

A Fill-rate Service level model for Integrated Network design and Inventory Allocation Problem

ISyE Technical Report J-09-02

Divya Mangotra, Jye-Chyi Lu, Yu-Chung Tsao
H. Milton Stewart School of Industrial and Systems Engineering

Abstract

Today's global economy relies heavily on transportation and warehousing to deliver goods. Outsourcing has made the distribution networks complex with several layers of national (import) distribution centers (NDCs) and regional distribution centers (RDCs) between suppliers and retailers. More distribution nodes in the supply chain also means more inventory stockpiles. We present an integrated facility location and inventory allocation problem for designing a distribution network with multiple NDCs and retailers. The key decisions are where to locate the regional distribution centers and how much inventory to hold at the different nodes of the distribution network such that the total network cost is minimized. A Type-II (fill-rate) service level measure is used in this analysis for modeling the safety stock inventory and the inventory cost analysis is based on the continuous review batch ordering policy. Our analysis shows that the type of service measure used affects the network design.

Keywords: Supply chain design; Inventory; Fill-rate Service level; Facility location; Continuous approximation model.

1 Introduction

Most global supply chain network consist of several NDCs (national distribution centers) shipping goods to multiple RDCs (regional distribution centers) who in turn satisfy the end customer demand at the retailers. The total logistics cost for operating a distribution network is the sum of facility location cost, inbound and outbound transportation cost, and inventory holding cost. We present a distribution network with several layers of RDCs and NDCs and solve for the optimal network design and inventory policy. Safety stock inventory is modeled using a *Type-II* service level measure. We show that the choice of service level measure impacts the inventory policy and network design.

In this paper, a continuous approximation approach is presented to model the integrated facility location and inventory allocation problem using a fill-rate service level (*Type-II*) approach. Facility location and inventory allocation decision are interrelated. One of the key cost components for the facility location problem is the transportation cost which depends on the inventory replenishment frequency at the different facilities. Similarly, the inventory allocation problem models the inventory cost at the distribution center in terms of the demand assigned to it. This requires information on which retailers are assigned to which distribution center. Even though the dependence between facility decisions and inventory policy is well understood, most literature deals with these problems seperately (see Daskin, 1995; Drezner, 1995; Ganeshan, 1999; Deuermeyer and Schwarz, 1981).

A *Type-II* service level of ρ at a DC location implies that exactly ρ proportion of demand at this location will be filled from stock during a given replenishment interval. Thus, a type-II service level reflects both the stock-out event and the amount backordered. Fill rate is typically considered a more relevant measure of service compared to Type-I as it enables the DC location to estimate what fraction of demand was converted to sales or equivalently what

was the expected number of units backordered/lost during the replenishment interval. Type-I service level simply measures the number of stock-outs in each period without accounting for the exact number of units short in an order. Though more relevant, Type-II service level is less widely used in research due to the complex form of backorder quantity term which makes it hard to model it. There are a few selected papers in the area of multi-echelon inventory that model service levels in terms of fill-rate (Deuermeyer and Schwarz, 1981; Schwarz et al., 1985; Ganeshan, 1999).

Our work is an extension of the work by Mangotra et al. (2009) on integrated facility location and inventory allocation problem. In their study, the authors present a continuous approximation technique for solving the problem under a *Type-I* service level policy for modeling the safety stock in the network. The key difference between this problem and the problem presented in the paper by Mangotra et al. (2009) lies in the representation of the safety stock term. Under the *Type-II* service level assumption, the re-order costs no longer have a closed form expression and a solution to the problem is generated using a two-stage hierarchical approach. Using a hierarchical approach (see van Houtum, 2006), the inventory decisions for the different types of inventories are made in two stages. While the cycle stock decisions are made in the first stage, along with the facility location decisions, to determine the optimal re-order batch size, the second stage decisions focus on the safety stock levels (and hence the re-order points).

The objective function in this paper minimizes the total logistic costs expressed as a sum of the inventory, facility and transportation costs, and meets the desired service level requirement at each inventory stocking level. Our analysis shows that the type of service level measure has an impact on the network design and inventory decisions.

2 Literature Review

Our paper integrates several independently studied problems in areas of facility location, inventory allocation and data approximation for large scale problems. In this section we provide a detailed review of existing literature for each area.

Facility location and Allocation Decision

This stream of literature deals with questions on how many DCs to open, where to locate them and how to assign retailers to these DCs. Daskin et al. (2002) and Shen et al. (2003) in their work propose a set-covering model to consider location and allocation policies for a DC-retailer network with risk pooling, and showed that this problem can be solved efficiently when the DC demand is assumed to be Poisson or deterministic. Shu et al. (2005) presents an extension to this model by considering a more general demand distribution while Shen and Qi (2007) solves the problem by relaxing the constant proportionality assumption between mean and variance of demand. Shen (2005) and Shen and Daskin (2005) study the location-allocation problem for a multi-commodity supply chain as well as customer service quality measure.

One-Echelon Facility Location and Inventory Policy Decisions

Nozick and Turnquist (1998) present an extension of the fixed-charge facility location model, defined in Daskin (1995), by including an inventory cost term in their DC location problem. Safety stock is approximated by a linear regression function of the number of DCs. Nozick and Turnquist (2001) in another work add the service responsiveness and uncertainty in delivery time to the DC to their model. Service responsiveness is defined in terms of stock-outs and time-based delivery. The stock-outs are incorporated in the safety stock function while the time-based delivery constraint is modeled explicitly as coverage distance. Snyder et al. (2007) model a stochastic location and inventory policy problem with risk pooling and develop a Lagrangian relaxation based exact algorithm to solve it. Their goal is to determine DC locations, assign retailers to DCs, and set inventory levels at DCs to minimize the total expected cost. Miranda and Garrido (2004) model an integrated capacitated facility location problem (CFLP) and inventory control decisions. The solution methodology in their work involves a lagrangian relaxation and the sub-gradient method. In another work,

Two-Echelon Integrated Network Design and Inventory Policy Decisions

Teo and Shu (2004) study an integrated logistic network problem which models inventory cost for multiple echelons of stocking locations. They proposed a set-covering model to design a two-echelon warehouse-retailer network under deterministic retailers demands. Their problem is to determine the warehouses locations, allocate retailers to warehouses, and inventory decisions for the warehouses and retailers. The computational results in their study apply to a small problem instance with 20 warehouses and 100 retailers. This model does not include either the demand or the supply uncertainty. Demand variability is added to the previous model in the study by Romeijn et al. (2007).

Mangotra et al. (2009) presents a continuous approximation (CA) model for solving the integrated facility location and inventory allocation problem. Their model takes a nonlinear form and solution techniques are developed using the theory of nonlinear programming. The numerical study in their work suggests that the total network cost is significantly lower in the case of the integrated model as compared with the non-integrated model.

Data Approximation models for facility location decision

Newell (1973) in a seminal paper presents data approximation techniques for warehouse location problem. Geoffrion (1976) studies a similar problem when the warehouse serves demand that is distributed uniformly over a plane. Erlenkotter (1989) extends the work by Geoffrion (1976) and Newell (1973), with more detailed expressions for the production cost in a General Optimal Market Area (GOMA) model. Dasci and Verter (2001) use the continuum approximation technique to study a production and distribution design problem. In their work, they explicitly model the facility cost including both the operational and acquisition cost components, however, inventory costs are excluded in their study. Burn et al. (1985) propose an analytic method for their distribution network problem with a single supplier and multiple customers. Their model uses the spatial density of customers to minimize the inventory and transportation cost for freight. Langevin et al. (1996) present an extensive review of continuous approximation models that have been developed for freight distribution

problems.

Fill-rate service level policy

There are only a selected few papers in the area of multi-echelon inventory that model service levels in terms of fill-rate (Deuermeyer and Schwarz, 1981; Schwarz et al., 1985; Ganeshan, 1999). A Type-II service level (*Fill-rate*) is recognized as the true service measure as it measures exactly how much of the demand was met. However, it is complicated to model it. The type-1 service level can be modeled using relatively simpler expressions and is hence appears widely in the inventory literature as a preferred service measure in determining the inventory policy.

In this paper we present an integrated facility location and inventory allocation problem. Our work is an extension of the work by Mangotra et al. (2009) that studies the same problem under a *Type-I* service level. We model this problem under a *Type-II* service level and propose a hierarchical solution to solve it. Our analysis shows that the type of service measure used affects the network design and inventory policy decisions.

3 Fill-rate model for inventory at RDC and NDC

The network under study in this paper has several NDCs (national distribution centers) serving demand across multiple RDCs (regional distribution centers) who in turn satisfy the end customer demand at the retailers. We model the objective function in our problem in terms of the fixed facility location cost, inbound and outbound transportation cost, and inventory holding cost. The goal is to find a network design and inventory allocation solution that minimizes the objective function and ensures that the fill-rate service level constraint is met at each RDC and NDC. A *Type-II* service level of ρ at a DC location implies that the DC would be able to satisfy exactly ρ proportion of demand during a given replenishment interval.

We first analyze the different costs to understand what is the impact of service level on each of them. Facility costs and transportation costs (both inbound and outbound) do not depend on the service measure, however the expected inventory is dependent on the service level. This is because the expected inventory is a function of both the cycle stock and the safety stock, and the safety stock term is a function of the *Type-II* service level measure.

For estimating the total facility Cost, consider a fixed rent F_r paid for opening and operating each RDC. The total facility cost $TF(x)$ is given by multiplying the facility cost of opening each RDC with the number of RDCs; namely,

$$TF(x) = F_r N_r(x) \tag{1}$$

We consider two components for the transportation cost-outbound and inbound costs. For the RDC, the outbound cost is the cost of shipping goods to the retailers located within its influence area. Inbound cost is the cost of sending shipments from the NDC to the RDC.

For the NDC, the outbound cost is the same as inbound cost for the RDC. The inbound cost from the outside supplier is not modeled explicitly at the NDC. Instead this cost is factored in the re-order cost at the NDC. Each transportation cost component consists of a fixed cost and a variable cost. The fixed component of cost can be associated with managing the fleet, drivers, etc. The variable cost is the cost per item.

Let C_f be the fixed cost per inbound shipment and C_v be the variable cost per item for each inbound shipment. Then the total inbound transportation cost, $TIT(x)$, is given by:

$$TIT(x) = (C_f + C_v Q_r(x)) \left(\frac{\xi E[D_r(x)]}{Q_r(x)} \right) N_r(x) \quad (2)$$

where $(C_f + C_v Q_r(x))$ is the transportation cost incurred in a single inbound shipment to a single RDC. The expected demand faced by RDC r is given by $E[D_r(x)]$ (see equation 5 presented later), ξ is the length of the planning horizon and $E[D_r(x)]/Q_r(x)$ is the expected number of inbound shipments to a single RDC during the planning horizon. $N_r(x)$ is the number of RDCs in the distribution network.

Let C_l be the delivery cost per mile per item and f_r be the constant that depends on the distance metric and shape of the RDC service region (see Daganzo, 1996; Dasci and Verter, 2001). Then the total outbound local delivery cost, $TOT(x)$, is given by

$$TOT(x) = C_l (f_r \sqrt{A_r(x)}) (\xi \lambda(x) \delta(x) R) \quad (3)$$

where R is the area of the distribution network, $A_r(x)$ is the influence area for RDC r , while $\lambda(x)$ is the demand rate at each store during the planning horizon and $\delta(x)$ is the store density function for $x \in A_r(x)$. The total customer demand during the planning horizon (ξ) in the service area R is given by $\int_R \xi \lambda(x) \delta(x) dx$. Since $\lambda(x) \delta(x)$ is a slow varying function of $x \in R$, we get $\int_R \xi \lambda(x) \delta(x) dx = \xi \lambda(x) \delta(x) R$. The average outbound distance traveled by each item is given by $f_r \sqrt{A_r(x)}$ (see Dasci and Verter, 2001).

The average inventory cost function in the objective function is expressed in terms of the average holding cost and the average re-order cost. Let Q be the re-order quantity at each RDC and Q_n be the re-order quantity at the NDC. While the average holding cost is a function of the cycle stock (re-order quantities Q and Q_n) and the safety stock, the average re-order cost depends only on the cycle stock.

For the *Type-I service* analysis with stock-out probability α_r and α_n , the safety stock at each RDC and the NDC is given by $Z_{\alpha_r} \sqrt{Var[D_{r_i,LT}]}$ and $Z_{\alpha_n} \sqrt{Var[D_{n,LT}]}$ (see previous study by Mangotra et al. (2009)). These expressions have a closed form that can be modeled into the objective function of the problem. Unlike the *Type-I service* analysis, fill-rate (*Type-II service*) analysis lacks the ability to express the safety stock term in a closed form. Let r_{r_i} and r_{r_n} be the re-order point for each RDC i and NDC n . Re-order point is defined as the level of inventory at which a new replenishment order, equal to the re-order quantity, is triggered. We present the safety stock terms as $r_i - E[D_{r_i,LT}]$ and $r_{r_n} - E[D_{n,LT}]$ in the problem formulation. In addition, there are two new constraints which link the expected number of backorders to the re-order point using the fill-rate measure. The derivation of these constraints is discussed next.

3.1 Fill-rate constraints

The order replenishment lead time for each RDC is a function of the transit lead time and the wait time (or delay in order fulfillment) in the event of a stock-out at the NDC. We assume that the transit lead time has a normal distribution with mean μ_r and variance σ_r^2 . The distribution for the additional wait time is hard to estimate and the random variable for the wait time is often replaced by its expected value W (see Deuermeyer and Schwarz, 1981; Ganeshan, 1999). In this analysis, we model the wait time with its expected value and use it along with the distribution for the transit lead time to get the distribution for the total order replenishment lead time.

Result 1: The wait time to process an order in the event of a stock-out at the NDC is given by W (using *Little's law* (Little, 1961), see Appendix).

$$W = \frac{(1 - \rho_n)Q_n}{Q} \frac{1}{(\lambda\delta R/Q)} \quad (4)$$

where $(1 - \rho_n)Q_n/Q$ is the expected number of backorders at the RDC for a given fill-rate (*Type-II service level*) ρ_n at the NDC. $(\lambda\delta R/Q)$ is the expected demand rate at the NDC.

The order replenishment lead time T has a normal distribution with parameters:

$$\begin{aligned} E[T] &= (\mu_r + W) \\ Var[T] &= \sigma_r^2 \end{aligned}$$

The demand distribution at each RDC in cluster C_{n_i} (discussed in Mangotra et al., 2009) is denoted by D_{r_i} and is a Poisson process with rate $\lambda\delta_i A_{r_i}$. The expected demand and variance of demand per cluster is given by:

$$E[D_{r_i}] = \lambda\delta_i A_{r_i} \quad (5)$$

$$Var[D_{r_i}] = \lambda\delta_i A_{r_i} \quad (6)$$

Let $(E[D_{r_i,LT}])$ and $(Var[D_{r_i,LT}])$ be the expected demand and variance of demand at the RDC in region C_{n_i} during its order replenishment lead time. Then, the following holds (proved by Mangotra et al., 2009)

$$\begin{aligned} E[D_{r_i,LT}] &= (\mu_r + W)E[D_{r_i}] \\ Var[D_{r_i,LT}] &= (\mu_r + W)Var[D_{r_i}] + \sigma_r^2 E[D_{r_i}]^2 \end{aligned}$$

The demand process at the NDC during its replenishment lead time can be approximated by a normal distribution (see Deuermeyer and Schwarz, 1981). The expected demand

($E[D_{n,LT}]$) and variance of demand ($Var[D_{n,LT}]$)¹ at the NDC during its order replenishment period is given by:

$$E[D_{n,LT}] = \frac{\sum_{i=1}^N \mu_n \lambda \delta_i C_{n_i}}{Q} \quad (7)$$

$$Var[D_{n,LT}] = \frac{\sum_{i=1}^N \mu_n \lambda \delta_i C_{n_i}}{Q^2} \quad (8)$$

Result 2 (Hopp and Spearman, 2000): Let X be the demand process during lead time at a location with mean θ and variance σ^2 . Further, $f(x)$ and $F(x)$ denote the probability density function and cumulative density function. If R is the re-order point for a (Q, r) policy, then the expected number of backorders at the location is given by

$$\begin{aligned} E[B(R)] &= \int_R^{\infty} (x - R) f(x) dx \\ &= (\theta - R)[1 - \Phi(z)] + \sigma\phi(z) \end{aligned} \quad (9)$$

where $z = (R - \theta)/\sigma$ and is a standard normal variable.

The demand process at the NDC during order replenishment lead time has a normal distribution and *Result 2* can be used to derive an expression for the expected number of backorders. Further we assume that the demand process at each RDC during order replenishment lead time has a normal distribution and use *Result 2* to get a similar expression for the expected number of backorders. The expected number of backorders at the RDC ($E[B_i]$) and the NDC ($E[B_n]$) are given by:

$$\begin{aligned} E[B_{r_i}] &= (E[D_{r_i,LT}] - r_i)[1 - \Phi(z_i)] + \sigma\phi(z_i) \quad \forall i \\ E[B_n] &= (E[D_{n,LT}] - r_n)[1 - \Phi(z_n)] + \sigma\phi(z_n) \end{aligned} \quad (10)$$

where $z_i = (r_i - E[D_{r_i,LT}])/\sqrt{Var[D_{r_i,LT}]}$ and $z_n = (r_n - E[D_{n,LT}])/\sqrt{Var[D_{n,LT}]}$

Note that in this work, it is assumed that the decision on what is the right value for fill-rate for each distribution node is made by the management. Let ρ_{r_i} (ρ_n) be the fill rate for each RDC i (NDC). We also assume that the re-order quantity at each RDC is equal to Q . We assume equality in the replenishment orders across all RDCs to make our problem more tractable. Furthermore, the goal of this study is to understand the impact on network design and inventory decisions under different service level policies. Modeling our problem under a non-equal re-order point would be an interesting topic for future studies.

Under the special case, $Q_i = Q \quad \forall i$, the expected number of backorders can be estimated as (see Ganeshan, 1999):

$$\begin{aligned} E[B_{r_i}] &= (1 - \rho_{r_i})Q \quad \forall i \\ E[B_n] &= \frac{(1 - \rho_n)Q_n}{Q} \end{aligned} \quad (11)$$

¹Note that both expected demand and variance of demand are in units of RDC re-order quantity Q

Using equations (10) and (11), the re-order point at each RDC i (NDC) can be estimated in terms of fill-rate at the RDC (NDC).

$$\begin{aligned} (E[D_{r_i,LT}] - r_i)[1 - \Phi(z_i)] + \sigma\phi(z_i) &= (1 - \rho_{r_i})Q \quad \forall i \\ (E[D_{n,LT}] - r_n)[1 - \Phi(z_n)] + \sigma\phi(z_n) &= \frac{(1 - \rho_n)Q_n}{Q} \end{aligned} \quad (12)$$

We can now define the optimization model for the integrated facility location and inventory allocation problem under a *Type-II* service level. Note that our model assumes that re-order quantity at the NDC is proportional to re-order quantity at the RDCs, i.e., $Q_n = kQ$

P^f minimize

$$\begin{aligned} \gamma(\mathbf{A}, Q, Q_n) &= \sum_{i=1}^N \left(\frac{C_{n_i}}{A_{r_i}} \right) F_r + \sum_{i=1}^N (C_f + C_v Q) \left(\frac{\xi \lambda \delta_i C_{n_i}}{Q} \right) \\ &+ \sum_{i=1}^N C_{lfr} \sqrt{A_{r_i}} \xi \lambda \delta_i C_{n_i} + R_r \sum_{i=1}^N \left(\frac{\xi \lambda \delta_i C_{n_i}}{Q} \right) \\ &+ \sum_{i=1}^N \left(\frac{C_{n_i}}{A_{r_i}} \right) h_r \left(\frac{Q}{2} + (r_i - E[D_{r_i,LT}]) \right) \\ &+ h_n \left(\frac{kQ}{2} + (r_n - E[D_{n,LT}]) \right) + R_n \sum_{i=1}^N \left(\frac{\xi \lambda \delta_i C_{n_i}}{kQ} \right) \end{aligned}$$

subject to

$$\begin{aligned} Q_{r_i} &\geq 0 \quad \forall r_i \\ A_{r_i} &\geq 0 \quad \forall r_i \\ k &\geq 2 \\ (E[D_{r_i,LT}] - r_i)[1 - \Phi(z_i)] + \sigma\phi(z_i) &= (1 - \rho_{r_i})Q \quad \forall i \\ (E[D_{n,LT}] - r_n)[1 - \Phi(z_n)] + \sigma\phi(z_n) &= (1 - \rho_n)k \\ Q_{r_i}, \frac{C_{n_i}}{A_{r_i}}, k &\in Z^+ \quad \forall r_i \end{aligned} \quad (13)$$

where the first three inequalities in equation set (13) are the non-negativity constraints, next two are the fill-rate constraints that give the optimal value of the re-order points and the last inequality is the integrality constraint. In addition, $\mathbf{A} = [A_{r_1}, A_{r_2}, \dots, A_{r_n}]$ is the size of optimal areas served by each RDC cluster as defined in the paper by Mangotra et al. (2009).

4 Solution Approach

We discuss a detailed solution approach in this section. Note that our approach is based on a two-phase approximation technique to solve the network design and inventory allocation

problem. In phase-I approximation a Grid Cover-Couple approach is used to partition the service region into sub-regions. In phase-II approximation, the continuous approximation technique is used to model the facility location and inventory allocation problem over each cluster within the NDC partition. For more detail about Phase-I and Phase-II approximations, please see Mangotra et al. (2009).

A closer look at the stationary point for the objective function $\gamma(\mathbf{A}, Q, k)$ reveals that it is a function of the re-order point at the RDCs and NDC, i.e., (r_i) and r_n . Furthermore, in order to calculate the re-order points, r_i and r_n for each RDC i and NDC n , values of (\mathbf{A}, Q, k) is needed (see inequalities (4) and (5) of the fill-rate problem P^f). We tackle this circular behavior between the decision variables by defining a hierarchical solution approach.

1. In the first stage, a value is fixed for each r_i, r_n and the stationary point for the objective function $\gamma(\mathbf{A}, Q, k)$ is calculated. The solution so obtained is adjusted to satisfy the integrality constraint (inequality 6 in equation set (13)) for $(C_{n_i}/A_{r_i}, Q, k)$. This gives a near optimal solution for problem P^f for a fixed value of r_i and r_n .
2. In the second stage of the problem, the optimal values of r_i and r_n are calculated for each RDC i and NDC n , using inequalities (4) and (5) from the equation set (13), for the values of (\mathbf{A}, Q, k) derived in the first phase. Using these new values of r_i and r_n , the first phase is solved again and the procedure is repeated till it converges.

The solution procedure for solving the problem in the first phase is similar to the one used in Mangotra et al. (2009). An iterative procedure is used to solve the partially relaxed version of the problem (ignoring the integrality constraint). Once a solution is obtained adjustments are made to incorporate integrality. The values of the optimal re-order points are calculated using the *Goolseek* tool in excel.

The objective function of problem P^f is nonlinear but it is shown to be convex over a certain region defined by inequalities given by (14), and is biconvex.

Result 3: $\gamma(\mathbf{A}, Q, k)$ is a convex function for values of (\mathbf{A}, Q, k) satisfying the following inequalities given by (14) (see appendix for proof).

$$\begin{aligned} & \left(\frac{\partial^2 \phi}{\partial Q^2} - \sum_{i=1}^N \frac{\partial^2 \phi / \partial Q \partial A_{r_i}}{\partial^2 \phi / \partial A_{r_i}^2} \frac{\partial^2 \phi}{\partial A_{r_i} \partial Q} \right) > 0 \\ & \left(\frac{\partial^2 \phi}{\partial Q^2} - \sum_{i=1}^N \frac{\partial^2 \phi / \partial Q \partial A_{r_i}}{\partial^2 \phi / \partial A_{r_i}^2} \frac{\partial^2 \phi}{\partial A_{r_i} \partial Q} \right) \frac{\partial^2 \phi}{\partial k^2} - \left(\frac{\partial^2 \phi}{\partial Q \partial k} \frac{\partial^2 \phi}{\partial k \partial Q} \right) > 0 \end{aligned} \quad (14)$$

Result 4: $\gamma(\mathbf{A}, Q, k)$ is a *biconvex* function for all values of Q and k , and values of \mathbf{A} satisfying equation (1) (see appendix for proof).

A partially unconstrained version is solved first and this solution is modified to get a near optimal solution for problem P^f . The partially unconstrained version of the problem ignores the integer value constraint (\mathbf{A}, Q, k) .

The stationary point of the objective function $\gamma(\mathbf{A}, Q, k)$ are given by:

$$A_{r_i} = \left(\frac{2F_r + h_r(Q + 2r_i)}{C_{l,f_r} \xi \lambda \delta_i} \right)^{2/3} \quad (15)$$

$$k = \frac{1}{Q} \sqrt{(2R_n) \left(\frac{\sum_{i=1}^N \xi \lambda \delta_i C_{n_i}}{(h_n - 2h_r(1 - \rho_n))} \right)} \quad (16)$$

$$Q = \sqrt{2 \left(\frac{(C_f + R_r + \frac{R_n}{k}) \sum_{i=1}^N \xi \lambda \delta_i C_{n_i}}{\sum_{i=1}^N \frac{h_r C_{n_i}}{A_{r_i}} + \frac{h_n k}{2} - h_r(1 - \rho_n)k} \right)} \quad (17)$$

Result 5: The stationary point of $\gamma(\mathbf{A}, Q, k)$ is a local minimum.

Proof. The stationary point obtained by solving equations (15), (16) and (17) satisfy the inequalities given by (14). Then, by theorem 3 (see appendix) the result follows. \square

Results 4 and 5 together imply that the stationary point of $\gamma(\mathbf{A}, Q, k)$ is indeed a cost minimizing solution of problem P^f .

For a fixed value of \mathbf{A} , an optimal solution (Q, k) for the unconstrained problem is obtained by simultaneously solving equations (16) and (17). The equations can be solved simultaneously using an iterative procedure and the solution generated is substituted in equation (15) to obtain an optimal value of \mathbf{A} and the procedure is repeated again with the value of \mathbf{A} . If the procedure terminates in finite time, then a stationary point is obtained. Next the stationary point is checked for compatibility with the inequalities given by (14). In case both the inequalities are satisfied, then the stationary point is an optimal solution for the partially unconstrained problem. From this solution an optimal solution to problem P^f is generated by forcing the integer value constraint (\mathbf{A}, Q, k) .

The steps for the iterative procedure are explained below:

1. Fix $k = 0$, $Q_k = 1$.
2. Calculate A_{r_i} , $i = 1, 2, \dots, N$, using equation (15).
3. Use the value of A_{r_i} , $i = 1, 2, \dots, N$, in equation (17) to get Q and calculate k using equation (16). Iterate between the values of Q and k till they converge.
4. If $Q = Q_k$, Stop go to step 5. Else $k = k + 1$, and $Q_k = Q^*$ repeat Step 2.
5. If all A_{r_i} are integers, go to step 6, else for all non-integer A_{r_i} get all possible combinations of $\lceil A_{r_i} \rceil$ and $\lfloor A_{r_i} \rfloor$. For each set of new A_{r_i} , get Q and k using step 3.
6. Adjust Q and k to get the nearest integer value. Evaluate the objective function at each set of values of A_{r_i} , Q and k . The set corresponding to the minimum value is the solution.

Table 1: Cost comparison for the different service measures-equal Q

	Type-I service model	Type-II service model
RDC	9	8
TI_r	3545200	2669890
TI_n	1042660	1095040
TF	90000	80000
TIT	2682650	2682650
TOT	858030	991731
TNC	8218541	7442170

Table 2: Inventory Parameters for Type-I service model

Q_n	27							
r_n	455							
zone	1	2	3	4	5	6	7	8
Q	2563	2563	2563	2563	2563	2563	2563	2563
r_r	10003	11068	3889	3110	2296	33809	5804	3083

5 Numerical illustration

For the numerical study in this chapter, the distribution network for a leading US retailer is considered. The entire US mainland has five sub-regions, namely, south-eastern, south-western, north-eastern, north western and mid-west. The distribution network has a total of five NDCs each serving one of the sub-regions. In this study we focus on the southeastern (SE) region, with the NDC is located at Savannah, Georgia, for generating results and developing insights. Furthermore, the SE region is partitioned into eight clusters using cluster analysis discussed in Mangotra et al. (2009).

Table 1 presents the results for the problem when a *Type-II* service level measure is used. In this analysis, a 99% service level is assumed at each RDC and a 75% service level at the NDC. These results support our claim that the choice of service level measure affects both the network design and inventory parameters. For the *Type-I* measure, the optimal network has 9 RDCs and for the *Type-II* the optimal network has 8 RDCs. The inventory parameters for each RDC and the NDC are also different under the different service level assumptions and are presented in tables 2 and 3.

The results in figure 1 show that the total network cost are lower for the *Type-II* service model. This happens because the average safety stock is significantly lower under the given set of input parameters used in this analysis (see appendix). A comparison of the safety stock and the total cost for each zone under the two different service level measures are given in figures 2 and 3.

Table 3: Inventory at RDC and NDC for a Type-II service model

Q_n	16							
r_n	388							
zone	1	2	3	4	5	6	7	8
Q	2978	2978	2978	2978	2978	2978	2978	2978
r_r	10327	24370	3659	2853	2030	38387	5694	2826

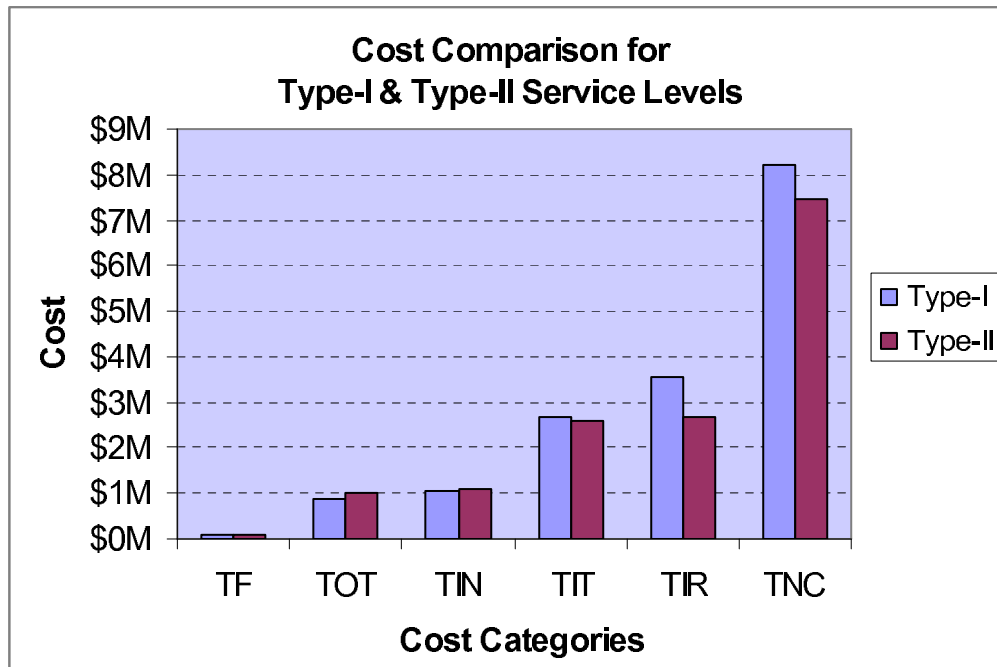


Figure 1: Cost Comparison by categories for Type-I and Type-II Service Levels

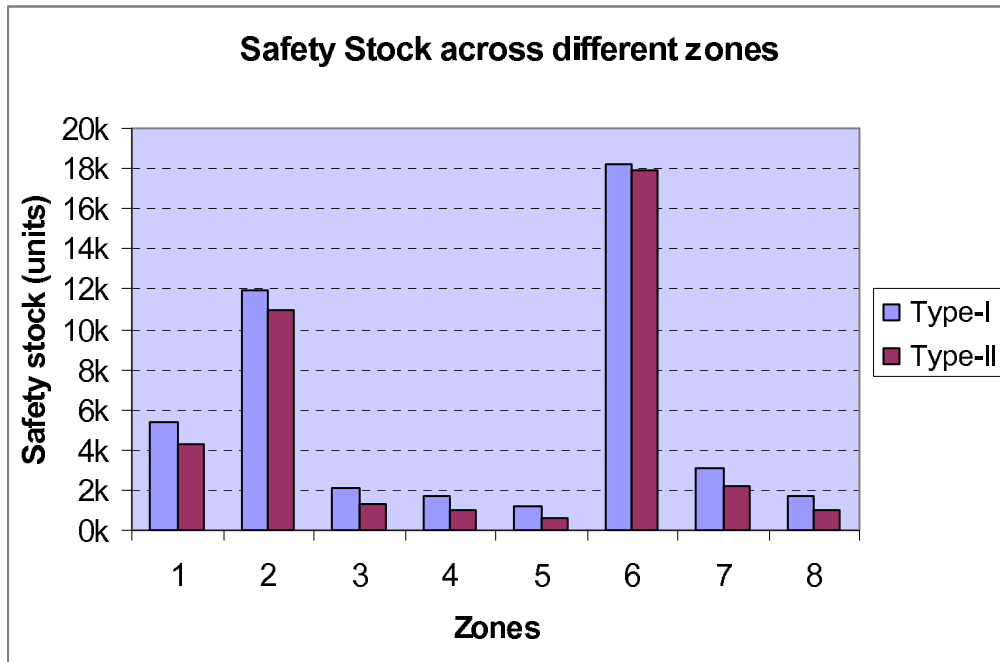


Figure 2: Safety Stock comparison across different zones

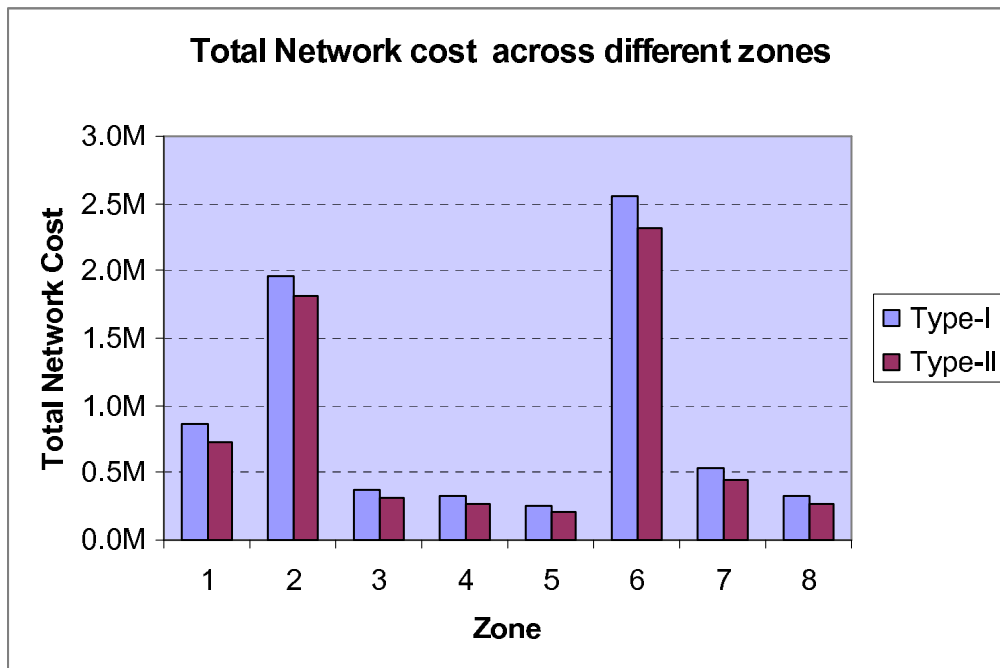


Figure 3: Total Network Cost comparison across different zones

Table 4: Parameter Data

F_r (\$/day)	10000	λ (units/day)	20
δ	0.006	C_l	0.009
ξ	360	C_f	400
f_r	0.3	C_v	0.6
R_r	100	R_n	10000
h_r	60	h_n	15
(μ_r, σ_r)	(4, 2)	(μ_n, σ_n)	(120, 10)
(α_r, Z_r)	(0.99, 2.326)	(α_n, Z_n)	(0.95, 1.645)

5.1 Conclusions

In this paper, a fill-rate (*Type-II service*) model is presented for the integrated facility location and allocation problem. This model differs from the *Type-I service* model in the way the safety stock term is modeled into the objective function. A hierarchical solution approach is proposed for solving the *Type-II service* model which requires an iterative solution procedure in the first phase and re-order point updates in the second phase. The numerical study is presented using data from a leading US retailer's south-eastern region which compares the network design and inventory results under the different service level models. Our results suggest that the choice of service level policy impacts the network design as well as inventory allocation decision.

This work is based on some simplified assumptions to make the problem tractable and enable us to derive insights. It would be interesting to relax some of them to match real-world scenarios, such as capacity limitations on DCs, multiple products and other inventory policies. In this analysis it is assumed that the NDC serves each cluster independent of demand behavior at other clusters in the sub-region. This can happen in real world when the NDC decides to review each cluster periodically. It would like interesting to see how using a combination of periodic review policy at each NDC and continuous review policy at each RDC would affect the network design and costs.

6 Appendix

6.1 Table 4 presents parameters used in the numerical example

6.2 Theorems from Optimization used in this work

Little's Law (Little, 1961): If n is the average number of customers in the system and τ is the average arrival rate, then the expected waiting time t is given by equation 18.

$$t = \frac{n}{\tau} \quad (18)$$

Theorem 1 (Bazaraa et al. (1993), pg. 96-97): Let

$$H = \begin{bmatrix} h_{11} & q^t \\ q & G \end{bmatrix}$$

where $q = 0$ if $h_{11} = 0$ and, otherwise, $h_{11} > 0$. Perform elementary Gauss-Jordan operations using the first row of H to reduce it to the following matrix in either case:

$$H = \begin{bmatrix} h_{11} & q^t \\ 0 & G_{new} \end{bmatrix}$$

Then, G_{new} is a symmetric $(n-1) \times (n-1)$ matrix, and H is positive semidefinite if and only if G_{new} is positive semidefinite. Moreover, if $h_{11} > 0$, then H is positive semidefinite if and only if G_{new} is positive semidefinite.

Theorem 2 (Bazaraa et al. (1993), pg. 91): Let S be a nonempty open convex set and let $f: S \Rightarrow E_1$ be twice differentiable on S . Then, f is convex if and only if the Hessian matrix is positive semidefinite at each point in S .

Theorem 3 (Bazaraa et al. (1993), pg 134): Suppose that $f: E_n \rightarrow E_1$ is twice differentiable at \bar{x} . If $\nabla f(\bar{x}) = 0$ and $H(\bar{x})$ is positive definite, then \bar{x} is a strict local minimum.

Next, we present proofs for Results (4) and (5). The first order conditions for deriving stationary point for function $\gamma(\mathbf{A}, Q, k)$ is given as:

$$\begin{aligned} \frac{\partial \gamma}{\partial A_{r_i}} &= \frac{C_l f_r \xi \lambda \delta_i C_{n_i}}{2A_{r_i}^{1/2}} - \frac{2C_{n_i} F_r + C_{n_i} h_r (Q + 2r_i)}{2A_{r_i}^2} = 0 \\ \frac{\partial \gamma}{\partial Q} &= -\frac{(\sum_{i=1}^N (C_f + R_r + R_n/k) \xi \lambda \delta_i C_{n_i})}{Q^2} \\ &+ \sum_{i=1}^N \left(\frac{h_r C_{n_i}}{2A_{r_i}} + \frac{h_n k}{2} \right) = 0 \\ \frac{\partial \gamma}{\partial k} &= \frac{h_n Q}{2} - (h_r (1 - \rho_n) Q) - \frac{\sum_{i=1}^N R_n \xi \lambda \delta_i C_{n_i}}{Q k^2} = 0 \end{aligned}$$

The Hessian matrix corresponding to the function γ is given by:

$$H = \begin{bmatrix} a_{1,1} & 0 & \dots & 0 & a_1 q & 0 \\ 0 & a_{2,2} & \dots & 0 & a_2 q & 0 \\ \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ \cdot & \cdot & \dots & \cdot & \cdot & \cdot \\ 0 & 0 & \dots & a_{n,n} & a_n q_n & 0 \\ qa_1 & qa_2 & \dots & qa_N & qq & qq_n \\ 0 & 0 & \dots & 0 & q_n q & q_n q_n \end{bmatrix}$$

where

$$\begin{aligned}
 a_{i,i} &= \frac{\partial^2 \phi}{\partial A_{r_i}^2}, & qq &= \frac{\partial^2 \phi}{\partial Q^2}, & q_n q_n &= \frac{\partial^2 \phi}{\partial k^2} \\
 a_i q &= \frac{\partial^2 \phi}{\partial A_{r_i} \partial Q}, & q a_i &= \frac{\partial^2 \phi}{\partial Q \partial A_{r_i}}, & q_n q &= \frac{\partial^2 \phi}{\partial k \partial Q_{r_i}} \\
 qq_n &= \frac{\partial^2 \phi}{\partial Q \partial k}, & a_i q_n &= \frac{\partial^2 \phi}{\partial A_{r_i} \partial k}, & q_n a_i &= \frac{\partial^2 \phi}{\partial k \partial A_{r_i}} \\
 & & i &= 1, 2, \dots, N
 \end{aligned}$$

Convex region for $\gamma(\mathbf{A}, Q, k)$

Using theorem 2, $\gamma(\mathbf{A}, Q, k)$ is convex *iff* the hessian matrix of γ is positive semidefinite. And from theorem 4.1, hessian matrix of γ is positive definite for values of (\mathbf{A}, Q, k) satisfying:

$$\begin{aligned}
 |G| &= \left| \begin{array}{cc} \left(\frac{\partial^2 \gamma}{\partial Q^2} - \sum_{i=1}^N \frac{\partial^2 \phi / \partial Q \partial A_{r_i}}{\partial^2 \gamma / \partial A_{r_i}^2} \frac{\partial^2 \gamma}{\partial A_{r_i} \partial Q} \right) & \frac{\partial^2 \gamma}{\partial Q \partial k} \\ \frac{\partial^2 \gamma}{\partial k \partial Q} & \frac{\partial^2 \gamma}{\partial k \partial k} \end{array} \right| > 0 \quad \text{and} \\
 &\left(\frac{\partial^2 \gamma}{\partial Q^2} - \sum_{i=1}^N \frac{\partial^2 \phi / \partial Q \partial A_{r_i}}{\partial^2 \phi / \partial A_{r_i}^2} \frac{\partial^2 \gamma}{\partial A_{r_i} \partial Q} \right) > 0
 \end{aligned}$$

We show that the function $\gamma(\mathbf{A}, Q, k)$ is biconvex using 2(a) and 2(b).

(2a) For a fixed value of (Q, k) , the function $\gamma(\mathbf{A}, Q, k)$ is convex in \mathbf{A} .

$$\begin{aligned}
 \frac{\partial \gamma^2}{\partial A_{r_i}^2} &= -\frac{C_l f_r \xi \lambda \delta_i C_{n_i}}{4A_{r_i}^{3/2}} + \frac{2C_{n_i} F_r + C_{n_i} h_r (Q + 2r_i)}{2A_{r_i}^3} \\
 &> 0 \Leftrightarrow A_{r_i} < \left(\frac{1}{2}\right)^{2/3} \left(\frac{2F_r + h_r (Q + 2r_i)}{C_l f_r \xi \lambda \delta_i} \right) \\
 &\text{which holds for } A_{r_i} = \left(\frac{2F_r + h_r (Q + 2r_i)}{C_l f_r \xi \lambda \delta_i} \right)
 \end{aligned}$$

□

(2b) For a fixed vector \mathbf{A} , the hessian matrix of $\gamma(\mathbf{A}, Q, k)$ is positive semidefinite.

$$\begin{aligned}
 \frac{\partial^2 \gamma}{\partial Q^2} &= 2 \left(\frac{(\sum_{i=1}^N (C_f + R_r + R_n / Q_n) \xi \lambda \delta_i C_{n_i})}{Q^3} \right) \\
 \frac{\partial^2 \gamma}{\partial k \partial Q} &= \sum_{i=1}^N \frac{R_n \xi \lambda \delta_i C_{n_i}}{Q^2 k^2} + \frac{h_n}{2} - h_r (1 - \rho_n) \\
 \frac{\partial^2 \gamma}{\partial k^2} &= 2 \left(\frac{\sum_{i=1}^N R_n \xi \lambda \delta_i C_{n_i}}{Q k^3} \right) \\
 \frac{\partial^2 \gamma}{\partial Q \partial k} &= \frac{\sum_{i=1}^N R_n \xi \lambda \delta_i C_{n_i}}{Q^2 k^2} + \frac{h_n}{2} - h_r (1 - \rho_n)
 \end{aligned}$$

$$\begin{aligned}
|H| &= \begin{bmatrix} \frac{\partial^2 \gamma}{\partial Q^2} & \frac{\partial^2 \phi}{\partial k \partial Q} \\ \frac{\partial^2 \gamma}{\partial Q \partial k} & \frac{\partial^2 \phi}{\partial k^2} \end{bmatrix} \\
|H| &= \left(2 \frac{\sum_{i=1}^N R_n \xi \lambda \delta_i C_{n_i}}{Q k^3} \right) \left(2 \frac{\sum_{i=1}^N R_n \xi \lambda \delta_i C_{n_i}}{Q k^3} \right) \\
&+ \left(2 \frac{\sum_{i=1}^N R_n \xi \lambda \delta_i C_{n_i}}{Q k^3} \right) \left(2 \frac{\sum_{i=1}^N (C_f + R_r) \xi \lambda \delta_i C_{n_i}}{Q^3} \right) \\
&- \left(\frac{\sum_{i=1}^N R_n \xi \lambda \delta_i C_{n_i}}{Q^2 k^2} + \frac{h_n}{2} \right) \left(\frac{\sum_{i=1}^N R_n \xi \lambda \delta_i C_{n_i}}{Q^2 k^2} + \frac{h_n}{2} \right) \\
&= 3 \left(\frac{\sum_{i=1}^N R_n \xi \lambda \delta_i C_{n_i}}{Q^2 k^2} \right)^2 + \left(\frac{\sum_{i=1}^N R_n \xi \lambda \delta_i C_{n_i}}{Q^2 k^2} \right)^* \\
&\quad \left[\frac{4}{Q^2 k} \sum_{i=1}^N (C_f + R_r) \xi \lambda \delta_i C_{n_i} - h_n \right] - \frac{h_n^2}{4} \\
|H| &\geq 0 \\
&\Leftrightarrow \frac{4}{Q^2 k} \sum_{i=1}^N (C_f + R_r) \xi \lambda \delta_i C_{n_i} \geq h_n \quad \text{and} \\
&\quad 3 \left(\frac{\sum_{i=1}^N R_n \xi \lambda \delta_i C_{n_i}}{Q^2 k^2} \right)^2 \geq \frac{h_n^2}{4}
\end{aligned}$$

Then by Theorem 2, we can say that for a fixed vector \mathbf{A} , the function $\gamma(\mathbf{A}, Q, k)$ is convex in (Q, k) . \square

References

- Bazaraa, M., Sherali, H., Shetty, C., 1993. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, Inc.
- Burn, L., Hall, R., Blumenfeld, D. E., Daganzo, C. E., 1985. Distribution strategies that minimize transportation and inventory cost. *Operations Research* 33 (3), 469–490.
- Daganzo, C. F., 1996. *Logistics Systems Analysis*. Springer, Berlin.
- Dasci, A., Verter, V., 2001. A continuous model for production-distribution system design. *European Journal of Operational Research* 129, 287–298.
- Daskin, M., 1995. *Network and Discrete Location: Models, Algorithms and Applications*. John Wiley and Sons, New York.

- Daskin, M., Coullard, C., Shen, Z. M., 2002. An inventory-location model: Formulation, solution algorithm and computational results. *Annals of Operations Research* 110, 83–106.
- Deuermeyer, B., Schwarz, L. B., 1981. In *Studies in the Management Sciences, Multilevel Production/Inventory Control Systems*. Vol. 16. North-Holland, Amsterdam, Ch. A Model for the analysis of System Service level in Warehouse/Retailer Distribution Systems: The Identical Retailer Case., pp. 163–193.
- Drezner, Z., 1995. *Facility Location: A Survey of Applications and Methods*. Springer-Verlag, New York, NY.
- Erlenkotter, D., 1989. The general optimal market area model. *Annals of Operations Research* 18, 45–70.
- Ganeshan, R., 1999. Managing supply chain inventories: A multiple retailer, one warehouse, multiple supplier model. *International Journal of Production Economics* 59, 341–354.
- Geoffrion, A. M., 1976. The purpose of mathematical programming is insight, not numbers. *Interfaces* 7, 81–92.
- Hopp, W., Spearman, M., 2000. *Factory Physics*. McGraw Hill.
- Langevin, A., Mbaraga, P., Campbell, J. F., 1996. Continuous approximation models in freight distribution: an overview. *Transportation Research B* 30 (3), 163–88.
- Little, J., 1961. A proof of the queuing formula: $L = \lambda w$. *Operations Research* 9, 383–387.
- Mangotra, D., Lu, J., Tsao, Y., Sharp, H., 2009. A continuous approximation approach for the integrated facility-inventory allocation problem. Tech. rep., Georgia Tech.
- Miranda, P. A., Garrido, R. A., 2004. Incorporating inventory control decisions into a strategic distribution network design model with stochastic demand. *Transportation Research Part E* 40, 183–207.
- Newell, G. F., 1973. Scheduling, location, transportation and continuum mechanics: some simple approximations to optimization problems. *SIAM Journal of Applied Mathematics* 25.
- Nozick, L., Turnquist, M., 1998. Integrating inventory impacts into a fixed-charge model for locating distribution centers. *Transportation Research-E* 34 (3), 173–186.
- Nozick, L., Turnquist, M., 2001. Inventory, transportation, service quality and the location of distribution centers. *European Journal of Operations Research* 129, 362–371.
- Romeijn, H. E., Shu, J., Teo, C., 2007. Designing two-echelon supply networks. *European Journal of Operational Research* 178, 449–462.
- Schwarz, L. B., Deuermeyer, B. L., Badinelli, R. D., 1985. Fill-rate optimization in a one-warehouse n-identical retailer distribution system. *Management Science* 31 (4), 488–498.

- Shen, Z., 2005. A multi-commodity supply chain design problem. *IIE Transactions* 37, 753–762.
- Shen, Z., Daskin, M., 2005. Trade-offs between customer service and cost in integrated supply chain design. *Manufacturing & Service Operations Management* 7, 188–207.
- Shen, Z., Qi, L., 2007. Incorporating inventory and routing costs in strategic location models. *European Journal of Operational Research* 179, 372–389.
- Shen, Z. M., Coullard, C., Daskin, M. S., 2003. A joint location-inventory model. *Transportation Science* 37 (1), 40–55.
- Shu, J., Teo, C., Shen, Z., 2005. Stochastic transportation-inventory network design problem. *Operations Research* 53 (1), 48–60.
- Snyder, L., Daskin, M., Teo, C., 2007. The stochastic location model with risk pooling. *European Journal of Operational Research* 179.
- Teo, C., Shu, J., 2004. Warehouse-retailer network design problem. *Operations Research* 52 (3), 396–408.
- van Houtum, G., 2006. Multi-echelon production/inventory systems: Optimal policies, heuristics, and algorithms, tutorials in *Operations Research*.