# Modeling Demand Propagation in Guaranteed Service Models ⋆

**Madeleine Löhnert** * **Jörg Rambau** **

* *DATEV eG, Nürnberg, (email: madeleine.loehnert@gmx.de)*
** *University of Bayreuth, 95440 Bayreuth, Germany (e-mail: joerg.rambau@uni-bayreuth.de)*

**Abstract:** In this paper, the class of guaranteed service models for multi-echelon inventory management is enhanced with explicit demand propagation. More specifically, the known mixed integer linear programming formulation for the guaranteed service model is refined by new variables and restrictions so that it describes the internal demand propagation exactly for linear demand bound functions. With this feature, aspects like outsourcing as well as decision-dependent stochastic demands at internal stock points can be expressed exactly. The relevance of the new model is shown in an illustrative example, where the new model is able to find a solution with almost 40% lower actual cost compared to the existing approximative model without explicit demand propagation.

*Keywords:* multi-echelon inventory management, mixed-integer linear programming, demand propagation, outsourcing, lost demand

## 1. INTRODUCTION

Inventory management in its basic form poses the following problem: Given a stock point with uncertain demand, when and how much supply has to be replenished in order to serve all customers at the lowest possible expected cost (inventory holding cost plus backordering cost)? If several stock points form an inventory network, the problem is called multi-echelon inventory management. Special cases occur when each stock point has a unique predecessor (divergent systems), a unique successor (convergent systems), or both (serial systems). Moreover, the backordering case (unsatisfied demand is served later) is distinguished from the lost-sales case (unsatisfied demand is lost). For more detailed background on multi-echelon inventory management, see de Kok and Fransoo (2003).

It is known that using individually optimal policies at all stock points rarely constitutes an optimal multi-echelon policy. However, the computation of an optimal multi-echelon policy is difficult in all but the most basic (famous) cases like the serial case with backordering and linear-plus-fixed ordering cost (Clark and Scarf, 1960). Over all, in the backordering case optimal policies could be derived more often than in the lost-sales case. The reason is as follows: Since the order volumes of downstream stock points depend on their ordering decisions to be optimized, demands at all upstream stock points are endogenous. This makes optimization more complicated. In the backordering case and divergent systems it can be assumed that all demands are propagated upstream in an unchanged manner. In the lost-demands case this is not true anymore, and the upstream demands can have intractable distributions, depending on the downstream ordering policies. Similar problems arise in the case of outsourcing opportunities or any other decision option that satisfies part of the demand outside the multi-echelon system to be optimized.

Various approaches to approximate optimal policies have been suggested. They can be divided into the Stochastic Service Model paradigm (SSM) and the Guaranteed Service Model paradigm (GSM), see Graves and Willems (2003) for a comparison and (Magnanti et al., 2006) for an approximate mixed-integer linear programming formulation (MILP) for the GSM.

This paper deals with the GSM branch of research. In the GSM, bounded demand is assumed in the sense that a *demand bound function* is known, i.e., for each node and for each duration there is a maximal value for the total demand presented to the node during any time interval with that duration. Based on this information, each stock point decides for each of its successors the maximal time between receiving an order and sending off the supply to its successor. Given this *guaranteed service time*, each successor stock point can predict the latest point in time when its order arrives at its inventory, given that the (transportation) delay between the stock points is deterministic (or bounded). From the demand bound function, minimal inventory levels can be computed that ensure that the guaranteed service times can be obeyed.

This approach uses two debatable assumptions: bounded demand and constant delays. The most common interpretation is: if anything goes wrong in practice (demand bounds are exceeded, unforeseen transportation delays occur), then some unmodeled *operational flexibility* is applied so that the service times can be guaranteed anyway. And here is a serious gap in the justification of the GSM paradigm: if that operational flexibility is, e.g., out-

⋆ Project of the Bayreuth research center for modeling and simulation (MODUS).

sourcing demand to an external emergency supplier, then the demand propagation inside the inventory network is changed, and the internal demand bound functions derived from simple upstream propagation would overestimate the upstream demand.

In this paper, the GSM is enhanced with exact demand propagation. This can not only be applied to the GSM with explicit outsourcing or lost demands but also to the recent Stochastic GSM (SGSM), that was developed in order to explicitly account for outsourcing in scenarios with different demand bound functions and expediting in scenarios with different delays. The corresponding two-stage stochastic MILP (2SMILP) was introduced by Rambau and Schade (2010) and further studied by Schade (2012) and Rambau and Schade (2014).

The main contribution of this paper is a new model, a 2SMILP, denoted by SGSM-DP, in which the demand propagation in the presence of outsourcing is captured exactly by explicit computation of demand bounds from the outsourcing decisions inside the model. In a small three-echelon example with five stock-points (three of which face exogenous demands) and three demand scenarios the new model improves the total expected cost by 40%. The downside is that the CPU time increases from around 10ms (SGSM) to slightly over 1s (SGSM-DP).

Remark: The problem of explicit demand propagation in the SGSM was preliminary investigated in the master's thesis by Löhnert (2016); the results in this paper are based on that thesis.

## 2. PROBLEM STATEMENT

Let $G = \big(N, A\big)$ be a divergent inventory network, i.e., $N$ is a set of *inventory nodes* and $A \subseteq N \times N$ is a set of *supply relations* so that for each node $j \in N$ there is at most one $i \in N$ with $(i, j) \in A$. Moreover, let $D$ be the leaves of the network, which correspond to the nodes with exogenous demands, called *demand nodes*. There are given deterministic *demand bound functions* $\phi_i(x_i)$ at all nodes $i$ that specify the maximal cumulative demands arriving in any time interval of length $x_i$. At the demand nodes these functions are exogenously given, and at upstream nodes they are accumulated from downstream demands. Since the network is divergent, this yields unique demand bound functions throughout. Moreover, there are *delays* $L_i$ for the transports of supplies to nodes $i$. There are marginal *holding costs* of $h_i > 0$ for inventory at node $i$. At each demand node $i$ a duration $\bar{s}_i^{\text{OUT}}$ specifies how long its end customers are prepared to wait (without incurring a backordering cost!) for an order to be fulfilled.

The idea of a guaranteed service model is to *impose* that the whole system delivers to all end customers in time. Then, no backordering cost is incurred, neither in demand nodes nor in internal nodes. Among all systems that can guarantee the service times for the end customers, a system is sought with minimal inventory holding costs for safety stock.

The independent decisions are *guaranteed (outgoing) service times* $s_i^{\text{OUT}}$ for all nodes $i$. The meaning is that node $i$ guarantees its successors that the time between receiving and delivering an order is at most $s_i^{\text{OUT}}$. The *guaranteed ingoing service times* $s_i^{\text{IN}}$ are dependent auxiliary variables used to simplify the formulation of some constraints. They mean that node $i$ need not wait any longer than $s_i^{\text{IN}} + L_i$ for an order to arrive. The dependent variable $x_i$ denotes maximal durations for which deliveries are taken from node $i$'s inventory. The material quantity $y_i$ is the inventory needed to be able to deliver from stock all orders that arrive during $x_i$ time units.

The original GSM MILP (Magnanti et al., 2006) reads as follows:

$$\min \sum_{i \in N} h_i y_i \tag{1}$$

s.t.

$$s_i^{\text{OUT}} - \bar{s}_i^{\text{OUT}} \leq 0 \qquad \forall i \in D \tag{2}$$

$$s_i^{\text{IN}} - s_j^{\text{OUT}} \geq 0 \qquad \forall (j, i) \in A \tag{3}$$

$$x_i - s_i^{\text{IN}} + s_i^{\text{OUT}} - L_i \geq 0 \qquad \forall i \in N \tag{4}$$

$$y_i - \phi_i(x_i) \geq 0 \qquad \forall i \in N \tag{5}$$

$$s_i^{\text{IN}}, s_i^{\text{OUT}}, x_i, y_i, q_i \geq 0 \qquad \forall i \in N \tag{6}$$

$$s_i^{\text{IN}}, s_i^{\text{OUT}}, x_i, y_i \in \mathbb{Z} \qquad \forall i \in N \tag{7}$$

The model is briefly explained in the following. The objective (1) measures the total holding cost per time unit for safety stock over all nodes. In particular, backordering costs are not considered. Constraint (2) bounds the guaranteed service times for each demand node by its customers' acceptable service times. Constraint (3) ensures that the ingoing service times at a node cannot be earlier than the outgoing service times of the supplier. Constraint (4) computes, from its given in- and outgoing guaranteed service times and the delays, the minimal amount of time units $x_i$ that node $i$ has to be able to deliver from its on-hand inventory. Constraint (5) transforms this quantity, by means of the demand bound functions, into a minimal necessary inventory level. Moreover, all variables are nonnegative, and all time units and quantities are considered integers. This corresponds to inventory networks storing expensive materials whose inventory levels are checked periodically, e.g., on a daily basis.

Note that this model is always feasible. For example, set $s_i^{\text{OUT}} := 0$. This has the meaning that each node must deliver the full demand to downstream nodes immediately. Then, setting $x_i := L_i$ and $y_i := \phi_i(x_i)$ for all $i$ leads to a feasible, yet most probably expensive solution. This solution is called the "all"-solution, because each node guarantees to deliver "all" orders immediately from stock.

Given a feasible GSM solution, a corresponding replenishment policy is straight-forward: just order in each node $i$ exactly the observed demand immediately. This order will arrive at most $x_i$ time units later. Until then, the available on-hand inventory $y_i$ is sufficient to keep the promised outgoing service time, as specified by $s_i^{\text{OUT}}$. Thus, each feasible solution leads to a feasible inventory management policy, and an optimal solution yields a policy that has minimal inventory holding costs for safety stock among all policies based on fixed guaranteed service times.

As discussed in the introduction, the bounded-demand assumption is very often interpreted as "usually fulfilled" so that in all remaining cases "operational flexibility" can be used to ensure the guaranteed service times. But what happens if for some time the demands exceed the demand bound functions and outsourcing is used as operational flexibility? Assume, the shortage in Constraint (5) is filled by outsourcing, and it shall be explicitly accounted for. Then it has to be specified how many pieces are ordered at a supplier outside the system. Denote by $q_i$ the outsourcing quantity at node $i$ at a marginal cost of $c_i > 0$. Then, this outsourcing decision can be incorporated into Constraint (5) as

$$y_i + q_i - \phi_i(x_i) \geq 0., \tag{5'}$$

(Throughout the paper, we will use primed equation numbers for constraints modified from earlier constraints.) The total cost is then increased by $\sum_{i \in N} c_i q_i$. The resulting model is denoted by GSM-o:

$$\min \sum_{i \in N} \big(h_i y_i + c_i q_i\big) \tag{1'}$$

s.t.

$$\begin{aligned}
s_i^{\mathrm{OUT}} - \bar{s}_i^{\mathrm{OUT}} &\leq 0 & \forall i \in D \quad & (2)\\
s_i^{\mathrm{IN}} - s_j^{\mathrm{OUT}} &\geq 0 & \forall (j,i) \in A \quad & (3)\\
x_i - s_i^{\mathrm{IN}} + s_i^{\mathrm{OUT}} - L_i &\geq 0 & \forall i \in N \quad & (4)\\
y_i + q_i - \phi_i(x_i) &\geq 0 & \forall i \in N \quad & (5')\\
s_i^{\mathrm{IN}}, s_i^{\mathrm{OUT}}, x_i, y_i, q_i &\geq 0 & \forall i \in N \quad & (6')\\
s_i^{\mathrm{IN}}, s_i^{\mathrm{OUT}}, x_i, y_i, q_i &\in \mathbb{Z} & \forall i \in N \quad & (7')
\end{aligned}$$

This model is always feasible by ignoring the outsourcing opportunity and using the "all"-solution.

Note, however, that now there is no clear way how to set the modified demand bound functions $\phi_i$ at internal nodes. Formally, all demand bound functions are given as exogenous data, which does not match reality anymore, since the demand propagated upstream is possibly reduced by the downstream outsourcing decisions $q_i$. Thus, in its current form, the GSM-o (and for that matter, also the original GSM) is only an approximate model that can be used if there is no frequent use of operational flexibility in form of outsourcing. This is usually ensured by setting the demand bound function to a large quantile ($> 90\%$) of some normally distributed stochastic demand. This quantile is called the *service level* and is the fraction of the actual demand that has to be fulfilled from inventory. A large service level can be justified at end customer nodes. However, at internal nodes it is not so clear why maintaining a large service level should be the most efficient policy.

The situation is worse in the SGSM. The SGSM considers several demand scenarios characterized by individual demand bound functions and adapts the outsourcing quantity to the observed demand scenario. Thus, large outsourcing quantities occur naturally in the high-demand scenarios of the SGSM. And in these scenarios, the demand propagation of the model is not exact.

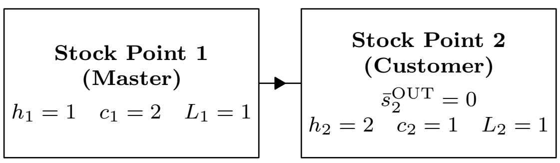


Fig. 1. An example network exhibiting the problem with demand propagation

Consider the following little example. The network together with the data except the demand bound functions is shown in Figure 1. Let the exogenous demand bound function at the demand node be $\phi_2(x) = x$. According to the usual a-priori demand propagation, this results in $\phi_1(x) = x$ as well. Consider a choice of $y_2 = 0, q_1 = 0$, and $s_1^{\mathrm{OUT}} = 0$, i.e., evaluate the cost of the good-looking decision where nothing is stored at the demand node (expensive marginal holding cost), where nothing is outsourced at the upstream node (expensive marginal outsourcing cost), and where all deliveries from the upstream inventory are immediate. In that case, Constraint (2) implies $s_2^{\mathrm{OUT}} = \bar{s}_2^{\mathrm{OUT}} = 0$, and, by Constraint (3), $s_2^{\mathrm{IN}} = 0$ holds as well. This results, by Constraint (4) in $x_2 = L_2 = 1$ and $x_1 = L_1 = 1$. (Note that all variables are required to be non-negative.) Since $y_2 = q_1 = 0$, it follows, by Constraint (5′), that $q_2 = y_1 = 1$. The incurred cost of this solution is then $c_2 + h_1 = 2$. In reality, however, the upstream demand bound function in the presence of complete outsourcing at the demand node is actually identical to zero. Thus, in reality, whenever $q_2 = 1$ we can afford $y_1 = 0$, leading to a cost of $c_2 = 1$ *for the same independent decisions*, only half as much as predicted by the model. With little more effort examples can be constructed where the optimality of decisions is assessed incorrectly by the GSM-o.

Summary: For scenarios in which frequent outsourcing is preferable a more explicit, consistent model is needed. And this is the modeling problem posed in this paper. In the following, the GSM-o is enhanced with outsourcing with explicit demand propagation for linear demand bound functions of the form $\phi_i(x_i) := \alpha_i x_i$. The exact handling of more general demand bound functions is subject of research in progress.

## 3. A GSM WITH OUTSOURCING AND DEMAND PROPAGATION

The construction works in two steps. First, a model with non-linear restrictions (GSM-NL) is derived, in which the number of pieces $n_i$ ordered at node $i$ per time unit by downstream nodes is explicitly computed. Second, the non-linear restrictions are linearized by means of extra binary variables with the so-called *big-M-method* (GSM-DP).

For this section it is assumed that $x_i = 0$ implies $q_i = 0$, which is with no loss of generality because each feasible solution with $x_i = 0$ and $q_i > 0$ can be modified by setting $q_i = 0$ maintaining feasibility at a reduced cost.

For the first step, consider the "sufficient inventory" inequality (5′) $y_i + q_i \geq \phi_i(x_i) := \alpha_i x_i$ for some $\alpha_i \geq 0$. For internal nodes, the demand shall be endogenously defined by propagation from downstream nodes. Thus, for internal nodes the *exogenous* demand rate per time unit $\alpha_i$ (*data*) must be replaced by an *endogenous* demand rate $n_i$ (a *variable*). This yields the *non-linear sufficient-inventory inequality*

$$y_i + q_i \geq n_i x_i \tag{5''}$$

for the required inventory at internal nodes.

In order to compute an upstream demand rate $n_i$ from downstream outsourcing quantities $q_j$ with $(i,j) \in A$, two cases are distinguished. If $x_j = 0$, then, by assumption, no outsourcing happens, i.e., $q_j = 0$, and, therefore, the unreduced demand rate is propagated upstream. Otherwise, the outsourcing quantity $q_j$ is used to cover a part of the demand at $j$ over a period of $x_j$ time units. Thus, because of linear demand bound functions, the demand per time unit upstream at $i$ is reduced by $\frac{q_j}{x_j}$. Summarizing this concept over all successors of node $i$, the *demand-rate balance equations* are obtained that are non-linear for all but the demand nodes:

$$n_i = \alpha_i \qquad \forall i \in D, \tag{8}$$

$$n_i = \sum_{j:(i,j)\in A} n_j - \sum_{\substack{j:(i,j)\in A \\ x_j\neq 0}} \frac{q_j}{x_j} \qquad \forall i \in N\backslash D. \tag{9}$$

The following non-linear GSM extension summarizes the model, which is denoted by GSM-o-NL:

$$\min \sum_{i\in N} \big(h_i y_i + c_i q_i\big) \tag{1'}$$

s.t.

$$\begin{aligned}
s_i^{\mathrm{OUT}} - \bar{s}_i^{\mathrm{OUT}} &\leq 0 & \forall i \in D \quad & (2)\\
s_i^{\mathrm{IN}} - s_j^{\mathrm{OUT}} &\geq 0 & \forall (j,i) \in A \quad & (3)\\
x_i - s_i^{\mathrm{IN}} + s_i^{\mathrm{OUT}} - L_i &\geq 0 & \forall i \in N \quad & (4)\\
y_i + q_i - n_i x_i &\geq 0 & \forall i \in N \quad & (5'')\\
n_i &= \alpha_i & \forall i \in D \quad & (8)\\
n_i - \sum_{j:(i,j)\in A} n_j + \sum_{\substack{j:(i,j)\in A \\ x_j\neq 0}} \frac{q_j}{x_j} &= 0 & \forall i \in N\backslash D \quad & (9)\\
s_i^{\mathrm{IN}}, s_i^{\mathrm{OUT}}, x_i, y_i, q_i, n_i &\geq 0 & \forall i \in N \quad & (6'')\\
s_i^{\mathrm{IN}}, s_i^{\mathrm{OUT}}, x_i, y_i, q_i &\in \mathbb{Z} & \forall i \in N \quad & (7')
\end{aligned}$$

Again, ignoring the outsourcing opportunity, i.e., imposing $q_i = 0$ combined with the "all"-solution (always deliver the full demand from inventory immediately) is a feasible solution for this model.

Next, the GSM-o-NL is linearized. The main idea is as follows: if $x_i$ were a constant in all restrictions, then the system of restrictions would be linear. Since $x_i$ is an integer, all possible cases ($x_i = k$, $k \in K$) can be classified by the possible values of $x_i$. The respective cases are then indicated by additional binary *multiple-choice* variables $z_{ik}$, exactly one of which has to be set to one, formally:

$$\sum_{k=0}^{K} z_{ik} = 1 \; \forall i \in N. \tag{10}$$

If $z_{ik} = 1$, then node $i$ is prepared to cover demand for $x_i = k$ time units from its inventory $y_i$. The maximal $k$ needed, i.e., the maximal time to wait for an order to arrive, can be estimated by the sum of all delays $L_j$ on a directed path to $i$ in the inventory network. Denote by $k_{\max}$ this upper bound on $k$, and let $K := \{0, 1, \ldots, k_{\max}\}$ be the set of all $k$ to be considered in the model. We can recover $x_i$ from the $z_{ik}$ formally as follows:

$$\sum_{k=0}^{K} k z_{ik} = x_i \; \forall i \in N. \tag{11}$$

Note that $k_{\max}$ multiplied by the sum of all end customer demand rates downstream of $i$ is an estimate for the maximal possible shortage that can be experienced in node $i$. Let $M_i^{big}$ denote this maximal possible shortage. Then $n_i x_i - (y_i + q_i) \leq M_i^{big}$ in any optimal solution. Constraint (5″) can therefore be expressed using the constant $k$ and $z_{ik}$ instead of the $x_i$ by the linear constraint

$$n_i k - (y_i + q_i) \leq (1 - z_{ik}) M_i^{big}. \tag{5'''}$$

Indeed: Whenever $x_i = k$, then $z_{ik} = 1$, and thus the right-hand side evaluates to zero. The quantity $n_i k$ is the maximal required amount of material in node $i$ during $k$ time units. This should not exceed the available material in node $i$, which is $y_i + q_i$. Whenever $x_i \neq k$, then $z_{ik} = 0$, and the right hand side evaluates to $M_i^{big}$, which is so large that the inequality becomes redundant.

Contraint (9) can be linearized using an additional variable $n_{ij}$ for the demand rate propagated from $j$ to $i$. The total demand arriving at node $i$ is consequently

$$n_i = \sum_{j:(i,j)\in A} n_{ij} \; \forall i \in N\backslash D. \tag{9'a}$$

It is required that the outsourcing quantities $q_j$ are exactly the difference between downstream demand at $j$ and the propagated demand from $j$ to $i$ accumulated over $k$ time units. This can be ensured by the conditional equation $q_j = (n_j - n_{ij})k$ whenever $z_{ik} = 1$. This conditional equation can be reformulated by using the same big-M technique as above separately for the implied conditional inequalities $q_j \leq (n_j - n_{ij})k$ and $-q_j \leq -(n_j - n_{ij})k$ whenever $z_{ik} = 1$. This results in the two linear inequalities

$$q_j - (n_j - n_{ij})k \leq (1 - z_{jk}) M_j^{big}, \tag{9'b}$$

$$-q_j + (n_j - n_{ij})k \leq (1 - z_{jk}) M_j^{big}. \tag{9'c}$$

If $k = 0$, the propagated demand rates have yet to be tied to the downstream demand rates. In that case $q_j = 0$, and

$$n_j - n_{ij} \leq q_j \; \forall (i,j) \in A, \tag{9'd}$$

enforces that the demand rate is propagated completely upstream. Note that for $k \geq 1$ this constraint is redundant, in particular valid, so that no "big-M" is needed.

The complete resulting model with demand propagation, denoted by GSM-o-DP, reads as follows:

$$\min \sum_{i\in N} \big(h_i y_i + c_i q_i\big) \tag{1'}$$

s.t.

$$\begin{aligned}
s_i^{\mathrm{OUT}} - \bar{s}_i^{\mathrm{OUT}} &\leq 0 & \forall i \in D \quad & (2)\\
s_i^{\mathrm{IN}} - s_j^{\mathrm{OUT}} &\geq 0 & \forall (j,i) \in A \quad & (3)\\
x_i - s_i^{\mathrm{IN}} + s_i^{\mathrm{OUT}} - L_i &\geq 0 & \forall i \in N \quad & (4)\\
n_i k - (y_i + q_i) & & & \\
- (1 - z_{ik}) M_i^{big} &\leq 0 & \forall i \in N, & \\
& & \forall k \in K \quad & (5''')\\
n_i &= \alpha_i & \forall i \in D \quad & (8)\\
n_i - \sum_{j:(i,j)\in A} n_{ij} &= 0 & \forall i \in N\backslash D \quad & (9'a)\\
q_j - (n_j - n_{ij})k & & & \\
- (1 - z_{jk}) M_j^{big} &\leq 0 & \forall (i,j) \in A, & \\
& & \forall k \in K \quad & (9'b)\\
-q_j + (n_j - n_{ij})k & & & \\
- (1 - z_{jk}) M_j^{big} &\leq 0 & \forall (i,j) \in A, &
\end{aligned}$$

$$\forall k \in K \quad (9'\text{c})$$

$$n_j - n_{ij} - q_j \leq 0 \quad \forall (i,j) \in A \quad (9'\text{d})$$

$$\sum_{k=0}^{K} z_{ik} = 1 \quad \forall i \in N \quad (10)$$

$$x_i - \sum_{k=0}^{K} k z_{ik} = 0 \quad \forall i \in N \quad (11)$$

$$s_i^{\text{IN}}, s_i^{\text{OUT}}, x_i, y_i, q_i, n_i, n_{ij} \geq 0 \quad \forall i \in N \quad (6'')$$

$$s_i^{\text{IN}}, s_i^{\text{OUT}}, x_i, y_i, q_i \in \mathbb{Z} \quad \forall i \in N \quad (7')$$

$$z_{ik} \in \{0,1\} \quad \forall i \in N, \ \forall k \in K \quad (12)$$

In the same way as above, the "all"-solution without outsourcing is feasible for this model.

This model can be extended to the stochastic model SGSM in the same way as the GSM-o was extended to the original SGSM. To this end, consider a finite set of demand scenarios $\omega \in \Omega$ with probabilities $p^\omega$. The demand bound functions at the demand nodes are now scenario-dependent linear functions and are given by $\alpha_i^\omega x_i$. The guaranteed service times $s_i^{\text{OUT}}$ and $s_i^{\text{IN}}$, the times to deliver from inventory $x_i$ and the inventory level decisions $y_i$ in the SGSM-DP are considered as *here-and-now decisions*, i.e., they have to be taken without knowing the realized scenario. The outsourcing quantities $q_i^\omega$ are considered as *wait-and-see decisions* that can be taken as soon as the scenario has realized. Consequently, also the propagated demand quantities $n_i^\omega$, $n_{ij}^\omega$ are both scenario dependent.

The resulting two-stage stochastic mixed-integer linear program with recourse, which is denoted by SGSM-DP, that minimizes expected total costs is obtained from the GSM-o-DP as follows: for each wait-and-see decision variable introduce a copy for each scenario $\omega \in \Omega$. All constraints that contain a wait-and-see decision variable are then required to hold for each scenario $\omega \in \Omega$. The result is summarized in the following model, where the equation numbers with a superscript "$\omega$" reflect a constraint that now has to hold in each scenario:

$$\min \sum_{i \in N} \bigl( h_i y_i + \sum_{\omega \in \Omega} p^\omega c_i q_i^\omega \bigr) \quad (1'')$$

s.t.

$$s_i^{\text{OUT}} - \bar{s}_i^{\text{OUT}} \leq 0 \quad \forall i \in D \quad (2)$$

$$s_i^{\text{IN}} - s_j^{\text{OUT}} \geq 0 \quad \forall (j,i) \in A \quad (3)$$

$$x_i - s_i^{\text{IN}} + s_i^{\text{OUT}} - L_i \geq 0 \quad \forall i \in N \quad (4)$$

$$n_i^\omega k - (y_i + q_i^\omega) - (1 - z_{ik}) M_i^{big} \leq 0 \quad \forall i \in N, \ \forall k \in K, \ \forall \omega \in \Omega \quad (5'''^\omega)$$

$$n_i^\omega = \alpha_i^\omega \quad \forall i \in D, \ \forall \omega \in \Omega \quad (8^\omega)$$

$$q_j^\omega - (n_j^\omega - n_{ij}^\omega) k - (1 - z_{jk}) M_j^{big} \leq 0 \quad \forall (i,j) \in A, \ \forall k \in K, \ \forall \omega \in \Omega \quad (9'\text{b}^\omega)$$

$$-q_j^\omega + (n_j^\omega - n_{ij}^\omega) k - (1 - z_{jk}) M_j^{big} \leq 0 \quad \forall (i,j) \in A, \ \forall k \in K, \ \forall \omega \in \Omega \quad (9'\text{c}^\omega)$$

$$n_j^\omega - n_{ij}^\omega - q_j^\omega \leq 0 \quad \forall (i,j) \in A, \ \forall \omega \in \Omega \quad (9'\text{d}^\omega)$$

$$n_i^\omega - \sum_{j:(i,j) \in A} n_{ij}^\omega = 0 \quad \forall i \in N \backslash D, \ \forall \omega \in \Omega \quad (9'\text{a}^\omega)$$

$$\sum_{k=0}^{K} z_{ik} = 1 \quad \forall i \in N \quad (10)$$

$$x_i - \sum_{k=0}^{K} k z_{ik} = 0 \quad \forall i \in N \quad (11)$$

$$s_i^{\text{IN}}, s_i^{\text{OUT}}, x_i, y_i \geq 0 \quad \forall i \in N \quad (6''')$$

$$q_i^\omega, n_i^\omega, n_{ij}^\omega \geq 0 \quad \forall i \in N, \ \forall \omega \in \Omega \quad (6'''^\omega)$$

$$s_i^{\text{IN}}, s_i^{\text{OUT}}, x_i, y_i \in \mathbb{Z} \quad \forall i \in N \quad (7'')$$

$$q_i^\omega \in \mathbb{Z} \quad \forall i \in N, \ \forall \omega \in \Omega \quad (7''^\omega)$$

$$z_{ik} \in \{0,1\} \quad \forall i \in N, \ \forall k \in K \quad (12)$$

Once more, the "all"-solution with no outsourcing is feasible; this time each node must set $y_i$ to the maximal demand over $x_i = L_i$ time units among all scenarios in order to be able to deliver from stock under all circumstances.

Now, in each scenario the individually specified outsourcing decisions lead to the exact endogenously derived upstream demand rates.

## 4. EXAMPLE

In the following the impact of the enhanced model SGSM-DP is illustrated for a still small example by Löhnert (2016). The example network together with all input data is displayed in Figure 2. For the SGSM, the internal demands in each scenario are approximated by the complete accumulated downstream demand. For the SGSM-DP internal demands are computed by the model. In the Table 1, the resulting cost of the original SGSM model by Rambau and Schade (2014) is compared to the cost of the new SGSM-DP model. Note that the SGSM's suggestions for outsourcing need only be implemented (and payed) in reality if there really is a shortage given the inventory levels. Since the SGSM overestimates demands in the presence of outsourcing downstream, some upstream outsourcing might turn out to be superfluous. Thus, two cost values are presented for the SGSM solution: the model cost and the actual cost. The actual cost of the original SGSM is computed by fixing the here-and-now decisions (service times $s_i$, durations $x_i$, and inventories $y_i$) of the SGSM in the SGSM-DP. The model cost and the actual cost of the SGSM is then compared with the model cost of the SGSM-DP. The computations were performed on a Dell VOSTRO 3450 Laptop with (Intel i5-2410M, 64bit, 2.30GHz, 4GB RAM, Windows 7 Professional) using the free MILP solver SCIP (Gamrath et al., 2016). It can be seen that the original SGSM's model cost (optimal objective function value of the SGSM) overestimates the actual cost of its strategic inventory decisions (objective function value of

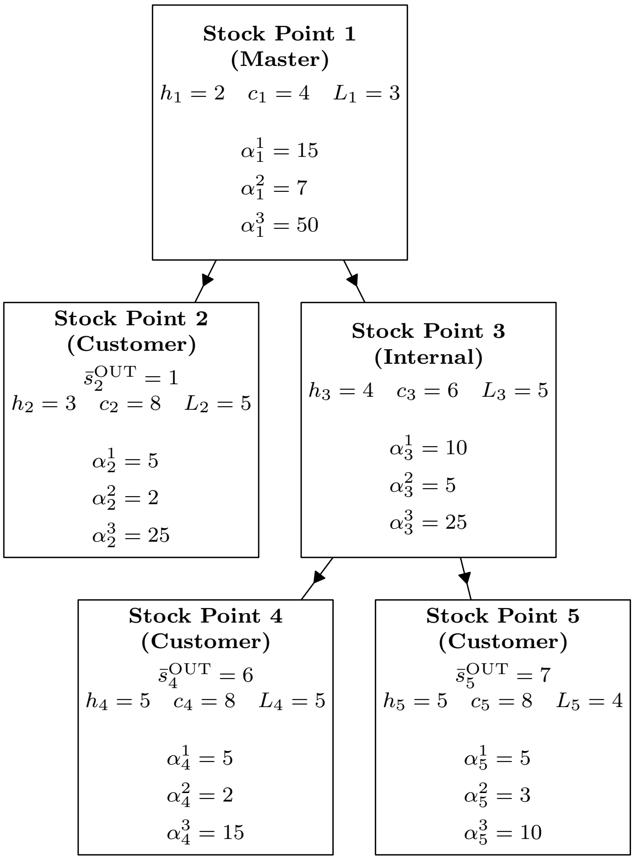


Fig. 2. An example network with high cost differences between SGSM and SGSM-DP

| | SGSM | SGSM-DP |
|---|---|---|
| # var's | 35 | 132 |
| # cons's | 37 | 605 |
| CPU time/s | $< 0.01$ | 1.29 |
| cost (actual cost) | 747 (567) | 410 |

Table 1. Results for the example instance (rounded)

the SGSM in the SGSM-DP) by over 30%. Moreover, because of the approximation of the internal demands, the SGSM's inventory decision is almost another 40% more expensive than the SGSM-DP's optimal solution with explicit demand propagation in each scenario. Admittedly, this example was constructed in such a way that differences become visible. However, there is no reason to assume that in more complicated real-life instances the difference is less pronounced.

## 5. CONCLUSIONS

The SGSM-DP can improve strategic inventory decisions: explicit accounting for outsourcing and demand propagation are advantages over all existing models in the GSM branch of research. By presenting an easy example with three echelons, three demand nodes, two other nodes, and three scenarios, we have shown evidence for the fact that the problems with demand propagation in GSM models are not neglectable. The advantages come at a cost: The SGSM-DP is computationally more demanding (by a factor of over 100 in the example!), and this will become more serious of an issue for real-life scale problems. Considering the fact that the model usually has to be solved separately for each product class handled by a network, short computation times can become vital.

This paper showed a *logically* sound way to exactly model demand propagation in GSM-type models for linear demand bound functions. It has yet to be investigated

- how the modeling approach can be generalized to, say, *piecewise linear* demand bound functions, and
- how large the *computation times* grow for more complicated supply chains in practice.

Some important distribution systems, like the real-world US spare-part distribution system of a large German automobile manufacturer (which motivated the SGSM research in the first place), are of moderate size (two echelons, a single root node, less than ten demand nodes) so that the SGSM-DP can probably be solved for them by directly feeding the model and the data to a solver. For more complicated multi-company production supply chains, taylor-made algorithmic methods from mixed-integer linear programming might be needed and should therefore be investigated.

## REFERENCES

Clark, A.J. and Scarf, H. (1960). Optimal policies for a multi-echelon inventory problem. *Management Science*, 6(4), 475–490.

de Kok, T.G. and Fransoo, J.C. (2003). Planning supply chain operations: definition and comparison of planning concepts. *Handbooks in operations research and management science*, 11, 597–675.

Gamrath, G., Fischer, T., Gally, T., Gleixner, A.M., Hendel, G., Koch, T., Maher, S.J., Miltenberger, M., Müller, B., Pfetsch, M.E., Puchert, C., Rehfeldt, D., Schenker, S., Schwarz, R., Serrano, F., Shinano, Y., Vigerske, S., Weninger, D., Winkler, M., Witt, J.T., and Witzig, J. (2016). The SCIP Optimization Suite 3.2. Technical Report 15-60, ZIB, Takustr.7, 14195 Berlin.

Graves, S.C. and Willems, S.P. (2003). Supply chain design: safety stock placement and supply chain configuration. *Handbooks in operations research and management science*, 11, 95–132.

Löhnert, M. (2016). *Das Stochastic Guaranteed Service Model in vielstufigen Lagernetzen.* Masterarbeit, Universität Bayreuth.

Magnanti, T.L., Shen, Z.J.M., Shu, J., Simchi-Levi, D., and Teo, C.P. (2006). Inventory placement in acyclic supply chain networks. *Operations Research Letters*, 34(2), 228–238.

Rambau, J. and Schade, K. (2010). The stochastic guaranteed service model with recourse for multi-echelon warehouse management. In *Proceedings of the International Symposium on Combinatorial Optimization (ISCO 2010)*, volume 36 of *Electronic Notes in Discrete Mathematics*, 783–790. Elsevier.

Rambau, J. and Schade, K. (2014). The stochastic guaranteed service model with recourse for multi-echelon warehouse management. *Mathematical Methods of Operations Research*, 79(3), 293–326.

Schade, K. (2012). *Bestandsoptimierung in mehrstufigen Lagernetzwerken.* Stochastische Optimierung. Springer.