IIE Transactions (2004) **36**, 1083–1097

ISSN: 0740-817X print / 1545-8830 online
DOI: 10.1080/07408170490500663

Strategic inventory deployment in the steel industry

BRIAN DENTON[1] and DIWAKAR GUPTA[2,*]

[1]*IBM Microelectronics, 1000 River Road, Essex Junction, VT 05452, USA*
E-mail: bdenton@us.ibm.com
[2]*Department of Mechanical Engineering, University of Minnesota, Minneapolis, MN 55455, USA*
E-mail: guptad@me.umn.edu

Received August 2001 and accepted May 2004

Integrated Steel Manufacturers (ISMs) perform all of the steps necessary to convert iron ore into finished products. As a result they are characterized by high capital expenditures and long cycle times. The custom nature of finished products and significant demand uncertainty explains why ISMs typically produce to order. However, recent increased competition from low cost *mini-mills* is causing some ISMs to compete by serving the needs of higher-paying customers who want exotic products, and faster and reliable deliveries. Consequently, ISMs are exploring the option of satisfying a portion of their demand by converting strategically placed semi-finished inventory into finished products, which helps to reduce both the time between order receipt and order dispatch, and its variability. In this study we propose a two-stage stochastic integer programming model that can be used to choose the semi-finished products that should be made to stock, and their target inventory levels. Properties of this model are exploited to develop a fast heuristic applicable to large-scale instances of the problem encountered in industry. Numerical experiments are used to validate the heuristic, and examples based on data from a particular ISM are used to illustrate important managerial insights.

1. Introduction

Industries in which products are highly customized tend to operate primarily according to a Make-To-Order (MTO) policy. However, it is frequently the case that some proportion of the production is planned according to a forecast of orders to reduce customer order lead times. Order lead times are quoted based on the estimated *cycle time*; the time between receiving order requests and the earliest time they can be delivered. This approach is common, but it carries the risk that forecast orders may not materialize. Strategies based on delayed differentiation and component commonality attempt to mitigate the risk by facilitating pooling of demand. These strategies significantly reduce inventory holding costs by positioning inventory upstream in the supply chain, where it can be applied to many different customer orders; for example, see Brown *et al*. (2000) for an application in the manufacture of semi-conductor devices, and Burman *et al*. (1998) for a similar application in printer assembly. In process industries, the degree of customization of finished products tends to be very high and there is a large (often continuous) range of possible semi-finished product designs that can be carried in stock. Therefore, the problem of positioning inventory for the purpose of reducing customer order lead times also includes the additional complexity of choosing which semi-finished designs to stock.

We describe next an instance of the problem of choosing semi-finished product designs and target inventory levels in the steel industry that has motivated our interest in this class of problems. Integrated Steel Manufacturers (ISMs) perform all of the steps necessary to convert iron ore into finished products. Their operations are characterized by high capital expenditures and long cycle times. Lately, they have started to face stiff competition from much less capital-intensive *mini-mills*. Mini-mills process scrap steel and enjoy a cost advantage in plain carbon-steel markets as well as significantly shorter cycle times. Realizing that their competitive strength lies in being able to make a large variety of high-quality products, many ISMs have positioned themselves in markets for more specialized/customized finished products where they enjoy greater pricing power. However, customers for such products often require reliable deliveries that are synchronized with their own production schedules. Therefore, these ISMs have observed an increase in product variety, and a simultaneous pressure to significantly reduce delivery lead time for a subset of their customers.

The pressure to reduce delivery lead times has motivated ISMs to change from a pure MTO production mode to a hybrid MTO/Make-To-Stock (MTS) mode. That means orders from customers who are willing to pay a premium for shorter lead times are satisfied by converting strategically placed semi-finished inventory into finished products.

*Corresponding author

The remaining customers continue to be served through the MTO mode, that is, the processing of their orders starts from the iron-ore stage. Stocking semi-finished inventory for all customers is not economical because of the custom nature of end-products and the high degree of demand uncertainty. Besides, the storage space necessary to store thousands of different product designs and the associated holding costs are prohibitive. However, for inventory designs with sufficiently high demand and/or margin, this is an effective approach to reduce cycle times. The approach is particularly appropriate since it is not economical to buffer demand and supply uncertainty through expensive capacity. In short, determining the position and amount of strategic inventory to hold is an important problem faced by ISMs.

ISMs often produce finished product in the form of flat-rolled steel coils, or *band* for short. Production of coils is achieved in two stages; (i) primary production; and (ii) finishing. Primary production refers to the conversion of raw materials (e.g., iron ore, coke and limestone) into band. Finishing operations, on the other hand, make surface and structural modifications to achieve customer specifications on an order (e.g., tin plating, painting, tube forming). In the schematic shown in Fig. 1, stages 1–3 comprise the primary operations, and stage 4 represents finishing operations.

There are broadly three categories of inventories in the manufacture of flat-rolled products. These are in the form of slabs, band and finished products. Slabs are the least differentiated and finished goods are fully differentiated products. Depending on their grade, slabs may be applicable to orders for many different types of finished products because their dimensions (width and weight) can be modified by later operations. For instance, the width of a slab can be reduced by *roughing*, a process by which pressure is applied to the sides of the slab during the rolling process. Band is more specialized, but inventory dimensions can still be modified. For example, weight can be reduced to meet intended order specifications by cropping. In contrast, typical finished inventory is applicable only to a single customer.

Positioning inventory at different staging points yields different potential reductions in cycle times. For instance, if a finished item is stocked for a particular customer then it is possible to reduce their cycle time to virtually zero (assuming no stockouts). Alternatively, positioning inventory at the slab stage has the potential to reduce cycle times by about 50% (from about 12 weeks to 6 weeks), and positioning it at the coil stage can reduce cycle times by about 75%. Placing inventory closer to the finished product results in lower cycle times and lower demand pooling benefits. On the other hand, placing inventory further upstream from the finished product stage results in greater pooling benefits at the expense of increased cycle times.

Uncertainty is a major factor affecting inventory planning at ISMs. We describe here the fundamental sources of uncertainty that influence the ISM's ability to match demand and supply. On account of having long cycle times ISMs often start production based on tentative customer orders. Firm customer commitments are received after the ISM has proceeded well into the manufacturing process. Deviations from original order amounts are not uncommon and cause shortages or excess inventory. Furthermore, the markets for specialized steel products are inherently volatile because they experience the brunt of variability amplification (bullwhip effect) due to their position at the beginning of many supply chains. The increase in variability as demand information travels upstream in a supply chain is called the bullwhip effect (see Lee *et al*. (1997) and Chen (2000) for addition discussion). Supply uncertainty results from yield losses at various points in the production process. For example, slabs may deviate from the desired grade and often such quality glitches can be ascertained only after casting has been completed. At this point in time, it is too late to adjust for the shortfall in supply on a short notice since other slabs are already scheduled in a carefully

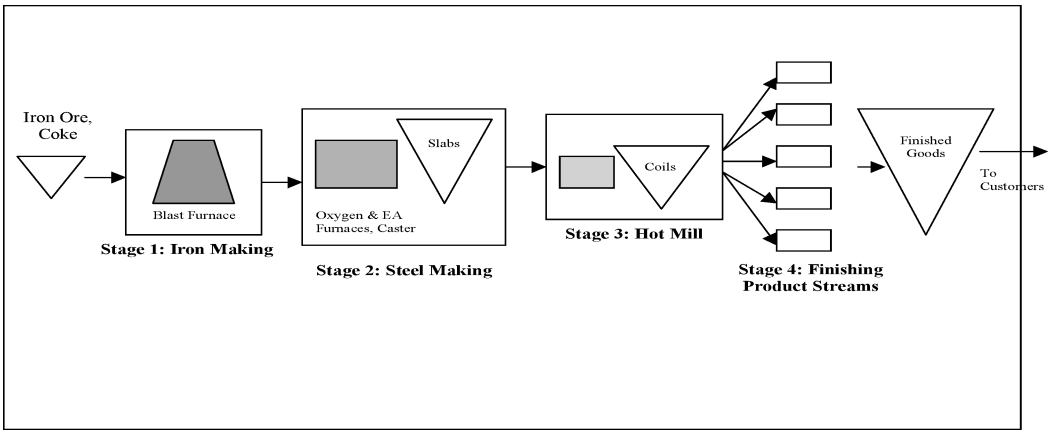

Fig. 1. The production process of an integrated steel mill. Primary inventory staging areas are slab, coil and finished goods. Finishing product streams are used to perform a variety of operations such as pickling, cold rolling, annealing, tempering, galvanizing, tinning, painting, and tube forming.

controlled production sequence. Another source of supply uncertainty arises when some slabs are purchased externally. In such cases yield losses are realized when suppliers send slabs that do not match order specifications.

The model we propose in this article can be used to address some strategic and operational planning decisions faced by ISMs when they attempt to operate a hybrid MTO/MTS system. These decisions are: (i) which customer segment should be served from the MTS mode; (ii) what should be the position and design of inventory of semi-finished products; and (iii) what should be their target production and inventory levels. The decisions are complicated by factors such as uncertainty in demand and production yield, product substitution, and finite storage space. We formulate the model as a two-stage stochastic linear program with binary first-stage decision variables, and develop a heuristic to solve real-world problem instances arising in the steel industry.

The main contributions of this article are as follows. We provide a formulation of a new problem of recent interest in the process industry that is motivated by discussions with senior managers in several different functional areas at a particular ISM. We show that an instance of the stochastic programming model we propose reduces to the p-median problem, which makes it a NP-complete problem. Structural properties of the deterministic equivalent problem and the second-stage recourse problem are presented that permit significant reduction in the number of discrete decision variables and lay the groundwork for efficient heuristics. We propose a heuristic that exploits the structure of the model, and provide a series of numerical experiments which establish its accuracy. Results for an instance of the problem from the steel industry (based on actual data) are used to motivate the economic importance of the model, and to illustrate managerial relevance.

This article is organized as follows. In Section 2 we provide a review of relevant literature. Section 3 contains model formulation, whereas some structural properties of the model can be found in Section 4. We propose a heuristic in Section 5 and report, in Section 5.1, a computational study of its accuracy. In Section 5.2, we also provide examples based on real data to illustrate the managerial importance of the model. Finally, a summary of the article can be found in Section 6.

2. Selected literature review

The problem of interest in this article is related to several different areas of literature including random yield, multi-product substitutable inventory, and stochastic fixed-charge network flow problems. We provide a brief overview of these topics here, and show links between our problem and previously studied problems.

A majority of related stochastic inventory literature deals with single-period (newsvendor) problems, and with two-product instances of the substitutable products problem. Single-period stochastic inventory models assume a two-stage decision process in which an initial inventory level is chosen, random supply and/or demand are observed, and inventory is subsequently allocated to demand. The simplest example is the newsvendor model for which there is a vast literature. We make no attempt here to summarize that literature and refer the readers to Porteus (1990) for a thorough review.

When demand and yield are deterministic, the problem of choosing which dimensional combinations to stock in order to minimize the combined stocking and substitution costs is also known as the *assortment problem*. This problem has been studied by several researchers (see Pentico (1988), and references therein). Ignall and Veinott (1969) first considered the multi-product inventory problem with downward (one-way) substitution and random demand. They focus on conditions under which a myopic ordering policy is optimal in a multi-period setting. Analytical results for two-product problems given perfect yield, are presented by McGillivray and Silver (1978) and Parlar and Goyal (1984). Bassok *et al*. (2000) consider solution methods for a two-stage stochastic linear programming formulation (2S-SLP) of large-scale multi-product problems with downward substitution. Gerchak *et al*. (1996) consider the two-product case in which there is also yield uncertainty. Studies of large-scale, single-period, multi-product substitutable inventory models that account for yield uncertainty have focused on applications to semi-conductor manufacturing. Bitran and Dasu (1992) study heuristics for a lot sizing model and Hsu and Bassok (1999) present an efficient algorithm for a 2S-SLP model that assumes a single lot size decision resulting in random yield of multiple products. Our model incorporates many aspects of these substitution models but also generalizes to cases other than downward substitution.

The importance of considering uncertain yield is well established. For example, Shih (1980) shows through an example that in the context of a newsvendor-type model, the cost impact of ignoring yield uncertainty can be nearly 5% of total costs. A comprehensive review of factors influencing yield randomness and related modeling approaches can be found in Yano and Lee (1995). For example, when each item in a lot may be defective with a constant probability, the number of good pieces is binomially distributed resulting in the so-called binomial yield model. However, a widely used model assumes *stochastically proportional* yield losses, i.e., the number of good items in a lot is the product of a random yield rate (with arbitrary distribution) and the lot size. We too assume stochastically proportional yield losses. Our model is applicable to large-scale problems involving multiple products with substitution.

The multi-product inventory problem with substitutable demand is related to the multi-location inventory problem with transshipment between locations studied by Karmarkar (1979), Robinson (1990), and others. These

problems are also related to the class of two-stage stochastic linear programming problems with network recourse in which the recourse problem constraints are of the network-flow type. Wallace (1986) studied specialized computational procedures for solving these types of problems. Our model is a generalization of the work of Wallace since it incorporates binary decision variables in the first-stage decision process. The first-stage decision variables represent the selection of supply and demand nodes that define the second stage network-flow problem.

There has been a growing interest in recent literature to model the problem of determining the optimal point of differentiation, subject to a service-level constraint (Lee and Billington, 1994; Lee, 1996; Garg and Tang, 1997; Lee and Tang, 1997; Graman and Magazine, 1998; Swaminathan and Tayur, 1998; Gupta and Benjaafar, 2004). These studies are also related to our model. However, in a majority of the cases, the authors' goal is to capture the benefits of inventory pooling when order-up-to-level inventory models are used to trigger replenishment. Only Swaminathan and Tayur (1998) consider design features of the semi-finished products as decision variables. Their model differs from ours in several ways. A key difference is that their model deals with an assembled product. The semi-finished products (computers) are called "vanilla" boxes and different designs are realized by choosing which components are assembled prior to demand realization. In the steel industry there is much greater product customization and the number of design choices is virtually infinite. Swaminathan and Tayur (1998) assume downward substitutions only, whereas we consider general substitutions. Furthermore, in our model we consider inventory storage space constraints whereas Swaninathan and Tayur (1998) model fixed costs.

Solution procedures for stochastic mixed-integer programming problems have been studied in several recent publications. For example, see Schultz *et al*. (1996) and Birge and Louveaux (1997, Chapter 8) for a review of general stochastic mixed-integer programming applications and solution methods, and Van der Vlerk (2000) for a recent bibliography. In this context our model can be classified as a *stochastic fixed-charge-network flow* problem (see Nemhauser and Wolsey, 1999, Chapter II.6 for discussion of the deterministic versions of these problems). Studies of the stochastic version are very limited. Louveaux and Peeters (1992) study a dual-based procedure for the stochastic uncapacitated facility location problem. Laporte *et al*. (1994) study exact solution procedures for a location problem with stochastic demands in which facility capacities (inventory levels) are chosen *a priori*. Another closely related work is by Rao *et al*. (2000). They study a multi-product inventory model with downward substitution and fixed setup costs. The differences between their model and ours are that they assume perfect yield, no storage constraints (rather, fixed setup costs), and take advantage of the downward substitution structure to propose simulation-based heuristics.

3. Model formulation and the hierarchical planning process

The model we will present can be used to address two types of decision-making processes for inventory deployment: (i) strategic planning decisions made on an infrequent basis (e.g., quarterly or biannually); and (ii) operational planning decisions made more frequently (e.g., weekly or monthly). The strategic planning problem, which we refer to as the Inventory Deployment Problem (IDP), can be described as follows. Given historical information about orders (demand volume and customer priorities), determine which orders should be served in the MTS production mode. Furthermore, given a known (but potentially large) set of inventory designs, determine which designs should be produced in the MTS mode to support the chosen customer orders, subject to a constraint on the total number of inventory designs that can be chosen. By default any orders not included in the MTS set are served by the MTO production mode. Orders that are planned to be served in the MTS mode, but for which there is insufficient planned production, are assumed to be satisfied by an alternate longer-cycle-time sourcing method (e.g., MTO production, outsourcing) and incur a shortage penalty. Excess production volume that cannot be utilized due to insufficient demand is assumed to incur a penalty representing the opportunity cost of reserving production capacity.

During the strategic planning phase, the ISM's goal is to choose those orders and inventory designs that support reliable high-volume slab-to-order allocation at an aggregate level during the horizon of the decision. The ISM then builds inventory of the chosen designs to mitigate the detrimental effects of supply and demand uncertainty. At this level of granularity, the entire planning period can be treated as a single period when accounting for the cost of supply–demand mismatches. Therefore, we propose a two-stage stochastic linear programming model in which demand and supply decisions across the strategic planning period are aggregated into a single second-stage period. For a case study supporting this assumption see Denton *et al*. (2003), in which a deterministic two-stage model was deployed at an ISM for determining slab design choices.

Operational planning of MTS orders involves periodically optimizing production to achieve target inventory levels (typically on a weekly or monthly basis), given that the type of inventory designs, and the orders served by them, are known. We refer to this problem as the Inventory-Level Problem (ILP).

At ISMs, operational planning periods for the MTS mode can reasonably be assumed to be independent for two reasons. First, due to long cycle times for slab production, and high dependence of production efficiency on sequencing and scheduling of the bottleneck resource (usually the continuous caster), shortages in one period cannot be recovered in the following period without substantial cost. Second, rescheduling to backfill a missed order

results in a domino effect that may cause many subsequent customer orders to be late. Thus, common practice is either to reschedule the order within the MTO mode (at substantial delay to the customer), or cover the order either by purchasing slabs/coils from another supplier, or by special processing of on-hand slabs/coils at higher cost. For example, a 42-inch wide in-stock coil can be cut to meet an order from a customer who requires a 30-inch wide coil, but since there is no market for the remaining 12-inch wide strip, this process generates substantial waste. Also, a typical ISM has ample capacity for finishing operations. Therefore, it is reasonable to assume that customized finishing can be accomplished in a short amount of time, and planning decisions in different operational planning periods are decoupled. Taking into account these realities of steel production and the need to keep the model tractable, we propose a model that assigns a fixed time-independent charge for each shortage, and shortages of MTS items do not carry over to the next period. Similar assumptions have been made to describe problems faced by other process-industry firms that segment their markets by lead time; see, for example, Carr and Duenyas (2000) for an example involving an automotive glass manufacturer. In the context of the proposed stochastic linear programming model, once the inventory-level (first-stage) decisions are made, the uncertainty in supply and demand is resolved and slab inventory is allocated optimally to realized customer orders in a second-stage linear program.

IDP decisions are long-range decisions that establish which inventory designs will be stocked, their target volumes, and the customers to be served in the MTS mode. The ILP, on the other hand, determines appropriate target inventory levels in each operational planning period, given uncertain demand and yield. In the following, we first formulate the IDP as a two-stage stochastic integer program in which the first stage corresponds to the selection of order-type, design, and production-level decisions, whereas the second stage corresponds to allocation decisions. We then argue that ILP can be obtained as a restriction of the IDP formulation with a set of preselected order-type and design choices, and an appropriate redefinition of model parameters and decision variables from a strategic to operational context. As a consequence of this observation, our efforts in the remainder of the article are focused on solving the IDP. In Section 4, we present the deterministic equivalent formulation for the IDP and in Section 5, we use the naturally existing hierarchical decomposition of decision variables, and certain properties of the IDP that we prove, to develop a heuristic solution procedure.

Both the IDP and ILP are formulated as two-stage stochastic linear programs. The second stages of the IDP and ILP are defined by the allocation of inventory designs to customer orders, over strategic and operational planning periods, respectively. This allocation depends on factors such as the grade of steel, the quality, and dimensions such as width and weight. Given a specific design for a piece of semi-finished inventory, and an order for a particular finished product, there is a set of rules used by planners to determine whether the application of the design to the order is permissible. The inventory allocation problem is then analogous to a *transportation problem* in which semi-finished inventories represent supply nodes, orders represent demand nodes, and there are some forbidden allocations.

3.1. *IDP model formulation*

The IDP problem has the network structure of a bipartite graph. Vertices in the graph can be partitioned into a set of potential supply nodes, $I = \{1, 2, \ldots, n\}$, that represent the set of design choices, and a set of potential demand nodes, $J = \{1, 2, \ldots, m\}$, that represent the different order choices. Edges between the supply and demand nodes indicate allowable allocations of supply to demand according to the application rules. We define the following additional notation:

c_i^{e} = per unit cost of having excess inventory of design i;
c_j^{s} = per unit cost of shortage for order-type j;
r_{ij} = additional revenue from cycle time reduction if design i is applied to order-type j;
c_i^{p} = additional per unit cost of producing design i in the MTS mode;
x_i = binary decision variable representing the decision to stock design i;
q_j = binary decision variable; $q_j = 1$ if order-type j is supplied from inventory, and 0 otherwise;
c = maximum number of permitted design choices;
w_i = production/procurement planned for design i;
y_{ij} = amount of order-type j supplied by design i;
a_{ij} = incidence parameter; $a_{ij} = 1$ if design i can be applied to order j and 0 otherwise;
s_j = shortage for order-type j;
e_i = excess production of design i;
U_i = random yield rate for design i;
D_j = random demand for order-type j;
ξ = random vector with yields, U_i, and demands, D_j, as components.

We use upper case for random variables, boldface for vectors, and Z_+ and $\Re_+$ to denote respectively the set of positive integers and positive real numbers. Similarly, $\Re$ denotes the set of all real numbers, and $B = \{0, 1\}$ is the set to which binary decision variables belong. The domains of the problem parameters are: $c \in Z_+$, $c_i^{\mathrm{e}} \in \Re$, $(c_j^{\mathrm{s}}, r_{ij}, c_i^{\mathrm{p}}, w_i, y_{ij}, s_j, e_i) \in \Re_+$, and $(x_i, q_j) \in B$. The random vector $\boldsymbol{\xi}$ has support $\boldsymbol{\Xi} \subseteq \Re_+^{m+n}$, probability distribution P, and finite first moments, $\bar{\boldsymbol{\xi}}$.

Total revenues and costs from serving customers in the MTO mode are assumed constant and therefore ignored in the model. The first-stage decisions are the design and order choices, $\mathbf{x} \in B^n$, and $\mathbf{q} \in B^m$, and the vector of planned production, $\mathbf{w} \in \Re_+^n$. In the first stage, there is a production cost, $\sum_{i=1}^{n} c_i^{\mathrm{p}} w_i$. In the second stage, there is a cost

$\sum_{i=1}^{n} c_i^e e_i$, for surplus production, and a cost $\sum_{j=1}^{m} c_j^s s_j$ for production shortages. The total additional revenue due to cycle time reduction from matching designs with demand is $\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} y_{ij}$. Thus, the complete second-stage objective function is:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} y_{ij} - \sum_{i=1}^{n} c_i^e e_i - \sum_{j=1}^{m} c_j^s s_j. \tag{1}$$

For convenience, the dependence of y_{ij}, e_i and s_j, on the realization of $\boldsymbol{\xi}$ has been suppressed in the notation.

There is a first-stage constraint on the number of available storage cells, c, of the form:

$$\sum_{i=1}^{n} x_i \le c, \tag{2}$$

which restricts the number of design choices but not the inventory levels. This is consistent with the approach used for storing high-volume slab designs at ISMs. High-volume designs are stored in *clone banks* in which slabs of identical dimensions are stacked in piles and several piles of identical slabs are stacked next to each other in a *cell*. Their heights are kept uniform by rotating the picking of slabs and as a result these piles are stable with much greater (effectively unlimited) heights. In contrast, low-volume storage systems have different types of slabs stacked in adjacent piles, making it difficult to maintain stability during picking if the stacks are high. Low-volume storage therefore has much shorter stacks. In short, the key storage constraint that is relevant for modeling design selection for high-volume slabs is the amount of space that can be allocated to clone banks, each of which can store very large quantities of identical slabs.

There are second-stage inventory balance constraints relating to the allocation of designs to demand of the form:

$$\sum_{j=1}^{m} a_{ij} y_{ij} + e_i = U_i w_i, \quad \forall\, i, \tag{3}$$

and

$$\sum_{i=1}^{n} a_{ij} y_{ij} + s_j = D_j q_j, \quad \forall\, j. \tag{4}$$

The first-stage binary decision variables, q_j, in Equation (4), reflect the fact that additional revenues and shortage-cost penalties are realized only for designs that are chosen to be covered via the MTS production mode. In other words there is no penalty associated with not being able to cover those customer orders for which no prior commitment has been made to have a shorter lead time. This is because customers perceive lead time variation on orders negatively and thus a shortage cost is applied to model the impact of customer dissatisfaction. The right-hand side of constraint (3) implies a stochastically proportional yield-loss model, which closely approximates yield losses in the steel industry.

The complete problem, assuming a *risk-neutral* firm, can be written as follows:

$$\max\{Z = -\mathbf{c}^p \mathbf{w} + Q(\mathbf{x}, \mathbf{q}, \mathbf{w})\}, \tag{5}$$

subject to

$$\sum_{i=1}^{n} x_i \le c, \tag{6}$$

$$\mathbf{x} \in B^n, \quad \mathbf{q} \in B^m, \quad \mathbf{w} \ge 0, \tag{7}$$

where $Q(\mathbf{x}, \mathbf{q}, \mathbf{w})$ is called the recourse function. It is the expected additional revenue earned, net of any shortage/overage costs, accruing from the inventory allocation decisions. In fact, $Q(\mathbf{x}, \mathbf{q}, \mathbf{w}) = E_{\boldsymbol{\xi}}\,[Q(\mathbf{x}, \mathbf{q}, \mathbf{w}, \boldsymbol{\xi})]$, where $Q(\mathbf{x}, \mathbf{q}, \mathbf{w}, \boldsymbol{\xi})$ is defined by:

$$Q(\mathbf{x}, \mathbf{q}, \mathbf{w}, \boldsymbol{\xi}) = \max\left\{ \sum_{i=1}^{n}\sum_{j=1}^{m} r_{ij} y_{ij} - \sum_{i=1}^{n} c_i^e e_i - \sum_{j=1}^{m} c_j^s s_j \right\}, \tag{8}$$

subject to

$$\sum_{j=1}^{m} a_{ij} y_{ij} + e_i = U_i w_i, \quad \forall\, i, \tag{9}$$

$$\sum_{i=1}^{n} a_{ij} y_{ij} + s_j = D_j q_j, \quad \forall\, j, \tag{10}$$

$$y_{ij} \le M a_{ij}, \quad \forall\,(i,j), \tag{11}$$

$$y_{ij} \le D_j x_i, \quad \forall\,(i,j), \tag{12}$$

$$y_{ij} \ge 0,\ s_j, e_i \ge 0, \quad \forall\,(i,j). \tag{13}$$

Note that problem (5)–(7) has complete recourse, i.e., it is feasible for any $(\mathbf{x}, \mathbf{w})$, due to the positive linear basis provided by $(\mathbf{e}, \mathbf{s})$ in constraints (9)–(10), and the fact that $U_i \ge 0, D_j \ge 0, \forall (i,j)$. Furthermore, randomness occurs only in the right-hand sides of constraints (9)–(12) and second-stage cost coefficients are deterministic. For each design, the amount of slab inventory is measured in tons, and y_{ij}s are treated as a continuous variable even though slabs are discrete units with an average weight of about 20 tons. This assumption is reasonable since the y_{ij}s are typically large (additional arguments in support of this assumption are provided in the next section). Note that in problem (5)–(7), we place no upper bound on the variables w_i since the total production in the MTS mode accounts for less than half of the ISM capacity and there are substantial production efficiencies associated with the high volume MTS mode. Although the solution methodology can easily accommodate an upper bound on the total MTS production, we have formulated the problem in a manner that is most consistent with the intended application.

We make the following assumptions about the objective function coefficients. Shortage costs, c_j^s, are nonnegative, i.e., it is never advantageous to incur a shortage.

The marginal revenues are such that if for some (i, j), $a_{ij} = 0$, then $r_{ij} = 0$ as well. Furthermore, $r_{ij} \geq \max\{-c_j^{\mathrm{s}} - c_i^{\mathrm{e}}, 0\}, \forall(i, j)$, i.e., it is never advantageous to choose not to allocate available supply of design i to order j if $a_{ij} = 1$ for some i. It is also assumed that first-stage procurement cost coefficients are such that $c_i^{\mathrm{p}} + c_i^{\mathrm{e}} > 0$, since otherwise it is trivially optimal to produce an infinite quantity of design i, and that for each design i there is an order-type j such that $c_j^{\mathrm{s}} + r_{ij} > c_i^{\mathrm{p}}$, since otherwise it is optimal not to produce design i at all.

For ISMs there are typically no significant fixed costs associated with choosing a particular design since the amount of space available in the slab yard is dedicated space for storing slab inventory. However, the storage cell constraint in Equation (6) plays a role similar to fixed costs through the implied opportunity cost associated with not choosing one of the other potential designs. Also, note that there is no explicit constraint forcing optimal production level $w_i^* = 0$ if $x_i = 0$. However, this is implied by constraints (9) and (12) since otherwise $e_i > 0$ and unnecessary excess costs are incurred with no additional rewards. Similarly, constraints (10) and (12) imply $q_j^* = 0$ if all x_i for which $a_{ij} = 1$ are zero, i.e., if no applicable inventory design has been chosen. Constraint (11) limits the application of design i to order-type j based on whether the application is feasible, i.e., $a_{ij} = 1$. In this constraint, M is chosen to be sufficiently large so that all demand can be met if design i is indeed selected.

3.2. *The ILP model and the hierarchical framework*

As mentioned at the start of Section 3, both IDP and ILP can be addressed by the common underlying mathematical model given in Equations (5)–(13). Essentially, when the design choice variables, x_i, and the order choice variables, q_j, are fixed to predetermined values, we obtain the ILP. Although the two models are structurally similar, in the context of operational planning the decision variables and model parameters take on a new meaning. For instance, the decision variables, w_i, are now interpreted as short-term production quantities required to achieve target inventory levels in an operational planning period (e.g., week). The second-stage decision variables e_i and s_j represent excess inventory holding cost, and the cost of not being able to satisfy demand due to a stockout in the operational period. Similarly the variables, y_{ij}, denote the amount of order-type j supplied by design i within an operational period.

The random variables defined in the model also take on a different meaning in the ILP. The yield variables, U_i, now corresponds to the uncertain yield rate within an operational planning period. Whereas the yield variables in the IDP are influenced by long-range factors, such as the adoption of new manufacturing processes and the quality of purchased slabs, the yield variables in the ILP depend on short-term factors, such as quality uncertainty in a particular production batch. Similarly the random demand variables, D_j, correspond to the short-term demand that arrives for a stocked design in the operational planning period. Since this represents a reduced aggregation of demand when compared to the IDP (note IDP aggregates demand over many operational planning periods), the demand variables in the ILP are typically more variable in relation to their means.

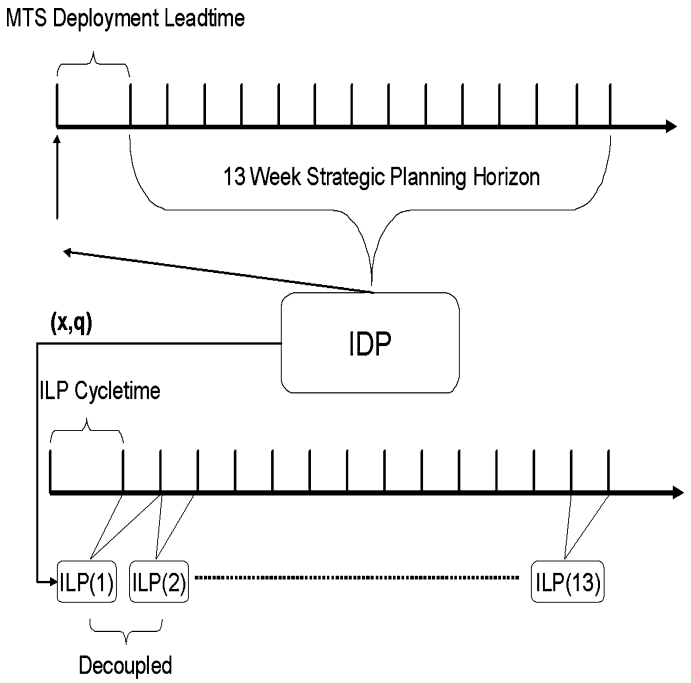

Fig. 2. Illustration of the application of the IDP and ILP within a hierarchical planning process.

The application of the IDP and ILP together forms a natural hierarchical planning architecture in which IDP establishes long-range strategic decisions, and ILP is used to establish shorter range operational planning objectives. Figure 2 is an illustration of an appropriate planning context for the two models. In this illustration IDP is used on a quarterly basis, and the second stage is a single 13-week period. The IDP design and order choices are determined and provided as input to the ILP model. Independent ILP problems are then solved, where each problem corresponds to decisions for a particular week of the 13 weekly periods.

4. Deterministic equivalent problem analysis

Since instances of the IDP encountered in practice have a very large number of decision variables (see Section 5 for details), we assume from this point forward that the support, Ξ, is a finite set of scenarios, $\boldsymbol{\xi}^k = (u_1^k, u_2^k, \ldots, u_n^k, d_1^k, d_2^k, \ldots, d_m^k)$, with associated probabilities p_k, $k = 1, \ldots, K$. (In Section 5.1 we discuss methods

for generating scenarios.) The deterministic equivalent of problem (5)–(7) can be written as:

$$\max \left\{ Z = -\sum_{i=1}^{n} c_i^{\mathrm{p}} w_i + \sum_{k=1}^{K} p^k \left[\max \left\{ \sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij} y_{ij}^k - \sum_{i=1}^{n} c_i^{\mathrm{e}} e_i^k - \sum_{j=1}^{m} c_j^{\mathrm{s}} s_j^k \right\} \right] \right\}, \tag{14}$$

subject to

$$\sum_{i=1}^{n} x_i \leq c, \tag{15}$$

$$\sum_{j=1}^{m} a_{ij} y_{ij}^k + e_i^k = u_i^k w_i, \quad \forall\, i, k, \tag{16}$$

$$\sum_{i=1}^{n} a_{ij} y_{ij}^k + s_j^k = d_j^k q_j, \quad \forall\, j, k, \tag{17}$$

$$y_{ij} \leq M a_{ij}, \quad \forall\, i, j, \tag{18}$$

$$y_{ij}^k \leq d_j^k x_i, \quad \forall\, i, j, k, \tag{19}$$

$$x_i, q_i \in \{0, 1\}, \quad \forall\, i, j, y_{ij}^k, s_j^k, e_i^k \geq 0, \quad \forall\, i, j, k. \tag{20}$$

In the remainder of this section, we present three propositions that establish characteristics of the IDP. These results are used later in the construction of the solution heuristic. Proofs of all propositions can be found in the Appendix.

Consider a relaxation of Equations (14)–(20) in which the right-hand sides of the subproblem constraints are replaced by their first moments (Huang *et al.*, 1977). Put differently, replace each set of subproblem constraints (16), (17) and (19), with the sum of the K rows weighted by their associated probabilities, p^k. This relaxation, known as the *mean-value problem*, significantly reduces the size of the deterministic equivalent problem by reducing the K subproblems to a single subproblem.

Proposition 1. *The mean-value relaxation of the IDP is equivalent to the p-median given in Equations* (21)–(25) *below*:

$$\max \left\{ Z = \sum_{i=1}^{n} \sum_{j=1}^{m} r_{ij} y_{ij} \right\}, \tag{21}$$

subject to

$$\sum_{i=1}^{n} x_i \leq c, \tag{22}$$

$$\sum_{i=1}^{n} y_{ij} \leq 1, \quad \forall\, j, \tag{23}$$

$$y_{ij} \leq x_i, \quad \forall\, i, \tag{24}$$

$$x_i, y_{ij} \in \{0, 1\}, \quad \forall\, i, j. \tag{25}$$

Thus, the special case of the IDP in which $K = 1$, corresponds to the p-median problem. The p-median problem has been studied extensively and it is well know to be NP-complete (see Garey and Johnson, 1979, p. 220).

Proposition 2. *For each fixed* $\mathbf{x} \in B^n$ *there is an optimal integer solution to a relaxation of the IDP in which the constraints* $\mathbf{q} \in B^m$ *are relaxed to* $\mathbf{q} \in \Re^m$, *and* $0 \leq q_j \leq 1\ \forall\ j$. *That is, the optimal* $\mathbf{q}^*$ *from the relaxed problem is such that* $\mathbf{q}^* \in B^m$.

Thus, by Proposition 2, we need not restrict $\mathbf{q}$ to be binary. This fact reduces the number of stage-1 binary decision variables from $m + n$ to n, and effectively increases the size of instances of IDP that can be solved reliably by exact methods. It is also important in the development of heuristics.

Recall that earlier in this section, we assumed that the y_{ij}s are continuous, even although production occurs in discrete units. The following proposition further supports this assumption by showing that at most c allocations may have non-integer values; the remaining are equal to the demand which are integer valued.

Proposition 3. *At most c orders will receive partial* (*non-integer*) *allocation of supply.*

Since in the type of application considered here the d_j^ks are typically large, and c is typically much smaller than m, the worst-case error from the approximation of continuous allocation of discrete items of inventory is expected to be small.

The heuristic proposed in the next section relies on decomposition of the scenario subproblems. It is therefore important to establish the computational complexity of the second-stage recourse problem. Since it has a network structure corresponding to the transportation problem, polynomial-time algorithms such as the primal dual or the flow augmentation algorithms, can be used to solve such problems (Nemhauser and Wolsey, 1999, Chapter I.3). In some special cases the transportation problem can be solved trivially using a greedy-type algorithm (Dietrich, 1990). The basic idea of Dietrich's algorithm is that arcs are ordered such that flows across each are sequentially maximized subject to the maximum of available supply and demand. The existence of such a sequence can be proved based on the fact that every basis for a transportation problem is triangular (Murty, 1983, Corollary 13.2, p. 382). However, identifying the sequence is not easy, except in certain special cases (see, for example, Hsu and Bassok (1999)). Unfortunately, for applications in the steel industry where substitution is not strictly downwards, these special circumstances do not apply.

5. Heuristic description and numerical experiments

The IDP is a difficult problem to solve due to its large size. The typical problem has approximately 10^7 design choices and 10^4 catalog items with positive demands. On average a single design may cover 10 different end items. Therefore, even if there is a single demand/yield scenario (in which case the problem reduces to the p-median problem), the resulting mixed-integer program has on the order of 10^8

continuous decision variables y_{ij}s. It is therefore not realistic to expect to obtain an exact solution, say by using a branch-and-bound approach. Given the similarities between the IDP and the well known p-median problem, and the fact that several good heuristics are available to solve the p-median problem (see Chapter 2 of Mirchandani and Francis (1990) for details), it is natural to ask if one of those heuristics can be adapted to solve the IDP. It turns out that the extremely large size of the IDP also limits that choice. Consequently, we propose a heuristic that decouples the first- and second-stage decision variables. The proposed approach is to initially approximate the first-stage decision variables using the p-median relaxation of the problem, and then compute the optimal second-stage decisions after fixing the first-stage decisions.

Decomposition Heuristic (DH):

Step 1. (Relaxation). Relax the IDP to the corresponding deterministic p-median problem. Apply the greedy-interchange heuristic (see Cornuejols *et al.* (1977) for a description) to determine a near-optimal solution, $\mathbf{x}^{\mathrm{d}}$.

Step 2. (Restriction). Restrict the IDP to $\mathbf{x} = \mathbf{x}^{\mathrm{d}}$. Solve the restricted IDP using the L-shaped decomposition method to obtain $\mathbf{q}^{\mathrm{d}}$ and $\mathbf{w}^{\mathrm{d}}$.

Recall from Proposition 2 that $\mathbf{q}$ need not be restricted to be binary variables when solving the IDP. Therefore, Step 2 is a stochastic linear program for which the L-shaped method can be used to obtain the optimal solution (for details, see Van Slyke and Wets (1969)). Next, we test the heuristic proposed above in a series of numerical experiments.

5.1. *Numerical experiments*

We report results from two different types of numerical experiments designed to test the decomposition heuristic. The first study compares the heuristic solutions to the IDP with the optimal solutions in a series of small-sized, randomly-generated test problems. The second study uses actual data from a particular ISM to demonstrate the use of the IDP and ILP models in realistic examples. Calculations in both studies are performed on a Sun Ultra 10 workstation with 128 MB Ram, the programming is done in C/C++, and the commercial solver CPLEX 4.0 is used both for solving mixed-integer programs and linear programs.

There is a wide range of possible problem structures that could arise in practice. We define the size of a problem in terms of c, n and m and then choose a number of randomly generated test instances for each problem size. The problem instances are generated as follows. From each of $i = 1, \cdots, m$ supply nodes, arcs are generated to each of $j = 1, \cdots, n$ demand nodes according to the outcome of a Bernoulli trial with success probability p^{a}. That is, for each supply–demand pair there is probability p^{a} that there is an arc between them and probability $1 - p^{\mathrm{a}}$ that there is no arc. For a particular instance of the problem, all supply–demand pairs have the same p^{a} which is obtained by sampling from the distribution $F(p^{\mathrm{a}})$. Larger probabilities imply greater substitutability of available supply.

Yield and demand were assumed to be i.i.d. in these experiments and sampled according to $U_i \sim \text{uniform}(a, b)$, $\forall\, i$ and $D_j \sim \text{normal}(\mu, \sigma^2)$, $\forall\, j$, to generate a set of K scenarios. The interval over which the uniform distribution is defined is such that $0 \leq a < b \leq 1$. Similarly, the mean (μ) and the standard deviation (σ) of the normal distribution are chosen to ensure non-negative values. The probability of having a negative demand realization becomes negligible when the coefficient of variation (the ratio μ/σ) is at least four. Additional procurement costs $\mathbf{c}^{\mathrm{p}}$ are assumed to be the same for all supply nodes and are fixed at one. Coefficients for additional revenues, shortage costs, and excess costs are all distributed as uniform(1, 4).

The results presented in Table 1 illustrate the accuracy of the heuristic for small randomly-generated test problems with respect to exact solutions obtained using the CPLEX mixed-integer solver. Relative errors are reported as $100 \times |\, (Z^* - Z)/Z \,|$. Column headings in the table are:

MV: Average relative error of the mean-value (p-median) solution compared to the exact solution.
ΔMV: Maximum relative error of the mean-value (p-median) problem compared to the exact solution among all randomly generated problem instances.
DH: Average relative error of the DH solution compared to the exact solution.

Table 1. Numerical results for randomly generated problem instances with $K = 25$, $U_i \sim U(0.8, 1)$, $\forall\, i$ and $D_j \sim N(10, 2)$, $\forall\, j$

$F(p^a)$	(c,n,m)	$\bar{MV}$	ΔMV	$\bar{DH}$	$\Delta\, DH$	*Avg. %*
$U(0.1, 0.3)$	(5, 10, 20)	9.43	15.65	0.31	2.09	86
	(5, 10, 30)	8.55	11.42	0.37	2.84	94
	(5, 20, 30)	2.77	4.67	1.88	7.53	90
	(5, 20, 40)	7.05	11.03	0.61	2.96	92
	(10, 20, 30)	5.50	6.35	0.76	1.65	95
	(10, 20, 40)	5.29	6.77	0.49	1.02	91
	(10, 30, 50)	3.59	4.79	0.18	0.47	88
$U(0, 0.4)$	(5, 10, 20)	10.18	14.79	1.14	4.92	88
	(5, 10, 30)	9.40	15.4	0.11	0.66	96
	(5, 20, 30)	7.06	9.03	0.80	2.60	92
	(5, 20, 40)	6.62	8.21	0.95	3.64	84
	(10, 20, 30)	6.06	7.19	0.47	1.21	98
	(10, 20, 40)	5.13	6.25	0.84	2.22	93
	(10, 30, 50)	3.84	4.71	0.26	1.23	94
$P(10, 0.25)$	(5, 10, 20)	59.74	78.62	0.28	1.93	94
	(5, 10, 30)	41.36	58.55	0.00	0.00	96
	(5, 20, 30)	34.84	47.56	1.89	7.69	88
	(5, 20, 40)	23.51	31.53	0.98	4.62	86
	(10, 20, 30)	33.46	50.91	0.77	2.08	89
	(10, 20, 40)	24.94	29.00	0.68	2.38	88
	(10, 30, 50)	20.58	27.88	0.49	1.14	87

ΔDH: Maximum relative error of the DH heuristic compared to the exact solution for all randomly generated problem instances.

Avg. %: Measure (in percent) of the number of designs chosen by the DH heuristic that are also in the optimal solution.

Average and maximum errors are determined from solutions to 20 randomly-generated problem instances. From here on forward, U stands for the uniform distribution and N for the normal.

In Table 1 results are presented for two different types of probability distributions: uniform (continuous), and Pareto (discrete). Two instances of the uniform distribution, $U(0.1, 0.3)$ and $U(0, 0.4)$, are picked. They have different variances but the same mean. The Pareto distribution is a truncated distribution of the form $P(10, 0.25) = C/k^{1.25}, k = 1, \ldots, 10$ where C is chosen such that the total probability sums to one. We chose the Pareto distribution as one of the test cases because it is significantly different from the uniform distribution, and because there is empirical evidence that the number of orders covered by inventory designs is Pareto distributed. In all examples we assume that the demand is normally distributed as $N(10, 2)$. However, for the test cases with the Pareto distribution we further assume that with probability 0.5 the demand is zero. This latter assumption has the effect of increasing demand variance. It simulates the effect of order cancelations, a major source of uncertainty for ISMs.

The results in Table 1 are quite favorable for the DH heuristic. The overall average error across all problem instances (420 in total) is 0.67% and the overall worst-case error is 7.69%. Also, in nearly 50% of the test problems the heuristic finds the optimal solution, and on average 90.5% of the designs chosen by the heuristic are also in the optimal solution. The average error is similar for the two uniformly distributed $F(p^a)$ test cases. For $U(0.1, 0.3)$ the average error over all the test cases is 0.66% and for $U(0, 0.4)$, it is 0.65%. Thus, the performance of the heuristics does not appear to be too sensitive to a change in the variance of p^a. The MV solution (obtained from solving the p-median relaxation) has an average relative error of 6.46% when yield rate is uniformly distributed. However, when $F(p^a) \sim P(10, 0.25)$, the average error is significantly larger at 34.06%. This can be attributed to order cancelations that cause significantly higher demand variability.

The effectiveness of the DH heuristic in general can be attributed to two factors: (i) variance reduction due to order pooling for high-volume designs; and (ii) the overlapping substitutability of designs for orders which leads to *chaining* (Jordan and Graves, 1995). Pooling of orders is a natural result of the fact that the p-median solution in Step 1 of the heuristic prefers high-volume designs which tend to be ones that cover many order types (see Denton *et al.* (2003) for further evidence). Chaining is supported by design substitution opportunities that exist due to the inherent flexibility of downstream manufacturing operations. Clearly, increased chaining is not a goal of the p-median model in Step 1. The highly substitutable nature of designs naturally results in choices in which substitution opportunities can be exploited in Step 2 of the heuristic.

To provide additional insight into the heuristic's performance, we examine the worst-case problem instance (the third row in the block with yield-rate distribution $P(10, .25)$) from Table 1. Note that c, n and m are 5, 20 and 30, respectively, and the maximum error is 7.69%. In this particular case, the DH solution and the optimal solution differ by a single design; that is, there is one design included in each set that is not in the other set. However, there are several factors that combine to cause the relatively large error. In both solutions, the design that is not in the other set of designs covers an order-type that is not covered by any other design. However, for the DH solution, this order-type has high shortage and excess costs of 3.92 and 3.59 respectively. In the optimal solution, in contrast, the shortage and excess costs are 1.86 and 2.8 respectively. This makes the cost of supply–demand mismatch much lower in the optimal solution. Another difference is that the optimal set of designs covers 11 order types for which the combined coefficient of variation is 0.57. In contrast, the DH design set covers only three order types and has a significantly higher coefficient of variation of 1.06. Having a larger number of order types allows for increased chaining opportunity in addition to the variance reduction caused by order pooling.

Since both MV and DH algorithms solve the p-median problem to choose designs, it is possible to compare the results from the MV solution, the DH solution, and the optimal solution and reach the following conclusion. The use of the mean-value (p-median) relaxation is a reasonable approach for choosing designs, however, it can result in significant error if it is used to determine planned inventory levels.

5.2. *A steel industry example*

The examples in this section illustrate the impact of uncertainty on inventory planning decisions for a set of slab designs actually stocked at an ISM. These designs are chosen by solving the p-median relaxation. Our experiments in this subsection also apply to another decision problem faced by ISMs. At least twice a year, ISMs shut down their furnaces for repair and maintenance. During this time, no new slabs are produced, but the hot mill and the finishing lines do not shut down. They are fed from a stockpile of slabs that is planned in advance of the shutdown, in order to maximize utilization of the hot mill and the finishing lines. The problem that ISMs face is to decide which slabs to stock and how much of each slab type to stock. Keeping more inventory reduces the chance of starvation during the shutdown, but also increases the chance of incurring unnecessary holding costs.

For reasons of maintaining confidentiality, no specific information about the policies or design choices of the ISM is revealed in the results reported below. However, the examples demonstrate the types of problems that can be studied using our model, as well as general trends and insights into the effects of uncertainty on inventory deployment decisions. We begin by explaining an approach for generating scenarios using historical order book data. The approach is based on discussions with senior managers at an ISM and captures what they view to be the major sources of uncertainty.

Using the historical data, we define two sets of orders: (i) a set of planned orders, J_p; and (ii) a set of possible *replacement* orders $J_s \cdot J_p$ corresponds to historical orders for which the processing-time-adjusted due dates fall during the chosen planning horizon in the previous year, and J_s corresponds to a set of likely replacements in the event customers change due dates or cancel orders. The set J_s consists of orders chosen randomly from periods that do not coincide with the chosen planning period. (Note that J_p could alternatively be based on a forecast of orders.) We denote the set of order types that arise in scenario k as J_p^k, and the corresponding actual order size for order-type j as d_j^k. The demand scenarios are generated by sampling from the sets J_p and J_s as explained in the following three-step process. (The $+/-$ operations below correspond to addition/removal of elements from a set.)

For $k = 1, \ldots, K$ do:

Step 0. Initialize J_s and J_p and set $d_j^k = d_j, \forall j \in J_p$, and $J_p^k = J_p$.

Step 1. For each $j \in J_p^k$, with probability p_j^s, perform the following interchange operations with a randomly drawn order from the order book, j_c:

$$J_p = J_p - j + j_c, \quad J_s = J_s + j - j_c.$$

Step 2. For all $j \in J_p$ set $d_j^k = (1 + \Delta_j)d_j$ where Δ_j is a sampled random deviate satisfying $\Delta_j > -1$.

The two parameters that cause uncertainty in demand are p_j^s and Δ_j. The parameter p_j^s causes perturbations in the set of orders that are likely to materialize in the chosen planning period, whereas the parameter Δ_j affects order quantities. For the purpose of generating scenarios, the parameters $p_j^s = p^s$ and $\Delta_j = \Delta$ are assumed the same for all order types. There is a single planned orders set, J_p, and a single replacement orders set J_s. However, the realized order scenarios, J_p^k, are distinct for each k. The length of the planning horizon chosen is such that there are approximately 10^3 orders in this length of time. To complete the generation of ξ^k, random yield rates are sampled from the appropriate yield distribution. Since actual yield data was not available for these computational experiments, we assumed uniformly distributed yield rates. Also, the parameter Δ is sampled from $N(\mu, \sigma^2)$, with a sufficiently large coefficient of variation to ensure a negligible chance of observing $\Delta < -1$.

In practice the calibration of the various cost coefficients is a difficult problem. For the purpose of numerical examples considered here it is assumed that shortage costs are identical for all order types, and that the excess and procurement costs are each the same for all slab designs. The marginal revenues, r_{ij}, depend on the relative differences between slab and order width and weight:

$$r_{ij} = \begin{cases} r - c^w\left(w_i^s - w_j^d\right) - c^m\left(m_j^d - m_i^s\right) & \text{if slab } i \text{ is applicable to order } j, \\ 0 & \text{otherwise,} \end{cases}$$

where w_j^d, w_i^s are the widths for order j and slab i, respectively, and m_j^d, m_i^s, are the weights for order j and slab i, respectively. For ISMs, when a slab is applicable to an order, it must have at least the order width and at most the order weight, i.e., $w_j^d \le w_i^s$ and $m_j^d \ge m_i^s$. This is because a wider slab can be reduced in width during hot rolling (within limits) by a process called roughing. Similarly, many customers accept somewhat lower-weight coils than the ordered weight. However, higher weights are typically not accepted on account of weight restrictions at customer loading docks. Coefficients c^w and c^m are calibrated with respect to the cold-application rules, such that the minimum reward for a feasible slab-to-order application satisfies $r_{ij} \ge 0$.

The numerical examples in Table 2 confirm the accuracy of the greedy-interchange heuristic for generating a solution to the p-median relaxation of the IDP for a real-world problem. In this table we compare the solutions obtained using the greedy-interchange heuristic with upper bounds generated by the Lagrangian dual to the p-median problem. The latter are generated using a standard subgradient approach

Table 2. Numerical results for several values of c for greedy, and greedy + interchange heuristics, and the solution of the Lagrangian dual (upper bound). Results are presented as the percentage of total order-tons that can be covered with c designs. The quantities in parentheses are the number of pairwise interchanges performed by the interchange heuristic

c	*Greedy heuristic*	*Greedy + interchange heuristic*	*Lagrangian dual*	*% gap*
10	22.751	23.342 (12)	23.473	0.56
20	33.130	33.144 (6)	33.699	1.65
30	40.964	40.979 (6)	41.581	1.54
40	46.562	46.685 (8)	47.522	1.76
50	50.756	50.770 (10)	50.793	0.04
60	54.340	54.902 (21)	55.769	1.55
70	57.425	57.988 (21)	58.577	1.00
80	60.059	60.620 (30)	61.143	0.85
90	62.439	62.974 (27)	63.488	0.81
100	64.628	65.198 (30)	65.736	0.82
150	73.450	73.794 (34)	74.608	1.09
200	79.707	79.833 (15)	81.158	1.63

Table 3. Computation times for solution of ILP with varying number of slab designs and demand scenarios

c/K	*25*	*100*	*300*	*500*
5	3	18	67	112
10	23	92	281	480
15	33	190	635	1251

(see for example, Nemhauser and Wolsey (1999)). Comparison with a Lagrangian dual upper bound (last column in Table 2) indicates that the combination of greedy and interchange heuristics typically finds a solution that is within 2% of optimality. Thus, it provides both fast and high-quality solutions for Step 1 of the decomposition heuristic for the applications considered here.

Some limited examples of solution times for the ILP solved in Step 2 of DH are illustrated in Table 3. The L-shaped method solves second-stage scenario subproblems to generate cuts that are applied to a master problem, where the latter is a relaxation of the ILP. Since the second-stage scenario subproblems are of the *network-flow* type, experiments are performed using the *CPLEX network-solver* option. The sample computation times illustrate the fact that when using the network-solver option, dependence on K for fixed c is nearly linear. Dependence on c for fixed K, on the other hand, is approximately quadratic. In general the solution times are roughly independent of changes to cost coefficients. Taking advantage of the network structure makes solution times between four and eight times faster for the problems in Table 3 compared to the CPLEX's primal-simplex solver.

We next report examples that demonstrate the managerial relevance of the model by uncovering new insights. Figure 3 is a plot of Z^* as a function of yield variance, with fixed mean. Yields are assumed uniformly distributed as $U_i \sim U(0.8 - \delta/2, 0.8 + \delta/2), \forall\, i$, and Z^* is plotted against δ. When yields are deterministic, optimal planned inventory levels are inversely proportional to the yield rate. Thus, a significant sensitivity to the first moment of the U_i is natural, especially when incremental production costs, c_i^{p}, are high. Our experiments shows that Z^* can also be sensitive

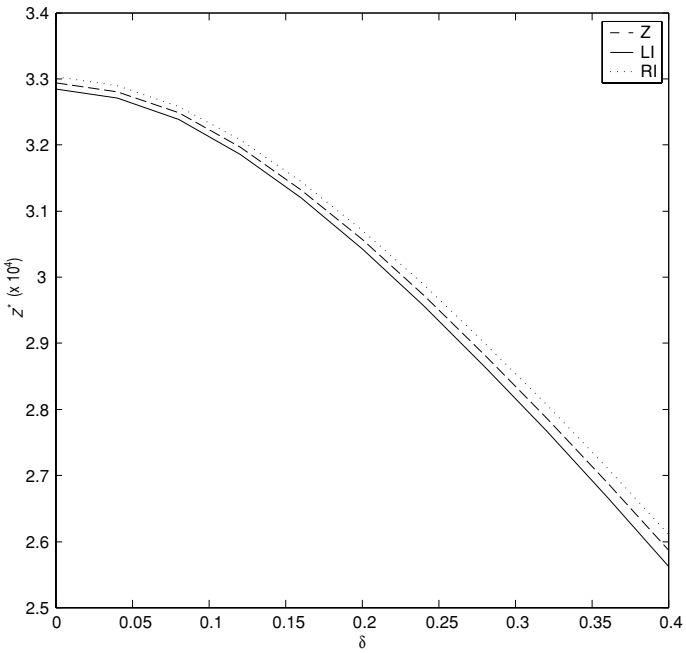

Fig. 3. Z^* as a function of $\delta = (b - a)$ for $U_i \sim U(a, b)$ for $n = 15$, $K = 500$, $p^{\mathrm{s}} = 0.1$, $\Delta \sim N(0, 0.1)$.

to changes in the second moments of the yield distribution. For example, in Fig. 3, Z^* is initially decreasing at an increasing rate; becoming roughly linear when $\delta > 0.1$. This sensitivity is further evidence that the mean-value solution would be a poor approximation when yield variance is high.

Historical data reveal that demand can be highly variable even for high-volume designs. Therefore, we studied examples that shed light on the cost of demand uncertainty. Results are compiled in Table 4 for several different choices of cost coefficients (for simplicity yields are assumed perfect in these examples). The table shows Z, the heuristic solution; σ_Z, a statistical estimate of its standard deviation; and LI and RI, the left and right limits of the 95% confidence intervals, respectively. The table also shows how uncertainty affects the ILP. The column WS in Table 4

Table 4. Numerical results for 15 slab designs, no yield losses, $K = 500$, $p^{\mathrm{s}} = 0.1$ and $\Delta \sim N(0, 0.1)$

c^p, c^s, c^e	Z	σ_Z	LI	RI	WS	MV	$EVPI$
(1, 4, 4)	64991.1	207.2	65397.3	64585.0	76707.7	63685.9	11716.6
(2, 4, 4)	44114.3	188.2	44483.4	43745.3	55811.4	42789.6	11697.0
(3, 4, 4)	23668.8	175.0	24011.8	23325.7	34915.0	21893.2	11246.2
(1, 4, 2)	67486.7	187.6	67854.6	67118.9	76707.7	65754.0	9220.9
(2, 4, 2)	45985.0	167.3	46313.0	45657.0	55811.4	44857.7	9826.3
(3, 4, 2)	25044.5	153.8	25346.1	24742.9	34915.0	23961.3	9870.5
(1, 4, −0.5)	73469.4	167.7	73798.2	73140.5	76707.7	68339.2	3238.3
(2, 4, −0.5)	49806.8	138.0	50077.3	49536.2	55811.4	47442.8	6004.6
(3, 4, −0.5)	27666.9	122.1	27906.3	27427.6	34915.0	26546.5	7248.0

Table 5. Lot sizes with respect to mean-value planned inventory levels for 10 slab designs, $U_i \sim U(0.8, 1.0)$, $\forall\, i$ and $K = 500$, $p^s = 0.1$ and $\Delta \sim N\,(0,0.1)$

(c^p, c^s, c^e)	$\frac{x_1^*}{\mu_2}$	$\frac{x_2^*}{\mu_2}$	$\frac{x_3^*}{\mu_3}$	$\frac{x_4^*}{\mu_4}$	$\frac{x_5^*}{\mu_5}$	$\frac{x_6^*}{\mu_6}$	$\frac{x_7^*}{\mu_7}$	$\frac{x_8^*}{\mu_8}$	$\frac{x_9^*}{\mu_9}$	$\frac{x_{10}^*}{\mu_{10}}$	$\frac{\sum x_i^*}{\sum \mu_i}$
(1, 4, 4)	1.00	1.01	0.91	0.91	1.03	1.01	2.25	0.93	0.98	0.89	1.01
(2, 4, 4)	0.94	0.99	0.90	0.91	1.35	0.94	1.54	0.94	1.42	0.87	0.99
(3, 4, 4)	0.92	0.96	0.88	0.89	1.33	0.92	1.51	0.92	1.40	0.85	0.97
(1, 4, 2)	1.03	1.04	0.98	0.96	1.08	1.11	1.33	0.93	1.05	1.04	1.04
(2, 4, 2)	1.01	1.01	0.89	0.91	1.04	1.02	1.39	0.97	0.88	0.88	1.00
(3, 4, 2)	0.96	0.98	0.91	0.96	1.03	1.11	1.79	0.88	0.99	0.76	0.98
(1, 4, −0.5)	1.09	1.15	1.09	1.09	1.44	1.59	1.30	1.18	1.20	1.41	1.15
(2, 4, −0.5)	1.02	1.09	0.94	1.06	1.18	1.05	1.22	1.08	1.17	1.06	1.06
(3, 4, −0.5)	0.98	1.02	0.94	0.94	1.28	1.01	1.19	1.01	1.22	0.75	1.02

represents the *Wait-and-See* solution, i.e., the solution obtained if all uncertainty is resolved prior to making the inventory-level decisions. Furthermore, MV and EVPI represent the Mean-Value solution and the *Expected Value of Perfect Information*, respectively. EVPI is the difference between WS and Z. It is a measure of the value of eliminating demand uncertainty. From the table it is clear that EVPI is typically large with respect to the optimal solution. Furthermore, results indicate that the mean-value solution is typically within 2 to 8% of optimal. This evidence shows that there are significant cost advantages associated with the use of the stochastic model for setting target inventory levels.

Table 5 shows how the ratio (x_j/μ_j) of the optimal and the mean-value planned inventory levels varies with respect to changing cost parameters. The mean-value planned inventory levels should be the same as the optimal planned inventory levels when both yield rate and demand are constant and equal to their respective means. Their ratio is therefore an indication of how much variability matters in making operational choices. It is useful to make such comparisons since ISMs commonly use the mean-value approach for choosing planned inventory levels. An interesting observation is that the total planned inventory is relatively insensitive to the introduction of uncertainty for a range of different cost parameters. Although intuitively it is expected that planned inventory levels should increase when yield rate and demand variability is introduced, the dependence is found to be generally weak. There are, however, some slab designs for which (x_j/μ_j) varies significantly (see, for example, x_7/μ_7).

6. Summary and conclusions

We presented a model for planned inventory deployment that is consistent with the ISM managers' view of key decision variables, cost drivers and uncertainties. We then established certain properties of the model that help reduce its computational complexity, and developed a heuristic suitable for solving large-scale instances of the problem similar to those encountered in practice. While specific conclusions may vary from one steel maker to another, our numerical experiments support the following general observations:

1. Choosing designs based on mean demand and yield rate typically results in optimal or near-optimal design choices. On the other hand, using the ILP mean-value problem to approximate inventory levels can result in significantly higher total costs compared to the optimum.
2. EVPI is in general high, indicating significant advantages associated with improving demand information within the supply chain. There is also significant dependence of total cost on yield rate variability. Both these observations underscore the importance of solving the stochastic version of the ILP.
3. Total inventory is relatively insensitive to the introduction of yield rate and demand uncertainty for a range of cost parameters; however, individual planned inventory levels can be strongly affected. From an ISM perspective, this means that the total storage space required to meet demand for inventoried slabs is fairly stable.
4. Order cancelations have a much more significant effect on inventory planning than order-size uncertainty.

As a final note we point out that the model studied in this article is applicable to general problems involving the design/configuration of transportation networks given uncertainty in supply and demand. It could be modified to account for other factors such as capacity constraints on total production and/or external purchases, or randomness in second-stage cost coefficients.

Acknowledgements

We are grateful to three anonymous referees for their helpful comments on an earlier version of this paper. This material is based in part upon work supported by the National Science Foundation under grant no. 9988721. Additional funding was provided by the Natural Sciences and Engineering Council of Canada (45904-98) via a research grant to DG.

References

Bassok, Y., Anupindi, R. and Akella, R. (2000) Single period multiproduct inventory models with substitution. *Operations Research*, **47**, 632–642.

Birge, J.R. and Louveaux, F. (1997) *Introduction to Stochastic Programming*, Springer-Verlag, New York, NY.

Bitran, G.R. and Dasu, S. (1992) Ordering policies in an environment of stochastic yields and substitutable demands. *Operations Research*, **40**, 999–1017.

Brown, A.O., Lee, H.L. and Petrakian, R. (2000) Xilinx improves its semiconductor supply chain using product and process postponement. *Interfaces*, **30**, 65–80.

Burman, M., Gershwin, S.B. and Suyematsu, C. (1998) Hewlett-Packard uses operations research to improve the design of a printer production line. *Interfaces*, **28**, 24–36.

Carr, S. and Duenyas, I. (2000) Optimal admission control and sequencing in a make-to-stock/make-to-order producton system. *Operations Research*, **48**, 709–720.

Chen, F. (2000) Quantifying the bullwhip effect in a simple supply chain: the impact of forcasting, lead times, and information. *Management Science*, **46**, 436–444.

Cornuejols, G., Fisher, M.L. and Nemhauser, G.L. (1977) Location of bank accounts to optimize float. *Management Science*, **23**, 789–810.

Denton, B., Gupta, D. and Jawahir, K. (2003) Managing increasing product variety at integrated steel mills. *Interfaces*, **33**, 41–53.

Dietrich, B.L. (1990) Monge sequences, antimatroids, and the transportation problem with forbidden arcs. *Linear Algebra and its Applications*, **139**, 133–145.

Garg, A. and Tang, C.S. (1997) On postponement strategies for product families with multiple points of differentiation. *IIE Transactions*, **29**, 641–650.

Garey, M.R. and Johnson, D.S. (1979) *Computers and Intractability: A Guide to the Theory of NP-Completeness*. W.H. Freeman and Company, New York, p. 220.

Gerchak, Y., Tripathy, A. and Wang, K. (1996) Coproduction models with random functionality yields. *IIE Tranactions*, **28**, 391–403.

Graman, G.A. and Magazine, M.J. (1998) An analysis of packaging postponement, in *Proceedings of the 1998 MSOM Conference*, University of Washington School of Business, Seattle, WA, pp. 67–72.

Gupta, D. and Benjaafar, S. (2004) Make-to-order, make-to-stock, or delay product differentiation?—a common framework for modeling and analysis. *IIE Transactions*, **36**, 529–546.

Hsu, A. and Bassok, Y. (1999) Random yield and random demand in a production system with downward substitution. *Operations Research*, **47**, 277–290.

Huang, C.C., Ziemba, W.T. and Ben-Tal, A. (1977) Bounds on the expectation of a convex function of a random variable: With applications to stochastic programming. *Operations Research*, **25**, 315–325.

Ignall, E. and Veinott, A. (1969) Optimality of myopic inventory policies for several substitute products. *Management Science*, **15**, 284–304.

Jordan, W.C. and Graves, S.C. (1995) Principles on the benefits of manufacturing process flexibility. *Management Science*, **41**, 577–594.

Karmarkar, U.S. (1979) Convex/stochastic programming and multilocation inventory problems. *Naval Research Logistics Quarterly*, **26**, 1–19.

Laporte, G., Louveaux, F.V. and Van Hamme, L. (1994) Exact solution of a stochastic location problem by an integer L-shaped algorithm. *Transportation Science*, **28**, 95–103.

Lee, H.L. (1996) Effective inventory and service management through product and process redesign. *Operations Research*, **44**, 151–159.

Lee, H.L. and Billington, C. (1994) Designing products and processes for postponement, in *Management of Design: Engineering and Management Perspectives*, Dasu, S. and Eastman, C. (eds.), Kluwer, Boston, MA, pp. 105–122.

Lee, H.L., Padmanabhan, V. and Whang, S.J. (1997) Information distortion in a supply chain: the bullwhip effect. *Management Science*, **43**, 546–558.

Lee, H.L. and Tang, C.S. (1997) Modeling the costs and benefits of delayed product differentiation. *Management Science*, **43**, 40–53.

Louveaux, F.V. and Peeters, D. (1992) A dual-based procedure for stochastic facility location. *Operations Research*, **40**, 564–573.

McGillivray, A. and Silver, E.A. (1978) Some concepts for inventory control under substitutable demand. *INFOR*, **16**, 47–63.

Mirchandani, P.B. and Francis, R.L. (1990) *Discrete Location Theory*, Wiley, New York, NY.

Murty, K.G. (1983). *Linear Programming*, J. Wiley, New York, NY.

Nemhauser, G.L. and Wolsey, L.A. (1999) *Integer and Combinatorial Optimization*, J Wiley, New York, NY.

Parlar, M. and Goyal, S. (1984) Optimal ordering decisions for two substitutable products with stochastic demands. *OPSEARCH*, **21**, 1–15.

Pentico, D.W. (1988) The discrete two-dimensional assortment problem. *Operations Research*, **36**, 324–332.

Porteus, E.L. (1990) Stochastic inventory theory, in *Handbooks in OR & MS, Vol. 2*, Heyman, D. P. and Sobel, M. J. (eds.), Elsevier, New York, pp. 605–652.

Rao, U.S., Jayashankar, M.S. and Zhang, J. (2000) A multi-product inventory problem with setup costs and downward substitution. Working Paper. Carnegie Mellon University., Pittsburgh, PA.

Robinson, L.W. (1990). Optimal approximate policies in multiperiod, multilocation inventory models with transshipments. *Operations Research*, **38**, 278–295.

Schultz, R., Stougie, L. and Van der Vlerk, M.H. (1996) Two-stage stochastic integer programming: a survey. *Statistica Neerlandica*, **50**, 404–416.

Shih, W. (1980) Optimal inventory policies when stockouts result from defective products. *International Journal of Production Research*, **18**, 677–685.

Swaminathan, J.M. and Tayur, S.R. (1998) Managing broader product lines through delayed differentiation using vanilla boxes. *Management Science*, **44**, S161–S172.

Van der Vlerk, M.H. (2000) Stochastic integer programming bibliography. Available at *http://mally.eco.rug.nl/biblio/stoprog.html* (date accessed 2000).

Van Slyke, R.M. and Wets, R.J.-B. (1969) L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal of Applied Mathematics*, **17**, 638–663.

Wallace, S.W. (1986) Solving stochastic programs with network recourse. *Networks*, **16**, 295–317.

Yano, C.A. and Lee, H.L. (1995) Lot sizing with random yields: a review. *Operations Research*, **43**, 311–334.

Appendix

Proof of Proposition 1. Let $\bar{\boldsymbol{\xi}} = (\bar{\mathbf{d}}, \bar{\mathbf{u}})$. Replacing random vector $\boldsymbol{\xi}$ with a single realization $\bar{\boldsymbol{\xi}}$, occurring with probability one, yields the mean-value problem. Given $\mathbf{x} \in B^n$, this problem has the following optimal first-stage decision variables:

$$q_j^* = \min\left\{\sum_{i=1}^{n} a_{ij}x_i, 1\right\} \quad \text{and} \quad w_i^* = \left(\frac{1}{\bar{u}_i}\right)\sum_{j\in C_i}\bar{d}_j x_i,$$

where $C_i = \{j \mid r_{ij} > r_{i'j}, \forall\, i'\}$. In other words, $q_j^* = 1$ if there is at least one design i that could cover order j, and the inventory level for design i, w_i^*, is equal to the sum of the demands (corrected for yield loss) for which design i has the highest reward r_{ij}. The optimal second-stage solution is

$s_j = 0\ \forall j$, $e_i = 0,\ \forall i$, and:

$$y_{ij}^* = \begin{cases} \bar{d}_j & \text{if } j \in \mathcal{C}_i, \\ 0 & \text{otherwise.} \end{cases}$$

Making the above substitutions for q_j^*, w_i^*, y_{ij}^* and transforming $y_{ij} \to \bar{d}_j y_{ij}$ in IDP:

$$\max \left\{ Z_R = -\sum_{i=1}^{n} \frac{c_i^{\mathrm{p}}}{\bar{u}_i} \sum_{j \in \mathcal{C}} \bar{d}_j x_i + \sum_{j=1}^{m} \sum_{i=1}^{n} r_{ij} \bar{d}_j y_{ij} \right\}, \quad \text{(A1)}$$

subject to

$$\sum_{i=1}^{n} x_i \le c, \quad \text{(A2)}$$

$$\sum_{i=1}^{n} a_{ij} \bar{d}_j y_{ij} = \sum_{j \in \mathcal{C}} \bar{d}_j x_i, \quad \forall i, \quad \text{(A3)}$$

$$\sum_{i=1}^{n} a_{ij} y_{ij} = \min \left\{ \sum_{i=1}^{n} a_{ij} x_i, 1 \right\}, \quad \forall j, \quad \text{(A4)}$$

$$y_{ij} \le M a_{ij}, \quad \forall (i,j), \quad \text{(A5)}$$

$$y_{ij} \le x_i, \quad \forall (i,j), \quad \text{(A6)}$$

$$x_i \in \{0,1\}, \quad \forall i, \quad y_{ij} \ge 0, \quad \forall (i,j). \quad \text{(A7)}$$

Substituting Equation (A3) into the first term of the objective function yields:

$$Z_R = \sum_{i=1}^{n} \sum_{j=1}^{m} \bar{r}_{ij} y_{ij}, \quad \text{(A8)}$$

where $\bar{r}_{ij} = r_{ij} \bar{d}_j - c_i^{\mathrm{p}} a_{ij} \bar{d}_j / \bar{u}_i$. Constraint (A4) implies $\sum_{i=1}^{n} a_{ij} y_{ij} \le 1$ and $\sum_{i=1}^{n} a_{ij} y_{ij} \le \sum_{i=1}^{n} a_{ij} x_i$, but constraint (A6) guarantees the latter. Thus, since $r_{ij} = 0$ if $a_{ij} = 0$ (by assumption) constraint (A4) can be replaced by $\sum_{i=1}^{n} y_{ij} \le 1$ and constraint (A5) can be relaxed to yield a problem equivalent to Equations (21)–(24) except for binary restrictions on y_{ij}. However, $y_{ij}^* \in B$ follows automatically from the total unimodularity of constraints in the p-median problem, given $\mathbf{x} \in B^n$ (Cornuejols *et al.* 1977). ■

Proof of Proposition 2. We consider the case in which $q_j^* > 0$ (otherwise $q_j^* = 0$ is integer valued) and show that it implies $q_j^* = 1$. Treating the LP relaxation of Equations (14)–(20) as a parametric program in $\mathbf{q}$ it can be rewritten as:

$$Z_{\mathbf{q}}^* = \max \left\{ Q(\mathbf{x}, \mathbf{w}, \mathbf{q}) \mid (\mathbf{x}, \mathbf{w}, \mathbf{y}, \mathbf{s}, \mathbf{e}) \in \mathcal{P}, q_j \in \{0,1\}, \right.$$
$$\left. \forall j, \sum_{i=1}^{n} a_{ij} y_{ij}^k + s_j^k = d_j^k q_j,\ \forall j, k \right\}, \quad \text{(A9)}$$

where $\mathcal{P} =$ ((15), (16), (18), (19), (20)) and the objective function are both independent of $\mathbf{q}$. Thus, taking the dual of Equation (A10) yields an LP with coefficients $d_j^k q_j$ appearing in the objective function but for which the constraint set is independent of $\mathbf{q}$. Since $d_j^k \ge 0, \forall (j,k)$ it follows that if $q_j^* > 0$ then:

$$\frac{\partial Z_q^*}{\partial q_j} > 0,$$

and therefore since Z_q is unconstrained in q_j, other than $0 \le q_j \le 1$, it follows that $q_j^* = j$. ■

Proof of Proposition 3. Every basis matrix of a transportation problem is triangular (Murty, 1983, Corollary 13.2, p. 382) and there exists an ordering of arcs such that greedy allocation (equivalently backward substitution) is optimal. Thus, if it is optimal to allocate any supply from a supply node to a demand node then it is optimal to allocate the maximum possible supply. Since there are at most c supply sources it follows that at most c orders can receive only partial allocation of supply. ■

Biographies

Brian Denton is a Senior Engineer in the Advanced Planning Systems department in IBM's Systems & Technology Group. Since joining IBM in 2001 he has been involved in the design and development of planning and scheduling systems for large-scale supply chain optimization problems in semiconductor and data storage device manufacturing industries. Prior to joining IBM he worked on OR/MS applications in the health care industry and in the steel manufacturing industry. His research interests are in the application and development of solution methodology for large-scale optimization problems arising in industry applications. He has a joint B.Sc. in Physics and Chemistry, an M.Sc. in Physics, and a. Ph.D. in Management Science from McMaster University.

Diwakar Gupta is a Professor in the Graduate Program in Industrial Engineering at the University of Minnesota. His research and teaching interests are in the area of stochastic models for production/inventory systems, supply chains, and health care delivery systems. Prior to joining the University of Minnesota, Diwakar held a Faculty position at the DeGroote School of Business, McMaster University. He received his Ph.D. from the University of Waterloo. His articles have appeared in journals such as *Management Science, Operations Research*, *IIE Transactions*, and *EJOR*. He is a Departmental Editor of *IIE Transactions—Scheduling and Logistics*, and an Associate Editor for the *International Journal of Flexible Manufacturing Systems*. He is also an Editorial Board Member of the *M&SOM* journal. More information on his current research projects and publications can be found by visiting the web page of his research laboratory (supply chain and operations research laboratory) at the URL: *http://www.me.umn.edu/divisions/ie/scorlab*.