A Continuous Approximation approach for the Integrated Facility Location-Inventory Allocation Problem

ISyE Technical Report J-09-01

Divya Mangotra, Jye-Chyi Lu, Yu-Chung Tsao
School of Industrial and Systems Engineering,
Georgia Institute of Technology, Atlanta, GA

Abstract

In today's business many companies have a complex distribution network with several national and regional distribution centers. In this paper, we study an integrated facility location and inventory allocation problem for designing a distribution network with multiple national distribution centers (NDCs) and retailers. The key decisions are where to locate the regional distribution centers (RDCs), how to assign retail stores to RDCs and what should be the inventory policy at the different locations such that the total network cost is minimized. We model our problem using a Type-I (probability of stock-outs) service level measure.

This paper presents a continuous approximation (CA) model for solving the problem described above. The model takes a nonlinear form and solution techniques are developed using the theory of nonlinear programming. The main contribution of this work lies in developing a refined CA modeling technique when the discrete data cannot be modeled by a continuous function. Our methodology is illustrated on a real life application of a leading US retailer. Numerical analysis suggests that the total network cost is significantly lower in the case of the integrated model as compared with the non-integrated model. It also shows that the regular CA approach leads to a solution which is inferior to the solution obtained by the modified CA approach.

Keywords: Supply chain design; Inventory; Facility location; Allocation; Continuous approximation model.

1 Introduction

Manufacturing outsourcing in the U.S. has never been stronger than it is today. Increased outsourcing has led to significant changes in the design of the retail distribution network. While the traditional distribution network had manufacturing plants supply goods to retail stores directly, off-shore manufacturing has increased the network's demand for transportation and warehousing to deliver the goods. When the goods arrive at seaports in the U.S., they must be consolidated by region at national (import) distribution centers. From these national distribution centers (NDCs), the goods are shipped to regional distribution facilities (RDCs), from which they are delivered to retail stores. These regional distribution facilities help pool risk by consolidating shipments from the import distribution centers (DCs). Most companies have complex distribution networks with several import and regional DCs. For example, Target, Inc. has 3 import warehouses, 22 regional distribution centers, and 1300 retail stores. Frito-Lay, Inc. operates its distribution network with 42 plants, one national DC, and 325 regional DCs (see Erlebacher *et al.* (13)).

Companies in the U.S. spend approximately \$14 billion per year on inventory interest, insurance, taxes, depreciation, obsolescence, and warehousing. Their logistics activities account for 15%-20% of the total cost of finished goods (Menlo, 2007). With such a huge inventory investment and the growing demand for warehouses, it is important to make optimal decisions for facility locations and inventory allocation in a supply chain.

Inventory is a key driver of supply chain performance, which is measured in terms of fill-rate and in-stock probability. It is for this reason that optimal inventory allocation along the different levels of the supply chain is important. This problem is known as the multi-echelon inventory problem (see Roundy (23); Deuermeyer *et al.* (11); Ganeshan (15) for a detailed review). This body of literature assumes that distribution centers have been located optimally between the manufacturer and the stores prior to the inventory decisions, and hence ignores the facility location cost. The facility location-allocation problem has been extensively studied (see Daskin (10); Drezner (12); Brandeau *et al.* (4)). However, this body of literature ignores inventory allocation decisions at the DCs.

The two key decisions of facility location and inventory allocation are dependent on each other. One of the key cost components for the facility location problem is the transportation cost, which depends on the frequency of inventory replenishment at different facilities. This replenishment frequency is a function of the inventory policy. Similarly, the inventory allocation problem models the inventory cost at the distribution center (DC) for a known value of demand served by each DC. This requires information regarding which retailers are assigned to which DC.

The interrelations between the facility location and inventory policy problems suggest that an integrated model with the facility, transportation costs, and inventory costs is needed to solve network design problems. In spite of this well understood dependence, most studies deal with these two problems individually because of the sheer complexity of integrating them. The integrated problem takes a non-linear form and the number of decision variables is enormous, making it computationally challenging to solve the problem for a real network. The studies by Teo *et al.* (26) and Teo *et al.* (22) are the only two papers we have found in the area of integrated logistic network design. Though these models are very detailed, the problem is only solved on a small scale and it does not account for supply uncertainty in the network.

Unlike previous studies, which use discrete models, this study adopts a continuous approximation (CA) approach to model the supply network design problem. The key idea under the CA approach is to define decision variables using continuous functions and hence reduce the complexity of the problem. While the CA approach does not determine the exact location of the distribution centers, it defines a service area for each distribution center in terms of circular influence areas. Earlier studies by Newell (19) and Dasci *et al.* (9) show that influence areas with central distribution nodes is a near optimal solution. The goal of this study is to provide logistics network planners with a high level solution for the integrated facility location and inventory allocation problem. To the best of our knowledge, this problem area is new, and our work presents the most detailed non-linear cost model. We propose a solution technique that preserves the interrelation between facility and inventory decisions.

Most papers in the area of network design use a discrete model to formulate the problem. Discrete models provide managers with optimal solutions, but their data and computational requirements increase tremendously as they try to capture real operational networks. Also, data reliability and model accuracy decrease as the amount of data increases. Continuous approximation could be a remedy to these shortcomings, as it requires less data to generate closed or near-closed form solutions (Dasci *et al.* (9)). Even though our modeling approach requires us to sacrifice the level of detail by replacing discrete variables with smooth functions, the resulting model more closely depicts real logistic networks.

The main purpose of this paper is to meet a three-fold goal: First, to highlight the importance of integrating facility location decisions with inventory decisions. We show that a non-integrated approach generates results that have a significantly higher total network cost compared to an integrated approach. Second, to present a solution approach, namely, continuous approximation, for solving the integrated facility location and inventory allocation problems with non-homogenous data. Third, propose a methodology for fine-tuning the continuous approximation technique when the input variables cannot be approximated by a smooth function.

This paper presents a continuous approximation model for solving the network design problem and answers the following questions: (1) which RDC locations should be open, (2) which retail store should be served from which RDC location, and (3) how much inventory should be held at the RDCs and the NDCs? The motivation for this approximation is that if we can cut down the size of data, then we can solve larger-scale problems to get some meaningful insights. Our solution defines the input data in terms of continuous functions and is capable of formulating these functions for a data set of any size.

The objective function proposed 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 requirements at each inventory stocking level.

2 Literature Review

Integrated network design and inventory policy decisions

There are several papers in the area of integrated facility location and single location inventory control. The research in this area considers distribution networks with a single plant serving multiple retailers. The distribution center (DC) location and inventory policy at the DC are both decision variables. It is assumed that each retailer has a variable demand process. Since the addition of inventory terms makes the objective function nonlinear, researchers have looked at approximations to linearize it.

Nozick *et al.* (20) approximated the safety stock cost at each DC by a linear regression function of the number of DCs, and uses this to estimate the inventory cost function. In their model, inventory is stocked at the DC and replenished using a one-for-one policy.

The fixed-charge facility location model defined in Daskin (10) uses the linear inventory cost function to determine the least cost set of DC locations. Nozick *et al.* (21) extended their previous model by adding service responsiveness and uncertainty in delivery time to the DC. Service responsiveness is defined in terms of stock-outs and time-based delivery. Stock-outs are incorporated in the safety stock function, while the time-based delivery constraint is modeled explicitly as coverage distance.

Shen *et al.* (25) studied a distribution network in which some of the retailers are allowed to act as distribution centers to achieve risk-pooling benefits in terms of inventory cost savings. Their problem determines which retailers should serve as DCs and how much inventory these stocking points should hold. The inventory model in their work is the continuous review (Q, r) model with a *Type-I service* constraint. They reformulate the nonlinear problem as a set-covering model, and propose a column generation algorithm that can solve the problem exactly for two special cases in $O(|n|^2 log|n|)$.

Miranda *et al.* (18) presented an integrated model for capacitated facility location problem (CFLP) and inventory control decisions. Their model determines the location of each distribution center based on the (Q, r) inventory policy at each DC location. Their solution methodology involves a lagrangian relaxation and the sub-gradient method. In another study, Erlebacher *et al.* (13) examined a distribution system design problem in which customer demand is distributed uniformly along a grid network. They proposed a two-stage heuristic procedure that fixes the number of DCs in the first stage to estimate DC demand, which the second stage then uses to estimate the number of DCs.

More recently, Teo *et al.* (26) studied an integrated logistic network problem that considers inventory cost for multiple echelons of inventory stocking locations. Their approach models the inventory cost at each DC and retailer. They use the convex inventory minimization function proposed by Roundy (23) along with transportation and facility costs, to formulate a MIP problem. They proposed a column generation technique to solve their model. Their solution is solvable in O(nlog(n)) time, and is within 2% of the optimal solution when the problem instance is small (20 warehouses and 100 retailers). However, their model does not include demand or supply uncertainty. In another study, Teo *et al.* (22) extended their previous model by adding safety stock terms to account for demand variability. In our work, we solve the integrated logistic design (i.e., facility location and inventory allocation) problem using two different approximation models for a scenario with 284 retail stores representing the southeastern region of a major US retail chains distribution network.

Data approximation for logistic network design

This line of research began to appear in the early 1970s in a seminal paper by Newell (19) that uses data approximation techniques for warehouse location problem. Geoffrion (16) studied a continuous model for warehouse location in which a warehouse serves demand that is distributed uniformly over a plane. Erlenkotter (14) used a General Optimal Market Area (GOMA) model to determine the optimal area served by a single production point when the demand is assumed to be distributed uniformly. This was an extension of a previous study by Geoffrion (16) and Newell (19), with more detailed expressions for the production cost. Rutten *et al.* (24) further refined the GOMA model by considering a distribution network and adding inventory cost terms. Burn *et al.* (5) studied a distribution network with a single supplier and multiple customers. They proposed an analytic method that uses the spatial density of customers to minimize the inventory and transportation cost of freight. Their work considers two different distribution strategies, direct shipping and peddling. Daganzo (8) presented continuum approximation techniques for the network design problem, focusing in particular on vehicle dispatch scheduling. Langevin *et al.* (17) presented an extensive review of continuous approximation models developed for freight distribution problems. Dasci *et al.* (9) studied a production and distribution design problem using the continuum approximation technique. Their work explicitly models facility costs by looking at the operational and acquisition cost components. The model presented in their work is an extension of a continuous approximation model for the facility design problem. However, their approach does not consider inventory costs.

Spatial Approximation

To the best of our knowledge, there are only two papers that study fine refinements of the continuous space models. These refinements are necessary when the underlying assumptions for the continuous models fail to hold. Blumenfeld *et al.* (3) studied logistics planning models that use continuous space models under general conditions, i.e., by relaxing the assumption of uniform density for stores. They developed an analytical framework for estimating the transportation costs for distributing goods from a single origin to multiple destinations, using clusters to account for dense customer destinations. These clusters can be analyzed as sub-regions of the main region. Wang et al. (27) studied spatial modeling and proposed smoothing techniques for non-homogeneous processes by considering details at different levels of the distribution network. Their work proposes fine refinements to the approximation models based on the level of detail captured by the data. We present a two-phase approximation technique to solve the integrated facility location and inventory allocation problem. The proposed approach captures the non-homogeneity of input parameters discussed by Blumenfeld *et al.* (3) and Wang et al. (27).

3 Total Network cost function

The network under study in this thesis is a three-level distribution system with retail stores at level zero meeting demand of the end customers. The regional distribution centers (RDCs) are located at level one to help consolidate shipments and pool risk. The national distribution centers (NDCs) are located at level two and they help consolidate shipments arriving from overseas manufacturers and deliver them to the RDCs. The goods flow from the facilities at the higher level to the facilities at the lower level until they reach level zero (see Figure 1). We will refer to this three-level network as a logistic network in the rest of the analysis.

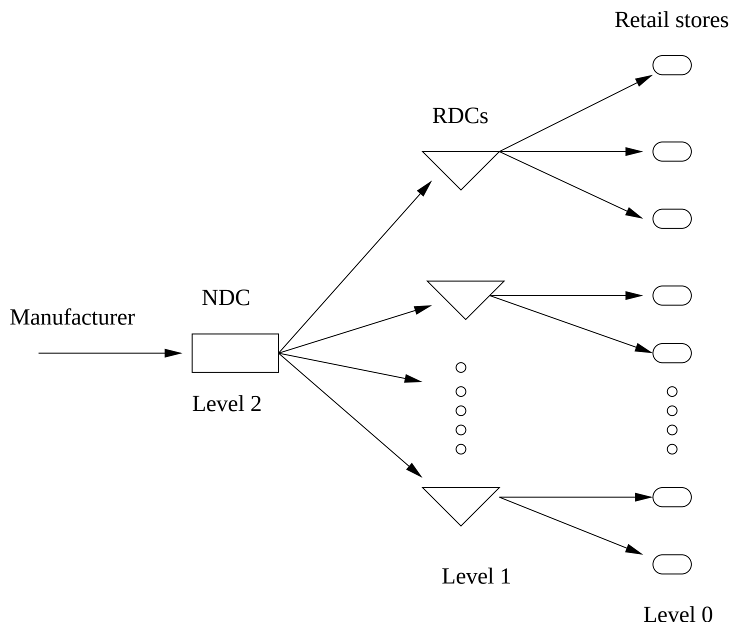


Figure 1: A Multi-Level Distribution Network

4 Assumptions

Before we model and solve the integrated facility location and inventory allocation problem, we want to make some assumptions around the network structure, demand pattern and inventory replenishment policies. Our assumptions help us frame the different logistics cost functions and hence create a practical model to analyze. These assumptions are needed to simply the complex logistics model without sacrificing our understanding of the problem. The simplified model structure also allows us to explore the different decision issues in detail. Most of our assumptions are consistent with those used extensively in the literature (see Ganeshan (15), Teo *et al.* (26), Dasci *et al.* (9)). The mathematical model in this paper is developed around the following assumptions.

1. The distribution network under study is an *arborescence network* (see Figure 1) in which each facility can serve multiple facilities in the lower level but can be served by only one facility from the upper level.

2. The location of the NDC is known and fixed.

3. Demand per unit time for each store is an independent and identically distributed Poisson process with rate λ.

4. Each product can be analyzed independent of other products. The demand for a single product is considered in our study.

5. The demand process at each RDC is a Poisson process as it is generated by the demand coming from the stores in its influence area. There is no reorder cost at the stores so that the demand at the store gets passed over to the RDC on a per item basis.

6. There is no lateral shipment of goods, i.e., movement of goods between facilities in the same echelon. Moreover, each facility serves its immediate lower echelon facilities via direct shipment.

7. We do not consider the pipeline inventory cost for units in transit from NDC to RDC or from suppliers to the NDC.

8. Each RDC's influence area is circular. It has been shown in the literature (Dasci *et al.* (9)) that the shape of the service area has little impact on the optimal solution. Moreover, each RDC is located in the center of the influence area.

9. The distances between the NDCs and the RDCs, and between the RDCs and the retail stores are calculated using the Euclidean norm.

10. The contraints from capacity limitations at the NDCs and the RDCs are not considered.

11. The inventory policy at the RDCs and the NDCs is a continuous review policy. Each RDC r implements the (Q_r, r_r) ordering policy, i.e., an order of Q_r units is placed everytime the *inventory position* [1] equals r_r. The NDC n implements (Q_n, r_n) policy, i.e., it orders Q_n units everytime its inventory position equals r_n.

12. Both RDC and NDC operate under a *Type-I* service level policy.

All the cost functions are modeled using a continuous approximation technique. The key idea under this technique is to express the entire distribution network in terms of smooth continuous functions. Let the distribution network under study be represented by a continuous service area R, and the discrete store locations be expressed as a spatial density function $\delta(x)$, $x \in R$. If the demand at the stores is expressed as a spatial density function $\lambda(x)$, $x \in R$, then the customer demand at each point $x \in R$ can be expressed as a product of the store density and the store demand density, and is given by $\lambda(x)\delta(x)$, $x \in R$. It is argued in Daganzo *et al.* (6) that if the customer demand is a slow varying function of x then the influence area of each RDC can be approximated by a circular region and it is a slow varying function of x. Influence area in this analysis is a region such that all the stores located within this region are served by the RDC located at the center. Let $A_r(x)$ be the influence area associated with RDC r. If we cover the entire area of the distribution network with circular influence areas of size $A_r(x)$, then the total number of RDCs ($N_r(x)$) is given by $\int_R (A_r(x))^{-1} dx$.

In our model, the components of the total network cost are calculated as follows:

[1] sum of on-hand and on-order inventory

- **Total Facility Cost**: A fixed rent, F_r, is 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,

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

- **Inbound Transportation and Outbound Delivery cost**: 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 reorder 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:

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

  where $(C_f+C_v Q_r(x))$ is the transportation cost incured in a single inbound shipment to a single RDC. The expected demand faced by RDC r is given by $E[D_r(x)]$, 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, Dasci *et al.* (8, 9)). Then the total outbound local delivery cost, *TOT(x)*, is given by

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

  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 *et al.* (9)).

- **Average Inventory cost for RDC**: Each RDC r orders in batches of $Q_r(x)$ and there is a reorder cost, $R_r(x)$, associated with each batch. The total reorder cost, $TR_r(x)$, for all the RDCs over the planning horizon is given by

$$TR_r(x) = N_r(x) \left(R_r(x) \right) \left(\frac{E[D_r]}{Q_r(x)} \right) \tag{4}$$

Average inventory at the RDC is given as the sum of the cycle inventory $Q_r(x)/2$ and safety inventory $(Z_{\alpha_r}\sqrt{Var[D_{r,LT}]})$. where $Var[D_{r,LT}] = \mu_r Var[D_r] + {\sigma_r}^2 E[D_r]^2$ and $E[D_{r,LT}] = \mu_r E[D_r]$, μ_r and ${\sigma_r}^2$ are the mean and the variance of the total order replenishment time respectively and α_r is the stock-out probability. Let h_r be the RDC inventory holding cost per item over the planning horizon ξ. Then the total RDC inventory holding cost, $TI_r(x)$, is given by

$$TI_r(x) = h_r N_r(x)\left(\frac{Q_r(x)}{2} + Z_{\alpha_r}\sqrt{Var[D_{r,LT}]}\right) + TR_r(x) \tag{5}$$

- **Average Inventory cost for NDC**: Each NDC n orders in batches of $Q_n(x)$ and there is a reorder cost, $R_n(x)$, associated with each batch. The reorder cost for each NDC, $TR_n(x)$, over the planning horizon is given by

$$TR_n(x) = R_n(x)\left(\frac{E[D_n(x)]}{Q_n(x)}\right) \tag{6}$$

where $E[D_n(x)]$ is the total expected demand at the NDC during the planning horizon and is given by $\xi\lambda(x)\delta(x)R$. Define the cycle inventory and safety inventory for the NDC as $Q_n(x)/2$ and $Z_{\alpha_n}\sqrt{Var[D_{n,LT}]}$ where $Var[D_{n,LT}] = \sum_r(\lambda(x)\delta(x)A_r(x))\mu_n)/(Q_r(x))^2$ (see Deuermeyer *et al.* (11)), μ_n is the expected lead time and α_n is the service level at the NDC. Let h_n be the inventory holding cost per item during the planning horizon ξ. Then the total NDC inventory holding cost is given by $TI_n(x)$ as

$$TI_n(x) = h_n\left(\frac{Q_n(x)}{2} + Z_{\alpha_n}\sqrt{Var[D_{n,LT}]}\right) + TR_n(x) \tag{7}$$

The cost expression derived in this section are in terms of each point x in the service region R. The total cost for the entire region is given by $\int_R(TNC(x))dx$, where $TNC(x)$ is the total network cost and is given by the sum of the facility, transportation and inventory cost functions. Each expression for the various cost components captures fine details of the network geometry. We can now define our integrated facility location and inventory allocation problem as

$$\text{minimize} \quad \int_R(TNC(x))dx = \int_R(TF(x) + TIT(x) + TOT(x) + TI_r(x) + TI_n(x))dx \tag{8}$$

s.t.

$$\begin{aligned}
\sum_r A_r(x) &= R && (9)\\
Q_n(x) &\geq 0 \ \forall \ n && (10)\\
Q_r(x) &\geq 0 \ \forall \ r && (11)\\
A_r(x) &\geq 0 \ \forall \ r && (12)\\
Q_n(x), \quad Q_r(x), &\quad A_r(x) \in Z^+ && (13)
\end{aligned}$$

where $Q_n(x)$, $Q_r(x)$, $A_r(x)$ are the decision variables in this problem. Equation (9) is the area coverage constraint. It ensures that the entire service region is covered by the sum of the RDC influence areas. Equation (10), (11) and (12) are the non-negativity constraints for the decision variables. Equation (13) guarantees integer values for $Q_n(x)$, $Q_r(x)$ and $A_r(x)$.

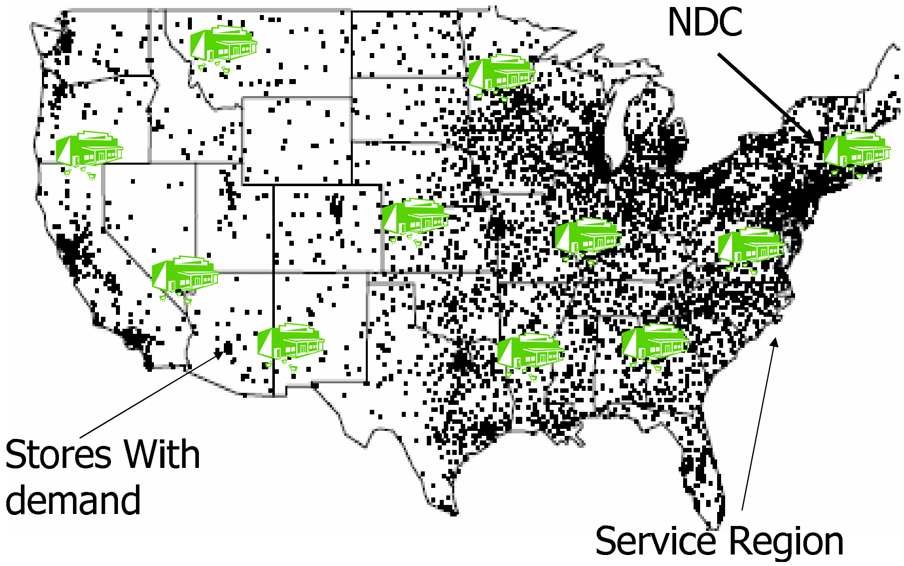
Note that any feasible solution, (A_{r_i}, Q, Q_n) for the optimization problem defined above should be strictly greater than 0. However, adding the equality condition in constraints (10), (11) and (12) does not change the solution. It follows from the observation that when $(A_{r_i}, Q, Q_n) = (0, 0, 0)$, the value of the objective function explodes (tends to infinity). Any feasible solution to the optimization problem above will be away from (0, 0, 0), so adding this point to the constraint set does not change the nature of the problem.

5 Solution Methodology - Two-phase approximation model

We now describe a two-phase approximation technique that used to solve the facility location and inventory allocation problem modeled in the previous section. Two-phase approximation in a extension to the continuous approximation (CA) approach (see Daganzo (8)). This extension is applicable when discrete data cannot be approximated with a smooth function as seen in the distribution network under study in this work (see Figure 2). The distribution network given in figure 2 shows the store locations for a leading automotive company in US. Clearly the store density in this figure is a non-homogenous Poisson process. This violates the *slow varying* property for the input function that is key for the analysis using the CA technique. A more detailed analysis of the store density data suggests that there are smaller areas over which these functions are smooth. Thus the main idea for a two-phase approximation method is to divide the network into smaller regions over which the discrete variable can be modeled using the slow varying functions. In phase-I the network is divided into smaller regions such that the distribution of store density over these sub-regions satisfy the *slow varying* property. The problem is modeled over the sub-regions using the cost functions described in ***section 3** and it is solved using the CA approach in phase-II.

5.1 Phase-I approximation: NDC Service Area and Grid Cover-Couple Approach

A *Grid Cover-Couple* approach is used to partition the service region into sub-regions. Suppose there are n NDCs in the service region. It is reasonable to assume that the total demand over the service region R is distributed equally amongst the NDCs. This problem of assigning equal demands to each NDC is a special case of the classic *Transportation*



Source: A leading Automobile company's Distribution network

Figure 2: Example of a Supply chain.

problem (see Appendix). The costs in this problem are modeled in terms of the distance from the NDC and a solution can be obtained by a greedy heuristic.

Let $(A_1, A_2, ...A_n)$ be the areas corresponding to the NDC partitions obtained after solving the assignment problem. The next step is to design a grid cover for each of these NDC sub-regions. It is this grid-cover that helps divide each NDC partition into regions with slow varying functions. A mesh of equal sized squares is designed to cover each NDC partition. The geometry of the square-mesh is an important decision and needs to satisfy the following conditions: 1) the smallest level of detail is captured at the *county*[2] level and 2) within each grid square the demand is slow varying. A trial and error method is used to choose a feasible size for the grid, e.g., we can look at all the county level demands and choose a county with the most variable demand. A square grid cover is designed for this county such that the store density within each grid is nearly constant. Choose the size of this square grid to form a grid cover for the entire NDC partition. This idea is illustrated in Figure 3. Note that a density can be assigned to each square on the grid because the store density for each county is known and county is the lowest level of detail captured by this grid-cover model.

Within each NDC partition, there are grids and each grid has a density associated with it. The grids with similar densities can be clustered together to form areas over which the store density function is *slow varying*. In order to form the clusters, a *tolerance limit* for similarity needs to be specified. The *tolerance limit* defines the amount of

[2]The term county is used in 46 of the 50 states of the United States for the tier of state government authority immediately below the statewide tier and above the township tier, in those states that subdivided counties into civil townships.

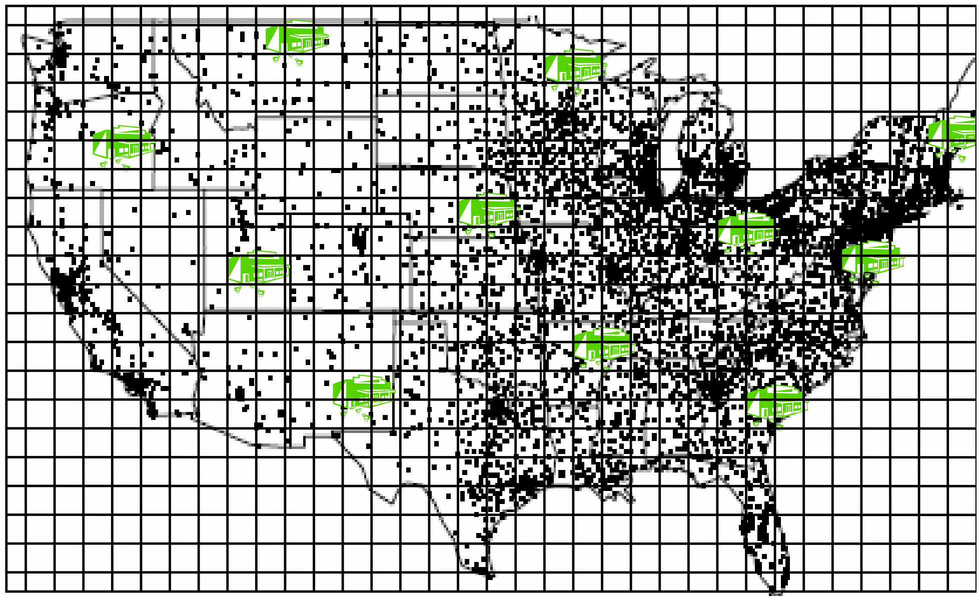


Figure 3: Grid Cover

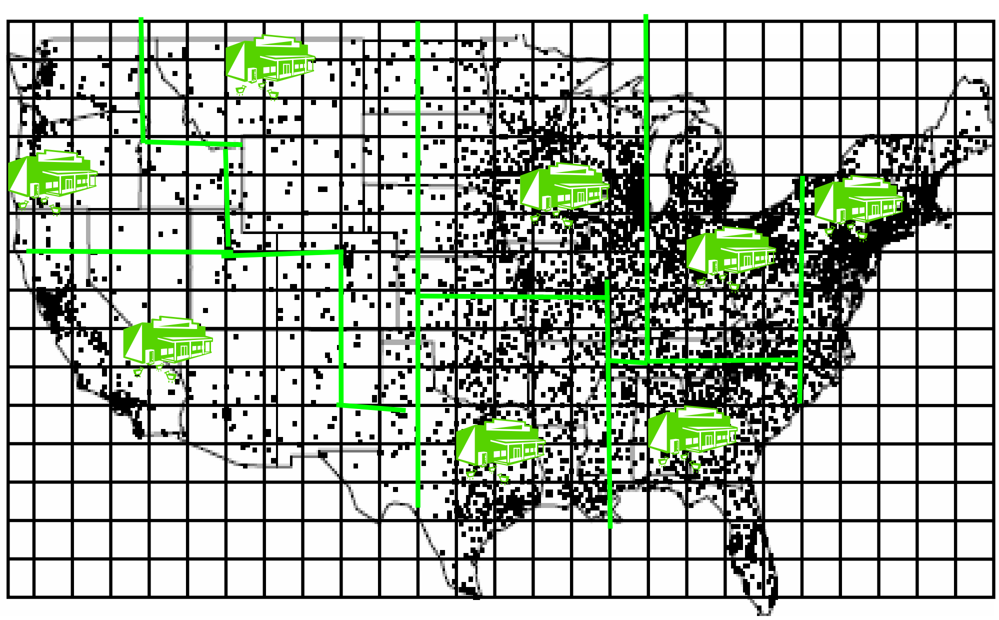


Figure 4: NDC sub-region.

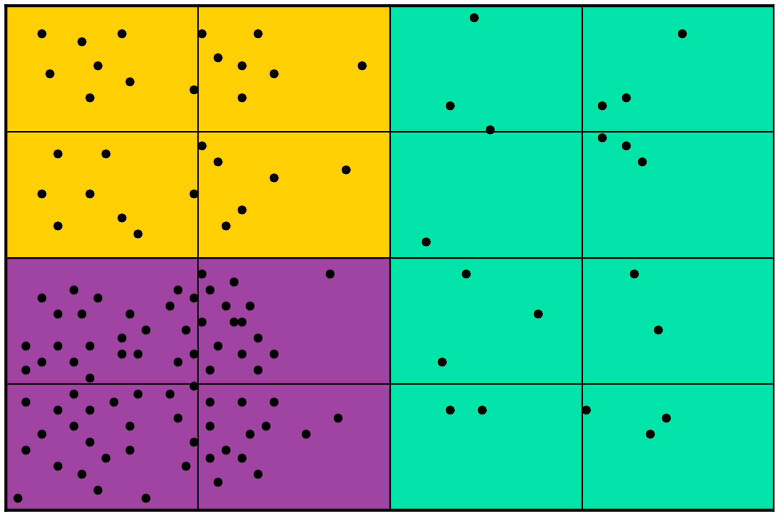


Figure 5: Coupling.

variability in the store density data that is acceptable while treating them as similar. Let ϵ be the desired tolerance limit. This means that the grids with density at most ϵ apart are considered similar. Choice of ϵ depends on the store density pattern in the existing distribution network. Using the tolerance limit the entire NDC sub-region is covered with clusters. Figure 4 and Figure 5 illustrate this idea. Clusters $(C_{j_1}, C_{j_2}, ...C_{j_i})$ exist within each NDC region A_i such that the store density is nearly constant over each cluster.

5.2 Phase-II approximation: RDC Influence Area using CA approach

The *phase-I* approximation divides the service region R into NDC partitions (sub-regions) $(A_1, A_2, ...A_n)$ and each partition A_i has clusters $(C_{j_1}, C_{j_2}, ...C_{j_i})$ with slow varying demand. The CA technique can be used to model and solve the facility location and inventory allocation problem over each cluster within the NDC partition. The optimization model developed in chapter 3 is used for modeling the total logistic costs in each cluster. The solution to this optimization problem will give the size of the circular influence area for each RDC (see Figure 6 for an illustration) and the optimal values of (Q, r) parameters for the RDC and the NDC. Further, using the size of the optimal influence area along with the information on the area for each cluster, the total number of RDCs in each cluster can be calculated. The total number of RDCs in the entire NDC partition is obtained by summing over the number of RDCs in each cluster.

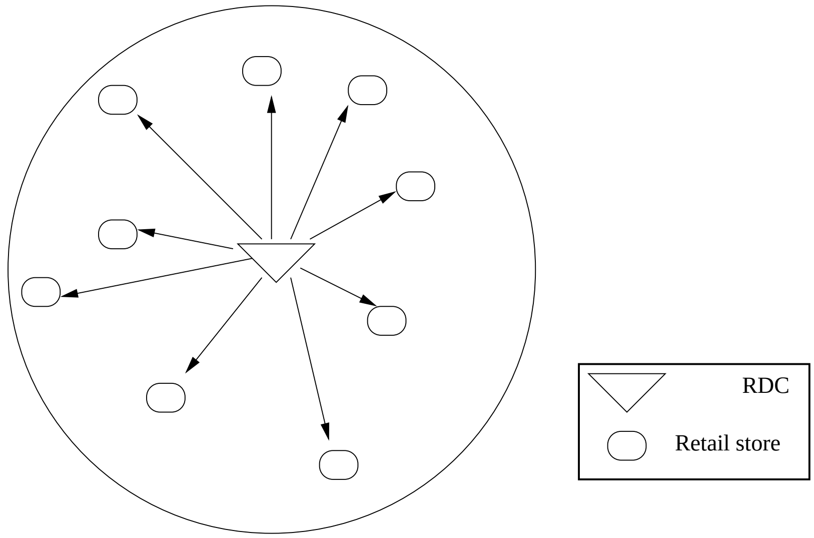


Figure 6: Influence area for a RDC

5.3 Continuous Approximation Model

Let us focus on a given NDC partition, say A_n, and suppose $(C_{n_1}, C_{n_2}, ...C_{n_N})$ be the clusters within A_n that are obtained using the grid cover-couple approach. Let $A_{r_i}(x)$ be the size of the influence area for each RDC in cluster C_{n_i}. The integrated facility location and inventory allocation problem is given by P(1):

P(1) Minimze

$$\begin{aligned}
TNC(x) &= \sum_{i=1}^{N}\left(\frac{C_{n_i}}{A_{r_i}(x)}\right)F_r + \sum_{i=1}^{N}(C_f + C_v Q_{r_i}(x))\left(\frac{\xi\lambda(x)\delta_i(x)C_{n_i}}{Q_{r_i}(x)}\right) \\
&+ \sum_{i=1}^{N}\left(C_l f_r\sqrt{A_{r_i}(x)}\xi\lambda(x)\delta_i(x)C_{n_i} + R_r\left(\frac{\xi\lambda(x)\delta_i(x)C_{n_i}}{Q_{r_i}(x)}\right)\right) \\
&+ \sum_{i=1}^{N} h_r\left(\frac{C_{n_i}}{A_{r_i}(x)}\right)\left(\frac{Q_{r_i}(x)}{2} + ss_{r_i}(x)\right) \\
&+ \frac{R_n}{Q_n}\sum_{i=1}^{N}(\xi\lambda(x)\delta_i(x)C_{n_i}) + h_n\left(Q_n(x) + ss_n(x)\right)
\end{aligned}$$

subject to

$$\begin{aligned} Q_{r_i}(x) &\geq 0 \ \forall \ r_i \\ A_{r_i}(x) &\geq 0 \ \forall \ r_i \\ A_{r_i} &\leq C_{n_i} \ \forall \ r_i \\ Q_n(x) &\geq 0 \\ Q_{r_i}(x), \frac{C_{n_i}}{A_{r_i}(x)}, Q_n(x) &\in Z^+ \end{aligned} \tag{14}$$

$$\begin{aligned} \text{where } ss_{r_i}(x) &= Z_{\alpha_{r_i}} \sqrt{\mu_r \left(\lambda(x)\delta_i(x)A_{r_i}(x)\right) + (\sigma_r)^2(\lambda(x)\delta_i(x)A_{r_i}(x))^2} \\ ss_n(x) &= Z_{\alpha_n} \sqrt{\left(\mu_n \sum_{i=1}^{N} \frac{\lambda(x)\delta_i(x)C_{n_i}(x)}{(Q_{r_i}(x))^2}\right)} \end{aligned} \tag{15}$$

Note that the problem P(1) is nonlinear in the objective function. Also the objective function does not exhibit any convex or concave behavior. The expressions for safety stock at the RDC and the NDC ($ss_r(x)$, $ss_n(x)$), and the reorder cost term at the NDC make the objective function hard to evaluate. It is possible, however, to define a lower bound on the TNC(x) function which makes the $ss_r(x)$ term linear. We will show that even with the presence of the nonlinear term it is possible to decompose the problem for each cluster and get a solution.

Remark: Since each cluster within a given NDC partition has slow varying demand, we can ignore the dependance of all continuous function on parameter x. For the rest of this study, the variables are represented as A_{r_i}, Q_{r_i}, Q_n, λ and δ.

For analyzing the problem it is assumed that the RDCs within the same cluster C_{n_i} order the same quanity Q_{r_i} from the NDC. However, different RDCs in different clusters can order different quantities. A key challenge in the case of unequal Q_{r_i} is to how to define Q_n in terms of Q_{r_i}. In this case, Q_n is defined as $f(Q_{r_i})$. As an initial guess, $f(Q_{r_i})$ can be defined as $\sum_{i=1}^{N} Q_{r_i}$. A lower bound is obtained for the problem using result 4.1. The objective function of the lower bound problem P^U is non-linear in $A_{r_i}(x)$, Q_{r_i} and Q_n but it is possible to decompose the problem over the N clusters.

Result 4.1: A lower bound on the TNC(x) function in problem P(1) can be obtained by using the following relation.

$$\sqrt{\mu_r\lambda(x)\delta_i(x)A_{r_i}(x) + (\sigma_r)^2(\lambda(x)\delta_i(x)A_{r_i}(x))^2} > \sigma_r(\lambda(x)\delta_i(x)A_{r_i}(x)) - \tfrac{\mu_r}{2\sigma_r}$$

[Proof of Result 4.1]: Follows from the *monotone* property (see appendix) where $a = \mu_r, \ b = (\sigma_r)^2, \ x = \lambda(x)\delta_i(x)A_{r_i}(x)$

Problem P^U: Minimize

$$
\begin{aligned}
\tau(\mathbf{A}, \mathbf{Q}, Q_n) &= \sum_{i=1}^{N} \left(\frac{C_{n_i}}{A_{r_i}}\right) F_r + \sum_{i=1}^{N} (C_f + C_v Q_{r_i}) \left(\frac{\xi\lambda\delta_i C_{n_i}}{Q_{r_i}}\right) \\
&+ \sum_{i=1}^{N} C_l f_r \sqrt{A_{r_i}} \xi\lambda\delta_i C_{n_i} + R_r \sum_{i=1}^{N} \left(\frac{\xi\lambda\delta_i C_{n_i}}{Q_{r_i}}\right) \\
&+ \sum_{i=1}^{N} \left(\frac{C_{n_i}}{A_{r_i}}\right) h_r \left(\frac{Q_{r_i}}{2}\right) + \sum_{i=1}^{N} \left(\frac{C_{n_i}}{A_{r_i}}\right) h_r Z_{\alpha_{r_i}} \left(\sigma_r \lambda\delta_i A_{r_i} - \frac{\mu_r}{2\sigma_r}\right) \\
&+ h_n \left(\frac{Q_n}{2} + Z_{\alpha_n} \sqrt{\sum_{i=1}^{N} \mu_n \lambda\delta_i C_{n_i}}\right) + \sum_{i=1}^{N} \left(R_n \frac{\xi\lambda\delta_i C_{n_i}}{Q_n}\right)
\end{aligned}
$$

subject to

$$
\begin{aligned}
Q_{r_i} &\geq 0 \ \forall \ r_i \\
A_{r_i} &\geq 0 \ \forall \ r_i \\
Q_n &= f(Q_{r_i}) \\
Q_{r_i}, \frac{C_{n_i}}{A_{r_i}}, Q_n &\in Z^+ \ \forall \ r_i
\end{aligned}
$$

where $\mathbf{A}$ is the same n-dimensional row vector defined before and $\mathbf{Q}$ is the n-dimensional row vector defined by $\mathbf{Q} = [Q_{r_1}, Q_{r_2}, ..., Q_{r_N}]$. Note that for any value of $(\mathbf{A}, \mathbf{Q}, Q_n)$, the value of the objective function $\tau([\mathbf{A}, \mathbf{Q}, Q_n)$ obtained by solving P^U is strictly less than the value of the objective function TNC for the original problem $P(1)$.

The objective function $\tau(\mathbf{A}, \mathbf{Q}, Q_n)$ is analyzed for possible convex or concave behavior using properties of the hessian matrix (see Theorem 4.1 and 4.2).

5.4 Multi-variate Optimization

Before we present the analyis for the integrated facility location and inventory allocation problem, it is important to familiarize the reader with some definitions and theorems from multi-variate optimization (see (Bazaraa *et al.* (2)). These theorms will be used in the next two sections to understand the behavior of the objective function. It is important to understand the convex or concave behavior of the objective function and understand whether the stationary points correspond to the local or global optimum.

Definition 4.1: Let f be a twice differentiable function. Then the Hessian matrix of

f is given by (Bazaraa *et al.* (2), pg. 90):

$$H(\vec{x}) = \begin{bmatrix} \frac{\partial^2 f(\vec{x})}{\partial x_1{}^2} & \frac{\partial^2 f(\vec{x})}{\partial x_1 x_2} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_1 x_n} \\ \frac{\partial^2 f(\vec{x})}{\partial x_2 x_1} & \frac{\partial^2 f(\vec{x})}{\partial x_2{}^2} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_2 x_n} \\ . & . & \cdots & \\ . & . & \cdots & \\ \frac{\partial^2 f(\vec{x})}{\partial x_n x_1} & \frac{\partial^2 f(\vec{x})}{\partial x_n x_2} & \cdots & \frac{\partial^2 f(\vec{x})}{\partial x_n{}^2} \end{bmatrix}$$

Definition 4.2: Given a symmetric matrix A

$$A = \begin{bmatrix} a & b \\ c & d \end{bmatrix}$$

A is positive semidefinite *iff* $ad - bc \geq 0$

Theorem 4.1 (Bazaraa *et al.* (2), 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-Jordon 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)x(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 4.2 (Bazaraa *et al.* (2), pg. 91): Let S be a nonempty open convex set and let f: $S \Rightarrow E_1$ be twice differntiable on S. Then, f is convex if and only if the Hessian matrix is positive semidefinite at each point in S.

Theorem 4.3 (Bazaraa *et al.* (2), pg 134): Suppose that f: $E_n \longrightarrow E_1$ is twice differentiable at $\overline{x}$. If $\bigtriangledown f(\overline{x}) = 0$ and $H(\overline{x})$ is positive definite, then $\overline{x}$ is a strict local minimum.

5.5 Solution Procedure

The stationary point for the objective function τ satisfies the following equations (see appendix for derivation of the stationary point).

$$A_{r_i} = \left(\frac{2F_r + h_r Q_{r_i} - \frac{h_r Z_{\alpha_i} \mu_r}{2\sigma_r}}{C_l f_r \xi \lambda \delta_i} \right)^{2/3} \tag{16}$$

$$Q_n = \sqrt{\left(\frac{2R_n \sum_{i=1}^{N} \xi\lambda\delta_i C_{n_i}}{h_n}\right)} \tag{17}$$

$$Q_{r_i} = \sqrt{2A_{r_i}\left(\frac{(C_f + R_r + (R_n/Q_n))\,\xi\lambda\delta_i C_{n_i}}{C_{n_i} h_r}\right)} \tag{18}$$

A solution procedure for the partially unconstrained problem, one that ignores the integer value constraint ($Q_{r_i}, \frac{C_{n_i}}{A_{r_i}}, Q_n \in Z^+,\ i = 1,2,...,N$) and the linkage constraint between Q_n and Q_{r_i} ($Q_n = \sum_{i=1}^{N} Q_{r_i}$), is derived first. The solution generated using the iterative procedure is checked for compatibility with the convex region inequalities (Result 4.5). If the inequalities are satisfied then the solution is a near optimal solution for problem P^U, else a *Response Surface* method is used to generate a good solution for the problem.

The steps for the iterative procedure are explained below:

1. Fix $k = 0$, $\mathbf{Q}_k = [1,1,...,1]$.
2. Calculate A_{r_i}, $i = 1,2,..N$, using equation (16).
3. Use the value of A_{r_i}, $i = 1,2,..N$, in equation (17) to get $\mathbf{Q} = Q_{r_i}$ and calculate Q_n using (18). Iterate between the values of $\mathbf{Q}$ and Q_n till it converges.
4. If $\mathbf{Q} = Q_k$, Stop go to step 5. Else $k = k + 1$, and $\mathbf{Q}_k = \mathbf{Q}$ 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 $\mathbf{Q}$ and Q_n using step (3).
6. Adjust $\mathbf{Q}$ to get the nearest integer values. Adjust Q_n such that $Q_n = \sum_{i=1}^{N} Q_{r_i}$. Evaluate the objective function at each set of values of A_{r_i}, $\mathbf{Q}$ and Q_n. The set corresponding to the minimum value is the solution.

5.6 Response Surface Analysis

The optimal solution ($\mathbf{A}$, $\mathbf{Q}$ and Q_n) obtained by solving problem P^E abd P^U is substituted in the original objective function TNC to get a feasible solution for problem P(1). It would be interesting to see how the value of the objective function changes in the neighborhood of ($\mathbf{A}$, $\mathbf{Q}$ and Q_n). To carry out this analysis a statistical technique is used, and explained in detail. The *Response Surface* technique as a tool that is used to improve the feasible solution. The basic idea in this method is to perturb the values of all the decision variables around the optimal values obtained so far and generate a response curve for the original TNC function.

A factorial experiment is designed with $2N+1$ variables where N is the number of zones within a given NDC partition. There are N variables corresponding to the RDC influence area, N variables for the order quanity for the clusters and one variable for the unknown k. Since running a 2^{2N+1} experiment can get very time consuming and expensive, a fractional factorial experiment (FFE) of the form $2^{(2N+1)-p}$ is considered. In a FFE, $(2N+1)-p$ variables are fixed and these variables are used to generate the remaining p from them. An experiment is set up using this information for two levels- high (1) and low (-1). The experiment data is then transformed to match the original scale of the variables. The objective function is evaluated at each of the design points and we try to fit a regression model (linear or nonlinear) to it. This regression equation is an estimate of the Response surface. The nature of the surface is inspected by using the first and second order conditions (i.e., by taking the first and the second order derivatives) and an optimal value for the decision variables is calculated using this information.

5.7 Equal Reorder Quantity Q

It is a common practice in multi-echelon inventory studies to assume that the reorder quantity Q_{r_i} is the same across all retailers (see Deuermeyer *et al.* (11), Ganeshan (15)). For the case when $Q_{r_i} = Q$ holds, the reorder quantity at the warehouse, Q_n, is expressed as an integer multiple of Q. For the first part of the analysis (case 1), we assume that $Q_{r_i} = Q$ at all the RDCs and $Q_n = kQ$ for the NDC. The solution procedure for solving the problem under Equal Reorder Quantity case is similar to the one discussed under Unequal Reorder Quantity case.

The stationary point for the objective function $\phi(\mathbf{A}, Q, Q_n)$ is given by equations (19), (20) and (21) (for details see *appendix*):

$$A_{r_i} = \left(\frac{2F_r + h_r Q - \frac{h_r Z_{\alpha_i} \mu_r}{\sigma_r}}{C_l f_r \xi \lambda \delta_i} \right)^{2/3} \tag{19}$$

$$k = \frac{1}{Q} \sqrt{\left(\frac{2R_n (\sum_{i=1}^{N} \xi \lambda \delta_i C_{n_i})}{h_n} \right)} \tag{20}$$

$$Q = \sqrt{\left(\frac{\sum_{i=1}^{N} \left(C_f + R_r + \frac{R_n}{k} \right) \xi \lambda \delta_i C_{n_i}}{\sum_{i=1}^{N} \frac{h_r C_{n_i}}{2A_{r_i}} + \frac{h_n k}{2}} \right)} \tag{21}$$

6 Discussion

The integrated model is compared with the non-integrated model and the average model. The non-integrated model is the one where the facility location and inventory decisions are made in isolation of each other. The model is first solved for the optimal influence area using information on the facility location cost and the transportation cost. Using this value of the influence area in the inventory and transportation cost functions, the optimal inventory decisions are made.

The average model is where the entire distribution region is assumed to be a smooth continuous region. It is also assumed that the store density and demand density functions are smooth over this region. In this model, the store density function for the entire region is defined by the average value of individual store densities, i.e, each $\delta_i = \overline{\delta} = \sum_{i=1}^{N} \frac{\delta_i}{N}$. Then, $\sum_{i=1}^{N} C_{n_i}$ is replaced by R, size of the entire distribution network. The decision variables in this case are A_r, Q and k.

6.1 Stationary point for the non-integrated problem- equal Q case

$$A_{r_i} = \left(\frac{2F_r}{C_l f_r \xi \lambda \delta_i} \right)^{2/3}$$

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

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

6.2 Stationary point for the integrated model using averages

$$A_r = \left(\frac{2F_r + h_r Q - \frac{h_r Z_{\alpha_r} \mu_r}{\sigma_r}}{C_l f_r \xi \lambda \overline{\delta}} \right)^{2/3}$$

$$k = \frac{1}{Q} \sqrt{\left(\frac{2R_n}{h_n} \right) (\xi \lambda \overline{\delta} R)}$$

$$Q = \sqrt{\left(\frac{\left(C_f + R_r + \frac{R_n}{k} \right) \xi \lambda \overline{\delta} R}{\frac{h_r R}{2A_r} + \frac{h_n k}{2}} \right)}$$

The analysis for the integrated facility location and inventory allocation problem sheds light on some important issues. Some of the key observations are listed below.

Observation 1. Optimal size of the RDC influence area is a function of order up to level Q_r. Thus, it is important to incorporate the inventory decisions into the network

Table 1: Store density and Average distance data

	GA	FL	TN	AL	KY	VA	NC	SC
Store Density	0.0059	0.0039	0.0038	0.002	0.0052	0.0471	0.0041	0.0019

design problem. Since the decision variables do not have a closed form expression a numerical iterative procedure is used to get a solution.

Observation 2 It is assumed in the above analysis that each cluster can be analyzed separately. This can only happen when a NDC serving different clusters reviews their inventory position periodically. In this case, the inventory policy at the NDC is a periodic review (T, r, nQ) policy (1). The study of the NDC periodic review policy and its impact on the network design is left for the future work.

7 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. The numerical analysis of this section is carried out using data for the southeastern (SE) region with the NDC located at Savannah, GA. Table 1 gives the store density for the eight states served by the Savannah DC. Clearly there is a significant amount of variation in the store density data across states.

For a fixed value of the inventory parameters Q and Q_n, the number of RDCs increase (decrease) with an increase (decrease) in the store density. Similarly, for a fixed number of RDCs, increase or decrease of the store density changes the value of the inventory parameters Q and Q_n. Thus, the density function affects the facility, transportation and inventory cost.

7.1 Comparison between Integrated, Non-integrated and Average case

The results for the integrated, non-integrated and average version of the facility location and inventory allocation problem under equal Q_{r_i}s are presented in table 2 and figure 7. The integrated case is the one with the minimum value of total network cost. For this example observe that the TNC is 6.6% higher in the case of non-integrated problem and 44% higher for the average case (see Table 2). These results justify the need for a two-phase approxiation approach.

A focus on the total costs (i.e. the total network cost minus the total inventory cost at

Table 2: Comparison between the three cases for equal Q

	Case 1	Case 2	Case 3
	Integrated	Non-Integrated	Average
RDC	9	22	7
TI_r	3545330	4015240	4966090
TI_n	1058490	1058470	1319280
TF	90000	220000	70000
TIT	2682650	2994400	3916000
TOT	858030	494390	1608130
TNC	8234370	8782490	11879500
Q_r	2564	1640	3623
Q_n	27	42	24

the NDC) for the RDCs in each zone show an interesting trend. Although total network cost for the integrated case is less than that for the non-integrated case, there could be zones for which the later case yields a lower total cost. In particular, for this example zones 3, 4, 5 and 8 have a lower total cost in the non-integrated case (see figure 8). If this problem was modeled with a decentralized decision maker, then these zones have no incentive to participate in an integrated activity. This opens a new direction for our research where game theory can be used. This interesting research proposal is left for future work.

Observe that the safety stock at each store in the integrated case is greater than that for the non-integrated case (see figure 9). This may look counter intuitive initially. However, a careful inspection shows that each zone has fewer RDCs in the integrated case. As the number of RDCs increases, the safety stock at each RDC decreases. This result is quite unlike the Square Root law which says that the total inventory in a system is proportional to the square root of the number of locations at which a product is stocked (see Chopra *et al.* (7)). The reason for this is while applying the square root law we observe that reducing the number of RDCs reduces the risk by pooling demand variability. But in our model, both the demand and supply variability are taken into account at the RDCs. Reduction in the number of RDCs means more inbound shipments to each RDC and thus more supply variability. Hence when both types of variabilities are taken into account it is possible to see this reverse relation between the number of RDCs and safety stock. Thus, each zone in the integrated problem has a higher value for total safety stock and reorder point as compared to each zone under the non-integrated case. A zonewise comparison of the safety stock in each zone in presented in figure 10.

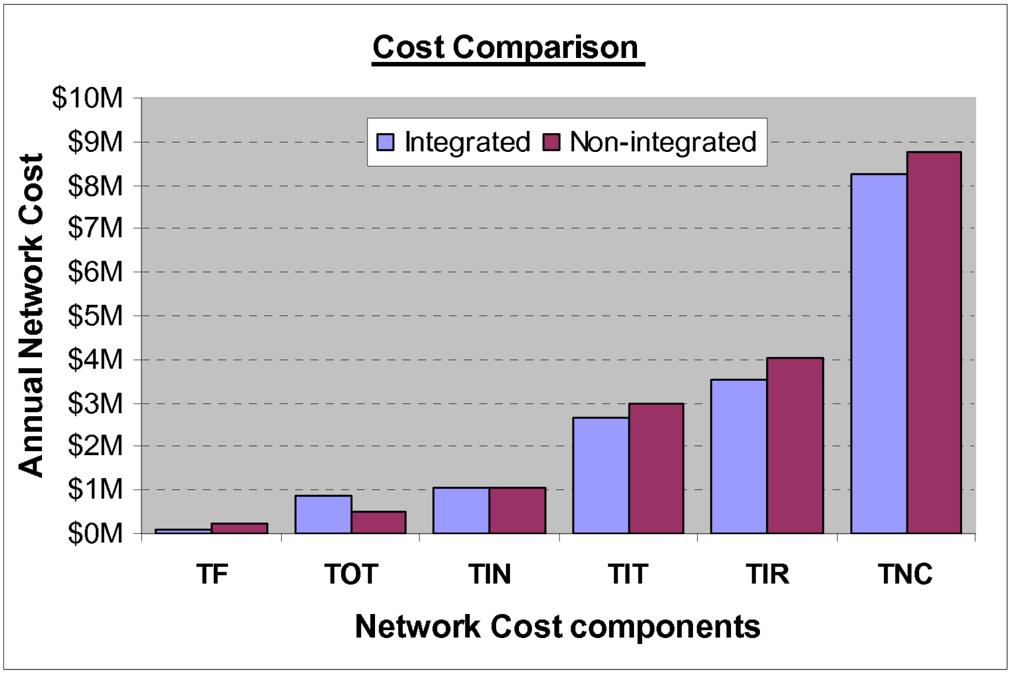


Figure 7: Total Network Cost

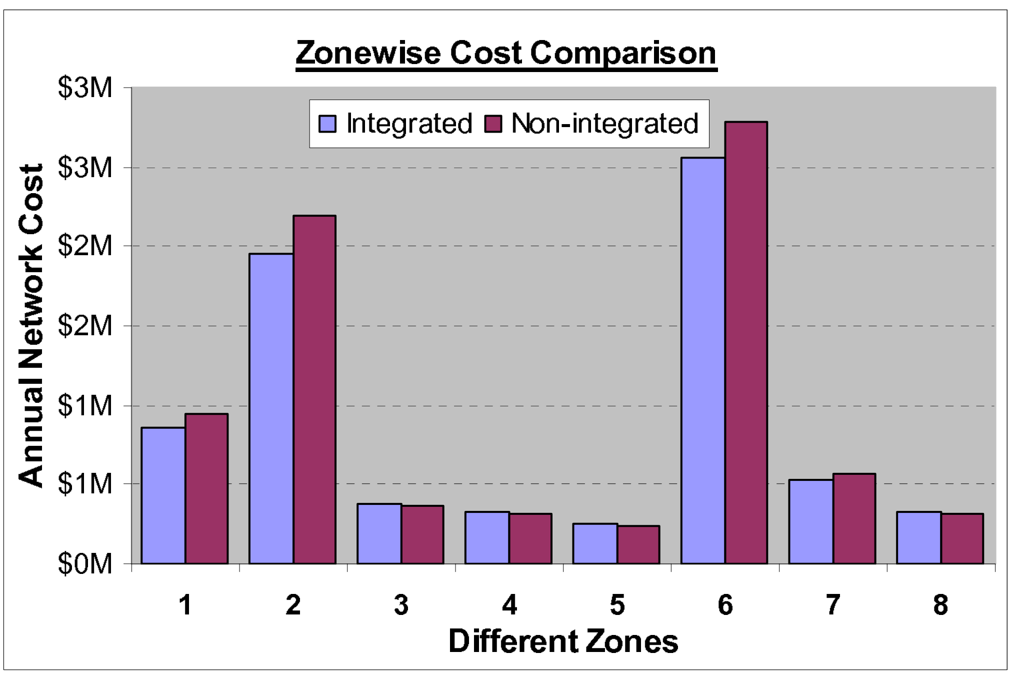


Figure 8: Total cost for each zone.

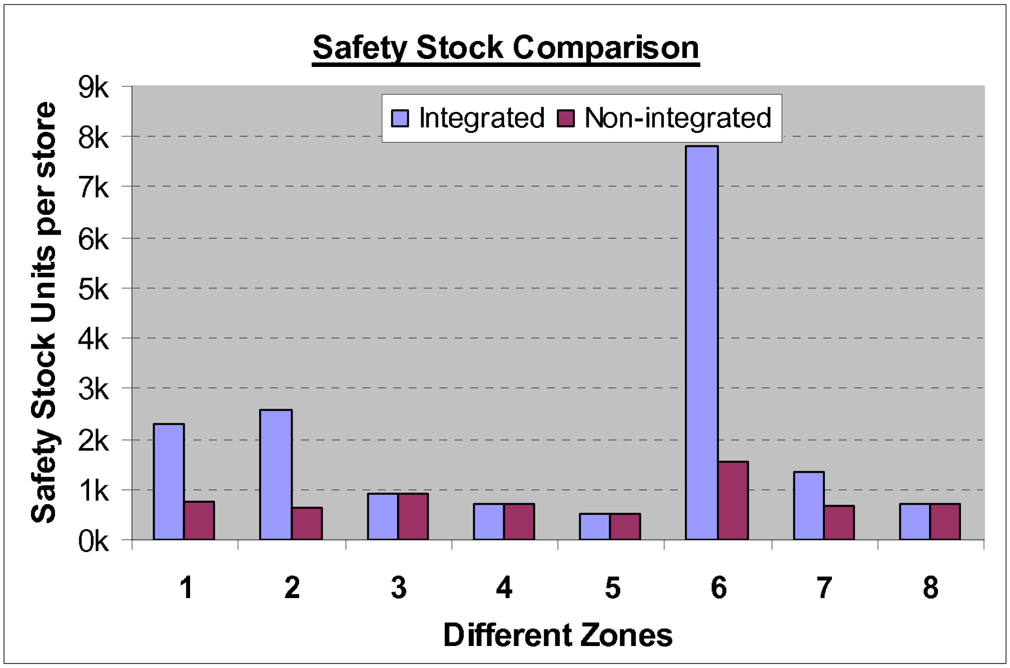


Figure 9: Safety stock for each store zone-integrated vs non-integrated

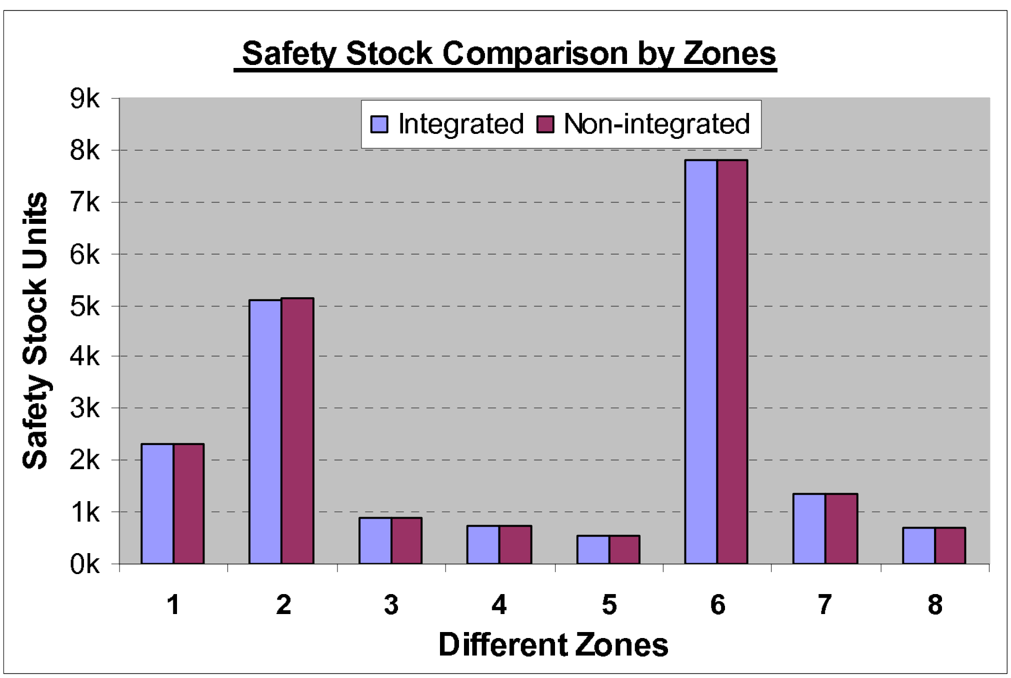


Figure 10: Safety stock for each zone-integrated vs non-integrated

8 Appendix

The following *monotone* property is used to replace the objective function with another function that is a lower bound on the original function.

Monotone property (see Appendix): If a and b are positive numbers and $x > 0$, then

$$\sqrt{ax + bx^2} > \sqrt{b}x - \frac{a}{2\sqrt{b}}$$

The following *monotone* property is used to replace the objective function with another function that is a lower bound on the original function.

Unequal Reorder Point

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

$$H = \begin{bmatrix} a_{1,1} & 0 & ... & 0 & a_1q_1 & 0 & ... & 0 & 0 \\ 0 & a_{2,2} & ... & 0 & 0 & a_2q_2 & ... & 0 & 0 \\ . & . & ... & . & . & . & ... & . & 0 \\ . & . & ... & . & . & . & ... & . & 0 \\ 0 & 0 & ... & a_{n,n} & 0 & 0 & ... & a_nq_n & 0 \\ q_1a_1 & 0 & ... & 0 & q_{1,1} & 0 & ... & 0 & q_{1,N+1} \\ 0 & q_2a_2 & ... & 0 & 0 & q_{2,2} & ... & 0 & q_{2,N+1} \\ . & . & ... & . & . & . & ... & . & \\ . & . & ... & . & . & . & ... & . & \\ 0 & 0 & ... & q_na_n & 0 & 0 & ... & q_{n,n} & q_{n,N+1} \\ 0 & 0 & ... & 0 & q_{N+1,1} & q_{N+1,2} & ... & q_{N+1,n} & q_{N+1,N+1} \end{bmatrix}$$

where

$$a_{i,i} = \frac{\partial^2 \tau}{\partial {A_{r_i}}^2}, \qquad q_{i,i} = \frac{\partial^2 \tau}{\partial {Q_{r_i}}^2}, \qquad q_{N+1,N+1} = \frac{\partial^2 \tau}{\partial {Q_n}^2}$$

$$a_iq_i = \frac{\partial^2 \tau}{\partial A_{r_i} \partial Q_{r_i}}, \qquad q_ia_i = \frac{\partial^2 \tau}{\partial Q_{r_i} \partial A_{r_i}} \qquad q_{N+1,i} = \frac{\partial^2 \tau}{\partial Q_n \partial Q_{r_i}}$$

$$q_{i,N+1} = \frac{\partial^2 \tau}{\partial Q_{r_i} \partial Q_n}, \quad a_iq_{N+1} = \frac{\partial^2 \tau}{\partial A_{r_i} \partial Q_n}, \quad q_{N+1}a_i = \frac{\partial^2 \tau}{\partial Q_n \partial A_{r_i}}$$

$$i = 1, 2..., N$$

First order conditions for finding the stationary point:

$$\begin{aligned}
\frac{\partial\tau}{\partial A_{r_i}} &= -\frac{1}{(A_{r_i})^2}\left[C_{n_i}F_r + \left(\frac{C_{n_i}h_rQ_{r_i}}{2}\right) - \left(\frac{h_rZ_{\alpha_i}\mu_r}{2\sigma_r}\right)\right] \\
&+ \frac{C_lf_r\xi\lambda\delta_iC_{n_i}}{2\sqrt{A_{r_i}}} = 0 \\
\frac{\partial\tau}{\partial Q_{r_i}} &= -\left[C_f + R_r + (R_n/Q_n)\right]\left(\frac{\xi\lambda\delta_iA_{r_i}}{Q_{r_i}^2}\right) + \left(\frac{h_rC_{n_i}}{2A_{r_i}}\right) = 0 \\
\frac{\partial\tau}{\partial Q_n} &= \frac{h_n}{2} - \frac{1}{Q_n^2}\frac{\sum_{i=1}^N R_n\xi\lambda\delta_iC_{n_i}}{Q_i} = 0
\end{aligned}$$

First and second order derivatives for the objective function $\tau([A_{r_i}], [Q_{r_i}], Q_n)$

$$\begin{aligned}
\frac{\partial\tau}{\partial A_{r_i}} &= -\frac{1}{(A_{r_i})^2}\left[C_{n_i}F_r + \left(\frac{C_{n_i}h_rQ_{r_i}}{2}\right) - \left(\frac{h_rZ_{\alpha_i}\mu_r}{2\sigma_r}\right)\right] \\
&+ \frac{C_lf_r\xi\lambda\delta_iC_{n_i}}{2\sqrt{A_{r_i}}} \\
\frac{\partial^2\tau}{\partial {A_{r_i}}^2} &= \frac{2}{(A_{r_i})^3}\left[C_{n_i}F_r + \left(\frac{C_{n_i}h_rQ_{r_i}}{2}\right) - \left(\frac{h_rZ_{\alpha_i}\mu_r}{2\sigma_r}\right)\right] \\
&- \frac{C_lf_r\xi\lambda\delta_iC_{n_i}}{4{A_{r_i}}^{3/2}} \\
\frac{\partial^2\tau}{\partial Q_{r_i}\partial A_{r_i}} &= \frac{-h_rC_{n_i}}{2{A_{r_i}}^2} \\
\frac{\partial^2\tau}{\partial Q_n\partial A_{r_i}} &= 0
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\tau}{\partial Q_{r_i}} &= -\left[C_f + R_r + (R_n/Q_n)\right]\left(\frac{\xi\lambda\delta_iA_{r_i}}{Q_{r_i}^2}\right) + \left(\frac{h_rC_{n_i}}{2A_{r_i}}\right) \\
\frac{\partial^2\tau}{\partial {Q_{r_i}}^2} &= 2\left[C_f + R_r + (R_n/Q_n)\right]\left(\frac{\xi\lambda\delta_iA_{r_i}}{Q_{r_i}^3}\right) \\
\frac{\partial^2\tau}{\partial A_{r_i}\partial Q_{r_i}} &= \frac{-h_rC_{n_i}}{2{A_{r_i}}^2} \\
\frac{\partial^2\tau}{\partial Q_n\partial Q_{r_i}} &= \frac{R_n\xi\lambda\delta_iC_{n_i}}{Q_n^2Q_i^2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial \tau}{\partial Q_n} &= \frac{h_n}{2} - \frac{1}{Q_n^2}\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Q_i} \\
\frac{\partial^2 \tau}{\partial {Q_n}^2} &= \frac{2}{Q_n^3}\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Q_i} \\
\frac{\partial^2 \tau}{\partial A_{r_i} \partial Q_n} &= 0 \\
\frac{\partial^2 \tau}{\partial Q_{r_i} \partial Q_n} &= \frac{R_n \xi \lambda \delta_i C_{n_i}}{Q_n^2 Q_i^2}
\end{aligned}$$

[Proof: $\tau(\mathbf{A}, \mathbf{Q}, Q_n)$ is a biconvex function]

(1) For a given value of Q_{r_i}, i = 1,2,..N, and Q_n, $\tau(\mathbf{A})$ is a convex function.

$$\begin{aligned}
\frac{\partial^2 \tau}{\partial {A_{r_i}}^2} > 0 \;\; &\Leftrightarrow \;\; \frac{2}{(A_{r_i})^3}\left[C_{n_i} F_r + \left(\frac{C_{n_i} h_r Q_{r_i}}{2}\right) - \left(\frac{h_r Z_{\alpha_i} \mu_r}{2\sigma_r}\right)\right] > \\
& \quad\;\; \frac{C_l f_r \xi \lambda \delta_i C_{n_i}}{4{A_{r_i}}^{3/2}} \\
&\Leftrightarrow \;\; A_{r_i} < 4^{2/3}\left[\frac{F_r + h_r Q_{r_i} - (h_r Z_{\alpha_i} \mu_r / \sigma_r)}{C_l f_r \xi \lambda \delta_i}\right]^{2/3}
\end{aligned}$$

which holds for all values of A_{r_i} satisfying the stationary condition

$$A_{r_i} = \left[\frac{F_r + h_r Q_{r_i} - (h_r Z_{\alpha_i} \mu_r / \sigma_r)}{C_l f_r \xi \lambda \delta_i}\right]^{2/3}$$

(1) For a given value of A_{r_i}, i = 1,2,..N, $\tau(\mathbf{Q}, Q_n)$ is a convex function.

$$H = \begin{bmatrix}
q_{1,1} & 0 & \ldots & 0 & q_{1,N+1} \\
. & q_{2,2} & \ldots & 0 & q_{2,N+1} \\
. & . & \ldots & . & . \\
. & . & \ldots & . & . \\
. & . & \ldots & q_{n,n} & q_{n,N+1} \\
q_{N+1,1} & q_{N+1,2} & \ldots & q_{N+1,n} & q_{N+1,N+1}
\end{bmatrix}$$

where H is the hessian matrix for $\tau(\mathbf{Q}, Q_n)$. Using theorem 4.1 and 4.2, $\tau(\mathbf{Q}, Q_n)$ is convex *iff* H is positive definite.

$$\text{H is positive definite } \textit{iff} \begin{vmatrix}
q_{N,N} & q_{N,N+1} \\
q_{N+1,N} & q_{N+1,N+1} - \sum_{i=1}^{N} \frac{q_{N+1,i}}{q_{i,i}} q_{i,N+1}
\end{vmatrix} > 0$$

$$\begin{aligned}
|H| &= q_{N,N}\left(q_{N+1,N+1} - \sum_{i=1}^{N}\frac{q_{N+1,i}}{q_{i,i}}q_{i,N+1}\right) - q_{N,N+1}q_{N+1,N} \\
&= 2\gamma\left(\frac{\xi\lambda\delta_N A_{r_N}}{Q_{r_N}^3}\right)\left[\frac{2}{Q_n^3}\frac{\sum_{i=1}^{N} R_n\xi\lambda\delta_i C_{n_i}}{Q_i}\left(1-\frac{R_n/Q_n}{\gamma}\right)\right] \\
&\quad \text{where } \gamma = (C_f + R_r + R_n/Q_n) \\
|H| &> 0 \quad \text{because } (R_n/Q_n)/\gamma < 1
\end{aligned}$$

Equal Reorder Point (1) First order conditions for deriving the stationary point for function $\phi(\mathbf{A}, Q, k)$

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

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

$$H = \begin{bmatrix}
a_{1,1} & 0 & ... & 0 & a_1q & 0 \\
0 & a_{2,2} & ... & 0 & a_2q & 0 \\
. & . & ... & . & . & . \\
. & . & ... & . & . & . \\
0 & 0 & ... & a_{n,n} & a_nq_n & 0 \\
qa_1 & qa_2 & ... & qa_N & qq & qq_n \\
0 & 0 & ... & 0 & q_nq & q_nq_n
\end{bmatrix}$$

where

$$\begin{gathered}
a_{i,i} = \frac{\partial^2\phi}{\partial {A_{r_i}}^2}, \qquad qq = \tfrac{\partial^2\phi}{\partial Q^2}, \qquad q_nq_n = \frac{\partial^2\phi}{\partial k^2} \\
a_iq = \frac{\partial^2\phi}{\partial A_{r_i}\partial Q}, \quad qa_i = \tfrac{\partial^2\phi}{\partial Q\partial A_{r_i}} \quad q_nq = \frac{\partial^2\phi}{\partial k\partial Q} \\
qq_n = \frac{\partial^2\phi}{\partial Q\partial k}, \quad a_iq_n = \tfrac{\partial^2\phi}{\partial A_{r_i}\partial k}, \quad q_na_i = \frac{\partial^2\phi}{\partial k\partial A_{r_i}} \\
i = 1, 2..., N
\end{gathered}$$

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

Using theorem 4.2, $\phi(\mathbf{A}, Q, k)$ is convex *iff* the hessian matrix of ϕ is positive semidefinite. And from theorem 4.1, hessian matrix of ϕ is positive definite for values of $(\mathbf{A}, Q, k)$ satisfying

[]

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

$$\left(\frac{\partial^2\phi}{\partial Q^2} - \sum_{i=1}^{N} \frac{\partial^2\phi/\partial Q\partial A_{r_i}}{\partial^2\phi/\partial {A_{r_i}}^2}\frac{\partial^2\phi}{\partial A_{r_i}\partial Q}\right) > 0$$

First and second order derivatives for the function $\phi(\mathbf{A}, Q, k)$

$$\begin{aligned}
\frac{\partial\phi}{\partial A_{r_i}} &= \frac{C_l f_r \xi\lambda\delta_i C_{n_i}}{2{A_{r_i}}^{1/2}} - \frac{2C_{n_i}F_r + C_{n_i}h_r Q - (C_{n_i}h_r Z_{\alpha_r}\mu_r/\sigma_r)}{2{A_{r_i}}^2} \\
\frac{\partial\phi^2}{\partial {A_{r_i}}^2} &= -\frac{C_l f_r \xi\lambda\delta_i C_{n_i}}{4{A_{r_i}}^{3/2}} + \frac{2C_{n_i}F_r + C_{n_i}h_r Q - (C_{n_i}h_r Z_{\alpha_r}\mu_r/\sigma_r)}{2{A_{r_i}}^3} \\
\frac{\partial\phi^2}{\partial Q\partial A_{r_i}} &= -\frac{C_{n_i}h_r}{2{A_{r_i}}^2} \\
\frac{\partial\phi^2}{\partial Q_n\partial A_{r_i}} &= 0 \\
\frac{\partial\phi}{\partial Q} &= -\frac{\sum_{i=1}^{N}(C_f + R_r + R_n/k)\xi\lambda\delta_i C_{n_i}}{Q^2} + \sum_{i=1}^{N}\left(\frac{h_r C_{n_i}}{2A_{r_i}}\right) + \frac{h_n k}{2} \\
\frac{\partial^2\phi}{\partial Q^2} &= 2\left(\frac{\sum_{i=1}^{N}(C_f + R_r + R_n/k)\xi\lambda\delta_i C_{n_i}}{Q^3}\right) \\
\frac{\partial^2\phi}{\partial A_{r_i}\partial Q} &= -\frac{h_r C_{n_i}}{2A_{r_i}^2} \\
\frac{\partial^2\phi}{\partial k\partial Q} &= \sum_{i=1}^{N}\frac{R_n\xi\lambda\delta_i C_{n_i}}{Q^2k^2} + \frac{h_n}{2}
\end{aligned}$$

$$\begin{aligned}
\frac{\partial\phi}{\partial k} &= \frac{h_n k}{2} - \frac{\sum_{i=1}^{N} R_n\xi\lambda\delta_i C_{n_i}}{Qk^2} \\
\frac{\partial^2\phi}{\partial k^2} &= 2\left(\frac{\sum_{i=1}^{N} R_n\xi\lambda\delta_i C_{n_i}}{Qk^3}\right) \\
\frac{\partial^2\phi}{\partial A_{r_i}\partial k} &= 0 \\
\frac{\partial^2\phi}{\partial Q\partial k} &= \frac{\sum_{i=1}^{N} R_n\xi\lambda\delta_i C_{n_i}}{Q^2k^2} + \frac{h_n}{2}
\end{aligned}$$

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

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

$$
|H| = \begin{bmatrix} \frac{\partial^2\phi}{\partial Q^2} & \frac{\partial^2\phi}{\partial k\partial Q} \\ \frac{\partial^2\phi}{\partial Q\partial k} & \frac{\partial^2\phi}{\partial k^2} \end{bmatrix}
$$

$$\begin{aligned}
|H| &= \left(2\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Qk^3}\right)\left(2\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{kQ^3}\right) \\
&+ \left(2\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Qk^3}\right)\left(2\frac{\sum_{i=1}^{N} (C_f + R_r) \xi \lambda \delta_i C_{n_i}}{Q^3}\right) \\
&- \left(\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Q^2k^2} + \frac{h_n}{2}\right)\left(\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Q^2k^2} + \frac{h_n}{2}\right) \\
&= 4\left(\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Q^2k^2}\right)^2 - \left(\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Q^2k^2}\right)^2 \\
&+ \left(2\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Qk^3}\right)\left(2\frac{\sum_{i=1}^{N} (C_f + R_r) \xi \lambda \delta_i C_{n_i}}{Q^3}\right) \\
&- \left(\frac{h_n^2}{4} + \frac{h_n \sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Q^2k^2}\right) \\
&= 3\left(\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Q^2k^2}\right)^2 + \left(\frac{\sum_{i=1}^{N} R_n \xi \lambda \delta_i C_{n_i}}{Q^2k^2}\right) * \\
& \quad \left[\frac{4}{Q^2k}\sum_{i=1}^{N}(C_f + R_r)\xi\lambda\delta_i C_{n_i} - h_n\right] - \frac{h_n^2}{4} \\
\text{For} \quad k &= \frac{1}{Q}\sqrt{\frac{2R_n \sum_{i=1}^{N} \xi \lambda \delta_i C_{n_i}}{h_n}} \\
|H| &= \frac{3h_n^2}{4} + \frac{k(C_f + R_r)h_n^2}{R_n} - \frac{h_n^2}{2} - \frac{h_n^2}{4} \\
&= \frac{k(C_f + R_r)h_n^2}{R_n} \\
&> 0 \quad \text{always}
\end{aligned}$$

References

[1] S. Ahire and C. Schmidt. A model for a mixed continuous review one-warehouse, n-retailer inventory system. *European Journal of Operational Research*, 92:69–82, 1996.

[2] M.S. Bazaraa, H.D. Sherali, and C.M. Shetty. *Nonlinear Programming: Theory and Algorithms*. John Wiley & Sons, Inc., 1993.

[3] D. E. Blumenfeld and M. J. Beckmann. Use of continuous space modeling to estimate freight distribution costs. *Transportation Research A*, 19A(2):173–187, 1985.

[4] M. Brandeau and S. Chiu. An overview of representative problems in location research. *Management Science*, 35:645–673, 1989.

[5] L.D. Burn, R.W. Hall, D. E. Blumenfeld, and C. E. Daganzo. Distribution strategies that minimize transportation and inventory cost. *Operations Research*, 33(3):469–490, 1985.

[6] C.F.Daganzo and G.F.Newell. Configuration of physical distribution networks. *Networks*, 16:113–132, 1986.

[7] S. Chopra and P. Meindl. *Supply Chain Management*. Prentice Hall; 2 edition, 2003.

[8] C. F. Daganzo. *Logistics Systems Analysis*. Springer, Berlin, 1996.

[9] A. Dasci and V. Verter. A continuous model for production-distribution system design. *European Journal of Operational Research*, 129:287–298, 2001.

[10] M. Daskin. *Network and Discrete Location: Models, Algorithms and Applications*. John Wiley and Sons, New York, 1995.

[11] B. Deuermeyer and L. B. Schwarz. *In Studies in the Management Sciences, Multilevel Production/Inventory Control Systems*, volume 16, chapter A Model for the analysis of System Service level in Warehouse/Retailer Distribution Systems: The Identical Retailer Case., pages 163–193. North-Holland, Amsterdam, 1981.

[12] Z. Drezner. *Facility Location: A Survey of Applications and Methods*. Springer-Verlag, New York, NY, 1995.

[13] S. Erlebacher and R. Meller. The interaction of location and inventory in designing distribution systems. *IIE Transactions*, 32:155–166, 2000.

[14] D. Erlenkotter. The general optimal market area model. *Annals of Operations Research*, 18:45–70, 1989.

[15] R. Ganeshan. Managing supply chain inventories: A multiple retailer, one warehouse, multiple supplier model. *International Journal of Production Economics*, 59:341–354, 1999.

[16] A. M. Geoffrion. The purpose of mathematical programming is insight, not numbers. *Interfaces*, 7:81–92, 1976.

[17] A. Langevin, P. Mbaraga, and J. F. Campbell. Continuous approximation models in freight distribution: an overview. *Transportation Research B*, 30(3):163–88, 1996.

[18] P. A. Miranda and R. A. Garrido. Incorporating inventory control decisions into a strategic distribution network design model with stochastic demand. *Transportation Research Part E*, 40:183–207, 2004.

[19] G. F. Newell. Scheduling, location, transportation and continuum mechanics: some simple approximations to optimization problems. *SIAM Journal of Applied Mathematics*, 25, 1973.

[20] L. Nozick and M. Turnquist. Integrating inventory impacts into a fixed-charge model for locating distribution centers. *Transportation Research-E*, 34(3):173–186, 1998.

[21] L. Nozick and M. Turnquist. Inventory, transportation, service quality and the location of distribution centers. *European Journal of Operations Research*, 129:362–371, 2001.

[22] H. E. Romeijn, J. Shu, and C. Teo. Designing two-echelon supply networks. *European Journal of Operational Research*, 178:449–462, 2007.

[23] R.O. Roundy. 98% effective integer-ratio lot-sizing for one warehouse multi-retailer systems. *Management Science*, 31:1416–1430, 1985.

[24] W.G.M.M Rutten, P.J.M. Van Laarhoven, and B. Vos. An extension of the goma model for determining the optimal number of depots. *IIE Transactions*, 33:1031–1036, 2001.

[25] Z. M. Shen, C. Coullard, and M. S. Daskin. A joint location-inventory model. *Transportation Science*, 37(1):40–55, 2003.

[26] C. Teo and J. Shu. Warehouse-retailer network design problem. *Operations Research*, 52(3):396–408, 2004.

[27] N. Wang and J.C. Lu. Multi-level spatial modelling and decision-making with application in logistics systems. Technical report, Georgia Institute of technology, 2006.