# Service Parts Inventory Control with Lateral Transshipment that Takes Time

Guangyuan Yang
Econometric and Tinbergen Institute, Erasmus School of Economics, Erasmus University Rotterdam, 3062 PA, Rotterdam, gyang@ese.eur.nl

Rommert Dekker
Econometric Institute, Erasmus School of Economics, Erasmus University Rotterdam, 3062 PA, Rotterdam, NL rdekker@ese.eur.nl

EI 2010-02

In equipment-intensive industries such as truck manufacturing, electronics manufacturing, photo copiers, and airliners, service parts are often slow moving items for which, in some cases, the transshipment time is not negligible. However, this aspect is hardly considered in the existing spare parts literature. We assess the effect of non-negligible lateral transshipment time on various aspects of spare parts inventory control. Furthermore, we introduce customer-oriented service levels by taking the uncommitted pipeline stocks into account. A case study in the dredging industry shows that lateral transshipment may lead to lower system performance, which supports the results from some recent studies. Furthermore, we find that considerable savings can be obtained when we include the uncommitted pipeline stocks in both base stock allocation and lateral transshipment decisions.

*Key words*: Inventory Control; Lateral Transshipment; METRIC; Customer-oriented Service Level

*History*: This paper was last revised on March 8, 2010.

## 1. Introduction

Research of service parts inventory control with lateral transshipment has been motivated by needs from various industries, including equipment-intensive industries such as truck manufacturing, electronics manufacturing, photo copiers, and airliners. Facing stochastic demand of critical spare parts, a multi-location inventory control system often allows movement of stock between locations at the same echelon level or even across different levels in order to fulfill customers' demand in time. Many of these critical spare parts are slow moving items for which, in some cases, air transport is

impossible or prohibitively expensive. For example, in dredging industry, critical spare parts usually weigh more than 4,000 kg which are way too costly to transport by air. The lateral transshipment time for these items can be more than 3 weeks, which is not negligible compared to lead times (around 7 weeks). Moreover, there are considerable amount of stocks in the pipeline between a depot and local bases due to the slow transportation. For instance, the average pipeline stocks can be as many as half of the average stocks on hand. As a result, the timing of transshipment and replenishment becomes an important factor in decision making (one would wait for a pipeline stock instead of asking for lateral transshipment if it takes less time for that pipeline stock being delivered to customers). To the best of our knowledge, this aspect is hardly considered in the existing spare parts literature.

Good customer-oriented performance measures are also lacking in the literature. The standard spare parts service levels in most literature, such as inventory availability, are supplier-oriented; whereas customers only observe deliveries with no delays and deliveries within a certain time span in case of delays. Some studies try to introduce more customer-oriented service levels by distinguishing the availability from different sources with different response times. However, this approach still emphasise operational processes of service suppliers, because it ignores that the pipeline stocks may arrive and be delivered to customers sooner than other emergency shipments. Particularly, customers do not care where the spare part is from as long as it is delivered in time.

Consequently, in this paper, we contribute to the existing literature by accessing the effect of non-negligible lateral transshipment time on various aspects of spare parts inventory control. Furthermore, we take pipeline stocks into account when we introduce the customer-oriented service levels here.

In the following section, a literature review is presented. In Section 3, in order to define the customer-oriented service levels and derive the expression of average inventory costs, a two-echelon inventory system with lateral transshipment, under central control, continuous review at all stock points and external stochastic demand at local service centers is formulated and analyzed firstly. Based on this analysis, in section 4, we minimize the average inventory cost subject to the service

level constraints. In section 5, we apply our models in a case study for a global market leader in the dredging industry. In the last part, we draw our conclusion.

## 2. Literature Review

In the past decades, a considerable amount of research has been dedicated to service parts inventory control with lateral transshipment. Most models in the literature (Lee, 1987; Axsäter, 1990; Alfredsson and Verrijdt, 1999; Diks and De Kok, 1996; Banerjee et al., 2003; Burton and Banerjee, 2005) assume that transshipment times are negligible. Nevertheless, different opinions on the effect of lateral transshipment on inventory system performance emerge from these studies.

Lee (1987) considers lateral transshipment in such a model. Demand occurs when there is a failure in a critical part, which is assumed to be a Poisson process. Failed parts are replaced by stocks on hand or lateral transshipment in case of stockout. He assumes that all bases are identical in a two echelon inventory system with continuous review base stock policy, such that, regardless of whether there are lateral transshipment among the bases, the demand at each base always follows Poisson processes with the same mean. Then, he evaluates approximately the portion of demand met by stock on hand and the portion of demand met by lateral transshipment based on three selecting rules for the source base: random selection, maximum stock on hand, and smallest number of outstanding orders. And he finds no significant difference in the performance of the three rules when all bases are identical. Finally, using this method, he concludes that lateral transshipment leads to substantial cost savings because less base stocks are needed at the bases. Axsäter (1990) relaxes the restrictive assumption of identical bases. He presents improved methods for approximating service levels by identifying the demand at local bases as the sum of the regular demand and overflow demand from other bases due to lateral transshipment. Alfredsson and Verrijdt (1999) extend Axsäter's model by allowing emergency shipment from a central warehouse and emergency shipment from a manufacturing facility such that no demand is backordered. They find that using both lateral transshipment flexibility and direct shipment flexibility results in significant cost reductions compared to using no supply flexibility at all.

However, some recent studies suggest that lateral transshipment could be beneficial only under certain conditions or even lead to worse inventory system performance. Diks and De Kok (1996) shows that lateral transshipment becomes advantageous only under conditions of "many retailers, high service levels, mean demands per period of the same size, and the central depot as close as possible to the supplier", since rebalancing the total net stock every review period incur additional costs. Banerjee et al. (2003) and Burton & Banerjee (2005) examine the effects of lateral transshipment under both policies based on availability and on inventory balancing (equalization) through a series of simulation experiments. They find neither of the two lateral transshipment policies are superior to a policy without lateral transshipment, since the additional delivery costs resulting from transshipment outweigh the benefits of avoiding retail level shortages.

On the other hand, some of the relatively few recent studies (Wong et al., 2005; Kutanoglu and Mahajan, 2009; Tagaras and Vlachos, 2002) consider the non-negligible lateral transshipment time in their models. Wong et al. (2005) studies repairable spare parts pooling in a multi-hub system for the airline industry. They include delayed lateral transshipment in their system performance approximation and optimal spare parts stocking level determination. Regarding the choice of the source for lateral transshipment, they use the closest neighbor rule as it is more acceptable in practice than the random choosing rule used by Axsäter (1990) and Alfredsson and Verrijdt (1999). They find that significant cost savings can be achieved by pooling the spare parts inventories via lateral transshipment.

Kutanoglu and Mahajan (2009) consider the spare parts arriving late due to non-negligible lateral transshipment time using time based service levels (the percentage of demand satisfied within a certain time span). They evaluate the service levels based on the availability from different sources (central and local warehouses) with different response times. However, these service levels ignore the pipeline stocks that may arrive and be delivered to customers sooner than lateral transshipment from other local warehouses or emergency shipment from the central warehouse, as a result, they are not true customer-oriented service levels. The authors develop a method enumerating over all possible stock profiles (stock levels across all local warehouses) with bounds to find the optimum

stock levels, which minimize the total cost subject to time based service level constraints. Then also conclude that lateral transshipment improves inventory system performance. Nevertheless, Tagaras and Vlachos (2002) points out lateral transshipment with non-negligible transshipment time may result in deterioration of the total group performance because the items transshipped are unavailable when they are in transit and cannot be used to satisfy demand at any locations. Unfortunately, they mainly focus on the sensitivity of the policy based on the variability within the demand distribution.

In this paper, we assess the effect of non-negligible lateral transshipment time on various aspects of spare parts inventory control. At first, we define the customer-oriented service levels based on the time spare parts are delivered to customers. Moreover, we take the pipeline stocks into account in our analysis of a two echelon inventory system with lateral transshipment. Due to the non-negligible lateral transshipment time, we also consider the timing of transshipment and the timing of replenishment in the lateral transshipment rule.

## 3. Model

In this section, we firstly describe our model in details. Then we analyze the model using the METRIC approximation (Sherbrooke, 1968). Furthermore, we validate our model by comparing our approximations with results from event-driven simulations.

### 3.1. Model Description

We consider a two-echelon system with the following properties:

- The system has a supplier that has infinite production capacity.
- The system has a central depot with finite base stock $S_0$, which is supplied from the supplier on a one-for-one basis over a constant lead time $L_0$. It also has multiple local service centers with base stocks $S_i$, which are replenished on a one-for-one, first-come-first served basis from the central depot, over constant lead times $L_i$. This base stock inventory control system is very common in practice for service parts, because of the high price and low demand characteristics of many of these items. The lead times are assumed constant because their distributions do not affect the system

performance (Alfredsson and Verrijdt, 1999), and moreover, companies prefer constant lead times models.

- Demand ($D_i$) occurs at local service center $i$, and is Poisson distributed with known average demand rate $\lambda_i$, independent of the demand at other service centers. If there are stocks available at the service center, the demand is fulfilled instantaneously with direct service. At the same time, a replenishment order is sent to the central depot. If there are no stocks available at the service center, and customers accept waiting for up to $T$ time units, the demand may be fulfilled either by a pipeline stock or a lateral transshipment from other neighboring service centers given the item can be delivered to the customer within $T$ time units. As lateral transshipment is usually more costly, we give priority to the uncommitted pipeline stock (the one not committed to other waiting customers). If the customer's waiting time for the uncommitted pipeline stock $W_i$ is no more than $T$, then the demand will be fulfilled by this pipeline stock; otherwise, we consider lateral transshipment as an option. For lateral transshipment at a local service center, the neighboring service centers are prioritized such that the one with shortest lateral transshipment time is considered first. The service center who delivers the item orders a new item from the depot. If none of the local service centers is able to satisfy the demand within the time span, the service center where the demand first occurred issues a backorder to the central depot and a substantial penalty cost is incurred.

More specifically, if stock out occurs at service center $i$, we only consider lateral transshipment when $W_i > T$. Let $LT_{ij}$ be the lateral transshipment time between service center $i$ and $j$. We assume $LT_{ij} = LT_{ji}$, and $LT_{ij} \neq LT_{ik}$ for all $i$, $j$ and $k$. Regarding the potential source bases of the lateral transshipment to service center $i$, we only consider the service centers in the set $B_i = \{k | LT_{ik} \leq T\}$, whose transshipment time $LT_{ik}$ is less than $T$.

• If $0 < W_i \leq T$, issue a back order and wait for an uncommitted pipeline stock.

• If $W_i > T$,

◇ If $IL_j^T > 0$ where $j \in B_i$, choose base $j$ such $LT_{ij}$ is the shortest; the source base issues a replenishment order from the central depot.

◇ Otherwise, issue a back order and wait for a replenishment.

### 3.2. Model Analysis

Since local service centers are replenished on a one-for-one, first-come-first served basis from the central depot, the demand at the central depot is consequently a superposition of these demands, $D_0 = \sum_i D_i$. Because these demands are independently Poisson distributed, $D_0$ is also a Poisson process regardless of whether there are lateral transshipment among the local service centers (Lee, 1987). Let $\lambda_0$ be the demand rate at the central depot, then $\lambda_0 = \sum_i \lambda_i$. Following the standard approach in Axsäter (2006), we have the distribution of inventory level at the central depot ($IL_0$):

$$P(IL_0 = j) = G(j, S_0, \lambda_0, L_0) \text{ where } G(j, S, \lambda, L) = \frac{(\lambda L)^{S-j}}{(S-j)!} e^{-\lambda L} \tag{1}$$

Given the distribution of the inventory level at the central depot, we can find its average inventory on hand ($EOH_0$), the average physical stock on hand which measures the capital tied up in the inventories, by

$$EOH_0 = EOH(S_0, \lambda_0, L_0) \text{ where } EOH(S, \lambda, L) = \sum_{j=1}^{S} j \frac{(\lambda L)^{S-j}}{(S-j)!} e^{-\lambda L} \tag{2}$$

The average number of back orders is the average units of inventories that have been requested but not yet delivered and can be calculated by

$$EBO_0 = EBO(S_0, \lambda_0, L_0) \text{ where } EBO(S, \lambda, L) = \lambda L - S + \sum_{j=1}^{S} j \frac{(\lambda L)^{S-j}}{(S-j)!} e^{-\lambda L} \tag{3}$$

Since $EBO_0$ can be interpreted as the average queue length and $\lambda_0$ can be interpreted as the average arrival rate to the queue, according to the well known *Little's formula* from queuing theory, the average delay at the central depot is $E(W_0) = EBO_0/\lambda_0$. This average delay is the same for all local service centers because of the Poisson demand at the central depot and the first-come-first-served assumption. Hence, the adjusted average lead-time for local service centers is the transportation time plus the average delay at the central depot: $\overline{L_i} = L_i + E(W_0)$.

At local service centers, we use the METRIC approximation (Sherbrooke, 1968), replacing the stochastic lead-time by its average, to evaluate the system performance. This approach is quite widely used as a good approximation, for instance, Axsäter (1990) uses METRIC when he approximates the fill rate and the proportion of demand met by lateral transshipment, and finds his

approximation is "very close to the results obtained by simulation"for both low and high proportions of demand met by lateral transshipment. Given this METRIC assumption, we have the distribution of the inventory level at service center $i$, $P(IL_i = j) = G(j, S_i, \lambda_i, \overline{L_i})$.

With the distribution of inventory level, we can evaluate the system performance by calculating average inventory on hand $EOH_i = EOH(S_i, \lambda_i, \overline{L_i})$, average number of back orders $EBO_i = EBO(S_i, \lambda_i, \overline{L_i})$.

Using METRIC, we can calculate the average pipeline stock, the average units of inventories that are in the delivering process from the central depot to service centers by

$$EPS_i = E(D_i(L_i)) = \lambda_i L_i \tag{4}$$

We define customer-oriented service levels as probability of direct service ($P^D$) and probability of service within $T$ time units ($P^T$), where $T$ is a response time to customers or customers' acceptable waiting time.

If there is no lateral transshipment or lateral transshipment time is negligible, then there is no time lag between a base fulfilling a request and a customer receiving the item. As a result, these two service levels equal the aggregated instantaneous fill rates and aggregated fill rates within $T$ time units.

According to Axsäter (2006), the local instantaneous fill rate $FR_i$ can be calculated by

$$FR_i = FR(S_i, \lambda_i, \overline{L_i}) \text{ where } FR(S, \lambda, L) = P(IL > 0) = \sum_{j=0}^{S-1} \frac{(\lambda L)^j}{j!} e^{-\lambda L} \tag{5}$$

The probability of direct service for all customers equal to the aggregated instantaneous fill rates at all service centers weighted by their demand rates.

$$P^D = \sum_i FR_i \lambda_i / \sum_i \lambda_i \tag{6}$$

When demand occurs at a local service center $i$ which has a non-positive inventory level, the service center has to decide whether to fullfill the demand by an uncommitted pipeline stock, based on the condition whether the customer's waiting time for the uncommitted pipeline stock is no

more than $T$. This condition depends on the timing of previous orders issued by the local service center to the central depot because of the constant lead time.

Due to the Poisson demand process, and the one-for-one, first-come-first served assumptions, in any small time interval, the arrival of a replenishment order is uniform distributed given the number of Poisson arrivals (Tijms, 2003). This leads to our theorem and corollary (for proofs see Appendix A and B), which can be applied later in the multi-echelon system.

THEOREM 1. *In a single echelon system, the probability that a customer at service center $i$ has to wait and waits no more than $T$ is*

$$P(0<W_i\leq T)=1-FR_i-\sum_{k=1}^{\infty}\sum_{j=0}^{k-1}\binom{S_i+k-1}{j}(\frac{T}{L_i})^j(1-\frac{T}{L_i})^{S_i+k-1-j}e^{-\lambda_i L_i}\frac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!} \tag{7}$$

*and the expected waiting time for the customer is*

$$E(W_i)=\sum_{k=1}^{\infty}\frac{kL_i}{S_i+k}e^{-\lambda_i L_i}\frac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!} \tag{8}$$

The fill rate within time span $T$, $FR_i^T$, can be calculated directly by summing up the local instantaneous fill rate and the probability that the demand is fulfilled by an uncommitted pipeline stock within the time span, which leads us to our Corollary.

Corollary 1. *In a single echelon system, the fill rate within the time span $T$ at service center $i$ is*

$$\begin{aligned} FR_i^T &= FR_i+P(0<W_i\leq T)\\ &= 1-\textstyle\sum_{k=1}^{\infty}\sum_{j=0}^{k-1}\binom{S_i+k-1}{j}(\frac{T}{L_i})^j(1-\frac{T}{L_i})^{S_i+k-1-j}e^{-\lambda_i L_i}\frac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!}\\ &= \textstyle\sum_{t=0}^{S_i-1}\frac{\lambda_i^t}{t!}(L_i-T)^t e^{-\lambda_i(L_i-T)}\\ &= FR(S_i,\lambda_i,L_i-T) \end{aligned} \tag{9}$$

The intuition is that if customers accept waiting for $T$ time units, then more demand can be fulfilled because pipeline stocks may arrive within $T$. This is equivalent to reducing the lead time from $L_i$ to $L_i-T$. A similar result has also been proved in Kruse (1981).

We apply Corollary 1 in the two-echelon system using the METRIC (Sherbrooke, 1968), and obtain the local fill rate within the time span $T$, $FR_i^T=FR(S_i,\lambda_i,\overline{L_i}-T)$. In case of no lateral transshipment, the probability of service within $T$ time units for all customers equal to the aggregated fill rates within $T$ at all service centers weighted by their demand rates.

$$P^T = \sum_i FR_i^T \lambda_i / \sum_i \lambda_i. \tag{10}$$

When we include lateral transshipment, since the transshipment time is not negligible, the probability of direct service does not equal to the aggregated instantaneous fill rate. Because part of the stocks on hand may be used to help other service centers with some delays, in other words, a service center not only fulfills the demand from its own customers but may also need to fulfill lateral transshipment requests from other service centers. Nevertheless, the probability of service within $T$ is still the same as aggregate fill rates within $T$, because lateral transshipment is only allowed when the transshipment time is less than $T$.

Moreover, the demand rates at the local bases change due to lateral transshipment. In order to approximate the new demand rates, we assume that the overflow demand streams, the demand not fulfilled at a service center but satisfied by the lateral transshipment from other service centers, are Poisson distributed and are independent of each other among the local service centers. We obtain the new demand rates by updating the overflow demand streams in each iteration until they converge. Alfredsson and Verrijdt (1999) and Reijnen et al. (2009) uses similar approximation and they find "excellent results"for identical service centers, and for nonidentical service centers, it "still performs very well".

The overflow demand of service center $i$ fulfilled by lateral transshipment from service center $j$ can be calculated by

$$OF_{ij} = \begin{cases} FR_j \lambda_i (1 - FR_i^T) & \text{if } LT_{ij} = min\{LT_{ik}, k \in B_i\}; \\ FR_j \lambda_i (1 - FR_i^T) \prod_{k \in B_{ij}} (1 - FR_k) & j \in B_i,\ k \in B_{ij}, \text{ where } B_{ij} = \{k | k \in B_i, LT_{ik} < LT_{ij}\}; \\ 0 & \text{otherwise.} \end{cases} \tag{11}$$

As a result, the adjusted demand rate for service center $i$ is:

$$\lambda_i^{ad} = \lambda_i + \sum_{j \in B_i} (OF_{ji} - OF_{ij}).$$

Obviously, these adjusted demand rates will affect the fill rates, which in turn will affect the overflow demands, leading to new adjusted demand rates. In our algorithm, in each iteration, we obtain a local fill rate within time span $T$, $FR_i^T$, for each service center. We then calculate the

difference ($\Delta_i$) between two consecutive $FR_i^T$s for all $i$. We continue the calculations until $\Delta_i < \varepsilon$ for all $i$, where $\varepsilon$ is a small positive number, say 0.0001.

Let the converged demand rate be $\lambda_i^*$ and the converged overflow demand be $OF_{ij}^*$, then the converged average inventory on hand is $EOH_i^* = EOH(S_i, \lambda_i^*, \overline{L_i})$; the converged average pipeline stock is $EPS_i^* = \lambda_i^* L_i$; the converged average number of back orders is $EBO_i^* = EBO(S_i, \lambda_i^*, \overline{L_i})$; the converged instantaneous fill rate is $FR_i^* = FR(S_i, \lambda_i^*, \overline{L_i})$; and the converged fill rate within $T$ time units is $FR_i^{T*} = FR(S_i, \lambda_i^*, \overline{L_i} - T)$.

In case of lateral transshipment with non-negligible transshipment time, the probability of direct service for all customers equals the fraction of requests fulfilled instantaneously minus the fraction of requests fulfilled by lateral transshipment which has time lags due to the transshipment time, weighted by the demand rates.

$$P_{lt}^D = \sum_i (FR_i^* \lambda_i^* - \sum_{j \in B_i} OF_{ji}^*) / \sum_i \lambda_i^* \tag{12}$$

Comparing Equation (6) and (12), we find that the probability of direct service may decrease because of lateral transshipment. In case of only two identical neighboring service centers, Equation (12) leads to a direct service level $(FR_1\lambda_1 + FR_2\lambda_2 - OF_{12} - OF_{21})/(FR_1\lambda_1 + FR_2\lambda_2)$, which is less than the service level in case without lateral transshipment. This effects does not exist if lateral transshipment time is assumed negligible.

The probability of service within time span $T$ for all customers in case of lateral transshipment equals the fraction of requests fulfilled within $T$ weighted by the demand rates.

$$P_{lt}^T = \sum_i FR_i^{T*} \lambda_i^* / \sum_i \lambda_i^* \tag{13}$$

In order to balance the trade off between service levels and cost, one has to calculate the average cost. Given the target service levels are met, the cost consists of holding cost for the stocks on hand at the central depot and all local service centers, carrying cost for the stocks in the pipeline between the central depot and all local service centers, and lateral transshipment cost for the stocks transshipped between all local service centers. In general, the holding cost per item per time unit

$hc$ includes storage cost, opportunity cost of capital tied up in stocks, insurance cost and costs associated with risk of deterioration or obsolescence; the pipeline stock cost per item per time unit $pc$ includes the same cost elements as holding cost except for storage cost; the lateral transshipment cost per item between service center $i$ and $j$ $lc_{ij}$ includes transportation cost, insurance cost and costs associated with risk of deterioration. Note that the expected number of stocks in the pipeline is affected by lateral transshipment because demand rate at each service center changes due to overflow demands. As a result, there is no equivalence between the optimal policy without pipeline stock cost and the optimal policy with it in Kranenburg and van Houtum (2007). Hence, we include the pipeline cost in the cost function.

$$C(\mathbf{S}) = hc(EOH_0 + \sum_i EOH_i^*) + pc(\sum_i EPS_i^*) + \sum_i \sum_{j \neq i} lc_{ij} OF_{ij}^*, \tag{14}$$

where $\{\mathbf{S}\} = \{S_0, S_i, S_j, ...\}$ and $S_0, S_i, S_j, ...$ are integers.

### 3.3. Model Validation

We evaluated our approximations by comparison with results obtained in an event-driven simulation. The test was conducted with three nonidentical service centers whose demand rates and lead times were different from each other. We conducted 100 independent simulations and calculated the average values of the simulated local instantaneous fill rates, local fill rates within time span at all service centers, as well as the average values of the probability of direct service and the probability of service within time span. Note that a single echelon system is a special case of our model, where there is no lateral transshipment and the central depot always has stock on hand. Accordingly, we used our exact evaluations for the single echelon system (see Theorem 1 and Corollary 1) to check the quality of our simulations, and found that the simulation errors were less than 0.5% and the standard deviations of the simulated values were less than 0.04 over 27 tested cases.

We then compared the results from our analysis with the results obtained from event-driven simulations for the two-echelon inventory system over 27 instances. We found that our approximations were very close to the results obtained from the simulation, whose relative deviations were less than 4% (for detailed results see Appendix C).

## 4. Optimization

We first consider the optimization problem minimizing average costs subject to the customer oriented service level constraints with lateral transshipment in the inventory control system.

$$Minimize_{\{\mathbf{S}\}} \quad C(\mathbf{S})$$

$$\text{subject to } P_{lt}^{D}(\mathbf{S}) \geq \beta^{D} \text{ and } P_{lt}^{T}(\mathbf{S}) \geq \beta^{T};$$

where $\beta^D$ and $\beta^T$ are the target service levels the inventory system has to achieve according to the contracts with customers, i.e. the target probability of direct service and the target probability of service within time span $T$.

The main trade-off in our inventory control system is that increasing base stock levels will increase the service levels, whereas it will increase the holding cost. Furthermore, including lateral transshipment affects not only the service levels at one local service center but also the service levels at other neighboring service centers. For instance, in case of one service center having sufficient base stocks, with neighboring service centers having no base stocks, including lateral transshipment leads to a lower probability of direct service and a lower probability of service within $T$ time units at the service center, but it increases the probability of service within $T$ at other service centers. As a result, the overall direct service levels may deteriorate and we may need more base stocks to fulfill the target probability of direct service requirement. Hence, we cannot use a decomposition approach to find the optimal solution for the inventory control system.

To solve the optimization problem, we use a complete enumeration over all possible stock profiles $\{\mathbf{S}\} = \{S_0, S_i, S_j, ...\}$ within a upper bound and a lower bound as in Kutanoglu and Mahajan (2009). We determine the upper bound of the base stock level at service center $i$ $S_i^{max}$ by finding the base stock level that achieves the target service levels when its demand rate is $\lambda_i + \sum_j \lambda_j$, and the central depot base stock is 0 which leads to a lead time of $L_i + L_0$. We then determine the lower bound of the total base stocks as in Kutanoglu and Mahajan (2009), assuming all the base stocks are pooled together at the central depot and they can be delivered to customers instantaneously when demand occurs at local service centers, which leads to a single echelon system with only one

base. Solving this system for the base stock level that achieves the target service levels, we can find the lower bound $S^{LB}$ for the total base stocks over the central depot and all service centers. Hence, the total base stocks can range from $S^{LB}$ to $\sum_i S_i^{max}$. We then enumerate all possible stock profiles $\{\mathbf{S}\}$ where the total base stocks are in the range $[S^{LB}, \sum_i S_i^{max}]$ and $0 \leq S_i \leq S_i^{max}$. For each stock profile $\{\mathbf{S}\}$, we check whether it satisfies the service level constraints and calculate the corresponding total inventory cost. The solution is the stock profile that satisfies the service level constraints with minimum total inventory cost.

On the other hand, when we exclude lateral transshipment, we can use the same approach for the optimization problem except for we set the overflow demand streams zeros, $OF_{ij} = 0$, for all $i$ and $j$.

## 5. Case Study

We apply our models in a case study of a manufacturer in the dredging industry, which builds dredging vessels and supplies equipment and control systems to customers worldwide.

Empirical data of a service part, an impeller, was collected and is used as input to our model. The demand figures presented below are realistic but fictitious for confidentiality reasons. An impeller is a rotating component (usually made of cast iron) of a centrifugal pump, which is used to move liquids through a piping system or for large discharge through smaller heads. This part is considered a service part, because in most cases it wears out faster than the pump casing. Therefore a damaged impeller has to be replaced with a new one to keep the pump running. Moreover, the impeller usually weighs more than 4,000 kg such that it is way too costly to transport it by air. Consequently, slow sea transport is needed, which takes much more time, leading to more pipeline stocks and non-negligible lateral transshipment time.

The company has a central depot in the Netherlands, which repairs all the broken impellers. The time required to repair an impeller is typically around 35 weeks, hence, the lead time $L_0 = 0.7$ years (35 out of 50 weeks). There are 3 operating service centers, located in Shanghai (SC1), Singapore (SC2), and Dubai (SC3) respectively. The lead time between the central depot and these service

centers is $L_1 = 0.16$ years (8 weeks), $L_2 = 0.14$ years (7 weeks), and $L_3 = 0.12$ years (6 weeks) respectively. In case of a stock out, a service center may request lateral transshipment from other service centers based on pre-specified rules as in Section 3.3. The lateral transshipment time is $LT_{12} = LT_{21} = 0.04$ years (2 weeks) between Shanghai and Singapore, $LT_{23} = LT_{32} = 0.06$ years (3 weeks) between Singapore and Dubai, and $LT_{13} = LT_{31} = 0.10$ years (5 weeks) between Shanghai and Dubai. Moreover, the time span of acceptable delay for customers is $T = 0.06$ years (3 weeks).

The annual demand rates at these service centers are $\lambda_1 = 20$ units, $\lambda_2 = 5$ units, and $\lambda_3 = 10$ units. Hence, the aggregated annual demand is $\lambda_0 = 35$ units. The target customer-oriented service levels are 90% probability of immediately delivery and 98% probability of delivery within 3 weeks.

The cost parameters for the impeller in this service network are estimated by the industrial expert. The holding cost is around 1,900 euros per unit per year; the pipeline carrying cost is 1,200 euros per unit per year; and the lateral transshipment cost is 1,800 euros per unit between Shanghai and Singapore, 2,100 euros per unit between Singapore and Dubai and 2,500 euros per unit between Shanghai and Dubai.

Using the algorithm in Section 4, we obtain the optimal base stock allocation $\{\mathbf{S}\}=\{25, 8, 3, 4\}$. The corresponding minimum cost is 26,743 euros per year (see Table 1). It breaks down to total holding cost, total pipeline inventory cost, and total lateral transshipment cost.

**Table 1 Total Inventory Cost break down**

| per year | Total Cost | Holding Cost | Pipeline Cost | Lateral Transshipment Cost |
|---|---|---|---|---|
| Euro | 26,743 | 19,971 | 6,112 | 660 |
| Percent | 100% | 74.7% | 22.9% | 2.5% |

As we can see in Table 1, the total holding cost takes the largest share of the total cost. However, the total pipeline cost accounts for 22.9% of the total cost, in fact, the total expected number of the pipeline stocks are more than half of the total expected stocks on hand. This is because the lead time of these slow moving items is long even though the demand rate is low. For instance, the demand rate at the service center in Shanghai is 20 units per year, one unit demand every 2.5 weeks on average, but its lead time is 8 weeks which is much longer. As shown in Figure 1, at the

service center in Shanghai, the expected pipeline stocks are quite close to its average inventory on hand; at the service center in Dubai, the expected pipeline stocks are just above 50% of the average inventory on hand. As a result, we cannot disregard these pipeline stocks in the inventory control policies.

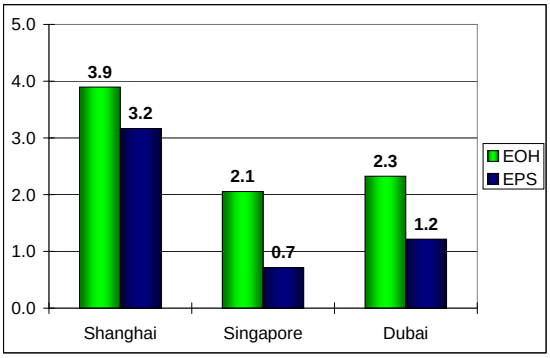


**Figure 1 Comparison between Expected Stocks-on-Hand (EOH) and Expected Pipeline Stocks (EPS)**

The uncommitted pipeline stocks can fulfill the demands that are not satisfied instantaneously, given they can be delivered to the customers within 3 weeks. In fact, among the demands which are not satisfied by direct services, 17.1% are fulfilled by the pipeline stocks as shown in Table 2. If we disregard these pipeline stocks, more lateral transshipment will be required in order to meet the target service levels, which will incur more cost.

**Table 2 Demands fulfilled within 3 weeks or beyond**

| per year | Total | within 3 weeks by PS | within 3 weeks by LT | beyond 3 weeks |
|---|---|---|---|---|
| Average # of items | 0.45 | 0.08 | 0.35 | 0.03 |
| Percent | 100.0% | 17.1% | 76.9% | 6.1% |

The probability of direct service we obtained at the optimal base stock allocation is 91.8%, and the probability of service within 3 weeks is 98.2%, which is slightly above the 90% and 98% targets. Comparing with the case where we do not consider lateral transshipment as in Table 3, we find that though the fill rates in Shanghai are improved because of lateral transshipment, the fill rates at all other service centers are reduced. As a result, lateral transshipment indeed reduces the direct service level of the inventory system as we pointed out in Section 3. Furthermore, the cost in case

with lateral transshipment is also higher than that in case without lateral transshipment. Hence, lateral transshipment leads to a lower system performance. This is contrary to cases where the lateral transshipment time is negligible.

**Table 3 Comparison of Performance Criteria without Lateral Transshipment and with Lateral Transshipment**

| | $P^D$ | $P^T$ | $FR_{SH}$ | $FR^T_{SH}$ | $FR_{SP}$ | $FR^T_{SP}$ | $FR_{DB}$ | $FR^T_{DB}$ | Total Cost |
|---|---|---|---|---|---|---|---|---|---|
| without LT | 0.927 | 0.982 | 0.937 | 0.988 | 0.929 | 0.972 | 0.908 | 0.975 | 26,077 |
| with LT | 0.918 | 0.982 | 0.940 | 0.989 | 0.926 | 0.971 | 0.904 | 0.974 | 26,743 |

We would like to remark that we considered a simple lateral transshipment rule in our model. In fact, in order to find the overall cost optimum, we need to enumerate over all possible lateral transshipment rules based on availability or on inventory balancing. For example, instead of supporting other service centers unconditionally, each service center reserves one unit base stock for its own customers. However, these lateral transshipment rules complicate the analysis even more, which are outside the scope of this paper.

In our investigation for this case study, we include the uncommitted pipeline stocks in our service levels evaluation which improve the service level within the time span. To illustrate the advantage of this approach, we compare our results from the case study with the results from the model which disregards the pipeline stocks. Note that the no-pipeline stock model is a special case of our model where the calculations of overflow demand streams, service levels and cost rates need to be modified.

Using the algorithm in Section 4 with modified service levels and cost rates calculations, we obtain the optimal base stock allocation for this model $\{\mathbf{S}\}=\{25, 10, 4, 5\}$. Hence, the total base stock requirement is 44 units, 4 units more than the requirement in case of including pipeline stocks in performance evaluation. As a result, the holding cost from this model (27,399 euros per year) is much larger than what we obtained (19,971 euros per year) before. Furthermore, the minimum total inventory cost obtained in this model, 33,678 euros per year, is much larger than that we obtained before.

Comparing our case study with the model which disregards the uncommitted pipeline stocks, we find that including these pipeline stocks will reduce the total inventory cost by 20.6%. The major contribution in the cost reduction is from the holding cost, because the uncommitted pipeline stocks increase the service levels and less base stocks are required to achieve the target service levels.

## 6. Conclusion

This paper accesses the effect of non-negligible lateral transshipment time on various aspects of spare parts inventory control, by analyzing a two-echelon service parts inventory control system with lateral transshipment. In this system, the transshipment time is not negligible, so we need to take the timing of transshipment and replenishment into account when we make lateral transshipment decisions. Moreover, we introduce the customer-oriented service levels, i.e. the probability of direct service and the probability of service within a certain time span, and find their analytical expressions based on the METRIC. We solve the optimization problem for the inventory base stock allocation by enumeration over all possible stock profiles.

In addition, our results from a case study for a market leader in dredging industry show that, for slow moving service parts whose lateral transshipment time is not negligible, lateral transshipment may lead to lower system performance. Furthermore, substantial cost savings can be obtained when we include the uncommitted pipeline stocks in base stock allocation and lateral transshipment decisions.

## Acknowledgments

This research has been made possible with support of Transumo (see www.transumo.nl). We also thank Ricardo van Gelder whose research helped us formulate our research. We appreciate the help and insights of Adriana Gabor, Jan Brinkhuis and Willem van Jaarsveld.

## Appendix A: Proof of Theorem 1

THEOREM 1 *In a single echelon system, the probability that a customer at service center $i$ has to wait and waits no more than $T$ is*

$$P(0 < W_i \leq T) = 1 - FR_i - \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \binom{S_i + k - 1}{j} (\frac{T}{L_i})^j (1 - \frac{T}{L_i})^{S_i+k-1-j} e^{-\lambda_i L_i} \frac{(\lambda_i L_i)^{S_i+k-1}}{(S_i + k - 1)!}$$

*and the expected waiting time for the customer is*

$$E(W_i) = \sum_{k=1}^{\infty} \frac{kL_i}{S_i+k} e^{-\lambda_i L_i} \frac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!}$$

*Proof of Theorem 1* Since lead time $L_i$ is constant, the order issued at time $t-L_i$ will be supplied to the local service center $i$ and become available to customers by $t$; and any orders issued after $t-L_i$ will not be available until after $t$. If there is no stock on hand when a customer arrives at time $t$ and sees $k-1$ $(k \geq 1)$ customers waiting in front of him, in other words, there are $S_i+k-1$ demands during $[t-L_i, t)$, then he has to wait for the pipeline stock corresponding to the $kth$ order issued after time $t-L_i$ due to the one-for-one first-come-first served assumptions. If this order is issued during $(t-L_i, t-L_i+T]$, then the corresponding pipeline stock will become available during $(t, t+T]$ and the customer's waiting time $W_i \in (0,T]$. Because of the memoryless property of the exponentially distributed inter-arrival times (denoted by $X_1, X_2, \cdots, X_k$, where $X_k$ is the inter-arrival time between $(k-1)th$ order and $kth$ order), we have

$$\begin{aligned}
&P(0 < W_i \leq T)\\
&= \textstyle\sum_{k=1}^{\infty} P(t-L_i < t-L_i+X_1+X_2+\cdots+X_k \leq t-L_i+T | D_i(L_i) = S_i+k-1) P(D_i(L_i) = S_i+k-1)\\
&= \textstyle\sum_{k=1}^{\infty} P(0 < X_1+X_2+\cdots+X_k \leq T | D_i(L_i) = S_i+k-1) P(D_i(L_i) = S_i+k-1)\\
&= \textstyle\sum_{k=1}^{\infty} \sum_{j=k}^{S_i+k-1} \binom{S_i+k-1}{j} (\frac{T}{L_i})^j (1-\frac{T}{L_i})^{S_i+k-1-j} P(D_i(L_i) = S_i+k-1)\\
&= \textstyle\sum_{k=1}^{\infty} \sum_{j=k}^{S_i+k-1} \binom{S_i+k-1}{j} (\frac{T}{L_i})^j (1-\frac{T}{L_i})^{S_i+k-1-j} e^{-\lambda_i L_i} \frac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!}\\
&= \textstyle\sum_{k=1}^{\infty} [1 - \sum_{j=0}^{k-1} \binom{S_i+k-1}{j} (\frac{T}{L_i})^j (1-\frac{T}{L_i})^{S_i+k-1-j}] e^{-\lambda_i L_i} \frac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!}\\
&= 1 - \textstyle\sum_{n=0}^{S_i-1} e^{-\lambda_i L_i} \frac{(\lambda_i L_i)^n}{n!} - \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \binom{S_i+k-1}{j} (\frac{T}{L_i})^j (1-\frac{T}{L_i})^{S_i+k-1-j} e^{-\lambda_i L_i} \frac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!}\\
&= 1 - FR_i - \textstyle\sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \binom{S_i+k-1}{j} (\frac{T}{L_i})^j (1-\frac{T}{L_i})^{S_i+k-1-j} e^{-\lambda_i L_i} \frac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!}
\end{aligned}$$

Note that this probability $P(0 < W_i \leq T)$ is equal to the probability that the demand is fulfilled by an uncommitted pipeline stock within the time span $T$ in this single echelon system.

The average waiting time for the customer is:

$$\begin{aligned}
E(W_i) &= \textstyle\sum_{k=1}^{\infty} E(W_i | D_i(L_i) = S_i+k-1) P(D_i(L_i) = S_i+k-1))\\
&= \textstyle\sum_{k=1}^{\infty} E(x_1+x_2+\cdots+x_k | D_i(L_i) = S_i+k-1) P(D_i(L_i) = S_i+k-1))\\
&= \textstyle\sum_{k=1}^{\infty} \frac{kL_i}{S_i+k} e^{-\lambda_i L_i} \frac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!}
\end{aligned}$$

□

## Appendix B: Proof of Corollary 1

Corollary 1. *In a single echelon system, the fill rate within the time span $T$ at service center $i$ is*

$$\begin{aligned} FR_i^T &= FR_i + P(0 < W_i \le T) \\ &= 1 - \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \binom{S_i+k-1}{j} (\tfrac{T}{L_i})^j (1-\tfrac{T}{L_i})^{S_i+k-1-j} e^{-\lambda_i L_i} \tfrac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!} \\ &= \sum_{t=0}^{S_i-1} \tfrac{\lambda_i^t}{t!} (L_i - T)^t e^{-\lambda_i (L_i - T)} \\ &= FR(S_i, \lambda_i, L_i - T) \end{aligned}$$

*Proof of Corollary 1*

$$\begin{aligned} & 1 - \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \binom{S_i+k-1}{j} (\tfrac{T}{L_i})^j (1-\tfrac{T}{L_i})^{S_i+k-1-j} e^{-\lambda_i L_i} \tfrac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!} \\ &= 1 - e^{-\lambda_i L_i} \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \tfrac{(S_i+k-1)!}{j!(S_i+k-1-j)!} (\tfrac{T}{L_i})^j (1-\tfrac{T}{L_i})^{S_i+k-1-j} \tfrac{(\lambda_i L_i)^{S_i+k-1}}{(S_i+k-1)!} \\ &= 1 - e^{-\lambda_i L_i} \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \tfrac{1}{j!} (\tfrac{T}{L_i})^j \tfrac{1}{(S_i+k-1-j)!} (1-\tfrac{T}{L_i})^{S_i+k-1-j} (\lambda_i L_i)^{S_i+k-1} \\ &= 1 - e^{-\lambda_i L_i} \sum_{k=1}^{\infty} \sum_{j=0}^{k-1} \tfrac{1}{j!} (\tfrac{T}{L_i})^j (\lambda_i L_i)^j \tfrac{1}{(S_i+k-1-j)!} (1-\tfrac{T}{L_i})^{S_i+k-1-j} (\lambda_i L_i)^{S_i+k-1-j} \\ &= 1 - e^{-\lambda_i L_i} \sum_{j=0}^{\infty} \tfrac{1}{j!} (\tfrac{T}{L_i})^j (\lambda_i L_i)^j \sum_{l=1}^{\infty} \tfrac{1}{(S_i+l-1)!} (1-\tfrac{T}{L_i})^{S_i+l-1} (\lambda_i L_i)^{S_i+l-1}, \quad \text{where } l = k-j \\ &= 1 - e^{-\lambda_i L_i} e^{\lambda_i W} \sum_{t=S_i}^{\infty} \tfrac{1}{t!} (1-\tfrac{T}{L_i})^t (\lambda_i L_i)^t, \quad \text{where } t = S_i + l - 1 \\ &= 1 - e^{-\lambda_i (L_i - T)} [e^{\lambda_i (L_i - T)} - \sum_{t=0}^{S_i-1} \tfrac{1}{t!} (1-\tfrac{T}{L_i})^t (\lambda_i L_i)^t] \\ &= \sum_{t=0}^{S_i-1} \tfrac{\lambda_i^t}{t!} (L_i - T)^t e^{-\lambda_i (L_i - T)} \\ &= FR(S_i, \lambda_i, L_i - T) \end{aligned}$$

□

## Appendix C: Simulation Results

In the simulations, we consider 27 cases with various lead times (L weeks), demand rates (lambda units) and base stocks (S units). We conducted 100 independent simulations for each test case and calculated the average values of the simulated local instantaneous fill rates, local fill rates within time span (0.1 weeks) at all service centers, as well as the average values of the probability of direct service (FR mean) and the probability of service within time span (FRT mean). We also calculated the average fraction of demand fulfilled with delays less than 0.1 weeks (Dw/D mean). Comparing these results with the evaluated values based on our analysis, we can see the relative deviations between the corresponding simulated and evaluated values are less than 0.5% as shown the the table below. We also check the quality of our simulation, and find that its performance is quite consistent with low standard deviations (less than 0.04) over all tested cases.

Table 4 Comparison between simulated results and evaluated values

| | | | Simulation | | | | | | Evaluation | | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| L | lambda | S | FR mean | FR std | FRT mean | FRT std | Dw/D mea | Dw/D std | FR | FRW | Dw/D |
| 0.3 | 3 | 3 | 0.935 | 0.021 | 0.979 | 0.010 | 0.044 | 0.015 | 0.937 | 0.977 | 0.040 |
| | | 4 | 0.986 | 0.009 | 0.997 | 0.003 | 0.010 | 0.007 | 0.987 | 0.997 | 0.010 |
| | | 5 | 0.998 | 0.004 | 0.9997 | 0.001 | 0.002 | 0.004 | 0.998 | 0.9996 | 0.002 |
| | 4 | 3 | 0.877 | 0.027 | 0.960 | 0.011 | 0.083 | 0.020 | 0.879 | 0.953 | 0.073 |
| | | 4 | 0.966 | 0.012 | 0.992 | 0.004 | 0.026 | 0.010 | 0.966 | 0.991 | 0.025 |
| | | 5 | 0.992 | 0.006 | 0.999 | 0.002 | 0.007 | 0.006 | 0.992 | 0.9986 | 0.006 |
| | 5 | 3 | 0.808 | 0.024 | 0.938 | 0.012 | 0.130 | 0.018 | 0.809 | 0.920 | 0.111 |
| | | 4 | 0.936 | 0.018 | 0.985 | 0.007 | 0.049 | 0.014 | 0.934 | 0.981 | 0.047 |
| | | 5 | 0.981 | 0.010 | 0.997 | 0.003 | 0.016 | 0.008 | 0.981 | 0.996 | 0.015 |
| 0.4 | 3 | 3 | 0.882 | 0.027 | 0.950 | 0.014 | 0.067 | 0.018 | 0.879 | 0.937 | 0.058 |
| | | 4 | 0.968 | 0.013 | 0.990 | 0.006 | 0.022 | 0.010 | 0.966 | 0.987 | 0.020 |
| | | 5 | 0.992 | 0.007 | 0.998 | 0.003 | 0.006 | 0.006 | 0.992 | 0.998 | 0.005 |
| | 4 | 3 | 0.778 | 0.031 | 0.903 | 0.017 | 0.125 | 0.020 | 0.783 | 0.879 | 0.096 |
| | | 4 | 0.922 | 0.021 | 0.974 | 0.010 | 0.052 | 0.015 | 0.921 | 0.966 | 0.045 |
| | | 5 | 0.978 | 0.012 | 0.995 | 0.004 | 0.017 | 0.009 | 0.976 | 0.992 | 0.016 |
| | 5 | 3 | 0.677 | 0.033 | 0.849 | 0.015 | 0.172 | 0.023 | 0.677 | 0.809 | 0.132 |
| | | 4 | 0.858 | 0.030 | 0.949 | 0.013 | 0.091 | 0.020 | 0.857 | 0.934 | 0.077 |
| | | 5 | 0.950 | 0.018 | 0.986 | 0.007 | 0.037 | 0.013 | 0.947 | 0.981 | 0.034 |
| 0.5 | 3 | 3 | 0.817 | 0.031 | 0.950 | 0.020 | 0.088 | 0.019 | 0.809 | 0.879 | 0.071 |
| | | 4 | 0.936 | 0.023 | 0.974 | 0.011 | 0.037 | 0.014 | 0.934 | 0.966 | 0.032 |
| | | 5 | 0.983 | 0.010 | 0.995 | 0.005 | 0.012 | 0.008 | 0.981 | 0.992 | 0.011 |
| | 4 | 3 | 0.679 | 0.038 | 0.836 | 0.022 | 0.157 | 0.023 | 0.677 | 0.783 | 0.107 |
| | | 4 | 0.856 | 0.029 | 0.938 | 0.014 | 0.082 | 0.020 | 0.857 | 0.921 | 0.064 |
| | | 5 | 0.948 | 0.018 | 0.981 | 0.008 | 0.033 | 0.012 | 0.947 | 0.976 | 0.029 |
| | 5 | 3 | 0.544 | 0.032 | 0.771 | 0.018 | 0.226 | 0.022 | 0.544 | 0.677 | 0.133 |
| | | 4 | 0.757 | 0.031 | 0.896 | 0.016 | 0.138 | 0.021 | 0.758 | 0.857 | 0.100 |
| | | 5 | 0.892 | 0.025 | 0.959 | 0.012 | 0.067 | 0.017 | 0.891 | 0.947 | 0.056 |

## References

Alfredsson, P. and Verrijdt, J., 1999. Modeling emergency supply fexibility in a two-echelon inventory system. *Management Science.* **45(10)**, 1416–1431.

Axsäter, S., 1990. Modelling emergency lateral transshipments in inventory systems. *Management Science.* **36(11)**, 1329–1338.

Axsäter, S., 2006. *Inventory Control.*, 2nd edition. International Series in Operations Research & Management Science. New York: Springer Science.

Banerjee, A., Burton, J. and Banerjee, S., 2003. A simulation study of lateral shipments in single supplier, multiple buyers supply chain networks. *International Journal of Production Economics.* **81-82**, 103–114.

Burton, J. and Banerjee, A., 2005. Cost-parametric analysis of lateral transshipment policies in two-echelon supply chains. *International Journal of Production Economics.* **93-94**, 169–178.

Diks, E.B. and De Kok, A.G., 1996. Controlling a divergent 2-echelon network with transshipments using the consistent appropriate share rationing policy. *International Journal of Production Economics.* **45**, 369–379.

Kranenburg, A. and van Houtum, G., 2007. Cost optimization in the (S-1, S) lost sales inventory model with multiple demand classes. *Operations Research Letters.* **35**, 493–502.

Kruse, K., 1981. Waiting Time in a Continuous Review (s, S) Inventory System with Constant Lead Times. *Operations Research.* **29(1)**, 202–207.

Kutanoglu, E. and Mahajan, M., 2009. An inventory sharing and allocation method for a multi-location service parts logistics network with time-based service levels. *European Journal of Operational Research.* **194(3)**, 728–742.

Lee, H., 1987. A multi-echelon inventory model for repairable items with emergency lateral transshipments. *Management Science.* **33(10)**, 1302–1316.

Reijnen, I.C., Tan, T. and van Houtum, G.J., 2009. Inventory planning for spare parts networks with delivery time requirements. Working paper, Eindhoven University of Technology, the Netherlands.

Sherbrooke, C., 1968. METRIC: A multi-echelon technique for recoverable item control. *Operations Research.* **16(1)**, 122–141.

Tagaras, G., Vlachos, D., 2002. Effectiveness of stock transshipment under various demand distributions and nonnegligible transshipment times. *Production and Operations Management.* **11(2)**, 183–198.

Tijms, H.C., 2003. *A first course in stochastic models.* Chichester: Wiley.

Wong, H., Cattrysse, D. and Van Oudheusden, D., 2005. Stocking decisions for repairable spare parts pooling in a multi-hub system. *International Journal of Production Economics.* **93-94**, 309–317.