s*, *S*) Policy		
	Ave. Inventory	Reorder	% Stockout	Ave. Inventory	Reorder	% Stockout
P-1	11.14	29	0	12.19	40	0
P-2	8.14	28	2	11.43	52	0
P-3	9.06	14	0	14.79	36	0
P-4	7.33	13	1	10.68	47	0
P-5	8.7	14	1	14.41	34	0
P-6	6.3	23	0	9.22	32	0

stochastic processes will allow us to formulate system equations that can embody the randomness and accommodate for various situations.

Continuous-time Markov chains and infinitesimal generator

In contrast to the discrete-time case, the transition probability $P_{ij}(t)$ of a continuous-time Markov chain is time-dependent. In this article, we consider homogeneous Markov chains. The transition matrix $P(t) = [P_{ij}(t)] \in R^{m\times m}$ satisfies a system of differential equations, known as the forward equation

$$\frac{dP(t)}{dt} = P(t)Q \qquad P(0) = I \tag{18}$$

where the identity matrix I provides the initial conditions; the matrix $Q = (q_{ij})$ is the so-called infinitesimal generator and is defined by

$$Q = \lim_{t\to 0+} \frac{P(t) - I}{t} = (q_{ij}) \in R^{m\times m} \tag{19}$$

The q_{ij} values satisfy: $q_{ij} \geq 0$ for $j \neq i$ and $q_{ii} = -\sum_{j\neq i} q_{ij}$ for $t \geq 0$. Representing the rates of changes of transition probabilities, the generator Q "generates" the continuous-time chain and plays a role similar to the one-step transition probability matrix in the discrete-time chain. For any real-valued function f on M and $i \in M$, the generator acts on f according to the following expression

$$Qf(\cdot)(i) = \sum_{i\in M} q_{ij}f(i) = \sum_{j\neq i} q_{ij}[f(j) - f(i)] \tag{20}$$

For additional properties of continuous-time, finite-state Markov chains, see, for example, Davis (1993), Ross (2000), and Yin and Zhang (1998).

The production system

Consider a manufacturing system that produces r different products. Let $\mathbf{u}(t) \in R^r$ denote the production rates that vary with time and are dictated by the random machine capacity. With the total surplus (the inventory/shortage level) $\mathbf{x}(t) \in R^r$ and the random demand rates $\mathbf{z}(t) \in R^r$, the system is given by a differential equation, which states that the rates of change of the surplus constitute the difference between the rates of production and the rates of demand. Our objective is to seek the optimal production rate, $\mathbf{u}^*(\cdot)$, to minimize a discounted cost function, subject to the system dynamics, the machine capacity $\alpha(t)$, and other operating conditions.

Assume the demand process $\mathbf{z}(\cdot) = \{\mathbf{z}(t) : t \geq 0\}$ to be a finite-state Markov chain having state space $Z = \{\mathbf{z}^1, \ldots, \mathbf{z}^d\}$. Considering the possible random breakdown and repair, we model the machine capacity by a continuous-time, finite-state Markov chain $\alpha(\cdot) = \{\alpha(t) : t \geq 0\}$ with state space $C = \{\alpha^1, \ldots, \alpha^c\}$. It should be noted that the term "machine" used herein has a broader sense than its literal meaning. A paper machine is certainly deemed a machine. Other pieces of equipment, such as a chemical reactor, are also considered machines. Breakdowns in a paper machine or web, frequently encountered in paper mills, are typical examples that cause random variations in production capacity. The two states, 1 and 0, in the state space $C = \{0, 1\}$ correspond to a machine's being up and down, respectively. Other possible situations would include, for instance, a machine deteriorates gradually in its throughput because of heavy usage and thus its condition is classified into several possible states, such as good, operable with minor deterioration, operable with major deterioration, and inoperable, thus $C = \{\alpha^1, \alpha^2, \alpha^3, \alpha^4\}$. At any given time the production capacity determines the set of all possible rates of production, $\mathbf{u}(t)$. For each state of the capacity, $\alpha^1 \leq \alpha^i \leq \alpha^c$, without loss of generality, denote the set of production rate constraints by $\mathcal{U}_i$

$$\mathcal{U}_i = \{\mathbf{u}(\cdot) = (u_1, u_2, \ldots, u_r)' \geq 0 : \mathbf{p}\cdot\mathbf{u} \leq \alpha^i\} \subset R^r \qquad i, j = 0, \ldots, c \tag{21}$$

where $\mathbf{p} = (p_1, \ldots, p_r) \geq 0$ is a given constant vector with each p_j ($j = 1, \ldots, r$) representing the amount of capacity needed to produce one unit of product j, where the prime symbol $'$ denotes the transpose. In the above, for a vector $\mathbf{x}$, by $\mathbf{x} \geq 0$, we meant that all of its components are nonnegative. Then the production rate at time t is subject to the constraint

$$\mathbf{u}(t) \in \mathcal{U}_{\alpha(t)} \tag{22}$$

The generators of the Markov chains $\alpha(\cdot)$ and $\mathbf{z}(\cdot)$ are given by $Q^c = (q^c_{ij}) \in R^{c\times c}$ and $Q^d = (q^d_{ij}) \in R^{d\times d}$, respectively. Thus for any functions ϕ on C and ψ on Z, we have

$$Q^c\phi(\cdot)(j) = \sum_{j_1\neq j} q^c_{jj_1}[\phi(j_1) - \phi(j)]$$
$$Q^d\psi(\cdot)(j) = \sum_{j_1\neq j} q^d_{jj_1}[\psi(j_1) - \psi(j)] \tag{23}$$

The production system is subject to a joint stochastic process, $\beta(t) = [\alpha(t), \mathbf{z}(t)]$, consisting of the capacity and demand pair. Observe that $\beta(\cdot)$ is also a Markov chain that has a state space of size $m = c\times d$

$$M = \{(\alpha^1, z^1), \ldots, (\alpha^c, z^1), \ldots, (\alpha^1, z^d), \ldots, (\alpha^c, z^d)\} = M_1 \cup \cdots \cup M_d \tag{24}$$

and a generator Q, an $m\times m$ matrix. Note that $M_i = \{(\alpha^1, \mathbf{z}^i), \ldots, (\alpha^c, \mathbf{z}^i)\}$.

Now the dynamic system is given by

$$\dot{\mathbf{x}}(t) = \beta(t)[\mathbf{u}(t), -I_d]' = \alpha(t)\mathbf{u}(t) - \mathbf{z}(t)$$
$$\mathbf{x}(0) = \mathbf{x},\ \beta(0) = [\alpha(0), \mathbf{z}(0)] = \beta \tag{25}$$

where I_d is the $d\times d$ identity matrix, and $\beta = (\alpha, \mathbf{z})$ and $\mathbf{x} \in R^r$ are, respectively, the initial state of the chain and the initial surplus level that is positive when it represents inventory and negative when it represents shortage. Define the cost functional as

$$J[\mathbf{x}, \mathbf{u}(\cdot), \beta] = E\int_0^\infty e^{-\rho t}\{h[\mathbf{x}(t)] + \zeta[\mathbf{u}(t), \beta]\}dt \qquad \beta \in M \tag{26}$$

where $\rho > 0$ is the discount rate, $\mathbf{u}(t) = [u^i(t), i \leq r]$ is the "normalized" production rate satisfying $0 \leq u^i(t) \leq 1$, $h(\cdot)$ is the holding cost, and $\zeta(\cdot)$ denotes the production cost. The expectation E is taken over both random machine capacity and random demand. Note that for notational simplicity, we have normalized $\mathbf{u}(\cdot)$ so that $u^i(t) \leq 1$, and the effective production rates are $\alpha(t)\mathbf{u}(t)$. Our goal is to find the optimal policy or the optimal production rate $\mathbf{u}^*(t)$, to minimize the objective function (Eq. 26), subject to dynamics described by Eq. 25, the capacity $\alpha(t)$, and other production constraints for the given initial conditions.

For each $\beta = M$, define the value function $v(\cdot, \beta)$ as the minimum of the cost over $\mathbf{u}(\cdot) \in A$, that is

$$v(\mathbf{x}, \beta) = \inf_{\mathbf{u}(\cdot)\in A} J[\mathbf{x}, \mathbf{u}(\cdot), \beta] \qquad \beta \in M \tag{27}$$

where A is the set of admissible controls. Because of the use of Markov chains, we have a total of m value functions, each corresponding to one of the m states of the Markov chain $\beta(\cdot)$. Using a dynamic programming approach, it can be shown that the value functions satisfy a system of partial differential equations known as HJB (Hamilton–Jacobi–Bellman) equations (see Fleming and Rishel, 1975; Sethi and Zhang, 1994)

$$\rho v(\mathbf{x}, \beta) = \min_{\mathbf{u}(\cdot)\in A}\{[\beta(\mathbf{u}, -I_d)' \cdot \nabla v(\mathbf{x}, \beta) + [h(\mathbf{x}) + \zeta(\mathbf{u}, \beta)]\} + Qv(\mathbf{x}, \cdot)(\beta) \qquad \beta \in M \tag{28}$$

where $\mathbf{x} \in R^r$, $\beta \in M$, $\mathbf{a} \cdot \mathbf{b}$ denotes the inner product of the vectors $\mathbf{a}$ and $\mathbf{b}$, and $\nabla f(\mathbf{x})$ is the gradient of f. The solution of Eq. 28 will lead to the optimal production rate we are looking for. Similar to many other controlled Markovian systems, however, the closed-form solution of the corresponding HJB equation is difficult or even impossible to obtain. Therefore, numerical solutions become necessary.

HJB equations and the numerical procedure

Adopting the numerical methods developed in Kushner (1990) (see also Yin and Zhang, 1998), we discretize Eq. 28 by discretizing the space R^r with grid $\Delta > 0$, which yields a discrete space R^r_Δ. Using Δ as the step size, we approximate the value function $v(x, \beta)$ by a sequence of functions $v^\Delta(x, \beta)$; and its partial derivatives $v_{x_j}(x, \beta)$ by the finite differences. Subsequently, we can rewrite the HJB Eq. 28 in terms of $v^\Delta(x, \beta)$ as

$$v^\Delta(\mathbf{x}, \beta) = \min_{\mathbf{u}\in\Gamma}\left(\rho + |q_{zz}| + |q_{\alpha\alpha}| + \sum_{i=1}^r \frac{|u_i - z_i|}{\Delta}\right)^{-1}\left\{\sum_{i=1}^r \frac{|u_i - z_i|}{\Delta} \times [v^\Delta(\mathbf{x}(x_i, +), \beta)I_{\{[u_i - z_i]\geq 0\}} + v^\Delta(\mathbf{x}(x_i, -), \beta)I_{\{[u_i - z_i]<0\}}] + h(\mathbf{x}) + \zeta(\mathbf{u})] + \sum_{\beta'\neq\beta} q_{\beta\beta'}v^\Delta(\mathbf{x}, \beta')\right\} \tag{29}$$

where

$$\mathbf{x}(x_j, +) = \mathbf{x} + \Delta e_j = (x_1, x_2, \ldots, x_{j-1}, x_j + \Delta, x_{j+1}, \ldots, x_r)'$$

$$\mathbf{x}(x_j, -) = \mathbf{x} - \Delta e_j = (x_1, x_2, \ldots, x_{j-1}, x_j - \Delta, x_{j+1}, \ldots, x_r)'$$

and $[u_j - z_j]$ denotes the jth component of the function $[\mathbf{u} - \mathbf{z}]$ and I_A is the indicator function of A given by $I_A = 1$ if A is true and 0, otherwise.

Observing that Eq. 29 can be expressed in the form

$$v^\Delta_{n+1}(\mathbf{x}, \beta) = \tau[v^\Delta_n(\mathbf{x}, \beta)]$$

we use the value iteration procedure to approximate the value function: Starting from an initial value of $\mathbf{x}$ and an arbitrary initial guess v^Δ_0, the procedure uses iterations until a certain convergence criterion is satisfied.

Hierarchical Approach

Because of the large state space of the joint stochastic process $\beta(t)$, we need to solve $m = c \times d$ HJB equations, which requires an intensive computation. In many cases, the computational requirements to obtain an optimal policy are staggering, to the point that a numerical solution becomes infeasible. This is the so-called curse of dimensionality (Bertsekas, 1976). In this work, we resort to a hierarchical approach for solution.

The hierarchical decision-making approach is one of the most important methods in dealing with large and complex systems. The basic idea is to reduce complex problems into manageable, approximate problems or subproblems, and to construct solution of the original problem from solutions of these simpler problems under the current setting. It is achieved by replacing some of the stochastic processes with their averages. Considering that in many manufacturing systems, the rates of changes of the random events involved are notably different (Yin and Zhang, 1998), Sethi and Zhang (1994) developed hierarchical approaches that lead to multilevel decisions. They also showed that such results are asymptotically optimal.

Timescale separation

In many manufacturing processes, the machine breakdowns and repairs take place much more frequently than the changes in demand. In pulp and paper mills, for instance, paper-machine breakdowns may occur once or twice per day, or once every 2 or 3 days, whereas the demand changes are much less frequent, usually to be considered once every week to every month. To reflect such differences in transition rates, we use timescale separation by introducing a small parameter $\varepsilon > 0$ into the system. We will use the singularly perturbed Markov chain techniques developed by Yin and Zhang (1998) to resolve the problems. Specifically, we assume that the generator of the chain, $\beta(t)$, is of the form $Q = Q^\varepsilon = (q^\varepsilon_{ij}) \in R^{m\times m}$, and that

$$Q^{\varepsilon} = \frac{1}{\varepsilon}\tilde{Q} + \hat{Q} \overset{\text{def}}{=} \frac{1}{\varepsilon}\begin{pmatrix} Q^c & & \\ & \ddots & \\ & & Q^c \end{pmatrix} + \begin{pmatrix} q_{11}^d I_c & q_{12}^d I_c & \cdots & q_{1d}^d I_c \\ \vdots & \vdots & \ddots & \vdots \\ q_{d1}^d I_c & q_{d2}^d I_c & \cdots & q_{dd}^d I_c \end{pmatrix} \quad (30)$$

where $\tilde{Q}$ and $\hat{Q}$ are of the same order of magnitude. Observe that the introduction of ε separates the system into two time-scales, in which $\tilde{Q}/\varepsilon$ dictates the chain's fast-changing part and $\hat{Q}$ governs its slowly varying part. Note that $\tilde{Q} = \text{diag}(Q^c, Q^c, \ldots, Q^c)$ and that $\hat{Q}$ is nothing but the Kronecker product $Q^d \otimes I_c$. Assume that Q^c is irreducible. Then the equilibrium distribution $\nu = (\nu_1, \nu_2, \ldots \ \nu_c)$ is the unique solution of

$$\nu Q^c = 0 \qquad \text{and} \qquad \sum_{i=1}^{c} \nu_i = 1 \quad (31)$$

To distinguish systems having significant rate differences, we designate the quantities involved by the small parameter ε in the problem formulation. That is, the chain, as well as the surplus $\mathbf{x}(t)$, the cost function $J[\mathbf{x}, \mathbf{u}(\cdot), \beta]$, and the value function $v(\mathbf{x}, \beta)$ in Eqs. 25–28 are all dependent on ε (because the chain is generated by Q^{ε}), and can be replaced by $\beta^{\varepsilon}(t)$, $\mathbf{x}^{\varepsilon}(t)$, $J^{\varepsilon}[\mathbf{x}, \mathbf{u}(\cdot), \beta]$, and $v^{\varepsilon}(\mathbf{x}, \beta)$, respectively.

Because ε is small, the Markov chain $\beta^{\varepsilon}(\cdot)$ jumps more frequently within the states in M_i and less frequently from M_i to M_j for $i \neq j$. Naturally, we aggregate all the states in M_i into a single state i. That is, we approximate $\beta^{\varepsilon}(\cdot)$ by an aggregated process, say, $\bar{\beta}^{\varepsilon}(\cdot)$ defined as $\bar{\beta}^{\varepsilon}(t)$ if $\beta^{\varepsilon}(t) \in M_i$. It has been shown (Yin and Zhang, 1998) that as $\varepsilon \rightarrow 0$, $\bar{\beta}^{\varepsilon}(\cdot)$ converges weakly to $\bar{\beta}(\cdot)$, a Markov chain having state space $\bar{M} = \{1, \ldots, d\}$ and generated by

$$\bar{Q} = \text{diag}(\underbrace{\nu, \ldots, \nu}_{d})\hat{Q}\ \text{diag}(\underbrace{1_c, \ldots, 1_c}_{d}) \in R^{d\times d}$$
$$\bar{M} = \{1, \ldots, d\} \quad (32)$$

where 1_c denotes a column vector of dimension c with all components being 1, and $\text{diag}(A^1, \ldots, A^l)$ denotes a block diagonal matrix having matrix entries A^i ($i = 1, \ldots, l$) of suitable dimensions. In our setup, $\text{diag}(\nu, \ldots, \nu) \in R^{d\times m}$, $\text{diag}(1_c, \ldots, 1_c) \in R^{m\times d}$. It can be shown that corresponding to the original problem, there is a limit problem. Moreover, the value functions $v^{\varepsilon}(\mathbf{x}, \beta)$, associated with the original problem, converge to the value functions of the limit problem under "suitable aggregation." More specifically, define $\Gamma_0 = \{(\mathbf{U}^1, \ldots, \mathbf{U}^d) : \mathbf{U}^i = (\mathbf{u}^{1i}, \ldots, \mathbf{u}^{ci})$, for $i = 1, \ldots, d\}$, and denote by A^0 the set of admissible controls for the limit problem. The limit problem is given by

$$P^0 : \begin{cases} \min J^0[\bar{\mathbf{x}}, U(\cdot), i] = E\displaystyle\int_0^{\infty} e^{-\rho t} G[\bar{\mathbf{x}}(t), U(\cdot), i]dt \\ \dot{\bar{\mathbf{x}}}(t) = \bar{f}[\bar{\mathbf{x}}(t), U(t), \bar{\beta}(t)],\ \mathbf{x}(0) = \mathbf{x},\ \bar{\beta}(0) = \bar{\beta} \in M = \{1, \ldots, d\} \\ v(\bar{\mathbf{x}}, i) = \inf_{U(\cdot)\in A^0} J^0[\bar{\mathbf{x}}, U(\cdot), i] \end{cases} \quad (33)$$

where

$$\bar{f}(\bar{\mathbf{x}}, U, i) = \sum_{j=1}^{c} \nu_j[(\alpha^j, \mathbf{z}^i)(\mathbf{u}^{ji}, -I_d)'] \qquad i \in \bar{M}$$

$$\bar{G}(\bar{\mathbf{x}}, U, i) = \sum_{j=1}^{c} \nu_j\{h(\bar{\mathbf{x}}) + \zeta[\mathbf{u}^{ji}, (\alpha^j, \mathbf{z}^i)]\} \qquad i \in \bar{M}$$

The associated HJB equations take the form

$$\rho v(\bar{\mathbf{x}}, i) = \min_{U(\cdot)\in A^0} \{\bar{f}(\bar{\mathbf{x}}, U, i)\cdot\nabla v(\bar{\mathbf{x}}, i) + \bar{G}(\bar{\mathbf{x}}, U, i)\} + \bar{Q}v(\bar{\mathbf{x}}, \cdot)(i) \qquad i \in \bar{M} \quad (34)$$

Note that in the above, we have used the index notation convention. Although the limit problem still involves stochastic processes, the total number of states of the limit Markov chain is much smaller than that of the original one. In fact, there are $m = d \times c$ states for the Markov chain $\beta^{\varepsilon}(t)$, whereas only d states for the limit $\bar{\beta}(t)$. Thus the total number of equations need to be solved in the limit HJB equations becomes d. It has also been shown that an optimal or a near-optimal decision of the limit problem is asymptotically optimal to the original problem when ε is small. Therefore we can use the optimal or near-optimal solution of the limit problem to construct controls for the original problem, which will result in near-optimal controls of the original problem. Interested readers are referred to Yin and Zhang (1998, Ch. 9) for proofs and more detailed discussion. Note that in obtaining the asymptotic properties, $\varepsilon \rightarrow 0$. When one uses the results in practice, ε is simply a small real number; $\varepsilon = 0.1$ is considered small enough.

Numerical procedures and asymptotic optimal solution

Using a hierarchical approach enables us to reduce the original stochastic problem to a simpler one. The next step requires constructing a procedure to solve the limit problem P^0. According to the numerical methods developed in Kushner (1990), we first discretize the value function of the limit problem $v(\mathbf{x}, \bar{\beta})$ on R^r by a sequence of functions $v^{\Delta}(\mathbf{x}, \bar{\beta})$ on R^r_{Δ} ($\Delta > 0$), and its partial derivatives $v_{x_j}(\mathbf{x}, \bar{\beta})$ by the corresponding finite-difference quotients. Then write the discretized version of the HJB equation, and rearrange it into an iteration form $v^{\Delta}_{n+1}(\mathbf{x}, \bar{\beta}) = \tau[v^{\Delta}_n(\mathbf{x}, \bar{\beta})]$. Starting from an initial value $\mathbf{x}$ and an arbitrary initial guess v^{Δ}_0, applying the value iteration procedure leads to the approximate solution of the limit HJB equation. Using the optimal or nearly optimal controls of the limit problem, we can then proceed to design controls for the original problems, yielding near optimality.

An application example

This section applies the hierarchical approach described above to a single-product, two-machine papermaking process. Using weekly customer demand and paper-machine operating data over an 82-week period collected from a large paper manufacturer, we obtain the generator Q^d and Q^c of the random demand and capacity processes.

In this example, we treat the demand process as a four-state Markov chain, $z(t) \in Z = \{z^1, z^2, z^3, z^4\}$. Because we are dealing with products measured by weight, we have "discretized" the data into four levels or states. Denote the total number of transitions from state i to state j during the entire 82-week period by n_{ij} ($i = 1, \ldots, 4$, $j = 1, \ldots, 4$); thus $\sum_i \sum_j n_{ij} = 81$. Denote the amount involved in this transition by s^l_{ij} ($1 \le l \le n_{ij}$). Therefore the amount of transitions from state i to state j in the entire 82-week totals $S_{ij} = \sum_{l=1}^{n_{ij}} s^l_{ij}$. Let $q_{ij} = S_{ij}/n_{ij}$. We construct the infinitesimal generator $Q^d = (Q^d_{ij})$ by designating q^d_{ij} as

$$q^d_{ij} = q_{ij} + q_{ii} \cdot \frac{q_{ij}}{\sum_{j \neq i} q_{ij}} \qquad i \neq j \tag{35}$$

and then let

$$q^d_{ii} = -\sum_{j \neq i} q^d_{ij} \tag{36}$$

The matrix so constructed satisfies $q^d_{ii} < 0$, $q^d_{ij} \geq 0$, and $\sum_j q^d_{ij} = 0$ for all i. The matrix Q^d in this example is

$$Q^d = \begin{pmatrix} -4.00 & 1.33 & 2.67 & 0 \\ 1.00 & -2.00 & 1.00 & 0.00 \\ 2.00 & 1.00 & -4.00 & 1.00 \\ 3.00 & 0.00 & 1.00 & -4.00 \end{pmatrix} \tag{37}$$

Consider two parallel machines each having capacities $\alpha_1(\cdot) \in \{0, \alpha^1\}$ and $\alpha_2(\cdot) \in \{0, \alpha^2\}$; thus the overall state space of the capacity is $C = \{(0, 0), (\alpha^1, 0), (0, \alpha^2), (\alpha^1, \alpha^2)\}$. Let the rates of a machine going down be λ_1 and λ_2, and the rates of resumption be μ_1 and μ_2, respectively; assume the infinitesimal generators of α_1 and α_2 are

$$Q^c_1 = \begin{pmatrix} -\mu_1 & \mu_1 & 0 & 0 \\ \lambda_1 & -\lambda_1 & 0 & 0 \\ 0 & 0 & -\mu_1 & \mu_1 \\ 0 & 0 & \lambda_1 & -\lambda_1 \end{pmatrix}$$

$$Q^c_2 = \begin{pmatrix} -\mu_2 & 0 & \mu_2 & 0 \\ 0 & -\mu_2 & 0 & \mu_2 \\ \lambda_2 & 0 & -\lambda_2 & 0 \\ 0 & \lambda_2 & 0 & -\lambda_2 \end{pmatrix} \tag{38}$$

To determine the generator for $\alpha(\cdot)$, we further assume that only one machine may change its state during a single transition. Therefore

$$Q^c_0 = \begin{pmatrix} -(\mu_1 + \mu_2) & \mu_1 & \mu_2 & 0 \\ \lambda_1 & -(\lambda_1 + \mu_2) & 0 & \mu_2 \\ \lambda_2 & 0 & -(\lambda_2 + \mu_1) & \mu_1 \\ 0 & \lambda_2 & \lambda_1 & -(\lambda_1 + \lambda_2) \end{pmatrix} \tag{39}$$

In this example, $\lambda_1 = 60/17$, $\lambda_2 = 24/5$, and $\mu_1 = \mu_2 = 120$. The matrix is expressed as follows

$$Q^c_0 = \begin{pmatrix} -240 & 120 & 120 & 0 \\ 3.53 & -123.53 & 0 & 120 \\ 4.8 & 0 & -124.8 & 120 \\ 0 & 4.8 & 3.53 & -8.33 \end{pmatrix} \tag{40}$$

Note the differences in the orders of magnitude between Q_d and Q^c_0. Using $\varepsilon = 0.01$, the generator $Q^{\varepsilon} = (\tilde{Q}/\varepsilon) + \hat{Q}$ of the joint Markov chain $\beta^{\varepsilon}(t) = [\alpha^{\varepsilon}(t), z(t)]$ is then obtained in accordance with Eq. 30 with $Q = Q^{\varepsilon} = (q^{\varepsilon}_{ij}) \in R^{m \times m}$ and $\varepsilon = 0.01$, where $\tilde{Q}$ and $\hat{Q}$ are of the same order of magnitude. Observe that the introduction of $\varepsilon = 0.01$ separates the system into two timescales, in which $\tilde{Q}/\varepsilon$ dictates the chain's fast-changing part and $\hat{Q}$ governs its slowly varying part. Subsequently, the generator $\bar{Q}$ of the "aggregated" Markov chain $\bar{\beta}(\cdot)$ was derived according to Eq. 32. Because we are dealing with a single product, all the vectors concerned become scalars. The holding cost and the production cost functions are $h(x) = 0.01x^+ + 0.7x^-$, where $x^+ = \max\{0, x\}$ and $x^- = \min\{0, -x\}$ for $x \in R$, and $\zeta(u, \beta) = 0.5u$, respectively. The control set is $\Gamma_0 = \{(U^1, \ldots, U^4) : U^i = (u^{1i}, \ldots, u^{4i}) = (0, u^{2i}, u^{3i}, u^{4i})$, for $i = 1, \ldots, 4\}$. The optimal control (u^{*ji}, $j = 1, \ldots, 4$, $i = 1, \ldots, 4$) for the limit problem then can be found by solving Eq. 33 numerically. Figure 2 displays the dependency of production rate u^{ji} and the corresponding value function v on surplus x, and capacity α^j when the demand is at its highest level z^4.

Using the optimal feedback control in the system Eq. 25 yields the optimal production rates and the resulted surplus corresponding to the given initial conditions as demonstrated in Figure 3. We assume the initial surplus level $x_0 = \mu$, which represents the average demand during the eighty-two weeks. Figure 3a shows the actual random demand during this 82-week period. Figure 3c provides the production rate to be taken; the corresponding movement of the surplus trajectory is presented in Figure3b.

The optimal control is of the threshold type (see Sethi and Zhang, 1994) or specified by turnpike sets. The idea can be

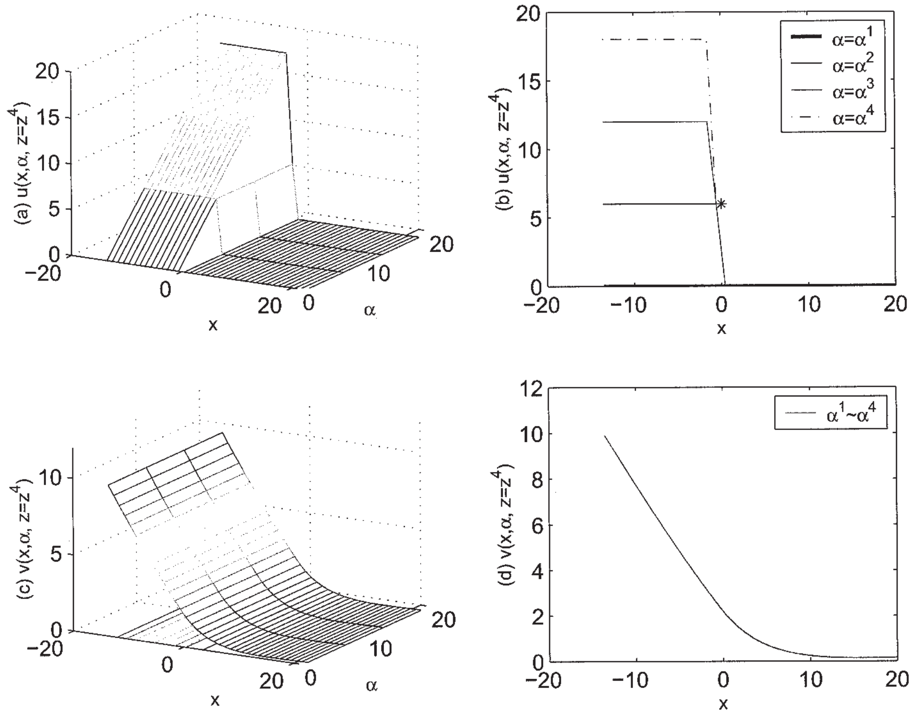

Figure 2. Dependency of the production rate (a and b) and the value function (c and d) on surplus and the capacity when $z = z^4$.

explained as follows: To reach point B from point A, one takes the turnpike as soon as possible and remains there as long as possible. In the current problem, the limit problem becomes one that is equivalent to finding optimal controls arising from random demand. Although the optimal control can be characterized by the turnpike sets, their closed-form solution is generally not available unless the Markov chain has only two states. Therefore numerical solution is needed.

Using u^{*ij}, the optimal control of the limit problem, we can construct a control for the original stochastic production system as

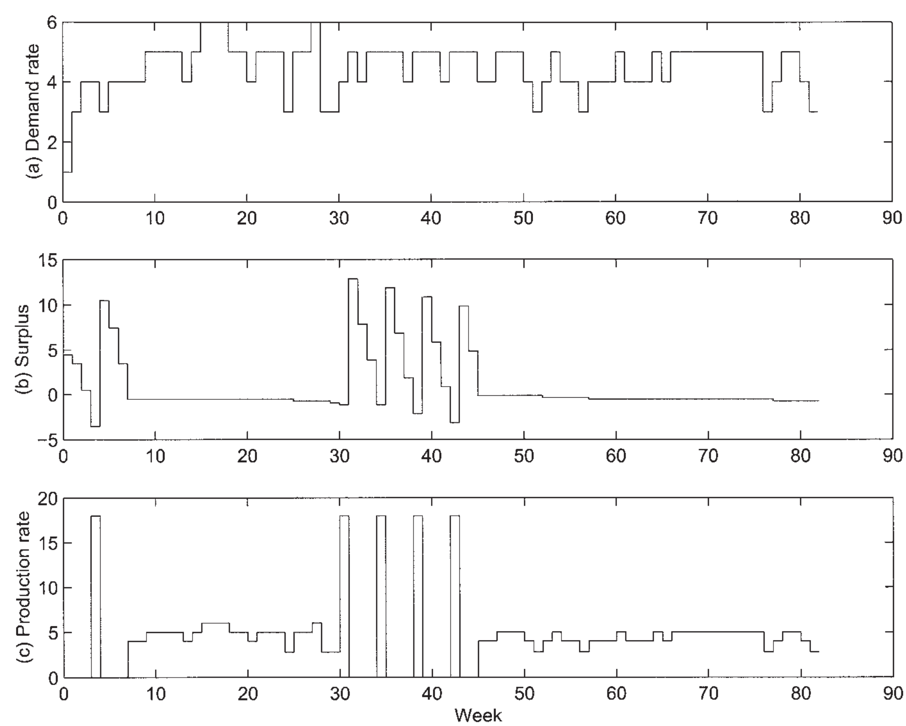

Figure 3. Trajectories of the random demand, surplus, and production rate.

$$u^{\varepsilon}(x, \alpha, z) = \sum_{i=1}^{4} \chi_{\{z=z^i\}} \sum_{i=1}^{4} \chi_{\{\alpha=\alpha^j\}} u^{*ji}(x, \alpha^j, z^i) \quad (41)$$

where χ_A is the indicator function of the set A with

$$\chi_A = \begin{cases} 1 & \text{if } x \in A \\ 0 & \text{otherwise} \end{cases}$$

It can be shown, as in Yin and Zhang (1998), that $u^{\varepsilon}(t) = u^{\varepsilon}[x^{\varepsilon}(t), \alpha^{\varepsilon}(t), z(t)]$ is asymptotically optimal. Given the fact that the hierarchical approach can significantly reduce the computation requirement, it is a promising approach to solve large-scale problems.

Discussion and Summary

Motivated by the needs for better inventory policies and production plans in the paper industry, this paper concerns modeling methods and numerical procedures for dynamic systems under uncertainties. We resort to stochastic models for solutions. Finite-state discrete-time and/or continuous-time Markov chains are used to describe the random behavior of the underlying systems. The optimal and/or near-optimal policies are obtained by stochastic optimization. To the best of our knowledge, this work represents the first attempt of using controlled Markovian dynamic systems and stochastic optimization in the paper industry.

Describing the inventory process by a finite-state, discrete-time Markov chain allows us to formulate a Markov decision process model that yields the optimal inventory policy. The key components of this sequential decision model are the set of states of the system, the set of possible decisions and their corresponding actions, the transition probabilities, and costs and rewards. The major steps in model formulation and solution procedure, including obtaining the state space of the Markov chain, designating the possible decisions and actions, calculating the transition probabilities, defining the cost structure, evaluating the cost function, and determining the optimal policy are discussed in detail. We simplify the decision-making procedure by discretizing the continuous demands into discrete levels. Such treatment is applicable to many cases in the chemical industry where the products are measured by their weights. An example of inventory control of finished paper products is presented for illustration. The optimal strategies obtained allow us to make production decisions sequentially throughout the life span of the process. Our calculation results have shown that the MDP models consistently provide better results than those of conventional models, especially for systems exhibiting high variability.

In planning and scheduling, the production system of interest contains continuous dynamics, intertwined with discrete event interventions, and is often referred to as a *hybrid* system. To quantify the random and jump behavior, we describe the system dynamics and formulate the random demand and random production capacity processes by finite-state, continuous-time Markov chains. Our objective is to obtain the optimal feedback policy that minimizes a properly defined expected cost. The minimum of such a cost over all admissible controls is called the value function that satisfies the Hamilton–Jacobi–Bellman equation. In general, because the closed-form solution of the HJB equation is impossible to obtain, we resort to a numerical procedure, which requires discretizing the HJB equations first and then applying an approximation algorithm.

It is observed that the computation needed in numerically solving such dynamic programming equations increases exponentially with the increase of the states. In many cases, the computational requirements to obtain an optimal policy are often staggering, to the point that a numerical solution becomes infeasible. This is the so-called *curse of dimensionality*, a phenomenon common in solving problems in stochastic control and optimization. To surmount this difficulty, we adopt a hierarchical approach to seek near-optimal solutions. The mathematical models and numerical algorithms are applied to a paper-manufacturing process. Using the real customer demand and production data collected from a large paper company, we obtain the near-optimal production rates under various situations. Examples are included to illustrate the model formulation and numerical procedure. Both original and hierarchical approaches are applied to several problems for comparison. As expected, the latter can significantly reduce the computation requirement. Of the several examples we worked on, depending on the state spaces of the corresponding Markov chains, the computation time required for solving the limit problems using the hierarchical approach ranges from 30 to 50% of the time needed for the original problem. Therefore it is a promising approach for large-scale systems. It should be noted that the hierarchical approach provides a near-optimal solution rather than the truly optimal solution. Nevertheless, from a practical perspective, the near-optimal solution is as good (useful) as the optimal solution.

Stochastic modeling and simulation have become frequently used and powerful tools in quantifying dynamic relations of sequences of random events and uncertainties. Although examples have been drawn only from the paper industry, the modeling methods and numerical procedure are applicable to other chemical and process industries.

Literature Cited

Applequist, G., O. Samikoglu, J. Pekny, and G. Reklaitis, "Issues in the Use, Design and Evolution of Process Scheduling and Planning Systems," *ISA Trans.*, **36**(2), 81 (1997).

Balasubramanian, J., and I. E. Grossmann, "Scheduling to Minimize Expected Completion Time in Flowshop Plants with Uncertain Processing Times," *Comput. Chem. Eng.*, **26**, 41 (2002).

Bassett, M. H., J. F. Pekny, and G. V. Reklaitis, "Using Detailed Scheduling to Obtain Realistic Operating Policies for a Batch Processing Facility," *Ind. Eng. Chem. Res.*, **36**(5), 1717 (1997).

Bertsekas, D., *Dynamic Programming and Stochastic Control*, Academic Press, New York (1976).

Bertsekas, D., *Dynamic Programming: Deterministic and Stochastic Models*, Prentice-Hall, Upper Saddle River, NJ (1987).

Boukas, E. K., H. Yang, and Q. Zhang, "Minimax Production Planning in Failure Prone Manufacturing Systems," *J. Optim. Theory Appl.*, **87**, 269 (1995).

Buffa, E. S., *Modern Production/Operations Management*, 6th Edition, Wiley, New York (1980).

Davis, M. H. A., *Markov Models and Optimization*, Chapman & Hall, New York (1993).

Fleming, W. H., and R. W. Rishel, *Deterministic and Stochastic Optimal Control*, Springer-Verlag, New York (1975).

Gupta, A., C. D. Maranas, and C. M. McDonald, "Mid-Term Supply Chain Planning under Demand Uncertainty: Customer Demand Satisfaction and Inventory Management," *Comput. Chem. Eng.*, **24**(12), 2613 (2000).

Hillier, F. S., and G. J. Lieberman, *Introduction to Operations Research*, 5th Edition, Holden-Day, Oakland, CA (1990).
Hopp, W. J., and M. L. Spearman, *Factory Physics—Foundations of Manufacturing Management*, 2nd Edition, McGraw-Hill, Boston (1996).
Kushner, H. J., "Numerical Methods for Stochastic Control Problems in Continuous Time," *SIAM J. Control Optim.*, **28**, 999 (1990).
Leiviskä, K., *Process Control*, Fapet Oy, Helsinki, Finland (1999).
Puterman, M. L., *Markov Decision Processes: Discrete Stochastic Dynamic Programming*, Wiley, New York (1994).
Ross, S., *Introduction to Stochastic Dynamic Programming*, Academic Press, New York (1983).
Ross, S., *Introduction to Probability Models*, Academic Press, New York (2000).
Sethi, S. P., and Q. Zhang, *Hierarchical Decision Making in Stochastic Manufacturing Systems*, Birkhäuser, Boston (1994).
Wilson, R. H., "A Scientific Routine for Stock Control," *Harvard Business Rev.*, **13**(1), 116 (1934).
Yan, H. M., G. Yin, and S. X. C. Lou, "Using Stochastic Optimization to Determine Threshold Values for Control of Unreliable Manufacturing Systems," *J. Optim. Theory Appl.*, **83**, 511 (1994).
Yin, G., and Q. Zhang, *Continuous-Time Markov Chains and Applications: A Singular Perturbation Approach*, Springer-Verlag, New York (1998).
Yin, G., and Q. Zhang, eds., "Mathematics of Stochastic Manufacturing Systems," Proc. 1996 AMS-SIAM Summer Seminar in Applied Mathematics, Lectures in Applied Mathematics, Am. Math. Soc., Providence, RI (1997).
Yin, K., G. Yin, and Q. Zhang, "Approximating the Optimal Threshold Levels under Robustness Cost Criteria for Stochastic Manufacturing Systems," Proc. IFAC Conf. of Youth Automation YAC '95, pp. 450–454 (1995).
Yin, K. K., H. Liu, and N. E. Johnson, "Markovian Inventory Policy with Application to the Paper Industry," *Comput. Chem. Eng.*, **26**(10), 1399 (2002).
Zipkin, P. H., *Foundations of Inventory Management*, McGraw Hill, Boston (2000).

Manuscript received Jan. 31, 2003, and revision received Feb. 11, 2004.