Efficient Mathematical Programming Model for Multi-Echelon Inventory Optimization based on the Guaranteed-Service Approach

V. G. Achkar[1], B. B. Brunaud[2] and I. E. Grossmann[3]*

[1]*Universidad Nacional del Litoral, Santiago del Estero 2829, Santa Fe, Argentina,*

[2]*Johnson & Johnson, 430 US Highway 22, Bridgewater, NJ 08807*

[3]*Carnegie Mellon University, 5000 Forbes St, Pittsburgh, PA 15213*

Keywords: Multi-echelon inventory optimization, safety stock, non-normal demand, MIQCP, piecewise.

Abstract. A Multi-Echelon Inventory Optimization (MEIO) model based on the Guaranteed-Service approach is presented for allocating safety stocks across a supply chain, to meet service levels at minimum cost. This paper builds on the previous work by Achkar et al. (2023), extending the approach to account for new advances. First, the Mixed-Integer Quadratically Constrained Program (MIQCP) reformulation is combined with a piecewise linear approximation to improve the computational efficiency, leading to significant reductions in computational time. In addition, the piecewise function yields an improved approximation for the fill rate targets. Furthermore, this work extends the GSM approach to represent non-normally distributed demands with high variability and proposes a different approach to account for stochastic lead times through a discrete function. The model is applied to several instances of a real-world case study, with more than 7300 product-location combinations, showing that optimal solutions can be obtained within few seconds of computational time.

1. INTRODUCTION

Inventory management plays a critical role in supply chain optimization (Grossmann, 2005; Papageorgiou, 2009; Snyder & Shen, 2019). In recent years, there has been growing interest in multi-echelon inventory optimization (MEIO) as it holds the potential to enhance efficiency and profitability of supply chains. MEIO aims to optimize safety stock levels across the entire network, striking a balance between inventory costs and the risk of stockouts. While excess inventory yields unnecessary holding

* grossmann@cmu.edu

costs, failing to meet the customer needs results in both loss of profit, and potentially the long term loss of customers (Jung et al., 2004a). Effective inventory management is mostly conditioned by the production system as well as by the demand and supply uncertainties (Rodriguez & Vecchietti, 2010). De Kok et al. (2019) provide a comprehensive review of MEIO models summarizing the extensive research conducted in this area. The authors highlight that multi-echelon inventory systems are still a very active area of research because of their complexity and practical relevance.

An approach in MEIO to determine safety stock levels is the Guaranteed-Service Model (GSM) (Graves & Willems, 2000; Simpson, 1958). From an optimization perspective, the GSM poses challenges due to the nonlinearity and nonconvexity of the problems involved. The previous contributions regarding solution methods developed for the GSM can be found in the work of Eruguz et al. (2016), where they are classified as follows: (i) dynamic programing solution approaches, (ii) mixed integer programming (MIP) approaches, and (iii) heuristics. Furthermore, other authors have explored the integration of the classic GSM with additional activities or approaches. You and Grossmann (2008a, 2008b, 2009) have developed models and algorithms that address inventory optimization and supply chain design under demand uncertainty, considering deterministic lead times and using the Cycle Service Level (CSL) as desired service levels. They reformulate the nonlinear problem as a mixed-integer nonlinear programming (MINLP) using a quadratic reformulation with decomposition algorithm based on a Lagrangean relaxation, piecewise linear approximations and a bi-criterion optimization approach. Additionally, Jung et al. (2008) introduce a simulation-optimization framework to estimate the optimal safety stock levels by employing a linearization of the model using an approximation of the concave function with the convex hull of straight lines. They present a case study involving two production sites and 8 warehouses supplying 10 final products to 30 sales regions. Moreover, Chu et al. (2014) propose a simulation-based optimization framework for solving multi-echelon inventory problems quantified by the fill rates, finding local optimal solutions within modest computational times. Recently, Achkar et al. (2023) presented a GSM model accounting for stochastic lead times and fill rates. They propose a Mixed Integer Quadratically Constrained Programming (MIQCP) model in conjunction with a second-order polynomial regression. The results demonstrated optimal solutions within modest computational times for cases involving up to 140 products and 4 locations.

In terms of demand modelling, the GSM approach assumes that it follows a stationary process that is normally distributed. However, in industrial practice, the demand often does not follow a normal distribution, with its coefficient of variation ($CV = \sigma/\mu$) being closer or even larger than 1. In such cases, it can be more attractive to use an alternative distribution that ensures nonnegative demand (Soares, 2013). In this work, we propose incorporating the gamma distribution as an alternative to represent demand for large CVs when the CSL is the service measure, with the fill rate being outside the scope of this work. The gamma distribution offers several advantages: it guarantees positive demand, and allows for a higher probability of very high demand compared to the normal distribution due to its right-skewness. There are few works in the literature that use Gamma distributions for safety stock setting (Beutel & Minner, 2012; Mirzaee, 2017; Moors & Strijbosch, 2002), but there is little work for extending the GSM model to account for nonnormal distributions for the demand. Mirzaee (2017) proposes an alternative approach to determine safety stocks by adjusting the service level, using the equivalent safety factor (*k*) that corresponds to the value of the empirical cumulative distribution function (*h(x)*).

This paper aims to build upon the previous work of Achkar et al. (2023) to include: (i) a solution method with improved computational efficiency, (ii) non-normal demand distributions, and (iii) an improved approach to account for stochastic lead times. The motivation for each point is as follows. First, their approach may become computationally expensive when attempting to find the optimal solution as the model size increases, particularly when the desired target is the fill rate. Second, it assumes normally distributed demand, which does not account for the highly variable demands often observed in real-world industrial settings, leading to underestimations of service levels. Third, the stochastic lead time extension of the proposed GSM, based on the work of Inderfurth (1993) can result in significant basestock levels on upstream nodes, and involves distinct calculations for internal and external demands.

To address the limitations mentioned above and to improve upon the previous model developed by Achkar et al. (2023) the present paper introduces several key enhancements:

(i) A new MIQCP model is introduced to enhance computational efficiency and improve estimation accuracy for fill rate targets. This model incorporates a piecewise linear approximation, which significantly reduces computational complexity compared to the previous polynomial regression. The new model achieves faster and more precise estimations, making it a valuable tool for optimizing fill rate targets.

(ii) An extension to considerate highly variable demands within the GSM framework. This is achieved by adjusting the safety factor, extending the approach of Mirzaee (2017). The model is adjusted to ensure the desired CSL in cases of large CVs, where the original distribution is unknown and the mean and standard deviation are available. The gamma distribution is included as an alternative distribution for these cases.

(iii) A novel approach for dealing with stochastic lead times is introduced, building upon the work of Minner (1998). A discrete function is introduced to address the challenges associated with lead time variability for divergent network topologies.

In summary, the extended GSM model includes: (i) the consideration of non-normally distributed demands, only when the target service measure is the Cycle Service Level (CSL), (ii) the inclusion of both stochastic demand and lead times, and (iii) the fill rate as an additional key customer service performance indicator, alongside the CSL. From an optimization perspective and to the best of our knowledge, this is the first work that proposes a quadratic reformulation to solve the nonlinear nonconvex problem in conjunction with a discrete function to include stochastic lead times and a piecewise linear approximation to extend the model for fill rate service levels. The MINLP model is reformulated as an MIQCP by exploiting the structure of the constraints of the base model. The piecewise linear approximation and the discrete function provide a more computationally efficient and precise approach. Examples for a real industrial system are presented to illustrate the application of the proposed model and its improved computational performance.

The outline of the paper is as follows. The problem statement and assumptions are described first, followed by the model formulation. Next, the solution approach is presented to include the solution strategies and the extension for non-normal distributions for the demand. The paper ends with the application of the model to illustrate its application to illustrative examples and real-world case studies. Conclusions are drawn in the final section.

2. PROBLEM STATEMENT AND ASSUMPTIONS

A supply chain is considered with a fixed design for locations $j \in J$ and a set of materials $p \in P$. Materials can be either raw materials or finished goods. The distribution routes within the supply chain are divergent and single-sourcing, meaning that a node holding a specific material p can only receive it from a single node and distribute it to one or more locations. It should be noted that the same node can receive many different materials $q \in P$ from other locations, but it should be a single sourcing route to supply each q on that node.

The demand at each node is independent and characterized by a stationary process, with a mean μ_{jp} and standard deviation σ_{jp}. A novelty in this work is that the statistical distributions of demand do not necessarily follow a normal distribution. Instead, the model includes demands with larger Coefficient of Variation (CV) by considering them to be gamma distributed. We propose this extension only for CSL, leaving the fill rate targets to be addressed in future work. Additionally, demand is assumed to be bounded as in the classic GSM, meaning that if in a certain time period it exceeds the bound, extraordinary measures are used to satisfy the excess demand. To propagate demand upstream, risk pooling assumptions as described in You and Grossmann (2009) are applied. Plant locations introduce a coefficient ϕ_{pq}, representing the bill of materials for product transformation. This coefficient depends on the relationship between materials and finished goods, and it is used in the demand propagation to determine raw materials demand parameters.

In terms of replenishment decisions, each stage operates under a periodic review inventory policy denoted as *(R,S)*, where R and S represent the review period and order-up-to level, respectively. A basestock level is maintained at each stage. Replenishment orders may be subject to a Minimum Order Quantity (moq_{jp}) requirement.

The core decision variable of the classic guaranteed service approach is the service time S_j. When a customer places an order of size $d_j(t)$ on node j at time t, the order will be fulfilled by time $t + S_j$, where S_j represents the guaranteed-service time of node j (Graves & Willems, 2000). Furthermore, each node j within the supply chain receives a service commitment from its upstream nodes $i \in J$. This commitment is known as the inbound service time (SI_{jp}) and is defined as SI_{jp}=max $\{S_{ip} \mid i : (i,j,p) \in A\}$, with A being the set of active arcs (i,j) for p. The Net Lead Time (NLT), shown in Figure 1, represents the period that is not covered by the guaranteed service time and must be covered with safety stock. It is calculated as $NLT_{jp} = SI_{jp} + LT_{jp} - S_{jp}$. For instance, if $NLT_{jp} = 0$, it indicates that the node operates under a make-to-order (MTO) policy, where no inventory is stored. Conversely, if $NLT_{jp} > 0$, there is a time period that needs to be covered with safety stock, indicating a make-to-stock (MTS) policy.

Regarding the lead times, they are associated for each location and material, and they are assumed to follow a normal distribution with mean LT_{jp} and standard deviation σ_{LTjp}. While the work of Achkar et al. (2023) introduces stochastic lead times based on the work by Inderfurth (1993), this paper proposes a novel extension inspired by Minner (1998). In the GSM, the lead times used in single-echelon systems are replaced by the net lead times in multi-echelon systems. In the present paper, it is assumed that the variable NLT_{jp} described above has a standard deviation σ_{NLTjp}, and its value depends on the specific NLT_{jp}. A new discrete function is proposed to determine the value of σ_{NLTjp}. This approach introduces an additional layer of complexity to the problem, as previous parameters now become variables in the model.

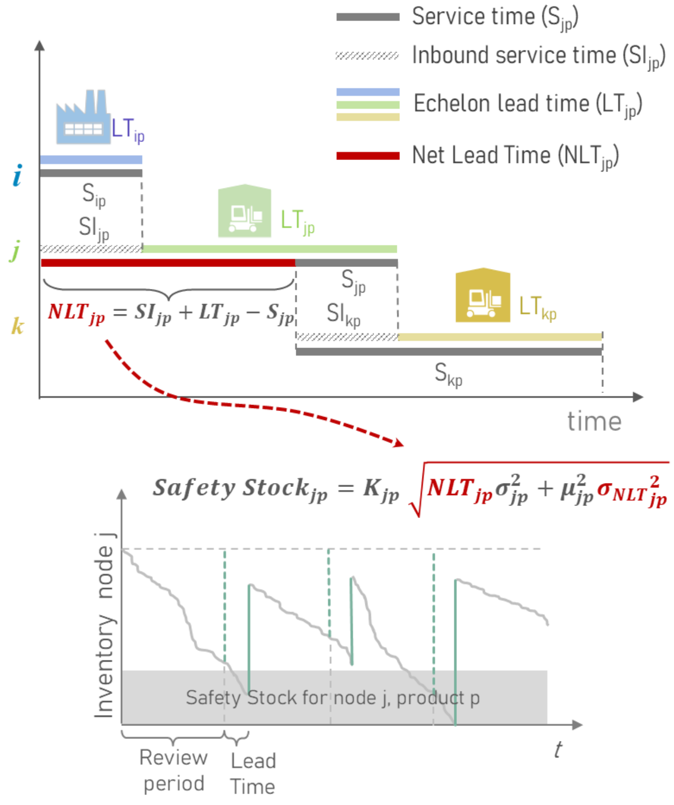

Figure 1. Guaranteed-service model elements

In terms of the service levels, a safety stock factor k_{jp} is associated with the Cycle Service Level (CSL) at each node in the supply chain. This factor represents the percentage of replenishment cycles during which stockouts do not occur. Alternatively, the modeler can also ask for a fill rate to be considered as a target service measure. The fill rate represents the proportion of demand that is successfully fulfilled out of the total demand placed. It should be noted that this option is applicable in the current paper only when the demand is assumed to follow a normal distribution. To incorporate fill rates into the model, the safety factor k_{jp} must become a variable (K_{jp}), and the loss function of the normal distribution $G(K_{jp})$ is required. The work by Achkar et al. (2023) uses a second-order polynomial regression to approximate this function. Although this approximation generally provides a good fit (R^2 = 0.98), slight differences in the approximation, particularly for larger safety factors, can lead to significant variations in the required safety stock levels. To address this limitation, the new approach utilizes a piecewise linear approximation, providing a better fit for the loss function $G(K_{jp})$.

Holding costs are incurred at all nodes within the supply chain. The objective is to find the guaranteed-service times (SI_{jp} and S_{jp}), and consequently how much safety stock to maintain, in order to minimize the total holding costs and satisfy the specified customer service level. The service times at the initial and the final nodes are given.

3. MODEL FORMULATION

The formulation of the model maintains certain variables related to Guaranteed-service Model (GSM). These variables include the guaranteed service time (S_{jp}) of a product p at a specific node j, the inbound service (SI_{jp}), and the net lead time (NLT_{jp}). Moreover, in this model, the traditional safety factor, which is an input parameter in the classic GSM, is transformed into a variable (K_{jp}) related to the service level.

The first set of constraints are derived from the classic GSM. Equation (1) defines the first inbound service time for the starting nodes in the network J^0, where si^0 is a given input. Equation (2) establishes the connection between the inbound guaranteed-service time SI_{jp} and the guaranteed-service time of upstream nodes S_{iq}. The set A consist of elements (i,j,p) indicating that there is an enabled route for material p from i to j. It should be noted that if the link is for the same product ($q = p$) and is a distribution link (from node i to node j), then $i \neq j$. Conversely, if node j is a plant location that produces p from q, then $q \neq p$ and $i = j$. To ensure feasibility, the service time S_{jp} is bounded by Eq. (3). Eq. (4) becomes active if there is a maximum accepted service time. The NLT_{jp} is defined in Eq. (5), with lt_{jp} being the lead time and r_{jp} the time period between reviews. In cases where nodes face external demand, the parameter $maxS_{jp}$ is set to 0, enforcing a make-to-stock (MTS) inventory policy for these demand nodes.

$$SI_{jp} = si^0_{jp} \qquad \forall\ j\ \in J^0, p \in P_j \qquad (1)$$

$$SI_{jp} \geq S_{iq} \qquad \forall\ (i,j,p) \in A, (q,p) \in \Phi, p \in P_j \qquad (2)$$

$$S_{jp} \leq SI_{jp} + lt_{jp} + r_{jp} \qquad \forall\ j\ \in J, p \in P_j \qquad (3)$$

$$S_{jp} \leq maxS_{jp} \qquad \forall\, j \in J, p \in P_j \tag{4}$$

$$NLT_{jp} \geq SI_{jp} - S_{jp} + lt_{jp} + r_{jp} \qquad \forall\, j \in J, p \in P_j \tag{5}$$

In the GSM, the CSL is typically used as the customer service performance indicator to determine safety stocks. However, when the desired target is the fill rate ($j,p \in F$), a constraint is introduced to calculate the equivalent CSL of a given fill rate value (fr_{jp}) (Chopra & Meindl, 2013). Eq. (6) establishes the relationship between the fill rate (fr_{jp}) and the safety factor, thereby determining the minimum required CSL that can meet the desired fill rate. The safety factor, as mentioned before, becomes a positive continuous variable, K_{jp}, and the aim is to find the lower required CSL that is able to meet the desired fill rate. In Eq. (6), the square root represents the "sigma-combination" formula, known from single-echelon models to determine the standard deviation of demand over the lead time, applying the random sum of two random variables representing the demand and the lead time (Drake, 1967; Silver et al., 1998). The mean demand and the standard deviation are represented by μ_{jp} and σ_{jp}, respectively. The supply variability parameters are NLT_{jp} and its standard deviation σ_{NLTjp}. The approach proposed to obtain σ_{NLTjp} is explained in subsequent paragraphs. $Q_{jp} = \max\{\mu_{jp}\, r_{jp}, moq_{jp}\}$ represents the average replenishment quantity of product p on location j, with moq_{jp} being the minimum order quantity for replenishment. $\Phi(K_{jp})$ and $\varphi(K_{jp})$ correspond to the cumulative distribution function and the probability density function of the standard normal distribution, respectively. These functions are utilized as part of the loss function of the normal distribution, denoted as $G(K_{jp}) = K_{jp}\, [1\text{-}\Phi(K_{jp})]\text{-}\ \varphi(K_{jp})$ (Axsäter, 2006).

$$fr_{jp} \leq 1 + \frac{\sqrt{NLT_{jp}\ \sigma_{jp}^2 + \mu_{jp}^2\ \sigma_{NLT_{jp}}^2}}{Q_{jp}} \Big(K_{jp}\big[1 - \Phi\big(K_{jp}\big)\big] - \varphi\big(K_{jp}\big)\Big) \tag{6}$$

$$\forall\, j \in J,\ p \in P_j,\ (j,p) \in\ F$$

On the other hand, if the service level target is CSL, the safety factor is given as an input k_{jp}, as in Eq. (7):

$$K_{jp} = k_{jp} \qquad \forall\, j \in J,\ p \in P_j,\ (j,p) \notin\ F \tag{7}$$

As mentioned before, stochastic lead times are introduced into the GSM by extending the approach for serial networks proposed by Minner (1998) to divergent networks, because it faces all the variability sources using safety stock. Other approaches such as Inderfurth (1993) can lead to large stock levels. To account for multi-echelon systems together with stochastic lead times, we define σ_{NLTjp} to represent the standard deviation of the NLT_{jp}. To obtain its value, we introduce a new discrete function. We assume that the variability of the lead time is pushed downstream if the NLT of node j for product p is not enough to cover the total replenishment time inherent to node j. In other words, if $NLT_{jp} \leq SI_{jp} + lt_{jp} + r_{jp}$, then, σ_{LTjp} should be propagated downstream. As the objective function is concave and there is an "all or nothing" optimal solution, we assume that σ_{NLTjp} accumulates either all lead time variabilities until stage j, or it pushes all of them to the following downstream nodes, as shown in Figure 2. In the left scenario, the plant operates under a make-to-order (MTO) policy, while the subsequent nodes i and k operate under make-to-stock (MTS) policies. Node j must incorporate both the lead time variability of node j itself ($\sigma^2_{LT_{jp}}$) plus the upstream node variability ($\sigma^2_{LT_{ip}}$) into its $\sigma^2_{NLT_{jp}}$ value. In the right case, nodes i and j operate under MTO policies, pushing all the variability to the most downstream node in the network. Consequently, the variance of the net lead time for node k is calculated as $\sigma^2_{NLT_{kp}} = \sigma^2_{LT_{ip}} + \sigma^2_{LT_{jp}} + \sigma^2_{LT_{kp}}$. The discrete function is depicted in Figure 3. Note that the "all or nothing" concavity property of the approach leads to a discretized function rather than a stepwise function, without incorporating any intermediate values for NLT_{jp}. This is because intermediate values are not considered as optimal alternatives in the optimization process.

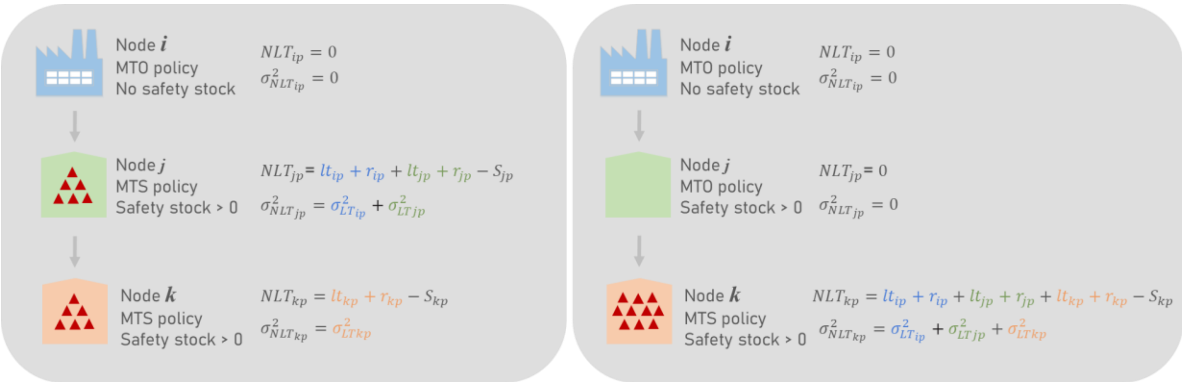

Figure 2: Representation of the NLT variance calculation for different nodes of the network depending on inventory policies (MTS or MTO).

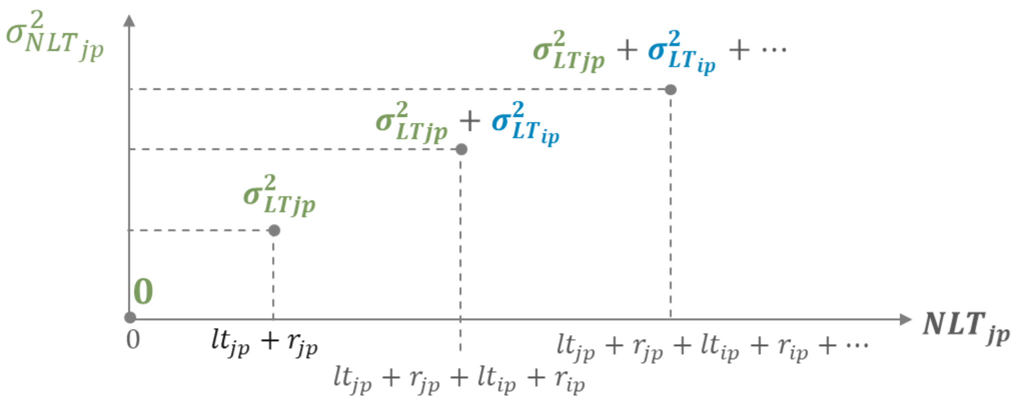

Figure 3. Discrete function to define $\sigma^2_{NLT_{jp}}$

A new positive continuous variable, X_{jp}, is proposed to represent $\sigma^2_{NLT_{jp}}$.The value for each step $s \in S_1$ is defined by the parameter d_{jps}. For the first step (s_1) of product p at node j, $d_{jps1} = 0$, and for s_2, $d_{jps2} = lt_{jp} + r_{jp}$. As NLT_{jp} increases, the bounds increase by adding the lead times and the review periods of upstream nodes with $d_{jps3} = lt_{jp} + r_{jp} + lt_{ip} + r_{ip}$, being i the upstream node. A new binary variable V_{1jps}

defines what the active step in Eq. (8) is according to the value of NLT_{jp}. Equation (9) ensures that only one step is active for each product on each location. Eq. (10) assigns the value of σ_{NLTjp} associated with the step s, given by the parameter c_{jps}, which represents the sum of lead time variances according to Figure 3. In the example, $c_{jps1} = 0$, $c_{jps2} = \sigma^2_{LT_{jp}}$, and $c_{jps3} = \sigma^2_{LT_{jp}} + \sigma^2_{LT_{ip}}$. It is worth to mention that this coefficient also includes the lead time variance of raw materials if they are involved in the production of product p.

$$NLT_{jp} = \sum_{s \in S_1} d_{jps} V_{1_{jps}} \qquad \forall j \in J, p \in P_j \tag{8}$$

$$\sum_{s \in S_1} V_{1_{jps}} = 1 \qquad \forall j \in J, p \in P_j \tag{9}$$

$$X_{jp} = \sum_{s \in S_1} V_{1_{jps}} c_{jps} \qquad \forall j \in J, p \in P_j \tag{10}$$

The previously presented discrete function fails to keep the concavity property of the GSM model when upstream lead times are larger than downstream lead times. The optimal solution can pick intermediate values for the service times and the NLT that avoids accumulating large lead times from upstream nodes. This is not correctly modelled, because nodes that have MTO policies must push their variabilities to some other location to meet their service levels. Therefore, Eqs. (11) and (12) are proposed using upper bounds with big-M to reinforce the “all or nothing” optimal decisions. If the service time S_{jp} becomes a positive value, a MTO policy is the only available option, forcing the NLT_{jp} = 0. Conversely, if the $NLT_{jp} > 0$, MTS policy with no delay ($S_{jp} = 0$) is the only available possibility.

$$S_{jp} \leq V_{2_{jps}} M \qquad \forall j \in J, p \in P_j \tag{11}$$

$$NLT_{jp} \leq (1 - V_{2_{jps}}) M \qquad \forall j \in J, p \in P_j \tag{12}$$

Finally, the objective function is to minimize safety stock holding cost as defined by Eq. (13), where h_{jp} is the coefficient that represents the holding cost for each material at each location. As Eq. (13) is concave, there is an “all or nothing” optimal solution, a stage either has no safety stock, or has sufficient safety stock to de-couple it from its downstream stages (Graves & Willems, 2000).

$$min \sum_{p} \sum_{j \in B_{jp}} h_{jp} K_{jp} \sqrt{NLT_{jp} \sigma^2_{jp} + \mu^2_{jp} \sigma^2_{NLT_{jp}}} \tag{13}$$

3.1 Solution Approach

The model presented in equations (1)-(13) represents a challenging problem known as a nonconvex mixed-integer nonlinear problem (MINLP). While these problems can in principle be solved with global optimization solvers (Sahinidis, 1996), the computational time required to find a global solution can be very expensive. Eq. (6) includes the normal distribution density and cumulative functions. Moreover, Eqs. (13) and (6) include concave functions that significantly impact the computational burden. To overcome these difficulties, we propose two solution strategies: (i) an exact mixed-integer quadratically constrained reformulation, and (ii) a piecewise linear approximation of Eq. (6) to replace the function $G(K_{jp})$.

First, we reformulate the MINLP as an MIQCP as in Achkar et al. (2023). Solvers like CPLEX and Gurobi can solve MIQCPs to global optimality quite effectively in reasonable computational times. To obtain the MIQCP reformulation, we introduce a new positive continuous variable, U_{jp}, that represents the square root, i.e. $U_{jp} \geq \sqrt{\tau_{jp}}$, as shown in Figure 4, with τ being the argument of the square root. We include Eq. (14) to define U_{jp}^2:

$$U_{jp}^2 \geq NLT_{jp}\ \sigma_{jp}^2 + \mu_{jp}^2\ X_{jp} \qquad \forall\ j\ \in J, p \in P_j \tag{14}$$

Hence, the objective function becomes the linear function given by Eq. (15):

$$min \sum_{p} \sum_{j \in B_{jp}} h_{jp}\ K_{jp}\ U_{jp} \tag{15}$$

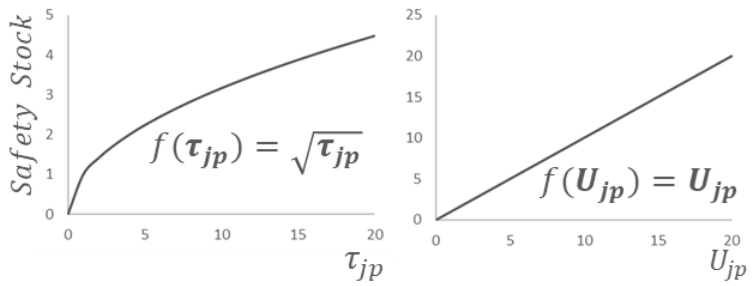

Figure 4. Reformulation from MINLP to MIQCP

Regarding $G(K_{jp}) = K_{jp}$ [1-$\Phi(K_{jp})$]- $\varphi(K_{jp})$ present in the fill rate constraint, we aim to improve the model tractability by replacing it with a piecewise linear approximation $Y(K_{jp})$, as shown in Figure 5 (A). This function is the same for all nodes on all locations, while demand is normally distributed. In the example (A), there are four breakpoints (p_1, p_2, p_3 and p_4) and three segments (s_1, s_2 and s_3). If the number of breakpoints increases, the precision of the estimation also does, but the computational efficiency decreases. For the current formulation we propose 8 breakpoints (p_1 to p_8) as shown in Figure 5 (B).

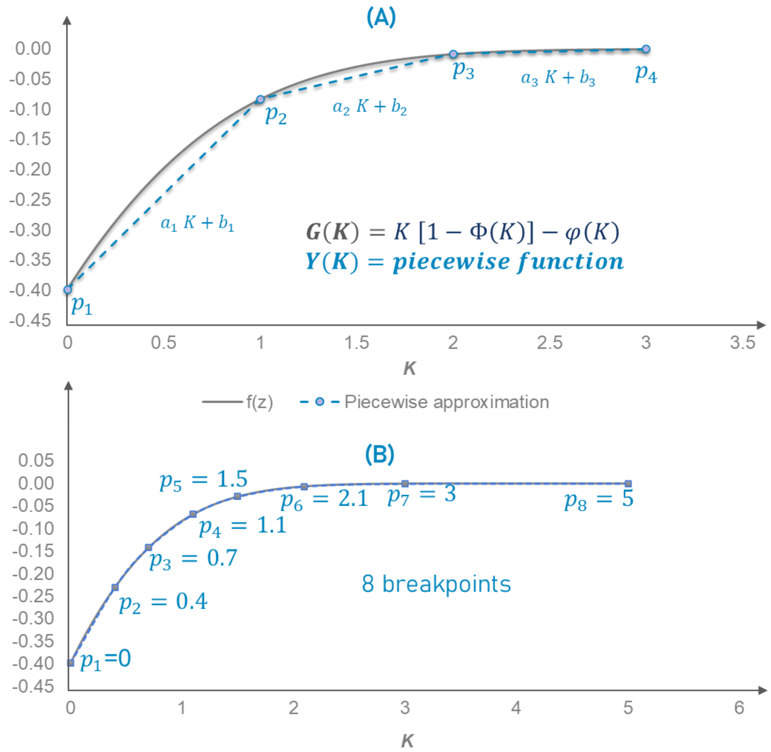

Figure 5. (A) Illustrative piecewise linear approximation, (B) Breakpoints selected for the current model

To introduce this approximation into the model, a continuous positive variable Y_{jp} is added to represent the piecewise linear function in Eq. (16). The continuous variable $\widetilde{K}_{jps}$ represents the value that the safety factor takes if the active segment is $s \in S_2$. As only one segment must be active, there is a binary variable W_{jps} that takes value 1 if segment s is the active segment, and 0 otherwise. The upper and lower bounds of each segment are given by Eqs. (17) and (18), with lb_{jps} and ub_{jps} being the breakpoints that define segment s. The condition of only one active segment is defined in (19). Equation (20) assigns the final value of the safety factor given by $\widetilde{K}_{jps}$ for active segment s to the variable K_{jp} that only depends of location j and material p.

$$Y_{jp} = \sum_{s \in S_2} \left(a_{jps} \widetilde{K}_{jps} + b_{jps}\, W_{jps} \right) \qquad \forall\, j\ \in J, p \in P_j \qquad (16)$$

$$lb_{jps} W_{jps} \leq\ \widetilde{K}_{jps} \qquad \forall\, j\ \in J, p \in P_j, s \in S_2 \qquad (17)$$

$$\widetilde{K}_{jps} \leq ub_{jps} W_{jps} \qquad \forall\, j\ \in J, p \in P_j, s \in S_2 \qquad (18)$$

$$\sum_{s \in S_2} W_{jps} = 1 \qquad \forall\, j\ \in J, p \in P_j \qquad (19)$$

$$\sum_{s \in S_2} \widetilde{K}_{jps} = K_{jp} \qquad \forall\, j \in J, p \in P_j \tag{20}$$

Finally, Eq. (6) is redefined including the new variables related to the piecewise functions:

$$fr_{jp} \leq 1 + \frac{U_{jp}}{Q_{jp}} Y_{jp} \qquad \forall\, j \in J,\; p \in P_j,\; (j,p) \in F \tag{21}$$

The new mathematical reformulation is a MIQCP given by Equations (1)-(5), (7)-(12), (14)-(21).

3.2 Extension for non-normal demand

In industrial practice, it is commonly observed that demand data histograms do not follow a normal distribution. Soares (2013) states that if the coefficient of variation (CV = σ/μ) of the demand is not much less than 1, there is a relatively high probability for negative demand when using the normal distribution. Consequently, simulating normally distributed demands with large CVs requires special treatment since demand cannot take negative values. If negative values are truncated to zero or removed from the analysis, the mean and standard deviation of the effective sample of demands differ from the original ones. As a result, the parameters used for safety stock and basestock calculations do not correspond to the actual data. Even if we have complete knowledge of demand realizations from the simulation and can calculate inventory targets based on these occurrences, achieving customer service levels becomes less effective as the coefficient of variation (CV) approaches 1. In other words, for distributions with large CVs, the GSM model predicts slightly lower Cycle Service Levels (CSLs) than expected, especially when CSL targets are closer to 100% as demonstrated in Table 2. This table presents the results of simulation experiments conducted using the Arena Software for a single-echelon network and a single product. Safety stocks and basestock levels ($B_{jp} = SS_{jp} + NLT_{jp}\ \mu_{jp}$) are calculated for different target CSLs: 90%, 93%, 96% and 99%. The focus is primarily on high service levels as they are the most commonly used targets in Multi-Echelon Inventory Optimization (MEIO). Demand is randomly generated following a normal distribution with CVs of 0.2, 0.5, 1, and 50. If a randomly generated number is negative, it is discarded, and a new number is generated until a positive value is obtained. The effective CVs of the demand obtained after truncation are 0.20, 0.46, 0.61, and 0.76. It is worth noting that representing demands with CVs larger than 0.8 becomes challenging using the truncated normal distribution, even for the case of an original CV = 50. For each combination of CV and target CSL, five replications of 10,000 periods each were run. A deterministic lead time equal to 1 and a period between reviews equal to 1 is assumed. Table 1 displays the mean CSL obtained from the five replications along with the 95% confidence interval for each experiment. It can be observed that the confidence intervals encompass the expected target for CV 0.2 and are close to the target for CV 0.5. However, for larger CVs, the intervals deviate from the target CSL as the target increases and the CVs get closer to 1. For CVs 0.6 and 0.7, all expected CSLs are more than 1 percentage point below the target.

Table 1: Mean CSL and Confidence intervals obtained from simulations

	Effective CSL Mean and Confidence Intervals from Simulations							
Original CV	0.2		0.5		1		50	
Effective CV	0.2		0.5		0.6		0.7	
Expected CSL	**Mean**	**CI**	**Mean**	**CI**	**Mean**	**CI**	**Mean**	**CI**
0.90	0.899	[0.898, 0.900]	0.894	[0.893, 0.896]	0.891	[0.888, 0.893]	0.890	[0.886, 0.893]
0.93	0.930	[0.929, 0.931]	0.925	[0.923, 0.926]	0.919	[0.918, 0.921]	0.915	[0.911, 0.919]
0.96	0.960	[0.958, 0.961]	0.955	[0.953, 0.956]	0.948	[0.946, 0.950]	0.944	[0.941, 0.946]
0.99	0.989	[0.989, 0.990]	0.987	[0.986, 0.988]	0.982	[0.981, 0.984]	0.978	[0.977, 0.979]

Although the assumption of normal distribution usually works relatively well also in these examples, with small deviations from the desired CSL, it can be more attractive to use an alternative distribution that ensures nonnegative demand. In this work, we propose incorporating the gamma distribution to represent demand for CSL targets, with fill rate being outside the scope of this work. The gamma distribution offers several advantages: it guarantees positive demand, can capture shapes similar to exponential and normal distributions, and allows for a higher probability of very high demand compared to the normal distribution due to its right-skewness. Additionally, it can effectively represent more volatile demand patterns with coefficient of variations (CVs) exceeding 1.

As the current mathematical model proposed in this work is based on the safety factor derived from the normal distribution, the objective of this section is to extend the GSM to address scenarios where demand exhibits large CVs, and the existing safety stock model fails to meet the target CSL. Few works in the literature use Gamma distributions for safety stock setting, and there is little work for extending the GSM model (Beutel & Minner, 2012; Mirzaee, 2017; Moors & Strijbosch, 2002)

Mirzaee (2017) proposes an alternative approach to determine safety stocks by adjusting the service level, using an equivalent safety factor (k_{jp}) that corresponds to the value of the empirical cumulative distribution function $H(x_{jp})$. In this work, our goal is to find an alternative method that ensures the desired customer service level in cases of large CVs, where only the mean and standard deviation of the demand are known. We propose employing the gamma cumulative distribution function $\Gamma(x)$ as an alternative to determine the safety factor k_{jp}. Following the approach of Mirzaee (2017), we adjust the safety factor using the standard normal cumulative density function (cdf), $\Phi(k_{jp})$, but the new safety factor comes from the inverse of the gamma distribution for the desired CSL, denoted $\Gamma_{jp}^{-1}(CSL_{jp})$, as depicted in Figure 6. The blue line corresponds to the normal distribution, with k_{Njp} being the safety factor obtained from the inverse of the normal cdf as in the classic GSM. On the other hand, the red line represents the gamma distribution for the same mean demand and standard deviation parameters, and k_{Gjp} is the safety factor obtained from the standardized inverse gamma cdf.

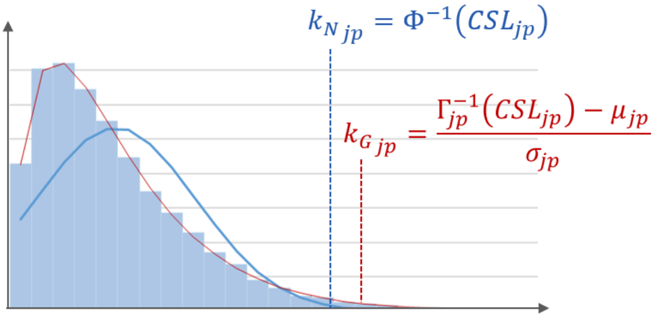

Figure 6: Proposed safety factors for normal and gamma distributed demands using the standard normal cdf

It is important to note that in certain situations, particularly for high service levels, k_{Gjp} may result in larger values. Conversely, there may be cases where k_{Gjp} is lower than the original safety factor k_{Njp} used for lower service levels, leading to a decrease in inventory levels. Since the original distribution of demand is unknown, we propose a new approach that selects the maximum value between the traditional safety factor k_{Njp} and the one obtained using the gamma distribution k_{Gjp}. This is represented in Eq. (22):

$$k_{jp} = max\left\{k_{N\,jp}, k_{G\,jp}\right\} = max\left\{\Phi^{-1}(CSL_{jp}), \frac{\Gamma_{jp}^{-1}(CSL_{jp}) - \mu_{jp}}{\sigma_{jp}}\right\} \qquad (22)$$

To evaluate the proposed approach, a simulation study is conducted using multiple datasets with demands following a gamma distribution. The CV for these datasets ranges from 0.6 to 2 We assume that CVs larger than 2 are related to highly erratic demands and the safety stock models may not be the appropriate ones to implement in those cases. Table 2 displays the results obtained for the CSL using the safety factors obtained from both the normal distribution (k_{Njp}) and the equivalent factor derived from the gamma distribution (k_{Gjp}), named N and G, respectively. This table presents the mean and the 95% confidence interval for different CSLs from 81% to 99%. We observe that k_{Gjp} performs better for larger CSL targets, while it yielded poorer results for lower service level values.

If the complete set of demand data were available, we could obtain the empirical distribution function *H(x)* for the dataset instead of using the gamma distribution. However, in this work, we assume that such a dataset is not available, and we only have access to the mean and standard deviations of the demand. Using Eq. (22), as shown in Table 2, we can observe that all CSL targets are satisfied. One drawback of this approach is that it may result in over-buffering in the case of very large CVs. In future work, we aim to extend this approach to incorporate fill rate targets. The proposed extension enables the MEIO model to accurately calculate safety stock levels for demands with large CVs, without compromising computational efficiency, as it serves as a pre-processing step to determine the parameter k_{jp} in Eq. (7). The MIQCP with the concavity property and model efficiency are conserved.

Table 2. Effective CSL for large CVs with original and safety factors k_{Njp} (N) and k_{Gjp} (G)

		CV = 0. 6		**CV = 0.8**		**CV = 1**		**CV = 2**	
Target CSL		**N**	**G**	**N**	**G**	**N**	**G**	**N**	**G**
0.81	**Mean**	**0.822**	0.803	**0.829**	0.799	**0.837**	0.796	**0.885**	0.782
	CI	[0.821, 0.824]	[0.802, 0.804]	[0.827, 0.831]	[0.799, 0.800]	[0.834, 0.840]	[0.792, 0.801]	[0.884, 0.886]	[0.779, 0.786]
0.84	**Mean**	**0.846**	0.836	**0.851**	0.834	**0.856**	0.830	**0.897**	0.822
	CI	[0.843, 0.849]	[0.834, 0.839]	[0.848, 0.854]	[0.832, 0.836]	[0.853, 0.859]	[0.828, 0.833]	[0.896, 0.899]	[0.818, 0.825]
0.87	**Mean**	**0.869**	**0.869**	**0.872**	0.867	**0.875**	0.864	**0.909**	0.861
	CI	[0.867, 0.871]	[0.867, 0.871]	[0.870, 0.874]	[0.865, 0.869]	[0.872, 0.877]	[0.861, 0.868]	[0.907, 0.912]	[0.858, 0.864]
0.90	**Mean**	0.892	**0.901**	0.893	**0.901**	0.893	**0.896**	**0.920**	0.897
	CI	[0.891, 0.894]	[0.900, 0.903]	[0.892, 0.894]	[0.900, 0.902]	[0.89, 0.896]	[0.893, 0.899]	[0.917, 0.922]	[0.893, 0.900]
0.93	**Mean**	0.917	**0.935**	0.916	**0.935**	0.914	**0.932**	**0.932**	**0.932**
	CI	[0.915, 0.918]	[0.933, 0.936]	[0.914, 0.917]	[0.934, 0.936]	[0.911, 0.916]	[0.930, 0.933]	[0.930, 0.934]	[0.930, 0.934]
0.96	**Mean**	0.943	**0.964**	0.940	**0.964**	0.938	**0.965**	0.946	**0.964**
	CI	[0.941, 0.944]	[0.963, 0.965]	[0.938, 0.941]	[0.962, 0.965]	[0.936, 0.939]	[0.964, 0.966]	[0.945, 0.948]	[0.962, 0.965]
0.99	**Mean**	0.974	**0.993**	0.970	**0.994**	0.969	**0.994**	0.966	**0.992**
	CI	[0.973, 0.976]	[0.993, 0.994]	[0.969, 0.972]	[0.993, 0.994]	[0.968, 0.97]	[0.994, 0.995]	[0.964, 0.968]	[0.992, 0.993]

4. APPLICATION AND RESULTS

4.1 Illustrative example

To better understand the results of the model, we propose to first solve the illustrative example presented in Achkar et al. (2023), depicted in Figure 7, with slight changes on demand and holding cost parameters. This case involves the production and distribution of a finished good (*SKU1*) produced from two raw materials (*Raw1* and *Raw2*) at a Plant. The finished good is then delivered to three retailers (R_1, R_2 and R_3) who satisfy external demand. The bill of materials (BOM) coefficients are $\phi_{Raw1,\,SKU1}$ = 1 and $\phi_{Raw2,\,SKU1}$ = 0.014. Table 3 provides the demand and lead time parameters, maximum service time, and unit holding costs for each material and location. The target service measure chosen is the fill rate, with a desired value of 97% for all products.

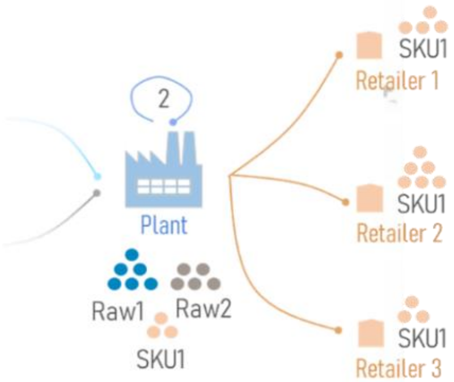

Figure 7: Illustrative example representation

Table 3: Illustrative example input data

Material		*Raw₁*	*Raw₂*	*SKU₁*	*SKU₁*	*SKU₁*	*SKU₁*
Location		*Plant*	*Plant*	*Plant*	R_1	R_2	R_3
Demand (units)	μ_{jp}	425,717	5,913	425,717	162,379	67,284	196,054
	σ_{jp}	116,671	1,620	116,671	48,714	40,370	98,027
(CV= σ_{jp} /μ_{jp})		0.27	0.27	0.27	0.30	0.60	0.50
Lead Time (weeks)	lt_{jp}	6	3	2	1	1	1
	σ_{LTjp}	1.9	0.7	0.0	0.3	0.6	0.4
Max Service Time S_{jp} (weeks)		-	-	-	0	0	0
h_{jp} ($/unit)		0.0586	0.0001	0.6	0.6	0.6	0.6

The advantage of using the MIQCP reformulation instead of the MINLP model regarding computational efficiency and obtaining the same global optimum has been shown in previous work Achkar et al. (2023). In this work, we solve the illustrative example comparing two approaches based on the MIQCP formulation. The first one is (a) the current approach proposed in this work, with a piecewise linear approximation. To provide a basis for comparison, we also solve the same example using (b) the polynomial regression proposed in Achkar et al. (2023), using a second-order polynomial regression in Eq. (6) ($h(K_{jp}) = a\, K_{jp}^{\,2} + b\, K_{jp} + c$), with a = -0.074700, b = 0.331986, c = -0.357195. Computational results are shown in Table 4. The tests were performed using Pyomo on an Intel Core i7 CPU with 12 GB RAM and 4 parallel threads, with Gurobi 9.5 as the QCP solver. In terms of approach (a), it is worth noting that the size of the model increases significantly due to the inclusion of constraints associated with the piecewise function. However, the computational time required to obtain the optimal solution is lower compared to approach (b).

Table 4: Computational Results

	(a) MIQCP with piecewise function (current approach)	(b) MIQCP with polynomial regression
CPU time (seconds)	0.17	0.25
Optimality GAP	0%	0%
Number of constraints	164	86
Number of binary variables	61	25
Number of integer variables	73	37
Number of continuous variables	85	43

Both the MIQCP formulation with the piecewise linear approximation (a) and the polynomial regression approach (b) successfully find the same optimal solution for the illustrative example. The obtained solution is presented in Table 5. The safety stock allocation decisions are the same for both approaches, deciding to push the safety stock of finished goods in the plant to downstream nodes, selecting a guaranteed service of 2 weeks for supplying the retailers. When comparing the total costs, approach (a) achieves a total cost of \$351,531, while approach (b) results in a total cost of \$358,046. The main factor contributing to the cost difference between the two approaches is the value of the safety factor K_{jp} obtained through the different approximation methods. As small differences in the safety factor can yield large differences in safety stock costs, the selection of the most accurate approximation method is of great importance.

In Table 6, a comparison of the precision between the two approaches is presented. This comparison is based on the fill rate obtained by using the safety factor K_{jp} from the piecewise linear approximation, and the regression approximation in the original Eq. (6). The results indicate that the piecewise linear approximation provides a better fit for approximating the loss function with smaller deviations from the target fill rate of 97%. This higher precision in approximation leads to a 2% reduction in the total cost. Therefore, for this example, the improved precision achieved through the piecewise linear approximation contributes to better accuracy in determining the safety factor and subsequently optimizing the cost, resulting in cost savings compared to the regression approximation.

Table 5: Illustrative example results

		(a) MIQCP with piecewise function (current approach)					(b) MIQCP with polynomial regression				
Material	Location	S_{jp}	SI_{jp}	K_{jp}	SS_{jp}	$Cost_{jp}$	S_{jp}	SI_{jp}	K_{jp}	SS_{jp}	$Cost_{jp}$
Raw_1	*Plant*	0	0	1.88	1,629,691	$ 95,451	0	0	1.63	1,409,022	$ 82,526
Raw_2	*Plant*	0	0	1.45	7,648	$ 1	0	0	1.44	7,583	$ 1
SKU_1	*Plant*	3	0	0	0	$ 0	3	0	0	0	$ 0
SKU_1	$Retailer_1$	0	3	0.79	94,563	$ 56,738	0	3	0.87	103,339	$ 62,004
SKU_1	$Retailer_2$	0	3	0.67	66,236	$ 39,742	0	3	0.74	73,508	$ 44,105
SKU_1	$Retailer_3$	0	3	1.14	266,000	$ 159,600	0	3	1.21	282,350	$ 169,410

Table 6: Precision of approximations

		(a) MIQCP with piecewise function (current approach)			(b) MIQCP with polynomial regression		
Material	Location	K_{jp}	Fill rate from Eq. (7)	Difference from target	K_{jp}	Fill rate from Eq. (7)	Difference from target
Raw_1	*Plant*	1.88	0.9765	0.0065	1.63	0.9557	-0.0143
Raw_2	*Plant*	1.45	0.9711	0.0011	1.44	0.9703	0.0003
SKU_1	*Plant*	0	-	-	0	-	-
SKU_1	$Retailer_1$	0.79	0.9709	0.0009	0.87	0.9745	0.0045
SKU_1	$Retailer_2$	0.67	0.9703	0.0003	0.74	0.9738	0.0038
SKU_1	$Retailer_3$	1.14	0.9707	0.0007	1.21	0.9746	0.0046

4.1 Larger-size examples

To assess the performance of the proposed approaches on larger cases, the previous example is expanded to include 4 finished goods and 6 raw materials (10 materials in total), as well as 20 finished goods and 120 raw materials (140 materials in total). The demand distributions are assumed to be normally distributed, and the target fill rate for all materials is set at 97%. The results of these tests are summarized in Table 7.

For the first extension with 10 materials, approach (a) achieves a total cost of $8,472,820, with a computational time of 0.3 seconds and an average error (difference between the obtained fill rate and the target) of 0.0054. On the other hand, approach (b) results in a total cost of $7,326,065, the computational time is 0.9 seconds, but with an average error of -0.0495. In this case, although the total cost using approach (b) is lower, on average, the safety factors proposed are not enough to meet the desired customer service level.

For the larger case with 140 materials, approach (a) finds the optimal solution within 1.27 seconds, with an average error of 0.0030. In contrast, approach (b) is not able to find the optimal solution in 1000 seconds, and obtains a solution with a 19% GAP. In addition, the total cost of approach (a) increases, but the average error presented in Table 7 indicates that approach (b) underestimates the service levels. This clearly shows the advantage of the model proposed in this work to significantly increase the computational efficiency while improving the solution accuracy.

In summary, the results show that approach (a) outperforms approach (b) in terms of computational efficiency and solution accuracy, even when dealing with larger cases.

Table 7: Extended illustrative example results for 10 materials

	Extension for 10 materials		Extension for 140 materials	
	(a) MIQCP with piecewise	(b) MIQCP with regression	(a) MIQCP with piecewise	(b) MIQCP with regression
CPU time (seconds)	0.25	0.90	1.27	1,000
Optimality GAP	0%	0%	0%	19%
Constraints	572	298	5793	3243
Binary variables	215	89	1896	720
Integer variables	257	131	2288	1112
Continuous variables	295	148	2745	1373
Total Cost	$ 8,472,820	$ 7,326,065	$ 22,743,468	$ 20,212,859
Avg. error for fill rate	0.0054	-0.0495	0.0030	-0.0210

Finally, to illustrate the application of the proposed solution strategy in a large and complex supply chain, a computational experiment is conducted based on a real-world case study involving 1400 products and 18 locations, as depicted in Figure 8. The numbers assigned to each node indicate the number of products that can be stored at that location, as different products follow different routes. In total, there are 7371 product-location combinations to consider for allocation decisions. Nodes with people icons represent locations that receive external demand. The demand for each product is stochastic, independent, and identically distributed, while lead times follow a normal distribution.

Figure 9 displays the computational time required to solve different instances of the problem, ranging from 100 to 1400 products. Two targets were considered: the orange line corresponds to the problem with a target customer service level (CSL) of 98% for all products, while the dashed blue line represents the results when a fixed fill rate (FR) target of 98% is set. It can be observed that real-world cases can be solved to optimality within a few seconds of CPU time. The computational burden increases when the fill rate is used as the target measure, as more constraints and variables become active. Table 8 provides detailed information on the model sizes and objective values for selected instances. When the target service level is the fill rate, there is a noticeable reduction in the total holding cost. This can be attributed to the fact that even though orders may not be completely fulfilled, partial fulfilment is taken into account when calculating the overall service level. In situations where the CV is relatively low and the demand is assumed to be normally distributed, the fill rate tends to be higher compared to the CSL. While occasional stockouts may still occur, they are typically of smaller magnitude (Chopra & Meindl, 2013; Peter L. King, 2011). For instance, in the illustrative example, the safety factor required to achieve a 97% fill rate is $K_{Retailer1,SKU1} = 0.79$ (Table 6). On the other hand, if the target were a 97% CSL, the safety factor would be $\Phi^{-1}(0.97) = 1.88$. This would result in a higher safety stock level and consequently higher costs.

It is worth mentioning that using BARON to solve the smallest instance for both the MINLP and MIQCP formulations failed to yield a feasible solution within 1000 seconds. This demonstrates that the proposed MIQCP reformulation solving with Gurobi yields order of magnitude improvements in computational efficiency.

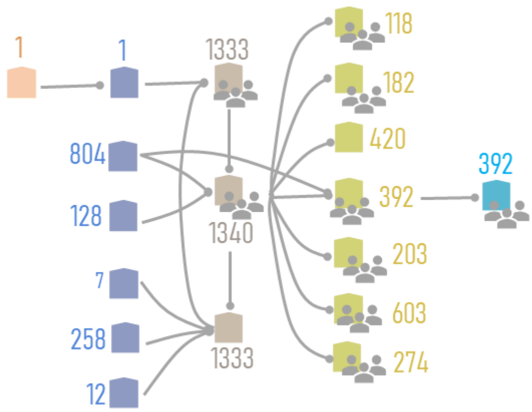

Figure 8. Case-study network

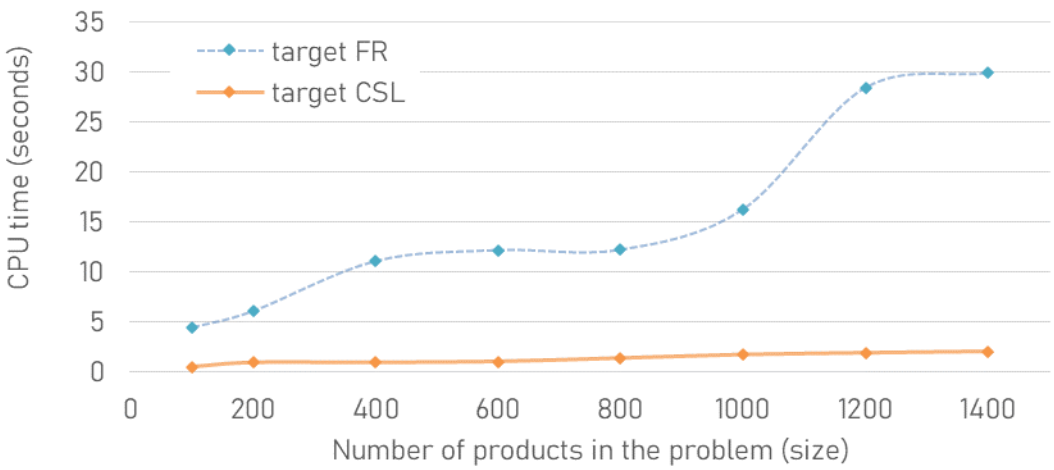

Figure 9. Instance size vs. computational time

Table 8. Model sizes and optimal solutions

	Items	Constraints	Binary variables	Integer variables	Continuous variables	Total holding cost ($)
Target 98% CSL	**200**	14 261	5 525	7 831	5 766	$ 2 025 853
	400	27 210	10 991	15 155	10 991	$ 2 936 325
	600	39 816	15 504	21 930	16 066	$ 4 526 933
	800	55 018	21 194	30 080	22 216	$ 8 626 338
	1000	66 471	25 412	36 146	26 836	$ 11 693 035
	1200	81 104	31 132	44 230	32 746	$ 16 412 021
	1400	86 381	34 783	49 525	36 856	$ 22 281 963
Target 98% FR	**200**	27 341	10 757	13 063	13 895	$ 1 009 487
	400	48 870	19 423	23 819	24 741	$ 1 548 778
	600	73 386	28 932	35 358	37 183	$ 2 246 342
	800	104 023	40 796	49 682	52 795	$ 5 129 822
	1000	124 596	48 662	59 396	63 203	$ 7 533 684
	1200	152 324	59 620	72 718	77 279	$ 11 155 372
	1400	171 249	66 781	81 523	86 891	$ 15 268 330

5. CONCLUSIONS

In this work, we have presented an optimization model based on the guaranteed-service approach to determine safety stocks in multi-echelon supply chains. The objective of this work is to provide a model that is both accurate and efficient in determining the appropriate safety stock levels. We build the on the model proposed in Achkar et al. (2023), by extending it in several important directions.

The initial MINLP problem has a concave objective function, and the loss function of the normal distribution in the fill rate constraint. The proposed solution strategy combines an MIQCP reformulation with a piecewise linear approximation to enhance computational efficiency. The results demonstrate that the reformulated model outperforms the original MINLP formulation and previous approaches based on polynomial regression. It is the first model to combine an MIQCP reformulation with piecewise linear functions to achieve improved computational efficiency in the guaranteed-service model. Additionally, the piecewise-linear approximation improves the precision of the safety stock setting by providing a more accurate estimation of the fill rate. This is of great relevance, since slight variations in service levels can have significant impacts on inventory costs.

Furthermore, this work extends the GSM to handle non-normally distributed demands with high variability. The extension to non-normal demand distributions with higher variability allows for the inclusion of a wider range of products, which is particularly relevant in pharmaceutical supply chains. It is demonstrated that safety stocks based on safety factors derived from the standard normal distribution may underestimate the customer service level, and the proposed adjustment including the gamma distribution provides a more accurate estimation of safety stock levels for highly variable demands.

Finally, the approach to handle stochastic lead times is modified from the one proposed in Achkar et al. (2023). The standard deviation of the net lead time in the present paper depends on the value of the net lead time itself, and a discrete function is proposed to determine the value of the standard deviation σ_{NLTjp}. By exploiting the concavity property of the objective function, a novel discrete function is proposed to obtain the value of σ_{NLTjp}.

Overall, the proposed model accurately represents real-world systems and enables the determination of optimal safety stock levels to meet target customer service levels while minimizing inventory costs. Illustrative examples and real-world cases from the pharmaceutical industry have been presented to demonstrate the applicability of the proposed formulation. The model can efficiently solve large-scale problems with over 7300 product-location combinations. Future work will focus on non-normal demand with fill rate targets, and responsive characteristics to address supply chain disruptions.

6. NOMENCLATURE

6.1 Sets

J	Set of locations
J^0	Subset of starting locations in the network
P_j	Subsets of products that can be stored at location *j*
A	Subset of routes segments (from node *i* to node *j*) enabled for material *p*
F	Set of locations that have materials with an active fill rate as a target
S_1	Set of steps related to discrete function for NLT variance
S_2	Set of segments related to piecewise linear approximation for fill rate constraint

6.2 Parameters

μ_{jp}	Mean of the total demand of material *p* in location *j*
σ_{jp}	Standard deviation of the total demand of material *p* in location *j*
lt_{jp}	Lead time/order processing time of material *p* in location *j*
$\sigma_{LT\,jp}$	Standard deviation of the lead time/order processing time of material *p* in location *j*
$\sigma_{NLT\,jp}$	Standard deviation of the NLT of material *p* in location *j*
h_{jp}	Holding cost of material *p* in location *j*
si_{jp}^0	Inbound service time for the source nodes in the network
ϕ_{pq}	Amount of material *p* required to produce material a unit of material *q*
$maxS_{jp}$	Maximum service time accepted for material *p* in location *j*
r_{jp}	Stock review period for material *p* in location *j*
moq_{jp}	Minimum Order Quantity of material *p* that location *j* must place

Q_{jp}	Replenishment order size of material *p* at location *j*
fr_{jp}	Fill rate level of material *p* at location *j*
k_{jp}	Safety factor associated with CSL of material *p* at location *j*
c_{jps}	Value assigned to NLT variance of material *p* at location *j* on step *s*
d_{jps}	Value of NLT represented in step *s* for material *p* at location *j*
a_{jps}	Slope of segment *s* associated to piecewise linear approximation for material *p* at location *j*
b_{jps}	Origin of segment *s* associated to piecewise linear approximation for material *p* at location *j*
lb_{jps}	Lower bound of NLT associated to piecewise approximation for segment *s*, material *p* at location *j*
ub_{jps}	Upper bound of NLT associated to piecewise approximation for segment *s*, material *p* at location *j*
$k_{N\,jp}$	Safety factor obtained from the standard normal cdf for material *p* at location *j*
$k_{G\,jp}$	Safety factor obtained from the standardized gamma cdf for material *p* at location *j*

6.3 Positive Variables

S_{jp}	Guaranteed service time within which location *j* will attend demand of material *p*
SI_{jp}	Inbound Guaranteed service time at location *j* of material *p*
NLT_{jp}	Net Lead time of material *p* at node *j*
U_{jp}	Variable used for quadratic reformulation on dependent demand net lead time formula
K_{jp}	Variable used to replace *k* input factor when the fill rate is introduced to determine safety stocks
X_{jp}	Variable defined to represent NLT variance of material *p* at node *j*
Y_{jp}	Variable that represents the value of part of the fill rate constraint resulting from piecewise approximation for material *p* at node *j*
$\widetilde{K}_{jps}$	Auxiliary variable that represents the value the safety factor for piecewise approximation for material *p* at node *j*, segment *s*

6.4 Binary Variables

$V_{1\,jps}$	Variable associated with discrete to define NLT variance
$V_{2\,jps}$	Variable associated with discrete to define NLT variance
W_{jps}	Variable associated with piecewise linear approximation

7. ACKNOWLEDGMENTS

The authors gratefully acknowledge the financial support from Johnson and Johnson and the Center for Advanced Process Decision-making (CAPD) from Carnegie Mellon University.

8. REFERENCES

Achkar, V. G., Brunaud, B. B., Perez, H. D., Musa, R., Mendez, C. A., & Grossmann, I. E. (2023). *Extensions to the Guaranteed Service Model for Industrial Applications of Multi-Echelon Inventory Optimization*. https://doi.org/10.48550/arXiv.2306.10961

Axsäter, S. (2006). A simple procedure for determining order quantities under a fill rate constraint and normally distributed lead-time demand. *European Journal of Operational Research*, *174*(1), 480–491. https://doi.org/10.1016/j.ejor.2005.01.037

Beutel, A. L., & Minner, S. (2012). Safety stock planning under causal demand forecasting. *International Journal of Production Economics*, *140*(2), 637–645. https://doi.org/10.1016/j.ijpe.2011.04.017

Chopra, S., & Meindl, P. (2013). Supply Chain Management. In *Pearson*. http://www.doiserbia.nb.rs/Article.aspx?ID=0013-32640670067A

Chu, Y., You, F., & Wassick, J. M. (2014). Hybrid method integrating agent-based modeling and heuristic tree search for scheduling of complex batch processes. *Computers & Chemical Engineering*, *60*, 277–296. https://doi.org/10.1016/j.compchemeng.2013.09.004

de Kok, T. (2019). Inventory Management: Modelling Real-life Supply Chains and Empirical Validity. *Foundations and Trends in Technology, Information and Operations Management*, *12*(4), 349–433. http://www.nowpublishers.com/article/Details/TOM-057

Drake, A. W. (1967). *Fundamentals of applied probability theory*. New York, McGraw-Hill.

Eruguz, A. S., Sahin, E., Jemai, Z., & Dallery, Y. (2016). A comprehensive survey of guaranteed-service models for multi-echelon inventory optimization. *International Journal of Production Economics*, *172*, 110–125. https://doi.org/10.1016/j.ijpe.2015.11.017

Graves, S. C., & Willems, S. P. (2000). Optimizing Strategic Safety Stock Placement in Supply Chains. *Manufacturing & Service Operations Management*, *2*(1), 68–83. https://doi.org/10.1287/msom.2.1.68.23267

Grossmann, I. E. (2005). Enterprise-wide optimization: A new frontier in process systems engineering. *AIChE Journal*, *51*(7), 1846–1857. https://doi.org/10.1002/aic.10617

Inderfurth, K. (1993). Valuation of Leadtime Reduction in Multi-Stage Production Systems. In *Operations Research in Production Planning and Control*. https://doi.org/10.1007/978-3-642-78063-9

Jung, J. Y., Blau, G., Pekny, J. F., Reklaitis, G. V., & Eversdyk, D. (2004a). A simulation based optimization approach to supply chain management under demand uncertainty. *Computers and*

Chemical Engineering, *28*(10), 2087–2106. https://doi.org/10.1016/j.compchemeng.2004.06.006

Jung, J. Y., Blau, G., Pekny, J. F., Reklaitis, G. V., & Eversdyk, D. (2004b). A simulation based optimization approach to supply chain management under demand uncertainty. *Computers & Chemical Engineering*, *28*(10), 2087–2106. https://doi.org/10.1016/j.compchemeng.2004.06.006

Minner, S. (1998). *Strategic Safety Stocks in Supply Chains*.

Mirzaee, A. (2017). *Alternative Methods for Calculating*. University of Missouri-Columbia.

Moors, J. J. A., & Strijbosch, L. W. G. (2002). Exact fill rates for (R, s, S) inventory control with gamma distributed demand. *Journal of the Operational Research Society*, *53*(11), 1268–1274. https://doi.org/10.1057/palgrave.jors.2601441

Papageorgiou, L. G. (2009). Supply chain optimisation for the process industries: Advances and opportunities. *Computers & Chemical Engineering*, *33*(12), 1931–1938. https://doi.org/10.1016/j.compchemeng.2009.06.014

Peter L. King. (2011). Understanding safety stock and mastering its equations. *APICS Magazine*. https://web.mit.edu/2.810/www/files/readings/King_SafetyStock.pdf

Rodriguez, M. A., & Vecchietti, A. (2010). Inventory and delivery optimization under seasonal demand in the supply chain. *Computers & Chemical Engineering*, *34*(10), 1705–1718. https://doi.org/10.1016/j.compchemeng.2010.02.009

Sahinidis, N. V. (1996). BARON: A general purpose global optimization software package. *Journal of Global Optimization*, *8*(2), 201–205. https://doi.org/10.1007/BF00138693

Silver, E. A., Peterson, R., & Pyke, D. F. (1998). *Inventory management and production planning and scheduling*. Wiley.

Simpson, K. F. . (1958). In-Process Inventories. *INFORMS*, *6*(6), 863–873.

Snyder, L. V., & Shen, Z. M. (2019). Fundamentals of Supply Chain Theory. In *Fundamentals of Supply Chain Theory*. Wiley. https://doi.org/10.1002/9781119584445

Soares, A. P. (2013). Inventory Control. In *Operations Research Management Science* (Vol. 53, Issue 9). Springer.

You, F., & Grossmann, I. E. (2008a). Mixed-Integer Nonlinear Programming Models and Algorithms for Large-Scale Supply Chain Design with Stochastic Inventory Management. *Industrial & Engineering Chemistry Research*, *47*(20), 7802–7817. https://doi.org/10.1021/ie800257x

You, F., & Grossmann, I. E. (2008b). Design of responsive supply chains under demand uncertainty. *Computers & Chemical Engineering*, *32*(12), 3090–3111. https://doi.org/10.1016/j.compchemeng.2008.05.004

You, F., & Grossmann, I. E. (2009). Integrated multi-echelon supply chain design with inventories under uncertainty: MINLP models, computational strategies. *AIChE Journal*, *59*(4), 420–440. https://doi.org/10.1002/aic.12010