INTEGRATED DECISIONS FOR SUPPLY CHAIN DESIGN AND INVENTORY ALLOCATION PROBLEM

A Thesis
Presented to
The Academic Faculty

by

Divya Mangotra

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
H. Milton Stewart School of Industrial and Systems Engineering

Georgia Institute of Technology
December 2007

INTEGRATED DECISIONS FOR SUPPLY CHAIN DESIGN AND INVENTORY ALLOCATION PROBLEM

Approved by:

Dr. Jye-Chyi Lu
Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Faiz Al-Khayyal
Co-Advisor
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Chen Zhou
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Paul Griffin
H. Milton Stewart School of Industrial and Systems Engineering
Georgia Institute of Technology

Dr. Mark Ferguson
College of Management
Georgia Institute of Technology

Date Approved: 4 October 2007

To my parents,

your love and support made it possible.

ACKNOWLEDGEMENTS

I would like to thank my advisor Dr. Jye-Chyi Lu for his support and encouragement which made this PhD journey an enjoyable experience. He has been a true mentor and inspiration, always guiding me to overcome challenges in both personal and professional life. I am also thankful to my co-advisor Dr. Faiz Al-Khayal for his help on technical details. His guidance and advice greatly contributed to the completion of this thesis. I am grateful to my committee members Dr. Mark Ferguson, Dr. Chen Zhou and Dr. Paul Griffin for their valuable suggestions that made me extend this thesis in several interesting directions.

A lot of friends and well-wishers have crossed my path during the last five years. It is hard to thank each of them individually here. I mention some very close friends and extend my gratitude to the rest. Thank you Kapil, Deepali, Shalini, Kuan, Eser, Hasan for your great company that helped me strike a balance between work and fun. Without you my stay in Atlanta would not have been so exciting. I would also like to thank my friends - Reena, Parul, Kriti and Arati for helping me escape Atlanta when I needed to and for being there for me always.

I express my gratitute to my parents, my brothers, Amit and Arjun, and my sister-in-law Shurobhi, for their love and patience. This PhD would not have been possible without the unconditional support from my parents and their faith in my decision. Finally, I thank my husband, Abhishek, for sharing every moment of happiness and frustration with me throughout this journey. For keeping my spirits high all the time and for making this dream come true.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

SUMMARY

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 the manufacturing plants supplying goods to the retail stores directly, the off-shore manufacturing has increased the network's demand for transportation and warehousing to deliver the goods. Thus, most companies have a complex distribution network with several import and regional distribution centers (DC).

In this thesis, we study an integrated facility location and inventory allocation problem for designing a distribution network with multiple national (import) distribution centers (NDC) and retailers. The key decisions are where to locate the RDCs and how much inventory to hold at the different locations such that the total network cost is minimized under a pre-defined operational rule for the distribution of goods. In particular, the inventory cost analysis is based on the continuous review batch ordering policy and the base-stock policy. Both Type-I (probability of stock-outs) and Type-II (fill-rate) service level measures are used in the analysis.

This thesis presents two different models for solving the integrated facility location-inventory allocation problem. The first model, continuous approximation (CA), assumes the distribution network to be located in a continuous region and replaces the discrete store locations with a store density function. The second model is a discrete representation of the problem as a mixed integer programming problem. Both the models take a nonlinear form and solution techniques are developed using the theory of nonlinear programming and linear reformulation of nonlinear problems.

The goal of the first part of the thesis is to model the problem using a modified

CA approach and an iterative solution scheme is presented to solve it. 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. In addition, the 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 is also shown that the regular CA approach leads to a solution which is inferior to the solution obtained by the modifed CA approach. Our analysis shows that the type of service measure used affects the network design.

In the second part of the thesis, the problem is modeled as a nonlinear mixed integer program and a linear reformulation solution technique is proposed to obtain a lower bound on the original problem. Computational results are presented for small problem instances. We conclude this part of the thesis by presenting an integrated model when a base stock inventory policy is used. A drop-decomposition heuristic is proposed to solve this problem.

CHAPTER I

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 the manufacturing plants supplying goods to the retail stores directly, the off-shore manufacturing has increased the network's demand for transportation and warehousing to deliver the goods. Warehouses are located in the network to allow for shipment consolidation and risk pooling (see Chopra *et al.* [10]). When the goods arrive at the seaports in the U.S. they have to be consolidated by region[1] before shipping them. This activity takes place at the import distribution centers. From the import distribution centers the goods are shipped to the local distribution facilities from which they are delivered to the retail stores. The regional distribution facilities help to pool risk by consolidating shipments from the import distribution centers (DCs). Most companies have a complex distribution network with several import and regional DCs, for example, Target, Inc. has three import warehouses, 22 regional distribition 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.* [16]).

Companies in the U.S. have been spending $14 billion per year on inventory interest, insurance, taxes, depreciation, obsolescence and warehousing. The logistics activity accounts for 15-20% of the total cost of finished goods (see Menlo [25]). With such a huge inventory investment and growing demand for warehouses, it is important to make optimal decisions for the facility locations and inventory allocation

[1]e.g., different regions within US are northeast, southeast, midwest, northwest and southwest

in a supply chain.

Facility location decisions are termed as *strategic decisions* in a supply chain as it is expensive to open or close a facility. The main reason for locating distribution facilities between the manufacturer(s) and retail stores is to help consolidate shipments and pool risk arising because of uncertainty in transportation time and demand (see Chopra *et al.* [10]). In order to make a decision on the optimal number of facilities to open, several trade-offs between the facilities, the inventory and the transportation costs should be considered. It is good to have many facilities serving the retail stores as it reduces the transportation costs and improves the service level. But having too many facilities could mean increased inventory and facility operating costs. There has been an extensive literature in the area of the facility location-allocation problem (see Daskin [14]). This stream of literature ignores the inventory decisions at the DCs.

Inventory allocation decisions are termed as *tactical* decisions as these decisions are made over shorter time intervals (e.g., weeks or months). Studies in the area of inventory allocation make decisions on what should be the optimal inventory policy at the store and the distribution centers. Inventory is a key driver of a supply chain's 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 and this problem is often known as the multi-echelon inventory problem (see Roundy [32], Deuermeyer *et al.* [15], Ganeshan [19]) for a detailed review). This stream of literature assumes that the 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 focus of this thesis is to study integrated facility location and multi-echelon inventory allocation problems. It is important to discuss the motivation for integrating the two problems. Most companies face the strategic decision of deciding on the number of distribution centers, their locations, and which customers to serve while maintaining acceptable service levels. 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. The replenishment frequency can be determined once the inventory policy is known. Similarly, the service level at the different facilities is a function of their inventory policy. The multi-echelon inventory 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. These interrelations between the two problems suggests that an integrated model with the facility, transportation and inventory costs is needed to solve a network design problem.

In recent years, there has been some research in the area of joint facility location and inventory problems. These studies consider inventory at the single echelon of regional distribution centers to account for demand uncertainty, and linear (Nozick *et al.*[30] or nonlinear (Shen *et al.* [34], Miranda *et al.* [26]) inventory cost functions are added to the facility location model. The impact of uncertainty in transportation in terms of distance and time is ignored in all of these research works. To the best of our knowledge, the study by Teo *et al.*[36] is the only paper that solves a joint model for the facility location and multi-echelon inventory problem. See Chapter II (page 9) for a detailed review of their work.

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.

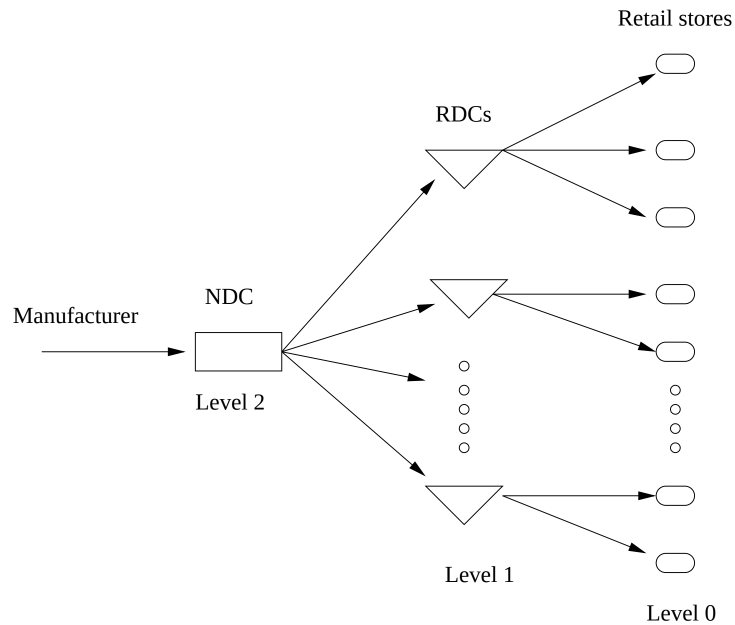

Figure 1: A Multi-Level Distribution Network

This study presents an integrated facility location and multi-echelon inventory allocation problem in a logistic network. The objective function in our problem 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. Service level requirements are modeled in terms of two common measures: stock-out probability and fill-rate. The service level measures are discussed in detail in Section 3.1. The key decisions in this problem are where to locate the RDCs, how to assign retail stores to them and what should be the inventory policy at the national and regional levels of DCs. It is reasonable to assume that the retail stores carry no inventory (e.g. most retail stores such as Wal-Mart carry negligible inventory).

Optimal inventory policy for a multi-echelon logistic network is a vast research area in itself. In this thesis, we restrict the focus on a special class of inventory policies, namely, the continuous review policy. With advancements in information technology, it is now possible for companies to keep track of inventories at all times.

Thus, continuous review policies are used in real-life inventory management. The first part of the thesis focuses on a batch order (Q, r) policy and the second part on a base stock (*s-1*, *s*) policy. The (Q, r) policy is often chosen by inventory stocking locations in order to maintain a stable order quantity in each period. Deuermeyer *et al.* [15], Schwarz *et al.*[7] and Ganeshan [19] analyze a single warehouse and multiple retailer distribution system where each location follows a continuous review (Q, r) policy. Having fixed the inventory policy, the decision then is what is the optimal value of the parameters (Q, r) (batch ordering policy) or s (base stock policy) .

The main purpose of this research is to attain a three-fold goal:

- First, to highlight the importance of integrating the facility location decisions with the inventory decisions. We show that the non-integrated problem generates results that have a significantly higher total network cost as compared to the integrated problem.

- Second, to present two solutions namely *continuous approximation for non-homogenous data* and *linear reformulation approximation* for solving the integrated facility location and inventory allocation problem.

- Third, propose a methodology for fine tuning the continuous approximation technique when the input variables cannot be approximated by a smooth function.

The rest of this thesis is organized as follows. Chapter 2 presents the relevant literature review. In chapter 3 we present a detailed model for the integrated facility location and inventory allocation problem. Chapters 4 states the assumptions used in this work and explains a detailed solution technique using the two-phase continuous approximation approach. In chapter 5, the problem is modeled using a fill-rate service measure. Chapter 6 discusses a numerical example and the managerial insights derived from it. A discrete model and solution for the integrated facility location and inventory allocation problem is presented in chapter 7. Chapter 8 discusses the problem under a base stock inventory allocation policy.

CHAPTER II

LITERATURE REVIEW

Past literature relevant to this study can be categorized into five areas - facility location-allocation, multi-echelon inventory policies, integrated network design and inventory policy decisions, data approximation for logistic network design, and spatial approximation.

Facility location problem

The studies in the area of facility location decisions focus on where to locate the distribution facilities and how to allocate stores to them. There has been an extensive literature in the area of facility location and store allocation problem (see Daskin [14]). This stream of literature terms the inventory decisions as tactical and ignores them in the model.

Multi-echelon inventory allocation problem

Multi-echelon distribution network is defined as a network with one warehouse and several retail stores. In these networks inventory is stored at the retail stores and the warehouse. The focus of the multi-echelon inventory problem is to make decisions on the optimal inventory policies at the different facilities that stock inventory. The break-through work in this area is attributed to Roundy [32] and Muckstadt *et al.* [27]. They show that the power-of-two policy is a lower bound on the cost of all policies. In the years that followed, multi-echelon inventory policies have been studied extensively (see Ahire *et al.* [1] for a comprehensive review). One-for-one ordering and batching ordering policies are the most commonly used policies in the literature on multi-echelon inventory problem. In the case of one-for-one policy it is possible to model the expected inventory and penalty cost expression exactly. This is one of the reasons why it is a commonly used model and the (Q, r) policy is an extension of the one-for-one policy. Deuermeyer *et al.* [15], Schwarz *et al.*[7], and Ganeshan [19] examine distribution networks where each facility follows a (Q, r) inventory policy and the retailers face a stationary Poisson demand. The key to solving the optimal inventory allocation problem in these studies has been to decompose the system into smaller sub-systems with their own demand arrival process. An approximate model is presented in Deuermeyer *et al.* [15] to calculate the system service levels, and Schwarz *et al.* [7] develops an optimization framework to maximize the system fill-rate subject to a safety stock budget constraint. More recently Ganeshan [19] develops an optimization framework to find the optimal value of (Q, r) parameters while minimizing the total logistics costs and meeting the desired fill-rate constraint. The problem in our study is modeled using the decomposition framework in the aforementioned studies. Our work differs from the work done by Ganeshan as we solve an integrated problem with multi-echelon inventory decisions between the RDCs and NDCs, and facility location decisions for the RDCs.

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. Locations for the distribution centers (DC) is a decision variable and so is the inventory policy at the DC. It is assumed that each retailer has a variable demand process. Since the addition of the inventory terms make the objective function nonlinear, researchers have looked at approximations to linearize it.

Nozick *et al.* [29] approximate the safety stock cost at each DC by a linear regression function of the number of DCs and use 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 linear inventory cost function is used in the fixed-charge facility location model defined in Daskin [14] to determine the least cost set of DC locations. Nozick *et al.* [30] extend 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. The stock-outs are incorporated in the safety stock function while the time-based delivery constraint is modeled explicitly as coverage distance.

Shen *et al.* [34] study 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 solves for which retailers should serve as DCs and how much inventory to hold at these stocking points. The inventory model studied in their work is the continuous review (Q, r) model with a *Type-I service* constraint. The nonlinear problem in their work is reformulated as a set-covering model. They propose a column generation algorithm that can solve the problem exactly for two special cases in $O(|n|^2 log|n|)$.

An integrated model for capacitated facility location problem (CFLP) and inventory control decisions is presented by Miranda *et al.* [26]. Their model solves for the location of each distribution center based on the (Q, r) inventory policy at each DC location. The solution methodology involves a lagrangian relaxation and the sub-gradient method. In another work, Erlebacher *et al.* [16] study a distribution system design problem with customers having demands distributed uniformly along a grid network. They propose a two-stage heuristic procedure that fixes the number of DCs in the first stage to estimate demands at the DCs and use these demands in the second stage to estimate the number of DCs.

More recently, Teo *et al.* [36] study an integrated logistic network problem in which they consider inventory cost for multiple echelons of inventory stocking locations. In their model, the inventory cost is modeled at each DC and retailer. They use the convex inventory minimization function proposed by Roundy [32] along with the transportation and facility costs to formulate a MIP problem. They propose a column generation technique to solve their model. Their solution is solvable in O(nlog(n)) time and it is within 2% of the optimal solution when the problem instance is small (20 warehouses and 100 retailers). This model does not include either the demand or the supply uncertainty. In another study, Teo *et al.* [31] extend the 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 problem with 284 retail stores representing the southeastern region of a major US retail chain's 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 [28] that uses data approximation techniques for warehouse location problem. Geoffrion [20] studies continuous model for warehouse location where a warehouse serves demand that is distributed uniformly over a plane. A General Optimal Market Area (GOMA) model is studied by Erlenkotter [17] to determine the optimal area served from a single production point when the demand is assumed to be distributed uniformly. It is an extension of the previous work by Geoffrion [20] and Newell [28], with more detailed expressions for the production cost. Further refinements to the GOMA model are studied by Rutten *et al.* [33] by considering a distribution network, and adding inventory cost terms. Burn *et al.* [8] study a distribution network with a single supplier and multiple customers. They propose an analytic method that uses the spatial density of customers to minimize the inventory and transportation cost for freight. Two different distribution strategies, direct shipping and peddling, are considered in their work. Daganzo [12] presents continuum approximation techniques for network designing problem in particular focusing on vehicle dispatching schedule. Langevin *et al.* [23] present an extensive review of continuous approximation models that have been developed for freight distribution problems. Dasci *et al.* [13] study a production and distribution design problem using the continuum approximation technique. Their work explicitly models the facility cost 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. They do not include inventory costs in their study.

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.* [6] study the logistics planning models that use continuous space models under general conditions, i.e., by relaxing the assumption of uniform density for stores. They develop an analytical framework for estimating transportation costs for distributing goods from a single origin to multiple destinations. In their work they form clusters to account for dense customer destinations. These clusters are then analyzed as sub-regions of the main region. Wang et al. [38] study spatial modeling and propose 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 details captured by the data.

We present a two-phase approximation technique to solve the integrated facility location and inventory allocation problem that captures non-homogeneity of input parameters discussed in the papers by Blumenfeld [6] and Wang [38]. Later, another solution approach is presented which is based on the idea of linearization of nonlinear terms in the discrete cost functions.

The model presented in this work considers an integrated logistic cost function with facility location costs, transportation costs, and multi-echelon inventory costs, and captures both the external and internal sources of variability. The following questions are answered: 1) which RDC locations should be open 2) which retail store should be served from which RDC location, and 3) how much inventory to hold at the RDCs and the NDCs, such that the total facility, inventory and transportation cost is minimized.

CHAPTER III

CONTINUOUS APPROXIMATION MODEL FOR THE INTEGRATED FACILITY LOCATION AND INVENTORY ALLOCATION PROBLEM

In this chapter, a detailed model is presented for the integrated facility location and inventory allocation problem. The underlying assumptions are an initial step towards creating a practical model and have been used extensively in the literature (see Ganeshan [19], Teo *et al.* [36], Dasci *et al.* [13]). Some of these assumptions would be relaxed to build more complex models.

3.1 Assumptions

1. The distribution network under study is an *arborescence network* (see Figure 2) 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. The stores maintain a minimal amount of inventory and it is ignored in this analysis. 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.* [13]) 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. Each RDC and NDC makes decentralized decisions on the inventory control policy.

    A multi-level inventory system could be centralized or decentralized. A centralized control policy is based on the inventory status of all the stocking facilities; e.g., a decision to ship a batch from the DC to a retailer should incorporate

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

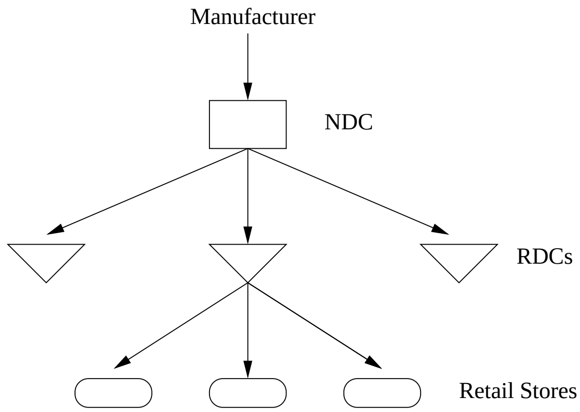

Figure 2: Arboresence Network.

information on the inventory at the other retailers. Centralized policies are very complicated to implement. Decentralized policies make decisions on the inventory control decision at each stocking location without any information on the inventory at the other locations.

In this analysis, a partly decentralized inventory policy is assumed. Even though each stocking location controls its own inventory, a centralized cost model is used in determining the optimal inventory for all stocking points. A centralized cost model takes the interrelationships between the upper echelon and lower echelon into account. For example, a higher base stock level at the upper echelon can reduce the lead time (in terms of stock-out delays) to the lower level.

3.2 Components of the Total Network cost function

This section models the facility location, transportation and inventory costs for the distribution network under study. Note that these cost functions can be categorized as *Mean-dependant* and *Variance-dependant*. *Mean-dependant* cost functions are modeled in terms of mean demand and/or mean travel time parameters while the *Variance-dependant* cost functions have variance of the demand and/or travel times parameters. The facility location cost function does not belong to either of the two categories. The transportation costs belong to the former category as it is influenced by the mean demand during the planning horizon. The inventory functions belong to both the mean-dependant and variance-dependant categories. Detailed analysis in Sections 3.2.3 and 3.2.4 shows how the inventory cost function depends on the variance in demand and travel time as well as on the mean demand and the mean travel times.

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.* [9] 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 $\int_R (A_r(x))^{-1} dx$.

3.2.1 Facility cost for RDC

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,

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

3.2.2 Average Transportation 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.

Average Inbound Transportation cost

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 $E[D_r(x)]$ (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.

Remark: The fixed inbound transportation cost can be interpreted as the cost associated with driver wages and trucking equipment maintenance. The variable cost on the other hand is the cost associated with the volume of goods ordered in each shipment. This can be interpreted as the cost associated with picking and packaging the goods in the warehouse, and loading the truck. There are several studies that consider the variable inbound cost component as a linear function of the volume per shipment (see Rutten *et al.* [33], Teo *et al.* [36]).

Outbound delivery cost

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.* [12, 13]). 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.* [13]).

3.2.3 Average Inventory cost for RDC

The research in the area of multi-echelon inventory allocation problems extensively uses continuous review policies (see Ahire *et al.* [1], Deuermeyer *et al.* [15], Ganeshan [19]). In a distribution network with one warehouse, multiple retailers, where both the warehouse and the retailers hold the inventory, assuming a continuous review policy allows for the approximation of the actual demand process at the warehouse. This approximation is needed as the demand process at the warehouse is a complex non-stationary renewal process as shown in Deuermeyer *et al.* [15]. It is for this reason that a continuous review policy is assumed at the RDCs and NDCs in this work as it facilitates the approximation of the demand process at the NDC. The demand process at the RDCs and NDCs is needed for the the analysis of the inventory cost functions.

Under a (Q, r) policy, each RDC places an order of Q units every time its inventory position (sum of on-hand and on-order inventory) is at or below r. The order quantity Q is known as the *cycle inventory* as it is used to satisfy the demand over the replenishment cycle. Reorder point r determines the inventory in stock to satisfy the demand during the replenishment lead time, i.e., the time between placement of a replenishment order and its arrival. It is a function of both the mean and variance of demand during the replenishment lead time. The optimal value of the reorder point is calculated by either considering the trade-off between holding and penalty cost or using a service level measure defined by the management. In real life it is hard to quantify the penalty cost associated with backorders and for this reason we focus on service level measures to estimate the reorder levels. The different service level measures are discussed next.

Service level

The two most common ways of defining the service level for a given inventory policy are:

1. Type-I service level: Proportion of order periods over which demand is fully met (i.e., when there is no stock-out);

2. Type-II service level: Proportion of demand satisfied immediately from inventory in stock.

An α value for *Type-I* service level implies that there is a $1-\alpha$ chance of stocking out in a given replenishment cycle. Thus, for a distribution center with a (Q, r) policy, if the demand over the lead time, $(D_{r,LT})$, is normally distributed with mean $E[D_{r,LT}]$ and variance $Var[D_{r,LT}]$, the cycle service level is given by:

$$Pr(D_{r,LT} \leq r) \quad = \quad \alpha \tag{4}$$

Similarly, a ρ value for *Type-II* service level implies that during a given replenishment interval ρ proportion of demand will be filled from stock. Thus, using the fill-rate or *Type-II* as a service measure allows for the estimation of fraction of demand that will be converted to sales or equivalently what was the expected number of units backordered/lost during the replenishment interval.

Even though fill-rate is recognized as the true service measure, it is the *Type-I* service that appears widely in the inventory literature (see Shen *et al.* [34], Miranda *et al.* [26], Teo *et al.* [36]). The reason for this is the relatively simpler expressions to model the *Type-I* service levels while determining the inventory policy. It is for the same reason that we first model the integrated facility location and inventory allocation problem using a *Type-I* service measure. Later in chapter V, a *Type-II* service measure model is discussed along with the challenges in implementing it in the problem.

Remark: There are two kinds of inventory considered in this analysis, namely, the average cycle stock ($Q/2$) and the average safety stock. Average cycle stock is the average amount of inventory in between the replenishment cycles. The more the cycle stock in each cycle, the fewer the reorders. The amount of cycle stock at a DC depends on the trade-off between the inventory holding cost at that DC and the fixed reorder cost for the DC every time it places a batch order. The reorder cost is a function of the *mean demand* and hence the cycle stock depends on the demand process. Average *safety stock* on the other hand is the inventory in stock to handle demand uncertainty during the replenishment cycle.

Figure 3 depicts the behavior of on-hand inventory over time when the demand process is uncertain. Reorders are triggered at points (a, b, c, d, e) when the on-hand inventory just falls below the reorder point. The vertical lines in the figure represent the order quantity Q. Note that the on-hand inventory at any point lies between $(Q + (ss + x))$ and $(ss - x)$ where ss denotes the safety stock and x is a random quantity that fluctuates around the safety stock level. Thus the average on-hand inventory is given by $(Q + (ss + x) + (ss - x))/2 = Q/2 + ss$.

Demand process at the RDC

The demand process at each retail store is Poisson with rate $\lambda(x)$. It is common

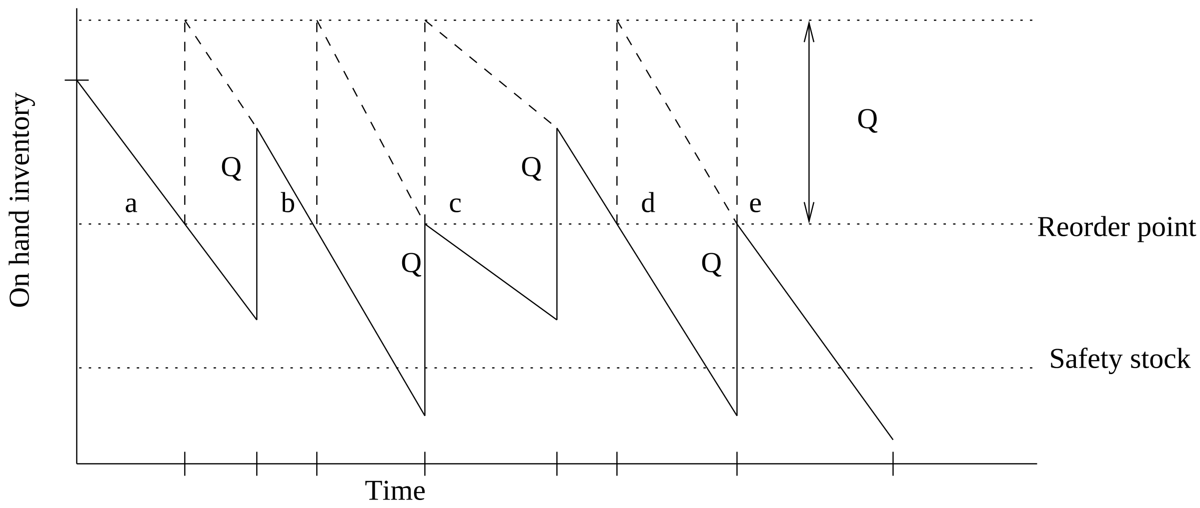

Figure 3: On-hand inventory versus time in the (Q, r) model, (a, b, c, d, e) are the reorder times

to assume a Poisson i.i.d demand process for a problem with multiple retail stores (Ganeshan [19], Nozick *et al.* [30]). Then the demand process at RDC r serving influence area $A_r(x)$ is Poisson with a demand rate given by the sum of the demand rates of individual stores within $A_r(x)$ (see Result 1).

Result 3.1: Let $X_1, X_2, ..., X_n$ be i.i.d Pois(λ_i) where $i = 1, 2, ...n$, then

$$\textstyle\sum_{i=1}^{n} X_i \quad \sim \quad \text{Pois}(\sum_{i=1}^{n} \lambda_i)$$

The number of stores in the region $A_r(x)$ is given by $\int_{A_r(x)} \delta(x)\text{dx}$. If $\delta(x)$ is a slow varying function of $x \in A_r(x)$ then $\int_{A_r(x)} \delta(x)\text{dx} = \delta(x)A_r(x)$. Since identical demand is assumed across the retail stores, the demand rate at RDC r is equal to the demand rate at each store within $A_r(x)$ times the number of stores, i.e., $(\lambda)(\delta(x)A_r(x))$. Let $D_r(x)$ be the demand process at RDC r, then we have $D_r(x) \sim \text{Pois}(\lambda(x)\delta(x)A_r(x))$. Using properties for a Poisson distribution, we define the mean and variance of D_r as:

$$E[D_r(x)] = \lambda(x)\delta(x)A_r(x) \quad Var[D_r(x)] = \lambda(x)\delta(x)A_r(x) \tag{5}$$

Next, we analyze the demand process at the RDC r during its order replenishment time LT. The order replenishment time has two main components- travel lead time and the order processing lead time. Let μ_r and ${\sigma_r}^2$ be the mean and variance of the travel lead time between the NDC and RDC r.

Service level at the RDC

Result 3.2: When the demand over lead time is normally distributed with mean $E[D_r, LT]$ and variance $Var[D_r, LT]$, we get

$$r_r = E[D_{r,LT}] + Z_{\alpha_r}\sqrt{Var[D_{r,LT}]} \tag{6}$$

Proof: See Appendix 3(c)

Equation 6 models the reorder point as a sum of the expected demand and the variability of demand during the lead time. $Z_{\alpha_r}\sqrt{Var[D_{r,LT}]}$ is the safety stock at the RDC.

Remark: Note that the safety stock for the RDC is not modeled as a function of safety stock at the NDC. It is reasonable because if a stock out happens at the RDC it loses customer demand irrespective of whatever the NDC has enough safety stock. Thus, it is assumed that the RDC is independently accounting for the total stock outs at the end customers and hence the interrelationship between the safety stocks are ignored.

Remark: The waiting time at a lower level facility due to stock-outs at the upper level facility is less than the travel lead time between the two levels of facilities (see Ganeshan [19]). This is ensured by the management by carefully selecting the service level α. Using this line of argument, the waiting time component can be ignored and only the travel lead time component is used in this analysis.

Thus, the total order replenishment time has a mean and variance given by μ_r and ${\sigma_r}^2$. The expected demand ($E[D_{r,LT}]$) and variance of demand ($Var[D_{r,LT}]$) at the RDC during its order replenishment period is given by the following (see Appendix

3(b)):

$$E[D_{r,LT}] = \mu_r E[D_r] \tag{7}$$

$$Var[D_{r,LT}] = \mu_r Var[D_r] + {\sigma_r}^2 E[D_r]^2 \tag{8}$$

We can now define the expected inventory cost at the RDC as a function of the holding cost h_r and reorder cost R_r. While holding cost depends on the cycle and the safety stock, the reorder cost depends only on the cycle stock.

Average Reorder cost for the 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

$$TR_r(x) = N_r(x)\left(R_r(x)\right)\left(\frac{E[D_r]}{Q_r(x)}\right) \tag{9}$$

Average inventory at the RDC is given as the sum of the cycle inventory $Q_r(x)/2$ and safety inventory ($Z_{\alpha_r}\sqrt{Var[D_{r,LT}]}$). In order to get the total average cycle plus safety inventory, we simply multiply by the total number of RDCs, $N_r(x)$, in the NDC service area, R where $N_r(x)$ is given by $\int_R (A_r(x))^{-1}dx$. We will discuss the design of the NDC service region under the discussion of the two-phase approximation method (see Section 4).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

$$TI_r(x) = h_r N_r(x)\left(\frac{Q_r(x)}{2} + Z_{\alpha_r}\sqrt{Var[D_{r,LT}]}\right) + TR_r(x) \tag{10}$$

where $E[D_r]$ is given by equation 5 and all the other variables have been defined above.

3.2.4 Average Inventory cost for NDC

Each NDC follows a continuous review policy and places an order of $Q_n(x)$ units every time its inventory position is at or below $r_n(x)$. $Q_n(x)$ is the cycle inventory that is used to satisfy the demand over the replenishment cycle. Reorder point $r_n(x)$ is a function of the safety inventory that is stocked to handle the demand uncertainty. Safety stock is modeled using information on demand during the order replenishment time. The order replenishment time for the NDC is simply the travel time between the manufacturer and the NDC. The inventory cost at the NDC is a function of the holding cost h_n and reorder cost $R_n(x)$. While holding cost depends on both the cycle and safety stock the reorder cost depends only on the cycle stock.

Demand process at the NDC

In order to estimate the demand process $D_n(x)$ at the NDC we need to account for all the orders from all the RDCs served by the NDC. Different order quantities and timing for each order creates a variable demand at the NDC. Infact the time between the demand occurrences at the NDC is a non-stationary renewal process and the following result is used in this analysis:

Result 3 (see Deuermeyer [15]. pp 173-174): The variance of the demand process (in units of RDC order Q_{r_i}s) at the NDC during its lead time, L_n, is given by:

$$Var[D_{n,LT}] = \sum_r \frac{(\lambda(x)\delta(x)A_r(x))L_n}{(Q_r(x))^2} \tag{11}$$

We can interpret this result as follows. The demand rate at RDC r is $\lambda(x)\delta(x)A_r(x)$ and each RDC r places an order with the NDC every time it has used Q_r units. Thus, the demand rate at the NDC in units of orders from RDC r can be approximated by $(\lambda(x)\delta(x)A_r(x))/Q_r(x)$ where $(Q_r(x), r_r(x))$ is the inventory policy at r. The demand rate during the lead time L_n is obtained by simply multiplying $(\lambda(x)\delta(x)A_r(x))/Q_r(x)$ with L_n. The total demand is obtained by taking the sum of the demand rate due to each individual RDC. When we take the variance over this new expression we get the desired result. Note that since the lead time for the NDC is stochastic in our analysis. As an initial guess we can estimate L_n with the expected lead time to the

NDC given by μ_n.

Average Reorder cost for the 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

$$TR_n(x) = R_n(x)\left(\frac{E[D_n(x)]}{Q_n(x)}\right) \tag{12}$$

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 $Var[D_{n,LT}]$ is given by equation (11) 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

$$TI_n(x) = h_n\left(\frac{Q_n(x)}{2} + Z_{\alpha_n}\sqrt{Var[D_{n,LT}]}\right) + TR_n(x) \tag{13}$$

where $Var[D_{n,LT}]$ is given by $\sum_r(\lambda(x)\delta(x)A_r(x))\mu_n)/(Q_r(x))^2$.

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

$$\text{minimize} \quad \int_R(TNC(x))dx = \int_R(TF(x) + TIT(x) + TOT(x) + TI_r(x) + TI_n(x))dx \tag{14}$$

s.t.

$$\sum_r A_r(x) = R \tag{15}$$

$$Q_n(x) \geq 0 \ \forall \ n \tag{16}$$

$$Q_r(x) \geq 0 \ \forall \ r \tag{17}$$

$$A_r(x) \geq 0 \ \forall \ r \tag{18}$$

$$Q_n(x), \quad Q_r(x), \quad A_r(x) \in Z^+ \tag{19}$$

where $Q_n(x)$, $Q_r(x)$, $A_r(x)$ are the decision variables in this problem. Equation (15) is the area coverage constraint. It ensures that the entire service region is covered by the sum of the RDC influence areas. Equation (16), (17) and (18) are the non-negativity constraints for the decision variables. Equation (19) 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 (16), (17) and (18) 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.

3.3 Appendix- Chapter III

Appendix 3(a)

Little's Law [24]: If n is the average number of customers in the system and τ is the average arrival rate of customers, then the expected waiting time t is given by equation 20.

$$t = \frac{n}{\tau} \tag{20}$$

Appendix 3(b)

Result: If $X_1,\ X_2,\ ...$ are independent and identically distributed, and if N is a nonnegative integer valued random variable independent of the Xs, then

1.
$$E\left[\sum_{i=1}^{N} X_i\right] = E[N]E[X]$$

2.
$$Var\left[\sum_{i=1}^{N} X_i\right] = E[N]Var(X) + (E[X])^2 Var(N)$$

Proof 1: We know that

$$E\left[\sum_{i=1}^{N} X_i\right] = E\left[E[\sum_{i=1}^{N} X_i|N]\right]$$

But

$$\begin{aligned} E\left[E[\sum_{i=1}^{N} X_i|N=n]\right] &= E\left[\sum_{i=1}^{n} X_i\right] \quad \text{by the independence of } X_i \text{ and N} \\ &= nE[X] \end{aligned}$$

Thus,

$$E\left[E[\sum_{i=1}^{N} X_i|N]\right] = NE[X]$$

and so we have

$$\begin{aligned} E\left[\sum_{i=1}^{N} X_i\right] &= E\left[NE[X]\right] \\ &= E[N]E[X] \end{aligned}$$

In our work, X_i is the demand processes at RDC r at a random time point i, the value of N is LT, i.e., the travel lead time interval, and $\sum_{i=1}^{LT} X_i$ is the demand process at the RDC during an interval of length equal to the lead time, LT. Thus, $\sum_{i=1}^{LT} X_i = E[LT]E[D_r]$ where D_r is the demand process per unit time at RDC r, and $E[LT]$ is the expected travel lead time. Result 1 can be re-stated as $E[D_{r,LT}] = \mu_r E[D_r]$ where $E[LT] = \mu_r$.

Proof 2: We have

$$Var\left(\sum_{i=1}^{N} X_i\right) = E\left[\left(\sum_{i=1}^{N} X_i\right)^2\right] - \left(E\left[\sum_{i=1}^{N} X_i\right]\right)^2 \tag{21}$$

If we condition on N, we get

$$E\left[\left(\sum_{i=1}^{N} X_i\right)^2\right] = E\left[E\left[\left(\sum_{i=1}^{N} X_i\right)^2 |N\right]\right]$$

Now we have the following identity,

$$E[Z^2] = Var(Z) + (E[Z])^2$$

So this gives us

$$\begin{aligned} E\left[\left(\sum_{i=1}^{N} X_i\right)^2 |N = n\right] &= Var\left(\sum_{i=1}^{N} X_i\right) + \left(E\left[\sum_{i=1}^{N} X_i\right]\right)^2 \\ &= nVar(X) + (nE[X])^2 \end{aligned}$$

where Z is a random variable. In particular for any N, we get

$$E\left[\left(\sum_{i=1}^{N} X_i\right)^2 |N\right] = NVar(X) + (NE[X])^2$$

Taking expectations on both sides we get

$$E\left[\left(\sum_{i=1}^{N} X_i\right)^2\right] = E[N]Var(X) + E[N^2](E[X])^2$$

Substiting this in equation 21 gives

$$\begin{aligned}
Var\left(\sum_{i=1}^{N} X_i\right) &= E[N]Var(X) + E[N^2](E[X])^2 - \left(E\left[\sum_{i=1}^{N} X_i\right]\right)^2 \\
&= E[N]Var(X) + E[N^2](E[X])^2 - (E[N]E[X])^2 \\
&= E[N]Var(X) + (E[X])^2(E[N^2] - (E[N])^2) \\
&= E[N]Var(X) + (E[X])^2 Var(N)
\end{aligned}$$

We use the same line of reasoning as discussed after result 1. $Var(X)$ is the variance of the demand process per unit time at RDC r and it is given by $Var[D_r]$. $Var(N)$ is the variability in the travel lead time and it is given by ${\sigma_r}^2$.

Appendix 3(c)

$$\begin{aligned}
& Pr({D_r}^{LT} < r_r) = \alpha_r \\
\Rightarrow \quad & Pr(\tfrac{D_{r,LT} - E[D_r,LT]}{\sqrt{Var[D_r,LT]}} < \tfrac{r_r - E[D_r,LT]}{\sqrt{Var[D_r,LT]}} = \alpha_r \\
\Rightarrow \quad & Pr(z < \tfrac{r_r - E[D_r,LT]}{\sqrt{Var[D_r,LT]}} = \alpha_r \text{ where } z \sim N(0,1) \\
\Rightarrow \quad & \Phi(\tfrac{r_r - E[D_r,LT]}{\sqrt{Var[D_r,LT]}} = \alpha_r \\
\Rightarrow \quad & \tfrac{r_r - E[D_r,LT]}{\sqrt{Var[D_r,LT]}} = \Phi^{-1}(\alpha_r) \\
\Rightarrow \quad & r_r = E[D_r, LT] + \Phi^{-1}(\alpha_r)\sqrt{Var[D_r, LT]} \\
\Rightarrow \quad & r_r = E[D_r, LT] + Z_{\alpha_r}\sqrt{Var[D_r, LT]} \text{ where } Z_{\alpha_r} = \Phi^{-1}(\alpha_r)
\end{aligned}$$

CHAPTER IV

SOLUTION METHODOLOGY: TWO-PHASE APPROXIMATION

This section describes a two-phase approximation technique used to solve the facility location and inventory allocation problem. Two-phase approximation in a extension to the continuous approximation (CA) approach (see Daganzo [12]). 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 4). The distribution network given in figure 4 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 Chapter 3 and it is solved using the CA approach in phase-II.

4.1 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.

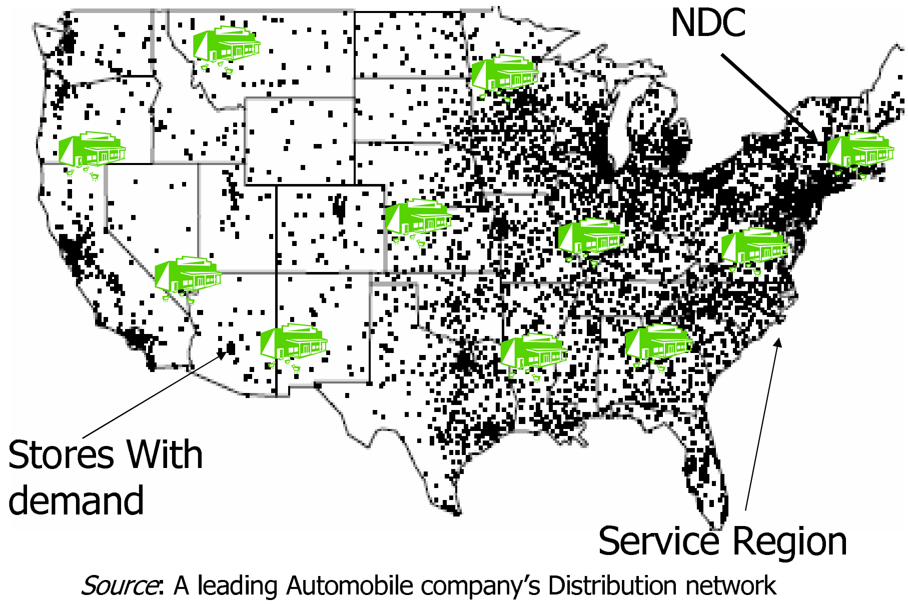

Figure 4: Example of a Supply chain.

This problem of assigning equal demands to each NDC is a special case of the classic *Transportation problem* (see Appendix 4(a)). 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*[1] 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.

[1] 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 sub-divided counties into civil townships.

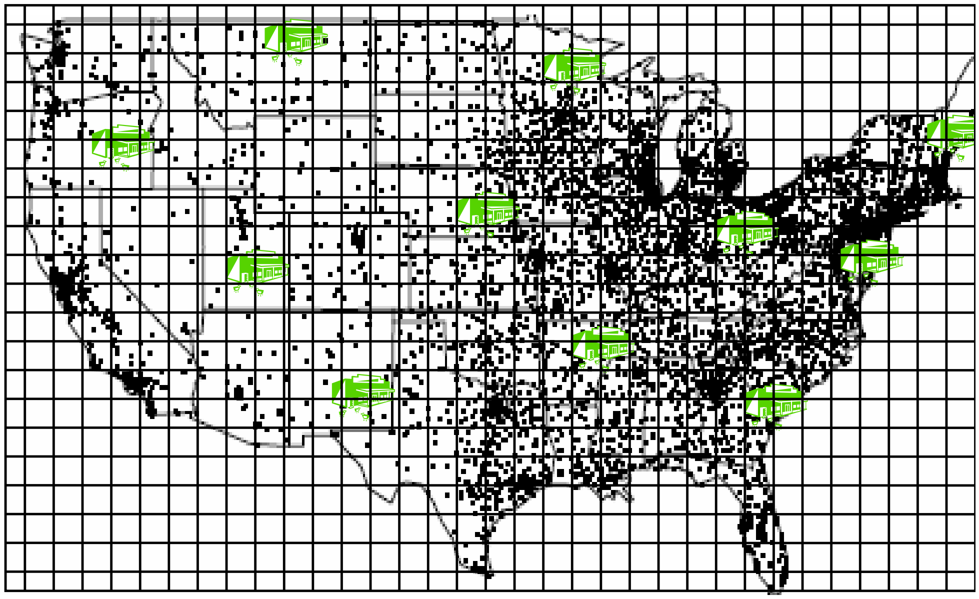

Figure 5: Grid Cover

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 5. 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 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 6 and Figure 7 illustrate this idea. Clusters $(C_{j_1}, C_{j_2}, ...C_{j_i})$ exist within each NDC region A_i such that the store

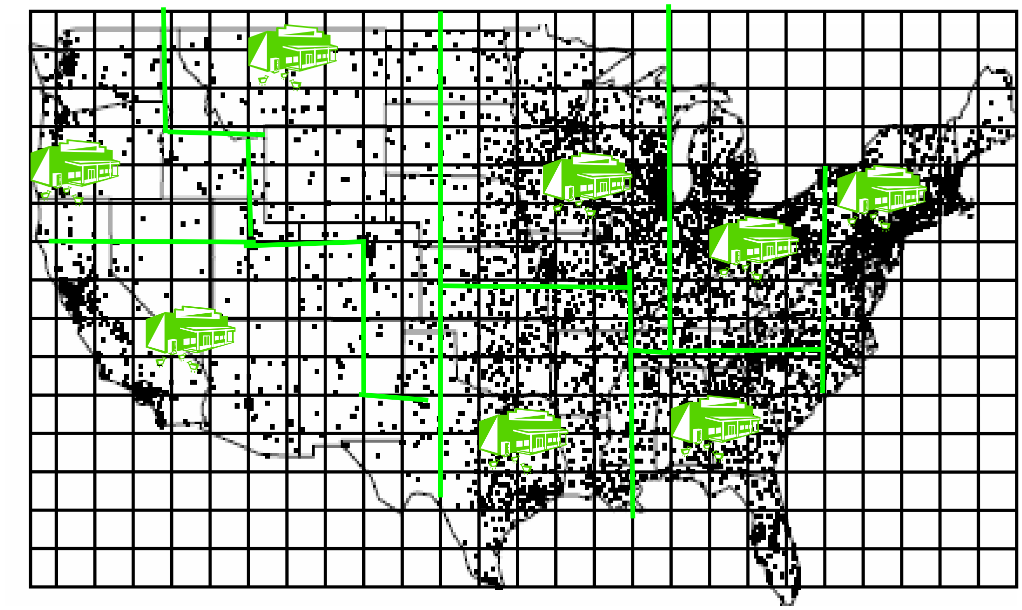

Figure 6: NDC sub-region.

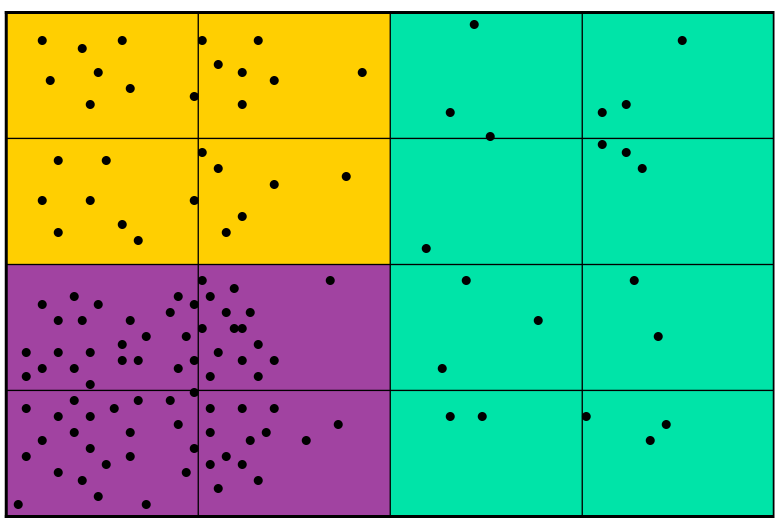

Figure 7: Coupling.

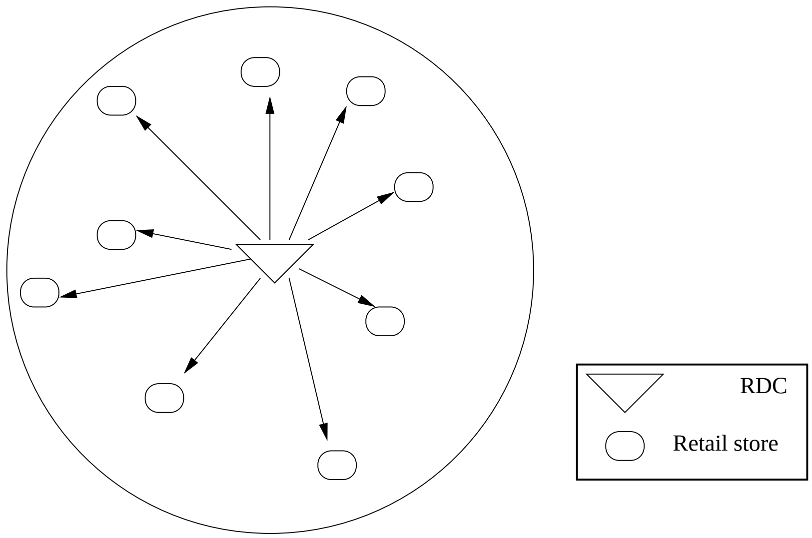

Figure 8: Influence area for a RDC

density is nearly constant over each cluster.

4.2 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 8 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.

4.3 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

$$
\begin{aligned}
Q_{r_i}(x) &\geq 0 \;\; \forall \;\; r_i \\
A_{r_i}(x) &\geq 0 \;\; \forall \;\; r_i \\
A_{r_i} &\leq C_{n_i} \;\; \forall \;\; r_i \\
Q_n(x) &\geq 0 \\
Q_{r_i}(x), \frac{C_{n_i}}{A_{r_i}(x)}, Q_n(x) &\in Z^+
\end{aligned}
\tag{22}
$$

$$
\begin{aligned}
\text{where } \; ss_{r_i}(x) &= Z_{\alpha_{r_i}}\sqrt{\mu_r\left(\lambda(x)\delta_i(x)A_{r_i}(x)\right) + (\sigma_r)^2(\lambda(x)\delta_i(x)A_{r_i}(x))^2} \\
ss_n(x) &= Z_{\alpha_n}\sqrt{\left(\mu_n\sum_{i=1}^{N}\frac{\lambda(x)\delta_i(x)C_{n_i}(x)}{(Q_{r_i}(x))^2}\right)}
\end{aligned}
\tag{23}
$$

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. For the special case when $Q_i = Q$ holds, the $ss_n(x)$ term becomes linear as well. Then the only nonlinear term in the objective function is the reorder cost at the NDC. 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. For this analysis, we start by focusing on the solution methodology for problem P(1) for the special case. Later, a solution for the original case with unequal Q_is is discussed.

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 4(b)): If a and b are positive numbers and $x > 0$, then

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

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 where $a = \mu_r$, $b = (\sigma_r)^2$, $x = \lambda(x)\delta_i(x)A_{r_i}(x)$ □

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.

4.3.1 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.* [5]). 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.* [5], 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.* [5], 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.* [5], 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.* [5], 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.

4.3.2 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.* [15], Ganeshan [19]). 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. Later, in the second part (case 2), the problem is solved for unequal Q_{r_i}s.

The original problem can be expressed by a modified problem P^E where $\phi(\mathbf{A}, Q, Q_n)$ is a lower bound on TNC for the case $Q_{r_i} = Q$. Note that $\mathbf{A}$ is a n-dimensional row vector defined by $\mathbf{A} = [A_{r_1}, A_{r_2}, ..., A_{r_N}]$.

P^E Minimze

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

subject to

$$\begin{aligned}
Q &\geq 0 \;\; \forall \;\; i = 1, 2, ..., N \\
k &\geq 2 \\
A_{r_i} &\geq 0 \;\; \forall \;\; i = 1, 2, ..., N \\
A_{r_i} &\leq C_{n_i} \;\; \forall \;\; i = 1, 2, ..., N \\
Q, \frac{C_{n_i}}{A_{r_i}}, k &\in Z^+
\end{aligned}$$

The stationary point for the objective function $\phi(\mathbf{A}, Q, Q_n)$ is given by equations (24), (25) and (26) (for details see *appendix* 4(c)):

$$A_{r_i} = \left(\frac{2F_r + h_r Q - \frac{h_r Z_{\alpha_i}\mu_r}{\sigma_r}}{C_l f_r \xi\lambda\delta_i}\right)^{2/3} \tag{24}$$

$$k = \frac{1}{Q}\sqrt{\left(\frac{2R_n(\sum_{i=1}^{N}\xi\lambda\delta_i C_{n_i})}{h_n}\right)} \tag{25}$$

$$Q = \sqrt{\left(\frac{\sum_{i=1}^{N}\left(C_f + R_r + \frac{R_n}{k}\right)\xi\lambda\delta_i C_{n_i}}{\sum_{i=1}^{N}\frac{h_r C_{n_i}}{2A_{r_i}} + \frac{h_n k}{2}}\right)} \tag{26}$$

Result 4.2: $\phi(\mathbf{A}, Q, k)$ is a convex function for values of ($\mathbf{A}$, Q, k) satisfying the following inequalities given by (27) (see appendix 4(c) for proof).

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

$$\left(\frac{\partial^2\phi}{\partial Q^2} - \sum_{i=1}^{N} \frac{\partial^2\phi/\partial Q\partial A_{r_i}}{\partial^2\phi/\partial {A_{r_i}}^2}\frac{\partial^2\phi}{\partial A_{r_i}\partial Q}\right)\frac{\partial^2\phi}{\partial k^2} - \left(\frac{\partial^2\phi}{\partial Q\partial k}\frac{\partial^2\phi}{\partial k\partial Q}\right) > 0 \qquad (27)$$

Result 4.3: The stationary point of $\phi\,(\mathbf{A}, Q, k)$ is a local minimum.

Proof. The stationary point obtained by solving equations (24), (25) and (26) satisfy the inequalities given by (27). Then by theorem 4.3 the result follows. □

Definition: A function f : **X** x **Y** → R is called *biconvex*, if f(x, y) is convex in y for fixed x ∈ **X**, and f(x, y) is convex in x for fixed y ∈ **Y**.

Result 4.4: $\phi\,(\mathbf{A}, Q, k)$ is a *biconvex* function for all values of Q and k, and values of **A** satisfying equation (24).

Proof. For a fixed value of (Q, k)

$$\begin{aligned}
\frac{\partial^2\phi}{\partial {A_{r_i}}^2} &= -\frac{C_l f_r \xi\lambda\delta_i C_{n_i}}{4{A_{r_i}}^{3/2}} + 2\left(\frac{C_{n_i}F_r + (C_{n_i}h_rQ/2) - (h_r Z_{\alpha_i}\mu_r C_{n_i}/2\sigma_r)}{{A_{r_i}}^3}\right) \\
&> 0 \Leftrightarrow A_{r_i} < (4)^{(2/3)}\left(\frac{2F_r + h_rQ - (h_rZ_{\alpha_i}\mu_r/\sigma_r)}{C_l f_r\xi\lambda\delta_i}\right)^{2/3} \\
&\text{which holds for} \quad A_{r_i} = \left(\frac{2F_r + h_rQ - \frac{h_rZ_{\alpha_i}\mu_r}{\sigma_r}}{C_l f_r\xi\lambda\delta_i}\right)^{2/3}
\end{aligned}$$

Similarly for a fixed value of vector **A**, the objective function $\phi(\mathbf{A}, Q, k)$ is convex in Q and k. It can be be shown that the hessian matrix corresponding to the objective function $\phi(\mathbf{A}, Q, k)$ at the fixed value of **A** is positive semidefinite.

$$|H| = \begin{vmatrix} \frac{\partial^2\phi}{\partial Q^2} & \frac{\partial^2\phi}{\partial k\partial Q} \\ \frac{\partial^2\phi}{\partial Q\partial k} & \frac{\partial^2\phi}{\partial k^2} \end{vmatrix} \geq 0 \quad \text{(see Appendix 4(c) for derivations)}$$

□

4.3.2.1 Solution Procedure

In this section a solution methodology for problem P^E is explained in detail. The objective function of problem P^E is nonlinear but it is shown to be convex over a certain region defined by inequalities given by (27), and is biconvex. A partially unconstrained version is solved first and this solution is modified to get a near optimal solution for problem P^E. The partially unconstrained version of the problem ignores the integer value constraint $(Q, \frac{C_{n_i}}{A_{r_i}}, k \in Z^+)$.

For a fixed value of $\mathbf{A}$, an optimal solution $(Q,\, k)$ for the unconstrained problem is obtained by simultaneously solving equations (25) and (26). The equations can be solved simultaneously using an iterative prodecure and the solution generated is substituted in equation (24) 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 (27). 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^E is generated by forcing the integer value constraint $(Q, \frac{C_{n_i}}{A_{r_i}}, k \in Z^+)$.

If the stationary point does not lie in the convex region of ϕ, a *Response Surface technique* is adopted to generate a good solution for the problem. This will be discussed in detail in section 4.4.

The steps for the iterative procedure are explained below:

1. Fix $k = 0$, $Q_k = 1$.

2. Calculate A_{r_i}, i = 1,2,..N, using equation 24.

3. Use the value of A_{r_i}, i = 1,2,..N, in equations 25 to get Q and calculate k using 26. 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.

4.3.3 Unequal Qs

For analyzing the problem under this case 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.

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 stationary point for the objective function τ satisfies the following equations (see appendix 4(d) for derivation of the stationary point).

$$A_{r_i} = \left(\frac{2F_r + h_r Q_{r_i} - \frac{h_r Z_{\alpha_i} \mu_r}{2\sigma_r}}{C_l f_r \xi \lambda \delta_i}\right)^{2/3} \tag{28}$$

$$Q_n = \sqrt{\left(\frac{2R_n \sum_{i=1}^{N} \xi \lambda \delta_i C_{n_i}}{h_n}\right)} \tag{29}$$

$$Q_{r_i} = \sqrt{2A_{r_i}\left(\frac{(C_f + R_r + (R_n/Q_n))\,\xi \lambda \delta_i C_{n_i}}{C_{n_i} h_r}\right)} \tag{30}$$

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).

***Result* 4.5**: $\tau(\mathbf{A}, \mathbf{Q}, Q_n)$ is convex on a region $L(\mathbf{A}, \mathbf{Q}, Q_n)$ defined by the following inequalities.

$$\frac{2}{{A_{r_i}}^3}\left(C_{n_i} + \frac{h_r C_{n_i} Q_{r_i}}{2} - \frac{h_r Z_{\alpha_i} \mu_r}{2\sigma_r}\right) - \frac{C_l f_r \xi \lambda(x) \delta_i(x) C_{n_i}}{4{A_{r_i}}^{3/2}} > 0$$

$$|H| = \begin{vmatrix} q_{N,N} - \frac{(q_N a_N)^2}{a_N a_N} & q_{N,N+1} \\ q_{N+1,N} & q_{N+1,N+1} - \sum_{i=1}^{N} \frac{q_{N+1,i}}{q_{i,i}} q_{i,N+1} \end{vmatrix} \geq 0$$

where

$$
\begin{array}{rlrlrl}
a_{i,i} = & \dfrac{\partial^2 \tau}{\partial {A_{r_i}}^2}, & q_{i,i} = & \frac{\partial^2 \tau}{\partial {Q_{r_i}}^2}, & q_{N+1,N+1} = & \dfrac{\partial^2 \tau}{\partial {Q_n}^2} \\
a_i q_i = & \dfrac{\partial^2 \tau}{\partial A_{r_i} \partial Q_{r_i}}, & q_i a_i = & \frac{\partial^2 \tau}{\partial Q_{r_i} \partial A_{r_i}} & q_{N+1,i} = & \dfrac{\partial^2 \tau}{\partial Q_n \partial Q_{r_i}} \\
q_{i,N+1} = & \dfrac{\partial^2 \tau}{\partial Q_{r_i} \partial Q_n}, & a_i q_{N+1} = & \frac{\partial^2 \tau}{\partial A_{r_i} \partial Q_n}, & q_{N+1} a_i = & \dfrac{\partial^2 \tau}{\partial Q_n \partial A_{r_i}}
\end{array}
$$

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

Result 4.6: $\tau(\mathbf{A}, \mathbf{Q}, Q_n)$is a biconvex function of [$\mathbf{A}$] and [$\mathbf{Q}$, Q_n]. For proof see appendix 4(d)

4.3.3.1 Solution Procedure

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. Again an iterative procedure, similar in idea to one formulated for the equal reorder quantity case, is used. 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 (28).

3. Use the value of A_{r_i}, i = 1,2,..N, in equation (29) to get $\mathbf{Q} = Q_{r_i}$ and calculate Q_n using (30). 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.

4.4 *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.

4.5 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.

4.5.1 Stationary point for the non-integrated problem - Equal Q

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

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

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

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.

4.5.2 Stationary point for the integrated model using averages

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

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

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

The analysis for the integrated facility location and inventory allocation problem sheds light on some important issues. Among the key observation, we have

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 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.

4.6 Appendix

Appendix 4(a)

Finding the NDC partitions requires solving the following optimization problem.

$$\text{Min} \quad \sum_{i=1}^{N}\sum_{j=1}^{N} d_{ij}X_{ij}$$
$$\text{Subject to}$$
$$\sum_{i=1}^{N} X_{ij} = 1 \quad \forall j$$
$$\sum_{j=1}^{N} X_{ij} = k \quad \forall i$$

where

$$X_{ij} = \begin{cases} 1 & \text{if retailer } j \text{ is assigned to NDC } i \\ 0 & \text{otherwise} \end{cases}$$

Suppose N is the total number of NDCs in the distribution network and m is the total retailers in the network. Then $k = m/N$ follows from the assumption that each NDC handles equal demand. Note that in this analysis each store has equal demand. Thus, assigning equal number of stores to each NDC guarantees that each NDC has equal demand.

Appendix 4(b)

Proof for Monotone Property.

$$\begin{aligned} ax + bx^2 &> ax + bx^2 - \frac{a^2}{4b} \\ &= \left(\sqrt{b}x - \frac{a}{2\sqrt{b}}\right)^2 \\ \Rightarrow \sqrt{ax + bx^2} &> \left(\sqrt{b}x - \frac{a}{2\sqrt{b}}\right) \end{aligned}$$

□

Appendix 4(c)

(1) First order conditions for deriving the stationary point for function $\phi(\boldsymbol{A}, Q, k)$.

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

□

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

$$
H = \begin{bmatrix}
a_{1,1} & 0 & ... & 0 & a_1 q & 0 \\
0 & a_{2,2} & ... & 0 & a_2 q & 0 \\
. & . & ... & . & . & . \\
. & . & ... & . & . & . \\
0 & 0 & ... & a_{n,n} & a_n q_n & 0 \\
q a_1 & q a_2 & ... & q a_N & qq & q q_n \\
0 & 0 & ... & 0 & q_n q & q_n q_n
\end{bmatrix}
$$

where

$$
\begin{aligned}
a_{i,i} &= \frac{\partial^2 \phi}{\partial {A_{r_i}}^2}, \quad qq = \tfrac{\partial^2 \phi}{\partial Q^2}, \quad q_n q_n = \frac{\partial^2 \phi}{\partial k^2} \\
a_i q &= \frac{\partial^2 \phi}{\partial A_{r_i}\partial Q}, \quad q a_i = \tfrac{\partial^2 \phi}{\partial Q \partial A_{r_i}} \quad q_n q = \frac{\partial^2 \phi}{\partial k \partial Q} \\
q q_n &= \frac{\partial^2 \phi}{\partial Q \partial k}, \quad a_i q_n = \tfrac{\partial^2 \phi}{\partial A_{r_i}\partial k}, \quad q_n a_i = \frac{\partial^2 \phi}{\partial k \partial A_{r_i}} \\
& \qquad i = 1, 2..., N
\end{aligned}
$$

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

.

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

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

$\square$

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

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

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

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

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

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

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

$\square$

Appendix 4(d)

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

$$H = \begin{bmatrix}
a_{1,1} & 0 & ... & 0 & a_1 q_1 & 0 & ... & 0 & 0 \\
0 & a_{2,2} & ... & 0 & 0 & a_2 q_2 & ... & 0 & 0 \\
. & . & ... & . & . & . & ... & . & 0 \\
. & . & ... & . & . & . & ... & . & 0 \\
0 & 0 & ... & a_{n,n} & 0 & 0 & ... & a_n q_n & 0 \\
q_1 a_1 & 0 & ... & 0 & q_{1,1} & 0 & ... & 0 & q_{1,N+1} \\
0 & q_2 a_2 & ... & 0 & 0 & q_{2,2} & ... & 0 & q_{2,N+1} \\
. & . & ... & . & . & . & ... & . & \\
. & . & ... & . & . & . & ... & . & \\
0 & 0 & ... & q_n a_n & 0 & 0 & ... & q_{n,n} & q_{n,N+1} \\
0 & 0 & ... & 0 & q_{N+1,1} & q_{N+1,2} & ... & q_{N+1,n} & q_{N+1,N+1}
\end{bmatrix}$$

where

$$\begin{gathered}
a_{i,i} = \frac{\partial^2 \tau}{\partial {A_{r_i}}^2}, \qquad q_{i,i} = \tfrac{\partial^2 \tau}{\partial {Q_{r_i}}^2}, \qquad q_{N+1,N+1} = \frac{\partial^2 \tau}{\partial {Q_n}^2} \\
a_i q_i = \frac{\partial^2 \tau}{\partial A_{r_i} \partial Q_{r_i}}, \qquad q_i a_i = \tfrac{\partial^2 \tau}{\partial Q_{r_i} \partial A_{r_i}} \qquad q_{N+1,i} = \frac{\partial^2 \tau}{\partial Q_n \partial Q_{r_i}} \\
q_{i,N+1} = \frac{\partial^2 \tau}{\partial Q_{r_i} \partial Q_n}, \quad a_i q_{N+1} = \tfrac{\partial^2 \tau}{\partial A_{r_i} \partial Q_n}, \quad q_{N+1} a_i = \frac{\partial^2 \tau}{\partial Q_n \partial A_{r_i}} \\
i = 1, 2..., N
\end{gathered}$$

First order conditions for finding the stationary point:

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

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

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

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

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

Proof: $\tau(\boldsymbol{A},\ \boldsymbol{Q},\ Q_n)$ *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.

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

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

$$
A_{r_i} = \left[\frac{F_r + h_rQ_{r_i} - (h_rZ_{\alpha_i}\mu_r/\sigma_r)}{C_lf_r\xi\lambda\delta_i}\right]^{2/3}
$$

(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.

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

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

$\square$

CHAPTER V

TYPE-II SERVICE LEVEL: FILL-RATE

Consider a specific replenishment order of Q units for a distribution center (DC) with a continuous review $(Q,\ r)$ policy . During the replenishment lead time, there are r units of inventory in stock. Fill-rate at the DC is then defined as the fraction of demand during the replenishment lead time that is filled from stock (see Ganeshan [19], Deuermeyer *et al.* [15]).

Fill rate is considered a more relevant measure of service as it enables the DC to estimate what fraction of demand was converted to sales or equivalently what was the expected number of units backordered or lost during the replenishment interval. Even though fill-rate is recognized as the true service measure, it is the cycle service level or *Type-I service* that appears widely in the inventory literature. The reason for this is relatively simpler expressions to model the type-1 service levels while determining the inventory policy. There are a selected few papers in the area of multi-echelon inventory that model service levels in terms of fill-rate (Deuermeyer *et al.* [15], Schwarz *et al.* [7], and Ganeshan [19]).

There are two kinds of inventory considered in this analysis, namely, the average cycle stock $(Q/2)$ and the average safety stock. Average cycle stock is the amount of inventory in between the replenishment cycles. The more the cycle stock in each cycle, the fewer the reorders. The amount of cycle stock at a DC depends on the trade-off between the inventory holding cost at that DC and the fixed reorder cost for the DC every time it places a batch order. The reorder cost is a function of the *mean demand* and hence the cycle stock depends on the mean of the demand process. Average *safety stock* on the other hand is the inventory in stock to handle demand uncertainty during the replenishment cycle. The amount of safety stock depends on the service level and and the demand process during the lead time.

In this chapter, a continuous approximation approach is presented to model the integrated facility location and inventory allocation problem using a fill-rate service (*Type-II service*) approach. The key difference beteen this problem and the problem presented in chapter IV lies in the representation of the safety stock term. A detailed discussion on the problem formulation is presented next. Later a hierarchical approach (see Houtum [37]) is discussed to solve the problem. The need for such an approach arises mainly due to the presence of the reorder point terms in the objective function and the inability to express them as a closed form expression. The key feature of the hierarchical approach is to make decisions for the two types of inventories in two levels. Cycle stock decisions are made at the first level, along with the facility location decisions, to determine the optimal reorder batch size, while the second level decisions focus on the safety stock levels (and hence the reorder points).

5.1 Fill-rate model for inventory

The objective function is modeled in terms of the fixed facility location cost, inbound and outbound transportation cost, and inventory holding cost. The goal is to minimize the objective function and ensure that the fill-rate service level constraint is met at each RDC and NDC.

The expressions for the total facility cost and the total expected inbound, and outbound transportation cost do not depend on the service measure. Thus, these expressions take the same form as in chapter IV. The expected inventory cost terms for the RDCs and the NDCs differ from those used in the previous chapter. This is because the expected inventory is a function of the cycle stock and the safety stock, and the safety stock term now depends on the *Type-II* service level measure.

The average inventory cost function in the objective function is expressed in terms of the average holding cost and the average reorder cost. While the average holding cost is a function of the cycle stock (reorder quantities Q and Q_n) and the safety stock, the average reorder 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 $Z_{\alpha_n}\sqrt{Var[D_{n,LT}]}$. These expressions had a closed form that could be modeled into the objective function of the problem. Unlike the *Type-I service* analysis, it is hard to express the safety stock term as a closed form expression in the case of fill-rate (*Type-II service*) analysis. Thus, the safety stock terms are written as $r_i - E[D_{r_i,LT}]$ and $r_{r_n} - E[D_{n,LT}]$. In addition, there are two new constraints which link the expected number of backorders to the reorder point using the fill-rate. The derivation of these constraints is discussed next.

5.1.1 Fill-rate contraints

The order replenishment lead time for each RDC is a function of the travel lead time and the wait time in the event of a stock-out at the NDC. The travel lead time has a normal distribution with mean μ_r and variance ${\sigma_r}^2$. 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 *et al.* [15], Ganeshan [19]). In this analysis, we model the wait time by it expected value and use it along with the distribution for the travel lead time to get the distribution for the total order replenishment lead time.

Result 5.1: 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 [24], see Appendix 3(a)).

$$W = \frac{(1-\rho_n)Q_n}{Q}\frac{1}{(\lambda\delta R/Q)} \tag{31}$$

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:

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

The demand distribution at each RDC in cluster C_{n_i} is denoted by D_{r_i} and is a Poisson process with rate $\lambda \delta_i A_{r_i}$.

$$E[D_{r_i}] = \lambda \delta_i A_{r_i}$$
$$Var[D_{r_i}] = \lambda \delta_i A_{r_i}$$

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 (see Appendix 3(b))

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

The demand process at the NDC during its replenishment lead time can be approximated by a normal distribution (see Deuermeyer *et al.* [15]). The expected demand $(E[D_{n,LT}])$ and variance of demand $(Var[D_{n,LT}])$ at the NDC during its order replenishment period is given by[1]:

$$E[D_{n,LT}] = \frac{\sum_{i=1}^{N} \mu_n \lambda \delta_i C_{n_i}}{Q} \tag{32}$$

$$Var[D_{n,LT}] = \frac{\sum_{i=1}^{N} \mu_n \lambda \delta_i C_{n_i}}{Q^2} \tag{33}$$

Result 5.2 (Hopp *et al.* [22]): 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

[1]Note that these are in units of RDC reorder quantity Q

density function and cumulative density function. If R is the reorder point for a (Q, r) policy, then the expected number of backorders at the location is given by

$$\begin{aligned} E[B(R)] &= \int_R^{\infty} (x-R)f(x)dx && (34) \\ &= (\theta - R)[1-\Phi(z)] + \sigma\phi(z) && (35) \end{aligned}$$

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 5.1* 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 5.1* to get a closed form 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:

$$\begin{aligned} E[B_{r_i}] &= (E[D_{r_i,LT}] - r_i)[1-\Phi(z_i)] + \sigma\phi(z_i) \quad \forall i \\ E[B_n] &= (E[D_{n,LT}] - r_n)[1-\Phi(z_n)] + \sigma\phi(z_n) && (36) \end{aligned}$$

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

In this work, it is assumed that there is a fixed value for the minimum fill-rate that is determined by the management. Let ρ_{r_i} (ρ_n) be the minimum fill rate for each RDC i (NDC). Then, for the special case when $Q_i = Q \quad \forall i$ the expected number of backorders can be estimated as (see Ganeshan [19]):

$$\begin{aligned} E[B_{r_i}] &= (1-\rho_{r_i})Q \quad \forall i \\ E[B_n] &= \frac{(1-\rho_n)Q_n}{Q} && (37) \end{aligned}$$

Using equations (36) and (37), the reorder point at each RDC i (NDC) can be estimated in terms of fill-rate at the RDC (NDC).

$$\begin{aligned} (E[D_{r_i,LT}] - r_i)[1-\Phi(z_i)] + \sigma\phi(z_i) &= (1-\rho_{r_i})Q \quad \forall i \\ (E[D_{n,LT}] - r_n)[1-\Phi(z_n)] + \sigma\phi(z_n) &= \frac{(1-\rho_n)Q_n}{Q} && (38) \end{aligned}$$

We can now define the optimizaton model for the integrated facility location and inventory allocation problem when a *Type-II* service measure is used.

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

subject to

$$
\begin{aligned}
Q_{r_i} \;&\geq\; 0 \;\; \forall \;\; r_i \qquad (39) \\
A_{r_i} \;&\geq\; 0 \;\; \forall \;\; r_i \\
k \;&\geq\; 2 \\
(E[D_{r_i,LT}] - r_i)[1 - \Phi(z_i)] + \sigma\phi(z_i) \;&=\; (1 - \rho_{r_i})Q \quad \forall i \\
(E[D_{n,LT}] - r_n)[1 - \Phi(z_n)] + \sigma\phi(z_n) \;&=\; (1 - \rho_n)k \\
Q_{r_i}, \frac{C_{n_i}}{A_{r_i}}, k \;&\in\; Z^+ \;\; \forall \;\; r_i
\end{aligned}
$$

where constraints (1), (2) and (3) are the non-negativity constraints. The fill-rate constraints are given by (4) and (5) and they give the optimal value of the reorder points. Constraint (6) is the integrality constraint.

5.1.2 Properties of the objective function

Result 5.3: $\gamma(\mathbf{A}, Q, k)$ is a convex function for values of ($\mathbf{A}$, Q, k) satisfying the following inequalities given by (40) (see appendix 5 for proof).

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

$$\left(\frac{\partial^2\phi}{\partial Q^2} - \sum_{i=1}^{N} \frac{\partial^2\phi/\partial Q \partial A_{r_i}}{\partial^2\phi/\partial {A_{r_i}}^2} \frac{\partial^2\phi}{\partial A_{r_i}\partial Q}\right) \frac{\partial^2\phi}{\partial k^2} - \left(\frac{\partial^2\phi}{\partial Q \partial k} \frac{\partial^2\phi}{\partial k \partial Q}\right) > 0 \tag{40}$$

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

$$A_{r_i} = \left(\frac{2F_r + h_r(Q + 2r_i)}{C_l f_r \xi \lambda \delta_i}\right)^{2/3} \tag{41}$$

$$k = \frac{1}{Q}\sqrt{(2R_n)\left(\frac{\sum_{i=1}^{N} \xi\lambda\delta_i C_{n_i}}{(h_n - 2h_r(1-\rho_n))}\right)} \tag{42}$$

$$Q = \sqrt{2\left(\frac{\left(C_f + R_r + \frac{R_n}{k}\right)\sum_{i=1}^{N} \xi\lambda\delta_i C_{n_i}}{\sum_{i=1}^{N} \frac{h_r C_{n_i}}{A_{r_i}} + \frac{h_n k}{2} - h_r(1-\rho_n)k}\right)} \tag{43}$$

Result 5.4: 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 (40). Then by theorem 4.3 the result follows. □

Result 5.5: $\gamma(\mathbf{A}, Q, k)$ is a *biconvex* function for all values of Q and k, and values of $\mathbf{A}$ satisfying equation (15).

5.2 Solution Approach

Note that the stationary point for the objective function $\gamma(\mathbf{A}, Q, k)$ depends on the reorder point (r_i) at each RDC. Further in order to calculate the values of r_is and r_n, the values of $(\mathbf{A}, Q, k)$ is needed (see constraints (4)and (5) of the fill-rate problem P^f). In order to tackle this circular behavior between the decision variables a hierarchical approach is used.

In the first phase, a value is fixed for r_is 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 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_is.

In the second phase of the problem, the optimal values of r_is and r_n are calculated, using equations (4) and (5) from the equation set (39), for the values of $(\mathbf{A}, Q, k)$ derived in the first phase. Using these new values of r_is, 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 the analysis of the *Type-I service* model-equal Q (chapter IV). 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 reorder points are calculated using the *Goolseek* tool in excel.

5.3 Discussion

In this chapter, 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. The solution approach used for solving the *Type-II service* model is a hierarchical approach with an iterative solution procedure in the first phase and reorder point updates in the second phase. The numerical study comparing the results for the two service models is presented in chapter (VI).

5.4 Appendix

(1) First order conditions for deriving the stationary point for function $\gamma(\boldsymbol{A}, Q, k)$.

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

□

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

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

where

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

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

Using theorem 4.2, $\gamma(\mathbf{A}, Q, k)$ 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

.

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

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

$\square$

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

$$\begin{aligned}
\frac{\partial\gamma}{\partial A_{r_i}} &= \frac{C_l f_r \xi\lambda\delta_i C_{n_i}}{2{A_{r_i}}^{1/2}} - \frac{2C_{n_i}F_r + C_{n_i}h_r(Q+2r_i)}{2{A_{r_i}}^2} \\
\frac{\partial\gamma^2}{\partial {A_{r_i}}^2} &= -\frac{C_l f_r \xi\lambda\delta_i C_{n_i}}{4{A_{r_i}}^{3/2}} + \frac{2C_{n_i}F_r + C_{n_i}h_r(Q+2r_i)}{2{A_{r_i}}^3} \\
\frac{\partial\gamma^2}{\partial Q\partial A_{r_i}} &= -\frac{C_{n_i}h_r}{2{A_{r_i}}^2} \\
\frac{\partial\gamma^2}{\partial k\partial A_{r_i}} &= 0
\end{aligned}$$

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

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

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

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

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

□

CHAPTER VI

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 the study in this chapter all the analysis is carried out using data for the southeastern (SE) region with the NDC located at Savannah, GA.

The two-phase approximation approach discussed in chapter 3 is used in the analysis. In *Phase-I*, the SE region is partitioned into clusters over which the demand function is slow varying. The SE region is made up of eight states. Information on the counties within each state and location of stores within each county is given. Using this the store density function for each county is calculated. Interestingly, the county level store density does not vary significantly for any given state. The average demand is the same for each store. Thus, each state is a homogeneous cluster on which the demand data is slow varying.

Next, as part of *Phase-II*, the problem is using the CA approach and a solution is obtained. Cost data used in this analysis in given in Appendix 6.4.

6.1 Effect of store density and cost parameters on the network design

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

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

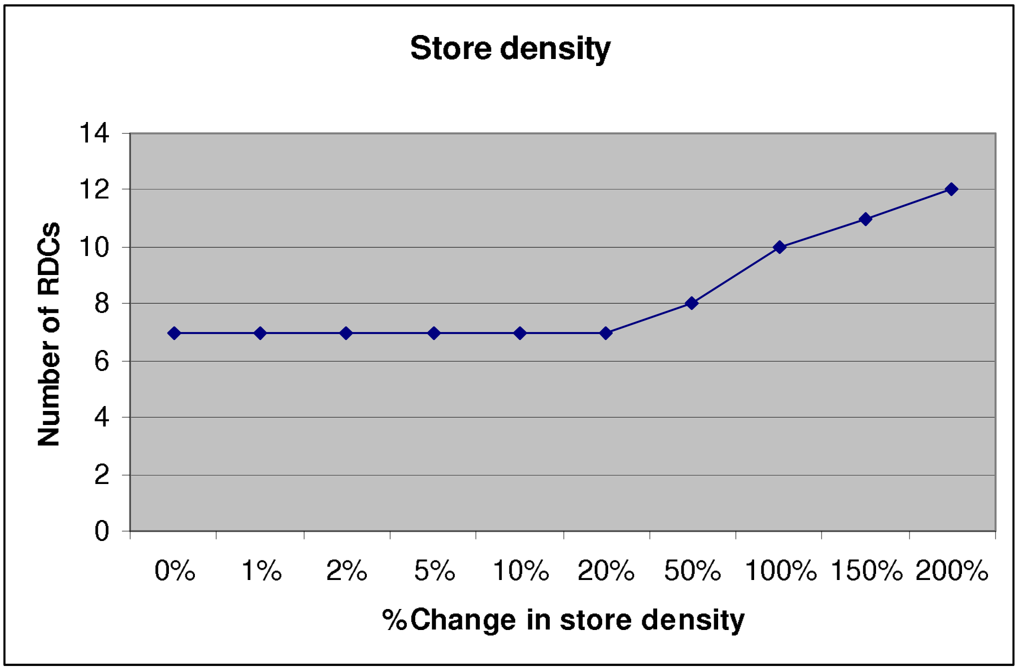

Figure 9: Store density Analysis

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.

From Figure 9, it is observed that the number of RDCs increase with an increase in the store density. This happens because an increase in the store density by a given factor increases the total network demand by the same factor. This can be easily seen from the expression for total demand given by $\lambda\delta_i C_{n_i}$ where λ is the demand at each store, δ_i is the store density for cluster with area C_{n_i}. An increase in the number of RDCs would mean a lot of money in terms of operational, transportation and inventory cost. Thus, it is important to be careful in estimating the store density to avoid any over or under estimation of the demand. A similar analysis is carried out to

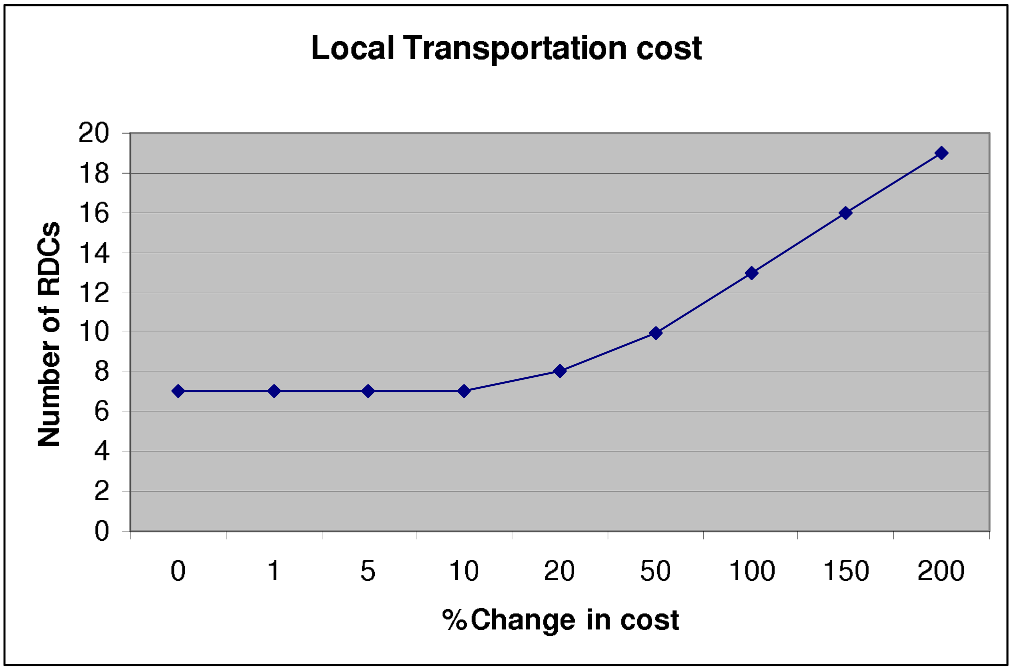

Figure 10: Local Delivery cost Analysis

study how the local delivery cost parameter effects the problem. Figure 10 shows that for a small change in the local delivery cost parameter c_l, the network design is not affected but large chnages in the cost can lead to a significant change in the network design in terms of number of RDCs (for fixed values of inventory parameters).

6.2 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 11. 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 the NDC) for the RDCs in each zone show an interesting trend. Although total network cost for the

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

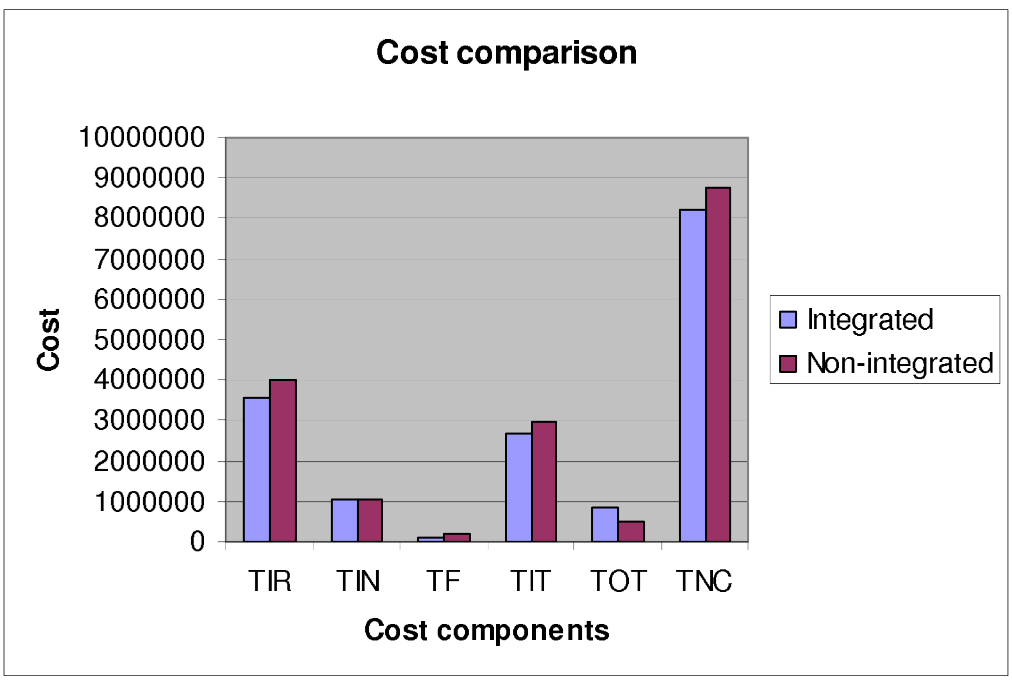

Figure 11: Total Network Cost

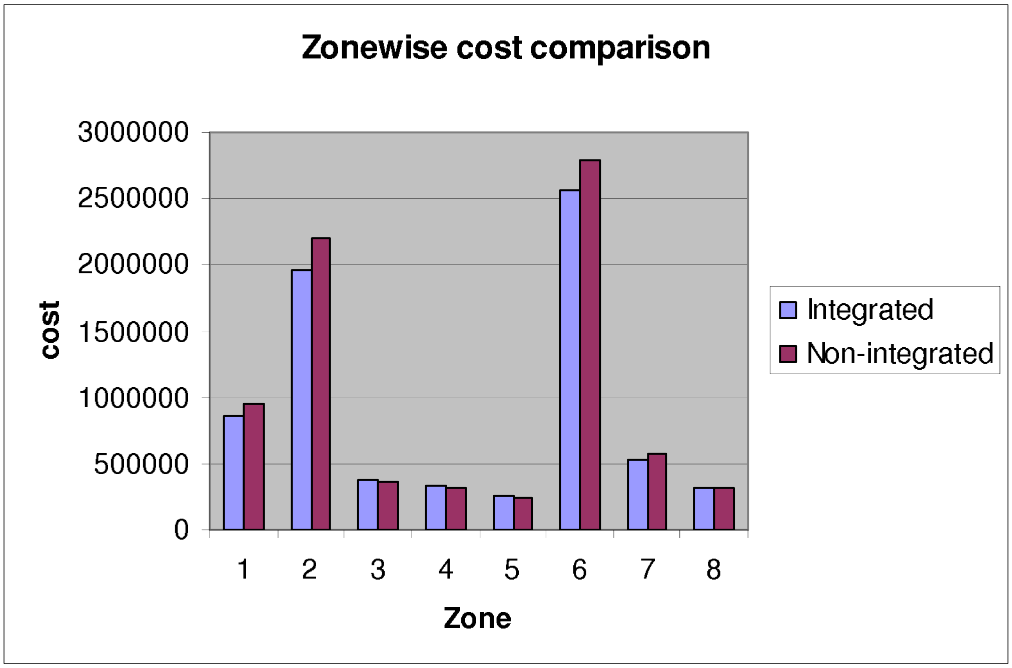

Figure 12: Total cost for each zone.

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 12). 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 13). 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.* [10]). The reason for this is while

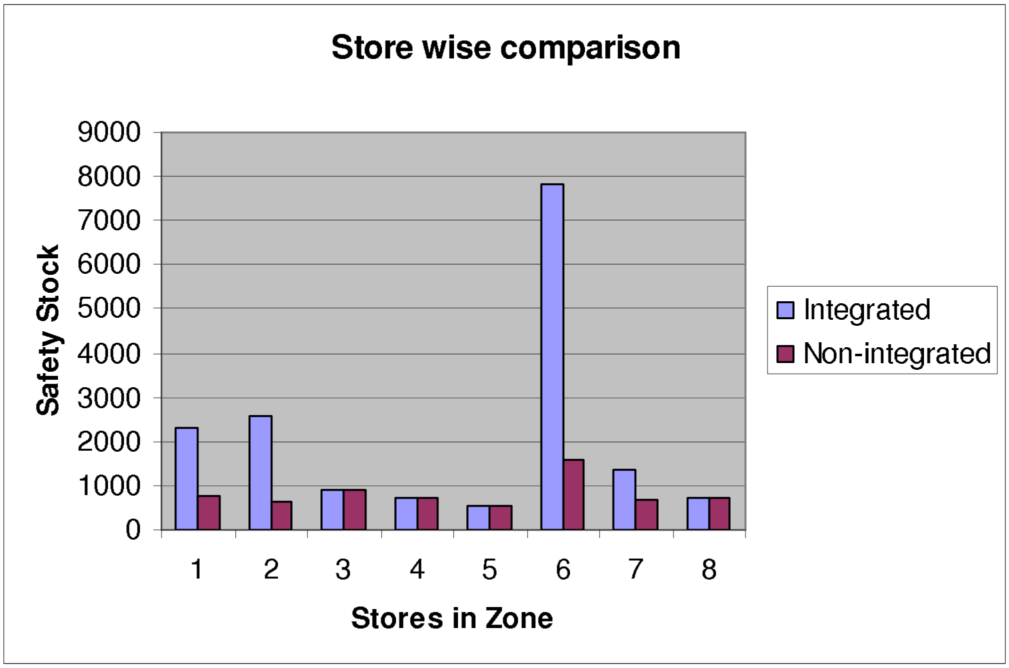

Figure 13: Safety stock for each store zone-integrated vs non-integrated

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 14.

Table 3 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. Clearly, using different service measures affect 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

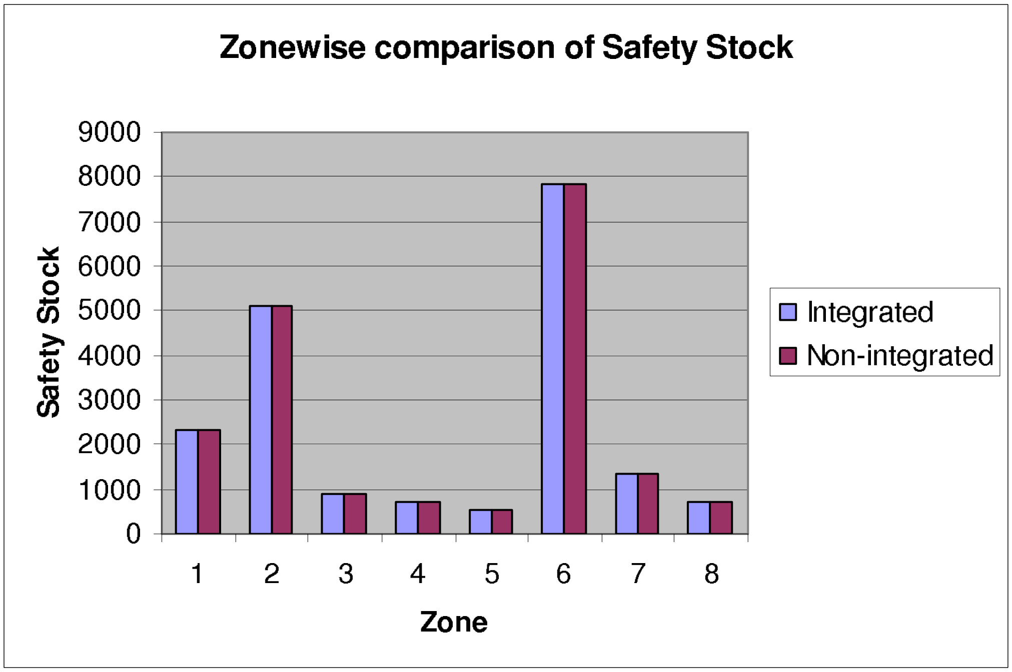

Figure 14: Safety stock for each zone-integrated vs non-integrated

each RDC and the NDC are given in tables 4 and 5.

The results (see figure 15) show that the TNC has a lower value for the *Type-II* service model. This happens because the average safety stock is significantly lower in this case. Figures 16 and 17 compare the safety stock and the total cost for each zone for the two types of service measures. Note that the order quantity Q_n and reorder point r_n for the NDC are in units of Q

Table 3: 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 4: 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

Table 5: Inventory Parameters for 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

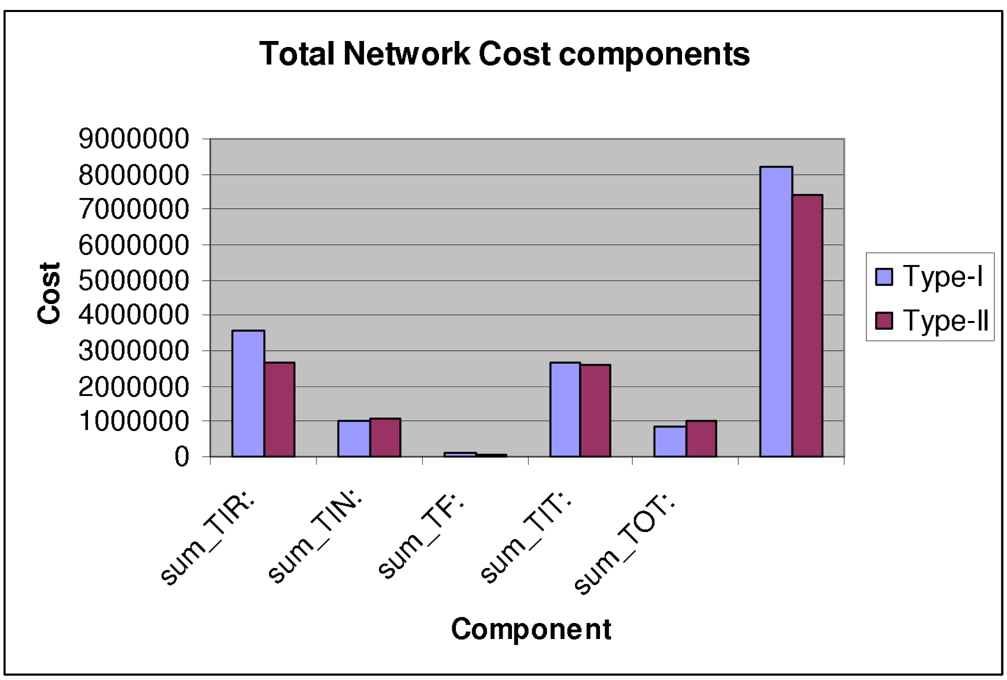

Figure 15: Total Network Cost

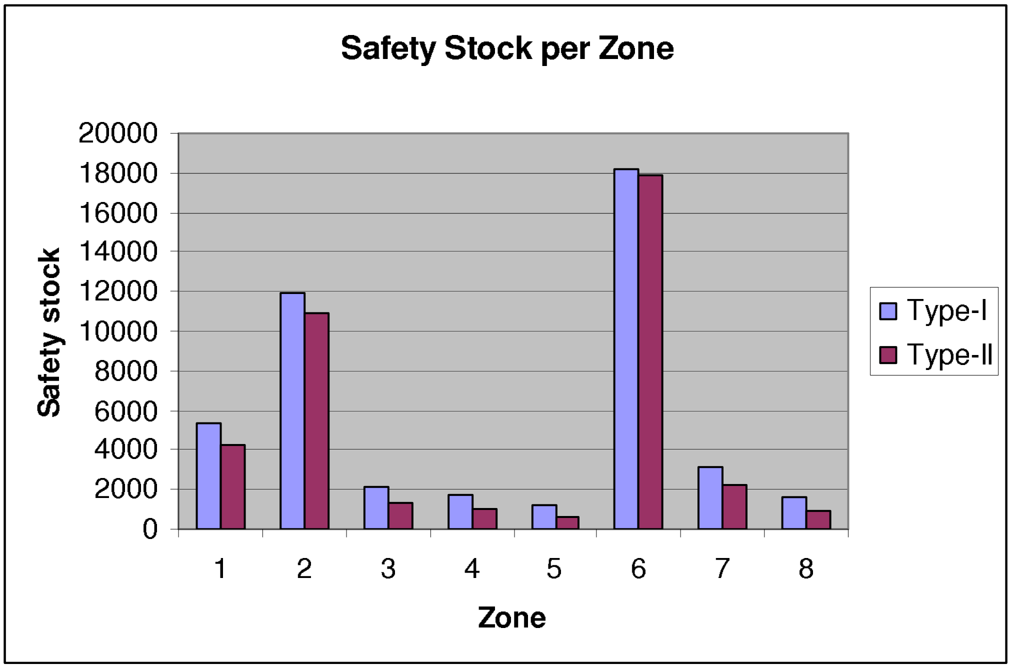

Figure 16: Safety Stock Analysis

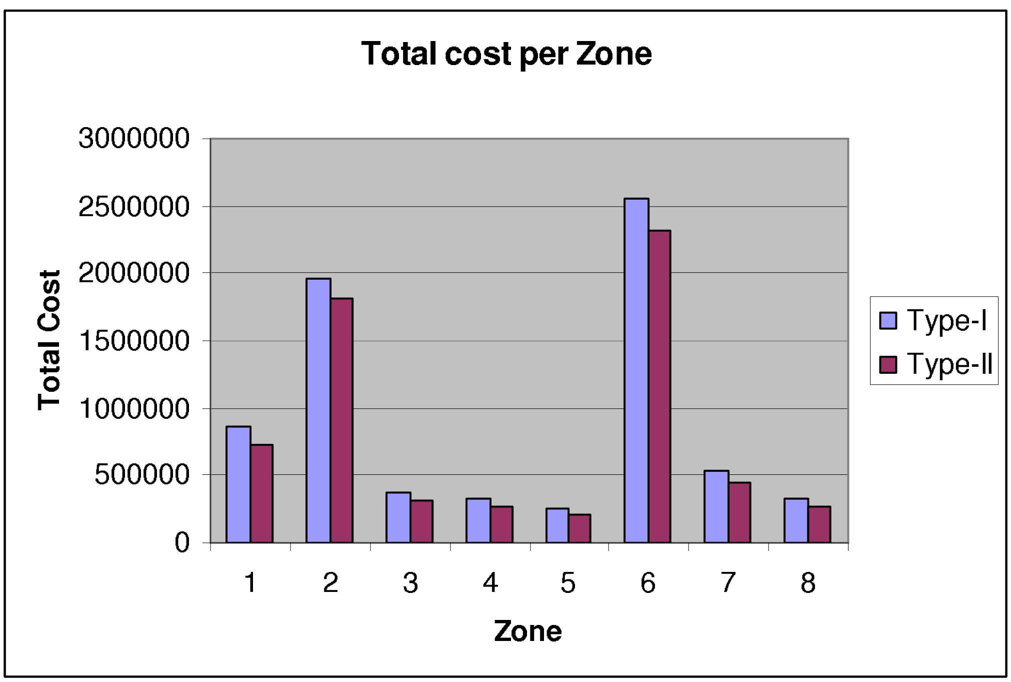

Figure 17: Total Cost per Zone Analysis

6.3 Conclusions

We propose a novel two-phase approximation method for solving the multi-echelon inventory allocation and network design problem. The method integrates facility costs, inventory costs, transportation costs, service levels and interrelated sources of variability. Using this method, problems with huge amounts of non-homogeneous demand data can be solved. Our model provides a powerful analysis tool in terms of studying potential changes in supply chain system due to the changes in the parameters. An example is discussed that shows that the network configuration and hence the key decision variables are affected by the level of details used in the analysis. In particular it is shown that good estimators of store density and local delivery cost are key to this analysis as they have a significant impact on the network costs.

Our results highlight the importance of integrating the facility location decision with the inventory policy decision. It is shown that the non-integrated problem generates results that have a significantly higher total network cost as compared to the integrated problem. We solve the integrated problem for two types of service measures- *Type-I* or stock-out probability and *Type-II* or fill-rate. The analysis shows that each problem produces a very different result.

It would be interesting to incorporate delivery lead time contraints in the model. In real life operations, most orders have a delivery time window, i.e., a lower and upper bound on the time it can take the order to arrive at a facility. It is an important service measure and can impact the network design.

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 in isolation without considering 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 continous review policy at each RDC would affect the network design and costs.

6.4 Appendix

Table 6: 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)

Table 7: Zonewise Cost Comparision for integrated problem with equal Q

Zone	TI_r	TIT	TOT	TF	TNC
1	414659	320654	112749	10000	858062
2	901572	709549	322781	20000	1953900
3	207089	124618	34037.8	10000	375745
4	180632	99631	33539.7	10000	323803
5	153003	73537.6	13190	10000	249731
6	1222830	1083920	248008	10000	2564750
7	272083	186001	59753.2	10000	527837
8	179733	98781.6	33971.8	10000	322486

Table 8: Zonewise Cost Comparision for non-integrated problem with equal Q

Zone	TI_r	TIT	TOT	TF	TNC
1	496220	355492	65095.5	30000	946808
2	1165210	786639	161390	80000	2193240
3	184665	138157	34037.8	10000	366860
4	157529	110456	33539.7	10000	311524
5	129192	81527.2	13190	10000	233909
6	1423780	1201680	110913	50000	2786370
7	300656	206209	42251.9	20000	569117
8	156607	109514	33971.8	10000	310092

CHAPTER VII

DISCRETE MODEL FOR THE INTEGRATED FACILITY LOCATION AND INVENTORY ALLOCATION PROBLEM

7.1 Introduction

It is only recently that researchers have started developing models that integrate the two classic supply chain problems of facility location and inventory allocation. Most of these works focus on the single echelon inventory allocation and facility location problem (see Miranda [26], Nozick *et al.* [30] and Erlebacher *et al.* [16]). A single echelon inventory allocation problem involves inventory decisions for the distribution centers in a network with manufacturing plants, distribution centers and retail stores. Teo *et al.* ([36], [31]) are the only studies that model a multi-echelon (two level) inventory allocation and facility location problem.

All the studies in the area of integrated network design and inventory decisions develop a discrete optimization model and propose a solution. While Erlebacher *et al.* [16], Nozick *et al.* [30] reduce the problem to a special form of facility location problem for which heurstics are known, Miranda *et al.* [26], Teo *et al.* ([36], [31]) use a column generation technique to solve the problem. Teo *et al.* [36] study an integrated logistic network problem in which they consider inventory cost for multiple echelons of inventory stocking locations. In their model, the inventory cost is modeled at each DC and retailer. They use the convex inventory minimization function proposed by Roundy [32] along with the transportation and facility costs to formulate a MIP problem. A column generation technique is used to solve their model in O(nlog(n)) time and the solution is within 2% of the optimal solution when the problem instance is small (20 warehouses and 100 retailers). This model does not include either the demand or the supply uncertainty. In another study, Teo *et al.* [31] extend the previous model by adding safety stock terms to account for demand variability.

In this section we look at a discrete optimization model of the integrated facility location and inventory allocation problem. It is a nonlinear problem and a *linearization* technique is proposed to solve it. Under the linearization technique the nonlinear terms are replaced with the linear terms, hence reducing the problem to a new form. In addition it is discussed that under certain properties of the the objective function and decision variables it is possible to have an exact reformulation. Exact reformulation means that the optimal solution for the linearized version of the problem is the same as the optimal solution for the original problem. The quality of the solution is analyzed by defining the optimality ratio. The chapter concludes with possible directions for future work.

7.2 Nonlinear Mixed Integer Model

The assumptions needed for modeling the discrete problem are similar to those listed in chapter III for the continuous model.

- The distribution network under study is an *arborescence network*. Further, the entire distribution network is partitioned into subregions such that each subregion is served by a single NDC.
- The location of the NDC is known and fixed.
- Demand per unit time for each store i is an independent and identically distributed Poisson process with rate λ_i.
- Each product can be analyzed independent of other products. The demand for a single product is considered in our study.

- 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.

- 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.

- The stores maintain minimal amount of inventory and it is ignored in this analysis. We do not consider the pipeline inventory cost for units in transit from NDC to RDC or from suppliers to the NDC.

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

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

- The inventory policy at the RDCs and the NDCs is a continuous review (Q, r) policy.

- Each RDC and NDC makes decentralized decision on the inventory control policy.

Parameters

$$
\begin{array}{rcl}
N & = & \text{Total number of store locations in the NDC subregion} \\
F_i & = & \text{Facility location cost at location } i \\
C_f & = & \text{Fixed cost component of the transportation cost between the NDCs and RDCs} \\
C_v & = & \text{Variable cost component of the transportation cost between the NDCs and RDCs} \\
 & & (\$/item) \\
C_l & = & \text{Delivery cost from the RDC to the store(\$/item/mile)} \\
h_i & = & \text{Inventory holding cost per item over the planning horizon at RDC } i \\
R_i & = & \text{Reorder cost for RDC } i \\
h_n & = & \text{Holding cost per item over the planning horizon at the NDC} \\
R_n & = & \text{Reorder cost for the NDC} \\
d_i & = & \text{Distance between the NDC and the RDC located at } i \\
Z_{\alpha_i} & = & \text{Critical value when in-stock probability for RDC } i \text{ is } \alpha_i \\
 & & \text{calculated using the standard normal table} \\
Z_n & = & \text{Critical value when in-stock probability for NDC is } \alpha_n \\
 & & \text{calculated using the standard normal table} \\
\mu_i & = & \text{Mean travel time between the NDC and RDC } i \\
{\sigma_i}^2 & = & \text{Variance of the travel time between the NDC and RDC } i \\
\mu_n & = & \text{Mean travel time between the manufacturer and the NDC} \\
\xi & = & \text{Length of the planning horizon (365 days)} \\
\lambda_i & = & \text{Demand rate for the retail store at location } i \\
i, j & \in & 1, 2...., N
\end{array}
$$

Decision Variables

$$X_i = \begin{cases} 1 & \text{if a RDC is opened at location } i \\ 0 & \text{otherwise} \end{cases}$$

$$Y_{ij} = \begin{cases} 1 & \text{if the store at location } j \text{ is served} \\ & \text{by the RDC at location } i \\ 0 & \text{otherwise} \end{cases}$$

Q = Reorder quantity at each RDC

k = Reorder quantity multiplicative factor for the NDC[1]

Original Optimization Problem - Equal Reorder quantity

P(1): Minimize

$$\begin{aligned} TNC \;=\; & \sum_{i=1}^{N} F_i X_i + \sum_{i=1}^{N} (C_f + C_v Q) \frac{\xi \sum_{j=1}^{N} Y_{ij}\lambda_j}{Q} + C_l \xi \sum_{i=1}^{N}\sum_{j=1}^{N} Y_{ij}\lambda_j d_{ij} \\ + \; & \sum_{i=1}^{N} h_i X_i \left(\frac{Q}{2} + \sum_{i=1}^{N} Z_{\alpha_i} \sqrt{\mu_i \left(\sum_{j=1}^{N} Y_{ij}\lambda_j\right) + {\sigma_i}^2 \left(\sum_{j=1}^{N} Y_{ij}\lambda_j\right)^2} \right) \\ + \; & \sum_{i=1}^{N} \left(R_i \frac{\xi \sum_{j=1}^{N} Y_{ij}\lambda_j}{Q} \right) + h_n \left(\frac{kQ}{2} + Z_{\alpha_n} \sqrt{\sum_{i=1}^{N}\sum_{j=1}^{N} Y_{ij}\lambda_j L_n} \right) \\ + \; & R_n \left(\frac{\sum_{i=1}^{N} \xi \sum_{j=1}^{N} Y_{ij}\lambda_j}{kQ} \right) \end{aligned}$$

subject to

$$Y_{ij} \le X_i \quad \forall i, j \quad (1)$$

$$\sum_{i=1}^{N} Y_{ij} = 1 \quad \forall i \quad (2)$$

$$1 \le Q \le \overline{Q} \quad (3)$$

$$1 \le k \le \overline{k} \quad (4)$$

$$X_i,\ Y_{ij} \in \{0, 1\} \quad (5)$$

$$k,\ Q \in Z^{+} \quad (6)$$

$$\overline{k}, \overline{Q}, F_i, h_i, h_n, C_f, C_v, C_l, d_i, d_{ij},$$

$$Z_{\alpha_i}, Z_{\alpha_n}, R_i, R_n, \xi, L_n, \mu_i, \sigma_i \text{ are known}$$

The first three terms in the objective function are the total facility location cost, the total inbound transportation cost from the NDC to the RDCs and the total delivery cost from the RDCs to the retail stores for all the items. The fourth and the fifth terms are the total average inventory cost and the average reorder cost for the RDCs. Similarly, the last two terms give the total average inventory cost for the NDC and the total average reorder cost for the NDC.

Constraint (1) ensures that a store would be served from RDC i only if there is a RDC located at i. Constraint (2) guarantees that each retailer is assigned to exactly one RDC. Constraint (3) and (4) give the upper and lower bounds for the NDC order quantity multiplier k and for the order quantity Q. Constraint (5) sets the decision variables Y_{ij} and X_i to be binary integers whereas constraint (6) sets all other decision variables k and Q to be positive integers.

Remark:The objective function is nonlinear with the following bilinear terms, QX_i, kQ, Y_{ij}/Q and $Y_{ij}/(kQ)$, and because of the presence of the square root terms in the numerator.

7.3 Linearization of the nonlinear objective function

The nonlinear terms in the objective function can be classified under two categories- 1) Square root, and 2) Bilinear terms. Unfortunately there is no single method that can handle these two forms of nonlinearity. So a stepwise approach is adopted where the square root nonlinearity is handled first followed by an approach to handle the bilinear nonlinearity. A detailed analysis for this stepwise approach is discussed in the following sections.

7.3.1 Square root nonlinearity

Result 7.1: If a and b are positive numbers and $x > 0$, then

$$\sqrt{ax + bx^2} \geq \sqrt{b}x - \frac{a}{2\sqrt{b}} \tag{44}$$

Proof: See Appendix 7(a)

Safety stock at RDC i

$$\sqrt{\mu_i \left(\sum_{i=1}^{N} Y_{ij}\lambda_j\right) \; + \; {\sigma_i}^2 \left(\sum_{i=1}^{N} Y_{ij}\lambda_j\right)^2} \;\; > \;\; \sigma_i \left(\sum_{i=1}^{N} Y_{ij}\lambda_j\right) - \frac{\mu_i}{2\sigma_i} \quad \text{using result 7.1}$$

where a = μ_i, b = ${\sigma_i}^2$ and x = $\sum_{i=1}^{N} Y_{ij}\lambda_j$

A lower bound $P_l(2)$ can be defined on the objective function $P(2)$ by replacing the square root terms with their lower bounds. The square root term corresponding to the safety stock at each RDC i is replaced by a new term using Result 7.1. The square root term for the NDC is a constant. This follows from the fact that each retailer j has to be assigned to exactly one RDC j. Thus, this term is replaced by a constant $\sqrt{\sum_{i=1}^{N} \lambda_j L_N}$.

Safety stock at the NDC

$$\sqrt{\left(\sum_{i=1}^{N}\sum_{j=1}^{N} Y_{ij}\lambda_j L_N\right)} \;\; = \;\; \sqrt{\left(\sum_{j=1}^{N}\sum_{i=1}^{N} Y_{ij}\lambda_j L_N\right)}$$

because λ_j , $L_N > 0$, sum by rows = sum by columns

$$\sqrt{\left(\sum_{i=1}^{N}\sum_{j=1}^{N} Y_{ij}\lambda_j L_N\right)} \;\; = \;\; \sqrt{\left(\sum_{j=1}^{N} \lambda_j L_N\right)}$$

because for each j $Y_{ij} = 1$ for some i

Thus, a lower bound to the original objective function (TNC) is obtained and is denoted by $\widehat{TNC}$.

$$\begin{aligned}
\widehat{TNC} &= \sum_{i=1}^{N} F_i X_i + \sum_{i=1}^{N} (C_f + C_v Q) \frac{\xi \sum_{j=1}^{N} Y_{ij}\lambda_j}{Q} + C_l \xi \sum_{i=1}^{N}\sum_{j=1}^{N} Y_{ij}\lambda_j d_{ij} \\
&+ \sum_{i=1}^{N} h_i \left(\frac{X_i Q}{2} + Z_{\alpha_i} \left(\sigma_i \sum_{j=1}^{N} Y_{ij}\lambda_j - \frac{\mu_i}{2\sigma_i} \right) \right) + \sum_{i=1}^{N} \left(R_i \frac{\xi \sum_{j=1}^{N} Y_{ij}\lambda_j}{Q} \right) \\
&+ h_n \left(\frac{kQ}{2} + Z_{\alpha_n} \sqrt{\sum_{j=1}^{N} \lambda_j L_n} \right) + \frac{R_n \left(\sum_{i=1}^{N} \xi \sum_{j=1}^{N} Y_{ij}\lambda_j \right)}{kQ}
\end{aligned}$$

7.3.2 Bilinear Nonlinearity

The original problem $P(1)$ has bilinear terms of the form QX_i, kQ, Y_{ij}/Q and $Y_{ij}/(kQ)$. In order to linearize these terms a reformulation technique, proposed by Al-Khayyal [2] and Al-Khayyal *et al.* [3], for solving bilinear programming problems is used. The reformulation technique uses convex and concave envelope for a bilinear function over a rectangular region. Each bilinear term is replaced by a new term along with four constraints.

Result 7.3 (Linear Reformulation): Given a nonlinear problem with terms of the form XY where X and Y are such that

$$\begin{aligned}
a &< X < b \text{ where } a > 0 \text{ and } b > 0 \\
c &< Y < d \text{ where } c > 0 \text{ and } d > 0
\end{aligned}$$

It is possible to replace any term of the form XY by a new variable Z where Z satisfies the following constraints (see equations 45), and the new problem is called the linearized approximation to the original nonlinear problem.

$$\begin{aligned}
Z &\leq aY + dX - ad \\
Z &\leq bY + cX - bc \\
Z &\geq aY + cX - ac \\
Z &\geq bY + dX - bd \qquad (45)
\end{aligned}$$

Remark: Let (X^*, Y^*, Z^*) be the values obtained by solving the linearized approximation of the nonlinear problem. If $X^* \in \{a, b\}$ or $Y^* \in \{c, d\}$, then $Z^* = X^*Y^*$ and the linearized problem would be an exact reformulation otherwise Z^* is treated as an approximation for X^*Y^*.

The nonlinear terms in product form, X_iQ and kQ, are considered first. The bounds on the optimal order quantity Q for each RDC can be estimated by considering the total possible demand that it can serve. It is known that N is the maximum number of possible RDC locations, λ_j is the demand per unit time at each store j and ξ is the length of the planning horizon (in days). RDC i would face maximum demand when all the stores are assigned to it. In this case the upper bound ($\overline{Q}$) on the order quantity for a RDC located at location i is given by $\sum_{i=1}^{N} \lambda_j \xi$. The minimum possible value of Q is given by 1. Similarly the lower bound of k is set to 1. For the upper bound, N the number of RDC locations can be a good starting guess.

Each of the nonlinear terms in product form is replaced by a new variable in the objective function, namely, $g_i = X_iQ$ and $u = kQ$. In doing so we add new constraints derived using the *Linear Reformulation* result.

$$X_iQ \;=\; g_i \;\text{ gives}$$

$$\begin{aligned} g_i &\leq \overline{Q}X_i \\ g_i &\leq Q + X_i - 1 \\ g_i &\geq X_i \\ g_i &\geq Q + \overline{Q}X_i - \overline{Q} \end{aligned}$$

$$kQ \;=\; u \;\text{ gives}$$

$$\begin{aligned} u &\leq k + \overline{k}Q - \overline{k} \\ u &\leq Q + \overline{Q}k - \overline{Q} \\ u &\geq Q + k - 1 \\ u &\geq k\overline{Q} + \overline{k}Q - \overline{Qk} \end{aligned}$$

Note that in this case ${g^*}_i = {X^*}_i Q^*$ but u^* will be treated as an approximation of $k^* Q^*$.

7.3.3 Denominator based nonlinearity

After taking care of the square root and product based nonlinearity, the next step is to look at the nonlinear terms Y_{ij}/Q and Y_{ij}/u in the denominator form. Two new variables β_{ij} and δ_{ij} are defined, and *linear reformulation* result is used to get the new set of constraints given below:

$$
\begin{array}{llcl}
\frac{Y_{ij}}{Q} = \beta_{ij} \text{ gives} & & & \\
D1 & \beta_{ij} & \leq & Y_{ij} \\
D2 & \beta_{ij} & \leq & \frac{1}{Q} + \frac{Y_{ij}}{\overline{Q}} - \frac{1}{\overline{Q}} \\
D3 & \beta_{ij} & \geq & \frac{Y_{ij}}{\overline{Q}} \\
D4 & \beta_{ij} & \geq & \frac{1}{Q} + Y_{ij} - 1
\end{array}
$$

$$
\begin{array}{llcl}
\frac{\beta_{ij}}{k} = \delta_{ij} \text{ gives} & & & \\
D5 & \delta_{ij} & \leq & \frac{\beta_{ij}}{2} \\
D6 & \delta_{ij} & \leq & \frac{\beta_{ij}}{\overline{k}} + \frac{1}{k} - \frac{1}{\overline{k}} \\
D7 & \delta_{ij} & \geq & \frac{\beta_{ij}}{\overline{k}} \\
D8 & \delta_{ij} & \geq & \frac{\beta_{ij}}{2} + \frac{1}{k} - \frac{1}{2}
\end{array}
$$

7.4 Final Optimization Model (FOM)

FOM: Minimize

$$
\begin{aligned}
TNC^F &= \sum_{i=1}^{N} F_i X_i + \sum_{i=1}^{N} C_f \xi \sum_{j=1}^{N} \beta_{ij}\lambda_j + \sum_{i=1}^{N} C_v \xi \sum_{j=1}^{N} Y_{ij}\lambda_j + C_l \xi \sum_{i=1}^{N}\sum_{j=1}^{N} Y_{ij}\lambda_j d_{ij} \\
&+ \sum_{i=1}^{N} h_i \left(\frac{g_i}{2} + Z_{\alpha_i} \left(\sigma_i \sum_{j=1}^{N} Y_{ij}\lambda_j - \frac{\mu_i}{2\sigma_i} \right) \right) + h_n \left(\frac{u}{2} + \sqrt{\sum_{i=1}^{N} \lambda_j L_n} \right) \\
&+ \sum_{i=1}^{N} \left(R_i \xi \sum_{j=1}^{N} \beta_{ij}\lambda_j \right) + R_n \left(\sum_{i=1}^{N} \xi \sum_{j=1}^{N} \delta_{ij}\lambda_j \right)
\end{aligned}
$$

subject to

$$
\begin{aligned}
Y_{ij} &\leq X_i \;\; \forall\, i, j \\
\sum_{i=1}^{N} Y_{ij} &= 1 \;\; \forall\, j \\
g_i &\leq \overline{Q} X_i \\
g_i &\leq Q \,+\, X_i \,-\, 1 \\
g_i &\geq X_i \\
g_i &\geq Q \,+\, \overline{Q} X_i \,-\, \overline{Q} \\
u &\leq k \,+\, \overline{k} Q \,-\, \overline{k} \\
u &\leq 2Q \,+\, \overline{Q} k \,-\, 2\overline{Q} \\
u &\geq 2Q + k - 2 \\
u &\geq k\overline{Q} \,+\, \overline{k} Q \,-\, \overline{Qk} \\
\beta_{ij} &\leq Y_{ij} \\
\beta_{ij} &\leq w + \frac{Y_{ij}}{\overline{Q}} - \frac{1}{\overline{Q}} \\
\beta_{ij} &\geq \frac{Y_{ij}}{\overline{Q}} \\
\beta_{ij} &\geq w + Y_{ij} - 1 \\
\delta_{ij} &\leq \frac{\beta_{ij}}{2} \\
\delta_{ij} &\leq \frac{\beta_{ij}}{\overline{k}} + \frac{1}{k} - \frac{1}{\overline{k}} \\
\delta_{ij} &\geq \frac{\beta_{ij}}{\overline{k}} \\
\delta_{ij} &\geq \frac{\beta_{ij}}{2} + \frac{1}{k} - \frac{1}{2}
\end{aligned}
$$

$$
\begin{aligned}
1 \leq k \quad & \leq \overline{k} \\
1 \leq Q \quad & \leq \overline{Q} \;\; \forall i \\
X_i,\; Y_{ij} \quad & \in \{0,\; 1\}
\end{aligned}
$$

$$g_i,\ u,\ v\ and\ w \in \aleph^+ \ \text{ where } \aleph^+ \text{ is the set of positive integers}$$

$$\beta_{ij},\ \delta_{ij} \in \Re^+ \ \text{ where } \Re^+ \text{ is the set of positive real numbers}$$

Let $(X^*{}_i,\ Y^*{}_{ij},\ g^*{}_i,u^*,\ \beta^*{}_{ij},\ \delta^*{}_{ij},\ k^*,\ Q^*,\ v^*,\ w^*)$ be an optimal solution for the *final optimization* model.

Remark: For any set $\overline{X} = (X_i,\ Y_{ij},\ k,\ Q_i)$ and the corresponding linearized set $\overline{X_L} = (X_i,\ Y_{ij},\ g_i,\ u_i,\ \beta_{ij},\ \delta_{ij})$ the following relation holds.

$$TNC^F(\overline{X_L}) < \widehat{TNC}(\overline{X}) < TNC(\overline{X})$$

Remark: Note that the linear version of problem obtained after applying the *linearization* approach would be exact if k was a binary decision variable. Since the range of values of k is continuous, the solution of the FOM gives a lower bound for the original problem. A branching method is used to get a good estimate of k and improve the lower bound. The lower and upper bound values of k are known. The branching method bisects the interval defined by the upper and lower bounds of k and optimizes the FOM over these intervals. The branching is repeated on the new intervals until the values of u^* and $\delta^*{}_{ij}$ are very close to the the values of k^*Q^*, $Y^*{}_{ij}/(u^*)$.

7.5 Numerical Illustration

For the numerical example we consider a region with 18 retail stores and a national distribution center (NDC). The locations of these stores and the NDC are fixed and known. The distances between the NDC and the stores is calculated using the euclidean norm (see appendix 7(b)). Further each retail store location is a candidate site for the regional distribution center (RDC). Thus, the maximum number of RDCs in this example can be 18. We used CPLEX to solve the *final optimization* model (FOM) and an optimal solution was obtained in 40 CPU seconds on a PC with a 2.66 GHz Pentium IV processor and 1-GB memory. Table 9 gives the optimal solution of

Table 9: Solution of the Linear Approximation Model

RDC location	Stores Allocation
4	1,3,5,6,8,10,15,17
11	2,4,7,9,11,12,13,14,16,18
Q^*	1000
k^*	14

Table 10: Objective function value

Optimal solution for FOM	Feasable solution for P(1)	Optimality Ratio
442493	506546	1.079

the FOM. In this problem lower and upper bounds for k are 2 and 20. For results on the bisection method for estimating k see table 11.

The objective function of the FOM is evaluated at the optimal solution. The optimal solution generated by the FOM is a feasible solution for the original problem P(1). The objective function of problem P(1) evaluated at this solution gives an upper bound. The ratio of the objective function values for the original nonlinear problem and the FOM is defined as the optimality ratio. In this example, the optimality ratio was found to be 1.079.

Table 11: Bisection Analysis for k

k	Q	TNC	TNC^F	Optimality Ratio
4	1000	666189	412493	1.615031668
8	1000	666189	412493	1.202567662
10	1000	514089	427493	1.146208545
12	1000	507189	442493	1.146208545
14	1000	506546	457493	1.078987959
16	1000	509814	472493	1.078987959

7.6 Conclusions

In this section, a *linear reformulation* solution technique is proposed for solving the integrated facility location and inventory allocation problem. The original problem has a nonlinear objective function with square root and bilinear terms. Since there is no single approximation technique that can handle the different forms of nonlinearity, a step by step method is proposed for approximating the objective function with a linear form. The square root terms are eliminated by defining a lower bound on the objective function. A *linear reformulation* method is defined that handles the remaining nonlinear terms by approximating them with linear terms and by imposing new constraints. The *final optimization* model (FOM) is linear in the objective function and the constraints, and can be solved using CPLEX. The optimal solution to the (FOM) defines a feasible solution for the original problem. The objective functions for the FOM and the original problem are evaluated at the optimal value and an optimality ratio is obtained.

When compared with the continuous approximation technique, the solution obtained using the *linear reformulation* gives exact cost expressions for the different components (facility, transportation and inventory). However, good estimates of upper and lower bounds are needed for Q and k in order to use the *linear reformulation* technique. Thus, it is not possible to compare one technique against another and say which is better. All we can say is that the continuous approximation technique aids in a quick sensitivity analysis of the problem while the *linear reformulation* technique gives a good solution with near accurate estimates of cost expressions.

As a future work, it would be interesting to compare the solution obtained by the *linear reformulation* technique with the exact solution. In order to get an exact solution we need to solve the original optimization problem for 2^{18} $(= 262144)$ instances. This is because the network can be designed in 2^{18} ways and we need to find the optimal inventory parameters (Q, k) for each of these instances.

Other possible directions for future work include formulating the integrated facility location and inventory allocation problem with multiple products. In this analysis the focus is only on one products but in the real world companies have to deal with multiple products.

7.7 *Appendix*

Appendix 7(a) Proof for Result 7.1

$$
\begin{aligned}
ax + bx^2 &> ax + bx^2 - \frac{a^2}{4b} \\
&= \left(\sqrt{b}x - \frac{a}{2\sqrt{b}}\right)^2 \\
\Rightarrow \sqrt{ax + bx^2} &> \left(\sqrt{b}x - \frac{a}{2\sqrt{b}}\right)
\end{aligned}
\tag{46}
$$

Appendix 7(b) Distance calculations

The data for the store locations specified the zip code associated with each location. This information was used to get the values of the longitudes and latitudes for each location. Using these values along with equation 47 we calculated the euclidean distances between all the store locations and the NDC location, and between every pair of store locations.

$$D(x1, x2) = 3963.0cos^{-1}[sin(a_1)sin(a_2) + cos(a_1)cos(a_2)cos(b_2 - b_1)] \tag{47}$$

where $(a_1,\ b_1)$ and $(a_2,\ b_2)$ are the latitudinal and the longitudinal coordinates associated with points $x1$ and $x2$ respectively.

CHAPTER VIII

BASE STOCK POLICY

A base stock policy is represented as a (*s-1*, *s*) policy in which a replenishment order is placed whenever the inventory position (inventory on-hand + inventory on-order + amount backlogged) is below *s*. This inventory control policy is studied extensively in the one warehouse multiple retailers distribution networks when dealing with low demand items, high holding costs and low ordering costs.

There are several studies in the area of multi-echelon base stock policy. Axsater [4] use a METRIC approach proposed by Sherbrooke [35] to characterize the demand process during lead time at each retailer. Under this approach, the successive lead times for each retailer are assumed to be independent, and Palm's theorem [11] is used to describe the outstanding orders at each retailer by a Poisson process. In another study, Axsater [4] present an exact formulation for the inventory holding cost and the backorder cost associated with each unit on order. A recursive procedure is used to evaluate the exact cost expressions. Graves [21] fit a negative binomial distribution and determine both the mean and variance of the number of outstanding orders.

Before we present a model for the integrated facility location-inventory problem for the base stock inventory control policy, the original distribution network needs to be modified. So far the distribution network under study consists of a single NDC and multiple retailers where the NDC acts like an import DC consolidating goods that are manufactured off-shore. Since there is uncertainty in the supply chain in terms of travel time, RDCs are located between the NDC and the retailers to provide risk pooling benefits in terms of inventory costs savings. In this network there are high reorder costs as manufacturing is done overseas. In order to facilitate the analysis using the base stock policy, a distribution network with operations involving low fixed cost costs and low demand has to be considered. Thus, a modified distribution network in with the NDC functions like a plant with infinite inventory and inventory is stored at the retailers and RDCs is considered. By solving the integrated facility location-inventory problem over the modified network, optimal parameters for the base stock inventory policy at the RDCs and retailers is determined along with the locations for the RDCs.

8.1 *Model*

Consider a distribution network with a single NDC and N retailers where each retailer location is a candidate location for the RDC. Inventory is stocked at the RDCs and retailers. The holding cost at the retailer j, $j = 1, 2, ..., N$ is h_j, and the holding cost at each RDC is h_r. When a demand arrives at the retailer it is either satisfied from the stock on-hand or backordered. Each location follows a base stock policy except for the NDC which is assumed to be a source of infinite inventory.

Some of the key assumptions in this analysis are listed below.

1. The demand at each retailer j is an independent Poisson process with rate λ_j.
2. The inventory holding cost at the retailer is more expensive than at the RDC, i.e., $h_r < h_j \quad \forall j$.
3. A service level is defined only for the retailers.
4. Inventory is managed at each location using a local control strategy under which a stocking location does not take into account the inventory positions at the other locations while deciding on its own policy.

Decision Variables

$$X_i = \begin{cases} 1 & \text{if a RDC is opened at location i} \\ 0 & \text{otherwise} \end{cases}$$

$$Y_{ij} = \begin{cases} 1 & \text{if the store at location j is served} \\ & \text{by the RDC at location i} \\ 0 & \text{otherwise} \end{cases}$$

$$s_i \; = \; \text{base stock inventory level at the RDC located at i}$$

$$s_j \; = \; \text{base stock inventory level at the retailer located at j}$$

Let B_i be the backorders and I_i be the on-hand inventory at RDC location i = *1,...,N*. Similarly, B_j be the backorders and I_j be the on-hand inventory at retailer location j = *1,...,N*. Then, we have

$$B_i = [D_i - s_i]^+ \quad \forall i = 1,2,...,N$$

$$I_i = [s_i - D_i]^+ \quad \forall i = 1,2,...,N$$

$$B_j = [B_{ij} + D_j - s_j]^+ \quad \forall j = 1,2,...,N$$

$$I_j = [s_j - (B_{ij} + D_j)]^+ \quad \forall j = 1,2,...,N$$

where D_i and D_j are the demand during lead-time at RDC i and retailer j. B_{ij} is the number of backorder for retailer j at RDC i. Note that the function $[x]^+$ is defined as $max(0, \text{x})$. In general, the distribution for B_{ij} is given by

$$Pr(B_{ij}(s_i) = k) = \sum_{L=k}^{\infty} \begin{pmatrix} L \\ k \end{pmatrix} (p_j)^k (1 - p_j)^{L-k} Pr(B_i = L)$$

Consider a retailer j that is assigned to RDC i. Let s_i be the base stock level at the RDC. Then the optimal base stock level s_j at retailer j for a desired fill-rate β_j is given by

$$s_j^*(s_i) \equiv min\{s_j : Pr(B_{ij} + D_j < s_j) \geq \beta_j\}$$

Similarly for a given set of retailers *1,2,...,*n_i assigned to RDC i, the optimal base stock level at RDC i is given by

$$s_i^*(s_1, s_2, ..., s_{n_i}) \equiv min\{s_i : Pr(B_{ij} + D_j < s_j) \geq \beta_j \quad \forall j \in 1,2,...,n_i\}$$

It is possible to define bounds on the base stock levels s_i and s_j for each RDC i and retailer j. The lower and upper bound for s_j is given by s_j^l and s_j^u. Similarly, s_i^u is the upper bound for the base stock level at RDC i and 0 is the lower bound. The derivation of these bounds follow next.

For a given RDC i and retailer j, each retailer will hold minimum stock when the RDC holds the maximum stock. Similarly, each retailer will hold maximum stock when the RDC holds the minimum stock. The minimum base stock level at the RDC can be 0 and the maximum can be ∞. Thus,

$$s_j^l(\infty) \equiv min\{s_j : Pr(B_{ij} + D_j < s_j) \geq \beta_j\}$$
$$s_j^u(0) \equiv min\{s_j : Pr(B_{ij} + D_j < s_j) \geq \beta_j\}$$

Similarly, the maximum stock at RDC i occurs when each retailer j assigned to it holds minimum possible stock. The minimum possible value of stock at any RDC is 0.

$$s_i^u(s_1^l, s_2^l, ..., s_{n_i}^l) \equiv min\{s_i : Pr(B_{ij} + D_j < s_j^l) \geq \beta_j \forall j \in 1,2,...,n_i\}$$

The expected inventory holding cost at RDC i and the retailers assigned to it are given by:

$$\boldsymbol{H}_j = h_j E[s_j - (B_{ij}(s_i) + D_j)]$$
$$\boldsymbol{H}_i = h_i E[s_i - D_i)]$$

In formulating the problem, the key cost components that need to be considered are the facility location cost, the inbound and outbound transportation cost, and the average inventory cost at the each RDC and the retailer. The expressions for the facility cost and the outbound transportation cost are the same as in the case of the batch ordering policy. For the inbound transportation cost between the NDC and each RDC, the fixed cost component is ignored (i.e., $C_f = 0$) under the base stock policy. The inbound transportation cost is modeled in terms of the distance between the NDC and each RDC. The optimization model for the integrated facility location and inventory allocation problem under the base stock policy is given by:

P^b: Minimize

$$\eta = \sum_{i=1}^{N} F_i X_i + \sum_{i=1}^{N}\sum_{j=1}^{N} \xi(C_v d_i + C_f d_{ij}) Y_{ij} \lambda_j + \sum_{i=1}^{N} \boldsymbol{H}_i X_i + \sum_{j=1}^{N} \boldsymbol{H}_j Y_{ij}$$

subject to

$$Y_{ij} \le X_i \quad \forall i, j \quad (1)$$

$$\sum_{i=1}^{N} Y_{ij} = 1 \quad \forall i \quad (2)$$

$$s_j^l \le s_j \le s_j^u \quad \forall j \quad (3)$$

$$0 \le s_i \le s_i^u \quad \forall j \quad (4)$$

$$Pr((B_{ij}(s_i) + D_j) < s_j) \ge \beta_j \quad \forall j \quad (5)$$

$$X_i,\ Y_{ij} \in \{0,1\}$$

$$s_i,\ s_j \in Z^+$$

where constraint (1) forces the assignment of each retailer to only an open RDC location. Each retailer has to be assigned to exactly one RDC location and this is achieved using constraint (2). Constraint (3) and (4) define the bounds on the base stock levels and constraint (5) is the fill-rate constraint for each retailer.

8.2 Solution methodology

A solution to the optimization problem P^b locates RDCs at one or more of the potential locations i = *1,2,...,N*. Also each retailer is assigned to exactly one RDC location. In addition, the base stock policy at each RDC i and each retailer is determined by the solution. In this section, we propose a solution procedure for solving the problem. Note that the distribution for B_{ij} as well as the expression for the expected inventory cost at the RDC and the retailer is hard to determine in general. However if a normal distribution is assumed for the lead time demand at each location, it is possible to simplify the objective function.

The solution approach for modeling the inventory cost is similar to the one proposed by Gallego *et al.* [18]. Under this approach, closed form expression for the average inventory and backorder level are derived using a normal demand distribution along all locations. Also the fill-rate constraint is modeled into the objective function in terms of the penalty cost function. For a single echelon model, for each service constraint there exists am equivalent representation in terms of backorder cost by setting the penalty cost (see Zipkin [39]) defined by $b = \frac{\beta}{1-\beta}h$ where h is the inventory holding cost. It has been shown that the same relationship may not hold for a multi-ecehlon model. Thus, we estimate $b = kh$, where k is a constant set by the management.

Applying these changes, the objective function in the original problem takes the following form:

$\tilde{P}^b$: Minimize

$$\begin{aligned} \eta &= \sum_{i=1}^{N} F_i X_i + \sum_{i=1}^{N}\sum_{j=1}^{N} \xi(C_v d_i + C_f d_{ij}) Y_{ij}\lambda_j \\ &+ \sum_{i=1}^{N}(min_{s_i}\{h_i E[I_i] X_i + \sum_{j=1}^{N}(b_j + h_j)\phi(z_j^*)\tilde{\sigma_j} Y_{ij}\}) \end{aligned} \tag{48}$$

The new constraint set is the same as before without the fill-rate constraint inequality. Note that it is still hard to optimize the problem in its current form. Thus, a heuristic approach is used to solve the problem. A *Drop-Decomposition* heuristic is used in this analysis. The key idea is to use the *Drop* heursitic (see Daskin [14]) along with the *Restriction-Decomposition* heuristic proposed by Gallego *et al.* [18]. While the *Drop* heursitic is used extensively in the facility location literature, the *Restriction-Decomposition* heuristic aids in finding a near-optimal base-stock policy for each RDC location-store assignment combination.

The *Decomposition* heursitic is a triple search heuristic. The base stock level for each RDC i is set at three levels: $s_i = 0$, $\mathrm{E}[D_i]$ and s_i^u, where $\mathrm{E}[D_i]$ is the demand at the RDC during its order lead time and s_i^u is the maximum possible value of the base stock level at the RDC. For each level an optimal value of the base stock level for each retailer is calculated using the following relationships.

$$
\begin{aligned}
s_j^*(s_i) \quad &= \quad \tilde{\mu_j} + z_j^* \tilde{\sigma}_j \quad \text{where} \\
\tilde{\mu_j} \quad &= \quad E[B_{ij}] + \mu_j \lambda_j \quad \text{see Appendix 8.5 for derivation} \\
\tilde{\sigma}_j^2 \quad &= \quad Var[B_{ij}] + \mu_j \lambda_j \quad \text{see Appendix 8.5 for derivation} \\
z_j^* \quad \text{solves} \quad &\Phi(z) = \frac{h_j}{h_j + b_j}
\end{aligned}
$$

The *Drop* heuristic begins by placing a RDC facility at each possible location. As the RDC facilities are consolidated, the total cost drops. Thus, at each iteration the *Drop* heuristic greedily removes facilities from the solution until it can no longer find one. Greedily selecting the node to be removed means that each node that is removed from the solution causes maximum reduction of the total cost.

At each iteration of the *Drop-Decomposition* heuristic, the total facility, transportation, inventory holding and backorder penalty cost is calculated. For calculating the inventory cost, the *Decomposition* heuristic is used. For each retailer, the optimal value of base stock level is calculated for three fixed levels of the RDC base stock level (0, $E[D_i]$, s_i^u). Here $E[D_i]$ is the expected demand during lead time at RDC i and s_i^u is the maximum possible base stock level for RDC i. The average inventory holding and penalty cost is calculated for each combination of RDC-retailer base stock level and the minimum value is chosen as the cost associated with holding inventory and backorders. Thus at each iteration of the *Drop-Decomposition* heuristic, the *decomposition* heuristic aids in calculating the inventory and penalty cost while the *drop* heuristic is used to greedily remove RDCs to reduce the total cost.

Table 12: Drop-Decomposition heuristic result

Iteration	Total Cost	Drop action	store assignment
1	243233		
2	229390	Remove DC at node 16	Add store 16 to RDC 8
3	215620	Remove DC at node 14	Add store 14 to RDC 8
4	201857	Remove DC at node 18	Add store 18 to RDC 8
5	188102	Remove DC at node 9	Add store 9 to RDC 8
6	174411	Remove DC at node 17	Add store 17 to RDC 8
7	160739	Remove DC at node 11	Add store 11 to RDC 8
8	147095	Remove DC at node 12	Add store 12 to RDC 8
9	133470	Remove DC at node 7	Add store 7 to RDC 8
10	120037	Remove DC at node 13	Add store 13 to RDC 8
11	106738	Remove DC at node 15	Add store 15 to RDC 8
12	93712	Remove DC at node 10	Add store 10 to RDC 8
13	81284	Remove DC at node 1	Add store 1 to RDC 8
14	68887	Remove DC at node 2	Add store 2 to RDC 8
15	57326	Remove DC at node 3	Add store 3 to RDC 8
16	45810	Remove DC at node 5	Add store 5 to RDC 8
17	34328	Remove DC at node 4	Add store 4 to RDC 8
18	22975	Remove DC at node 6	Add store 6 to RDC 8

8.3 Numerical Illustration

The *Drop-Decomposition* heuristics (DDH) proposed in the previous section is used to solve an example with 18 stores. The result obtained in each iteration of the heuristic is presented in table 12. The distance between each candidate RDC location and the NDC is given in table 14 (see appendix 8.5) and the distance matrix between any two locations is given in table 15 (see appendix 8.5). The final solution obtained using the DDH opened a RDC at candidate location 8 and assigned all stores to it.

Remark: It would be interesting to analyze the quality of this solution. An exact evaluation of the optimal solution for this problem requires complete enumeration. Enumeration itself is hard because of the exponential number of possibilities and the complex expressions for the inventory and penalty cost functions in the problem (see problem P^b). However the nature of this example makes exact enumeration easy with identical cost and demand attributes along the retailers. For any given RDC location

Table 13: Average Inventory cost at each RDC as a function of number of stores assigned to it

No. of stores	Total cost RDC i	Total cost at other RDCs	Average Inventory cost
1	111	1887	1998
2	210	1776	1986
3	307	1665	1972
4	403	1554	1957
5	498	1443	1941
6	593	1332	1925
7	688	1221	1909
8	782	1110	1892
9	876	999	1875
10	971	888	1859
11	1064	777	1841
12	1158	666	1824
13	1252	555	1807
14	1346	444	1790
15	1439	333	1772
16	1532	222	1754
17	1626	111	1737
18	1719	0	1719

i, it was observed that the average inventory holding and penalty cost increases as more stores are added to it (see table 13). However, as the RDCs are consolidated, the average inventory cost decreases. This decrease in the average inventory cost is more significant than the increase in the average inventory cost at RDC i. In this analysis the upper bound on the base stock level at each RDC is in the range of 14-52 units. It was estimated by calculating the least possible value of base stock level at each store. RDC i will have maximum base stock level when each store assigned to it maintains its base stock at minimum possible level.

Remark: The optimal solution of the facility location and store assignment problem for this example can be solved using CPLEX. The optimal solution opened a RDC at location 8 and assigned all stores to it. Also we know that the average inventory and penalty cost is minimized when all stores are assigned to a single RDC. Then it is plausible to say that the solution obtained by the DDH is indeed the optimal

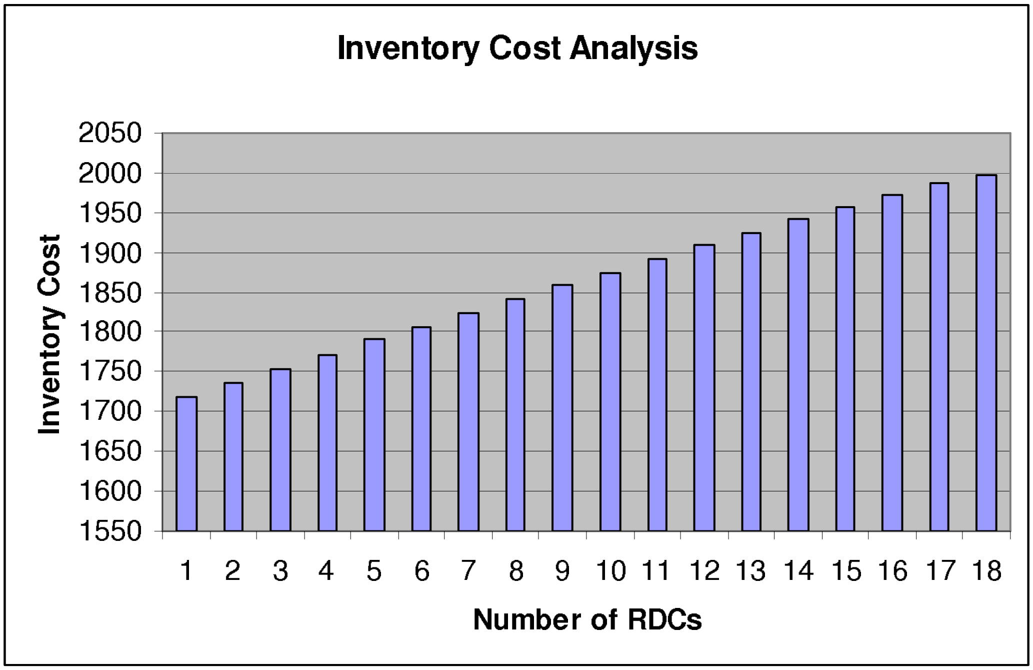

Figure 18: Linear Pattern for the average inventory cost

solution. Note that DDH is not gauranteed to produce an optimal solution in all instances and it would be interesting to explore this area as a future study.

Remark: For large problem instances it can get cumbersome to apply the DDH. One possible variation to the solution could be to analyze the shape of the average inventory cost function. It may be possible to express the inventory cost as a linear function of the number of RDCs. Infact in the previous example, the average inventory cost is indeed a linear function of the number of RDCs (see figure 18). The parameters of this linear regression model can be obtained using any statistical package such as MINITAB or Excel. Once the inventory cost is expressed as a function of the RDCs, the problem takes the form of an *uncapacitated facility location* problem and it can be solved using the available heuristics (see Daskin [14]). We leave this analysis for future work.

8.4 Conclusion

In this section, an integrated model for the facility location-inventory allocation problem is presented under a base stock control policy at the RDCs and the retailers. A *Drop-Decompostion* solution approach is presented to solve the problem. This approach uses two popular heuristics from the facility location (*Drop* heuristic) and the multi-echelon inventory allocation (*Decompostion* heuristic) literature.

When analyzing a multi-echelon inventory allocation problem, the main challenge lies in the estimation of the demand process at the upper echelon. A continuous review base stock policy makes this analysis simpler and is preferred in such analysis as the demand process at the DC can be easily estimated as the sum of the demand processes at the retailers. However this inventory policy can be applied only to a special class of distribution networks with negligible reorder costs and low demand.

Table 14: Distance between RDC location i and the NDC

RDC Location i	d_i		RDC Location i	d_i		RDC Location i	d_i
1	138		7	194		13	185
2	137		8	26		14	201
3	98		9	200		15	178
4	94		10	166		16	204
5	95		11	196		17	197
6	88		12	195		18	201

8.5 Appendix

$$
\begin{aligned}
E[B_i] &= (\phi(z_i) - z_i(1-\Phi(z_i)))\sqrt{\sum_{i=1}^{N}\mu_i\lambda_j Y_{ij}} \\
Var[B_i] &= [(z_i^2+1)(1-\Phi(z_i)) - z_i\phi(z_i)]\sum_{i=1}^{N}\mu_i\lambda_j Y_{ij} + E[B_i] - (E[B_i])^2 \\
E[I_i] &= (\phi(-z_i) + z_i(1-\Phi(-z_i)))\sqrt{\sum_{i=1}^{N}\mu_i\lambda_j Y_{ij}} \\
z_i &= \frac{s_i - \sum_{i=1}^{N}\mu_i\lambda_j Y_{ij}}{\sqrt{\sum_{i=1}^{N}\mu_i\lambda_j Y_{ij}}} \\
p_j &= \frac{\lambda_j Y_{ij}}{\sum_{i=1}^{N}\lambda_j Y_{ij}} \\
E[B_{ij}] &= p_j E[B_i] \quad \forall j \\
Var[B_{ij}] &= p_j(1-p_j)E[B_i] + (p_j)^2 Var[B_i] \quad \forall j \\
\tilde{\mu_j} &= E[B_{ij}] + \mu_j\lambda_j \\
\tilde{\sigma}_j^2 &= Var[B_{ij}] + \mu_j\lambda_j \\
E[B_j] &= (\phi(z_j) - z_j(1-\Phi(z_j)))\tilde{\sigma_j} \\
E[I_j] &= (\phi(-z_j) + z_j(1-\Phi(-z_j)))\tilde{\sigma_j} \\
z_j &= \frac{s_j - \tilde{\mu_j}}{\tilde{\sigma_j}}
\end{aligned}
$$

Table 15: Distance between RDC location i and store j

d_{ij}	1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	0	20	62	78	91	92	58	127	90	109	99	101	107	99	114	106	127	126
2	20	0	49	91	104	104	57	130	76	128	83	83	88	83	133	90	147	107
3	62	49	0	91	103	100	104	99	102	151	99	98	91	103	160	107	176	109
4	78	91	91	0	14	14	135	72	166	73	172	173	174	174	85	180	104	192
5	91	104	103	14	0	8	147	71	180	70	186	186	187	187	83	193	101	206
6	92	104	100	14	8	0	149	64	179	78	185	185	185	186	90	192	108	204
7	58	57	104	135	147	149	0	185	56	145	75	77	94	70	146	78	154	109
8	127	130	99	72	71	64	185	0	199	141	197	197	190	201	154	206	172	206
9	90	76	102	166	180	179	56	199	0	194	20	23	44	14	197	22	206	55
10	109	128	151	73	70	78	145	141	194	0	206	208	216	205	13	213	31	234
11	99	83	99	172	186	185	75	197	20	206	0	4	25	7	210	8	221	35
12	101	83	98	173	186	185	77	197	23	208	4	0	21	11	212	10	223	32
13	107	88	91	174	187	185	94	190	44	216	25	21	0	32	220	29	233	19
14	99	83	103	174	187	186	70	201	14	205	7	11	32	0	208	8	219	41
15	114	133	160	85	83	90	146	154	197	13	210	212	220	208	0	216	18	239
16	106	90	107	180	193	192	78	206	22	213	8	10	29	8	216	0	226	34
17	127	147	176	104	101	108	154	172	206	31	221	223	233	219	18	226	0	251
18	126	107	109	192	206	204	109	206	55	234	35	32	19	41	239	34	251	0

CHAPTER IX

FUTURE WORK

This work is based on some simplified assumptions in order to make the problem tractable and to enable us to derive meaningful insights. It would be interesting to relax some of the assumptions to match the 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 in isolation without considering the impact on the other clusters in the sub-region. This can happen in the 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 continous review policy at each RDC would affect the network design and costs.

It would be interesting to incorporate delivery lead time contraints in the model. In the real world distribution operations, most orders have a delivery time window, i.e., a lower and upper bound on the time it can take the order to arrive at a facility. It is an important service measure and can impact the network design.

Throughout this study, the integrated network design and inventory allocation problem is solved assuming a pull system. In a pull system, each facility analyzes its local inventory and decides on the order quantity. It would be interesting to model the problem under a push system where a central decision maker analyzes the inventory at all the facilities and makes the inventory allocation decisions.

REFERENCES

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

[2] Al-Khayyal, F. A., "Jointly constrained bilinear programs and related problems: An overview," *Computers and Mathematics with Applications*, vol. 19, no. 8, pp. 53–62, 1990.

[3] Al-Khayyal, F. and Falk, J., "Jointly constrained biconvex programming," *Mathematics of Operations Research*, vol. 8, no. 2, pp. 273–286, 1983.

[4] Axsäter, S., *Handbooks in Operations Research and Management Science, Vol. 4. Logistics of Production and Inventory.*, ch. Continuous review policies for multi-level inventory systems with stochastic demand. North-Holland, Amsterdam, The Netherlands, 1993.

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

[6] Blumenfeld, D. E. and Beckmann, M. J., "Use of continuous space modeling to estimate freight distribution costs," *Transportation Research A*, vol. 19A, no. 2, pp. 173–187, 1985.

[7] B.Schwarz, L., L.Deuermeyer, B., and Badinelli, R. D., "Fill-rate optimization in a one-warehouse n-identical retailer distribution system," *Management Science*, vol. 31, no. 4, pp. 488–498, 1985.

[8] Burn, L., Hall, R., Blumenfeld, D. E., and Daganzo, C. E., "Distribution strategies that minimize transportation and inventory cost," *Operations Research*, vol. 33, no. 3, pp. 469–490, 1985.

[9] C.F.Daganzo and G.F.Newell, "Configuration of physical distribution networks," *Networks*, vol. 16, pp. 113–132, 1986.

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

[11] C.Palm, "Analysis of the erlang traffic formula for busy signal assignment," *Ericsson Technics*, vol. 5, pp. 39–58, 1938.

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

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

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

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

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

[17] Erlenkotter, D., "The general optimal market area model," *Annals of Operations Research*, vol. 18, pp. 45–70, 1989.

[18] Gallego, G., Özer, O., and Zipkin, P., "Bounds, heuristics, and approximations for distribution systems," *Operations Research*, vol. 55, no. 3, pp. 503–517, 2007.

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

[20] Geoffrion, A. M., "The purpose of mathematical programming is insight, not numbers," *Interfaces*, vol. 7, pp. 81–92, 1976.

[21] Graves, S., "A multi-echelon inventory model for a repairable item with one-for-one replenishment," *Management Science*, vol. 31, pp. 1247–1256, 1985.

[22] Hopp, W. and Spearman, M., *Factory Physics.* McGraw Hill, 2000.

[23] Langevin, A., Mbaraga, P., and Campbell, J. F., "Continuous approximation models in freight distribution: an overview," *Transportation Research B*, vol. 30, no. 3, pp. 163–88, 1996.

[24] Little, J., "A proof of the queuing formula: L = λw," *Operations Research*, vol. 9, pp. 383–387, 1961.

[25] Menlo, W., "Logistics industry." http://www.menloworldwide.com/mww/en/aboutus/logistics_industry.shtml, 11/11/2007.

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

[27] Muckstadt, J. A. and Thomas, L. J., "Are multi-echelon inventory methods worth implementing in systems with low-demand-rate items," *Management Science*, vol. 26, no. 5, pp. 483–494, 1980.

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

[29] Nozick, L. and Turnquist, M., "Integrating inventory impacts into a fixed-charge model for locating distribution centers," *Transportation Research-E*, vol. 34, no. 3, pp. 173–186, 1998.

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

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

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

[33] Rutten, W., Laarhoven, P. V., and Vos, B., "An extension of the GOMA model for determining the optimal number of depots," *IIE Transactions*, vol. 33, pp. 1031–1036, 2001.

[34] Shen, Z. M., Coullard, C., and Daskin, M. S., "A joint location-inventory model," *Transportation Science*, vol. 37, no. 1, pp. 40–55, 2003.

[35] Sherbrooke, C., "Metric: A multi-echelon technique for recoverable item control," *Operations Research*, vol. 16, pp. 122–141, 1968.

[36] Teo, C. and Shu, J., "Warehouse-retailer network design problem," *Operations Research*, vol. 52, no. 3, pp. 396–408, 2004.

[37] van Houtum, G., "Multi-echelon production/inventory systems: Optimal policies, heuristics, and algorithms." Tutorials in Operations Research 2006.

[38] Wang, N. and Lu, J., "Multi-level spatial modelling and decision-making with application in logistics systems," tech. rep., Georgia Institute of technology, 2006.

[39] Zipkin, P., *Foundation of Inventory Management*. The Irwin McGraw-Hill Series, 2000.

VITA

Divya Mangotra was born on June 4, 1979 in India and graduated from St. Stephen's College, University of Delhi in India with a bachelor's degree in Mathematics in 2000. The following fall she was selected for a scholarship progam at the University of Cambridge to pursue masters in Mathematics. She joined the doctoral program in Industrial and Systems Engineering at Georgia Tech in 2002 and started research on the network design and inventory problems. The Logistics Institute at Georgia Tech selected her as a Global Logistic Scholar for the academic year 2005. She was awarded the General Motors scholarship for excellence in research in 2006. She has four years of research experience working on several industry projects. After completing her PhD studies, Divya joined Hewlett-Packard as an Analytical Consultant for the Strategic Planning and Modeling Group.