EFFICIENT DESIGN OF INBOUND LOGISTICS NETWORKS

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

In Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Zhijie Dong

August 2015

EFFICIENT DESIGN OF INBOUND LOGISTICS NETWORKS

Zhijie Dong, Ph. D.
Cornell University 2015

Logistics is a vitally important part of the economy, and it is now a $1.45 trillion industry in the United States representing 8.3 percent of GDP. Efficient design of routes and schedules for moving materials into manufacturing or assembly plants is a central part of inbound logistics operations. This dissertation builds on elements of traditional vehicle routing as well as broader elements of logistics planning. At the core of the process is a mathematical optimization termed capacitated clustering.

Two major categories of suppliers are analyzed in this research. The first supplier category includes suppliers with small quantities of materials, so daily pickups may not be required. A new approach is proposed that considers pick-up frequency and spatial design as joint decisions to minimize total logistics (transportation plus inventory) cost. The clustering-based optimization uses an approximation to the actual cost of a routing solution without actual route construction. The problem is shown to be analogous to a single-source fixed-charge facility location problem, and near-optimal solutions can be found using an efficient heuristic algorithm. Computational experiments show the effectiveness of how this model is formulated and a case study demonstrates that substantial total cost savings can be achieved in realistic applications.

A second category of suppliers ships moderately large volumes to a single plant but not enough to fill a truck themselves. One commonly used process is to have plant-

based collection routes on a daily basis that stop at multiple suppliers and return to the plant. The model developed here is formulated as a two-stage stochastic program, which includes uncertainty in the load quantities at suppliers and controls (either penalties or constraints) designed to improve the “regularity” of service to individual suppliers. Two adaptive decomposition heuristics are explored for solving the stochastic program in large scale, integer L-shaped method (ILSM) and progressive hedging (PH). An application to logistics operations in the automotive industry is used to demonstrate the effectiveness of the model and the PH solution method.

BIOGRAPHICAL SKETCH

Zhijie Dong was born on April 15, 1989 in Nanjing, Jiangsu Province, China. She graduated from Nanjing Jinling High School and obtained her Bachelor of Engineering degree at Nanjing University in 2011. Later that year, she earned a Master of Science degree in Civil Engineering at Columbia University in the City of New York. Zhijie came to Cornell University in August 2012 to pursue her PhD degree under the guidance of Professor Mark Turnquist in the area of transportation systems engineering, operations research and supply chain management. She worked as an Operations Research intern for CSX Transportation in 2014.

In addition to her academic activities and experiences, Zhijie is active both at campus and in the community. She served as the President of Cornell Chinese Scholars and Students Association and the Treasurer of Civil and Environmental Engineering Graduate Students Association. She is one of the winners for the Colman Leadership Award.

Zhijie met her husband, Ningmu Zou, on July 10, 2007 at Nanjing University, and they were married on July 10, 2014 at Cornell University.

To my parents and husband for their unconditional love, and to my advisor, Dr. Turnquist, for his tremendous support.

ACKNOWLEDGMENTS

First and foremost, I would like to thank my advisor, Dr. Mark Turnquist, for his continuous guidance and encouragement throughout my doctoral program. He is the one who led me into the world of transportation and system engineering. He has taught me how to be a good thinker, starting from discovering the research problems to determining appropriate solution methodologies, from shaping innovative research ideas to presenting findings effectively. He is always patient and helpful during my ups and downs in research. Beyond academia, he has provided valuable advice on my career development and been a tremendous support personally. I cannot thank Dr. Turnquist as well as his wife, Lynn Turnquist, enough.

My greatest appreciation also goes to my committee members, Dr. Linda Nozick and Dr. Huseyin Topaloglu. Dr. Nozick has provided insights in my research with her extensive knowledge about transportation modeling and optimization. Dr. Topaloglu is friendly and supportive in numerous ways from research to job-hunting. I have greatly benefited from my interactions with him, and really enjoyed his course in simulation.

I would like to extend my gratitude to Dr. Oliver Gao, Dr. Ricardo Daziano, and Dr. Francis Vanek for sharing their critical experiences with me and offering many pieces of constructive advice at different stages of my Ph.D. study. I am grateful to Dr. Jan Allen, Dr. Colleen McLinn, and Dr. Anne Poduska who invested tremendous effort in improving my academic writing skills.

Special thanks go to my Ph.D. fellows and friends, both at Cornell University and other universities, including Chen Wang, Lingfeng Cheng, Bangrui Chen, Haorao Zhao, Xi He, Kecheng Xu, Chao Zhang, Jie Yang, Yueran Zhuo, Haohua Wan, Qing Zhao, and Arash Beheshtian. I will miss the times we spend together working on projects, discussing research problems, preparing for job interviews, and hanging out. You have made my Ph.D. life at Cornell more enjoyable and unforgettable.

Last but not least, I want to thank my parents and my husband, Ningmu Zou. Nothing would have been possible without the unconditional love and support from you. Thank you very much for being there forever no matter what.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1

INTRODUCTION

1.1 Background

Logistics is a vitally important part of the economy. The International Monetary Fund reports that logistics costs account for about 12 percent of the world's gross domestic product (GDP) (Ballou, 2003). According to the 26th Annual State of Logistics Report (Council of Supply Chain Management Professionals, 2015), logistics is now a $1.45 trillion industry in the United States, and it has represented between 7.9 percent and 9.9 percent of GDP over the past decade (as shown in Figure 1-1).

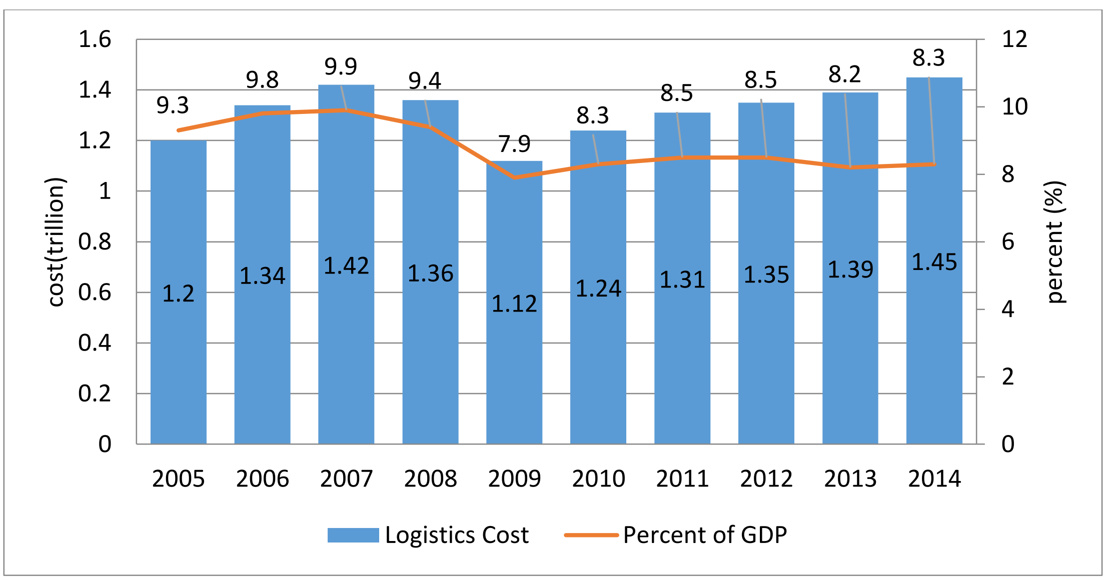


Figure 1-1. Logistics cost as a percent of GDP in U.S. from 2005 to 2014

The term, *logistics*, comes from late 18th and early 19th centuries in the context of military activities. Its modern application was initially developed from the automotive industry. The Council of Supply Chain Management Professionals (CSCMP) defines logistics as "part of supply chain management that plans, implements, and controls the efficient, effective forward and reverse flow and storage of goods, services and related information between the point of origin and the point of consumption in order to meet customers' requirements." The role that logistics plays in modern enterprises is as crucial as it is in wars –it can help maximize the efficiency of the production and distribution process and improve competitiveness of enterprises using limited resources, and it is considered to be the last frontier for cost reduction. With the increasing trend in economic globalization in recent years, the importance of logistics management has received considerable attention from various industries. Therefore how to manage logistics systems efficiently becomes a critical issue for almost all companies.

A logistics system consists of many components such as transportation, inventory, facility location, demand forecasting and information processing. Figure 1-2 demonstrates that transportation costs and inventory-carrying costs are the ones that most influence the performance of logistics systems, which account for 54.8% and 35.9% of the total logistics costs respectively (Council of Supply Chain Management Professionals, 2015). Transportation costs capture the cost of moving goods and inventory costs measure the cost of good storage. These two costs are required during the whole production and distribution processes, from manufacturing to delivery to the final consumers and returns. The trade-off between transportation and inventory costs was first recognized by Lewis *et al.* (1956) in a study of the application of air transportation. Because transportation costs and inventory costs are fundamentally important and visible, after that more and more related research has been conducted such as Blumenfeld *et al.* (1985), Horowitz and Daganzo (1986), and Daganzo and Newell (1993).

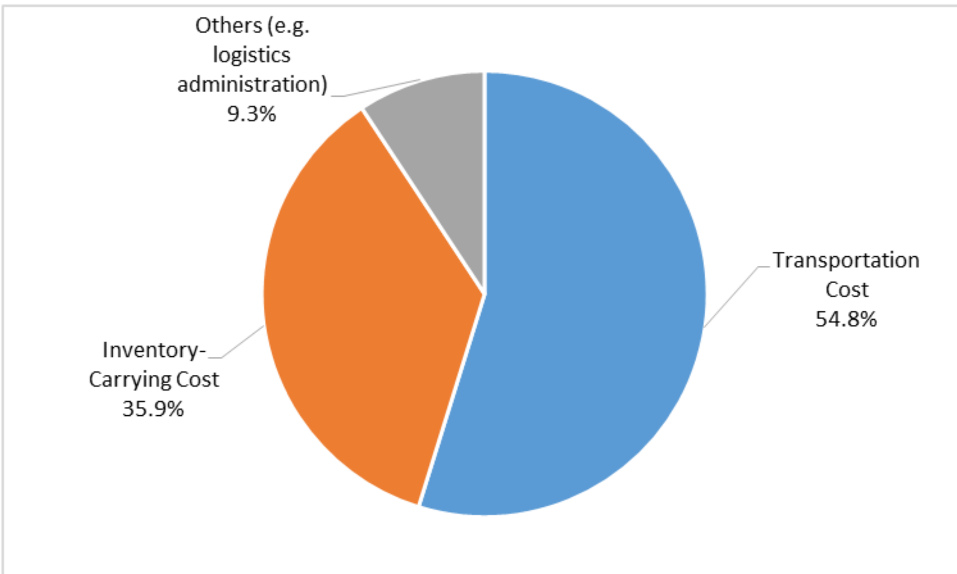


Figure 1-2. Components of logistics cost in 2014 in U.S.

Logistics operations can be categorized into two types – outbound logistics and inbound logistics. Outbound logistics refers to the transport, storage and delivery of goods from assembly plants or distribution centers to the end users (typically customers). Inbound logistics is a similar process but from suppliers to assembly/manufacturing plants. If we simplify the logistic model to four basic entities – plants, suppliers, customers, and distribution centers, Figure 1-3 shows the relationships between these two operations. Inbound and outbound operations integrate into the logistics systems, where supply chain managers seek to optimize the efficiency and reliability of distribution processes while minimizing transportation and inventory costs. Much logistics literature focuses on the outbound movements of products to customers. However, inbound logistics is also an important element of logistics operations, including a wide range of activities from supplier selection to the delivery of parts to assembly plants. Effectively managing the inbound moves of component parts from a large number of different suppliers is vital for controlling overall costs.

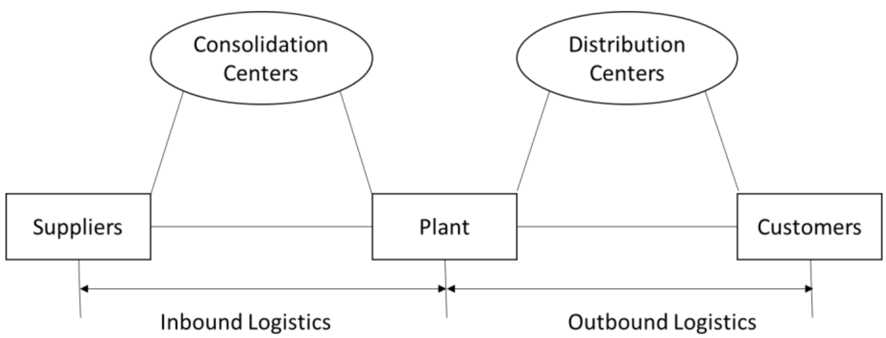


Figure 1-3. Two types of logistics operations – inbound and outbound

1.2 Logistics Network Planning

A logistics network involves many firms and locations (suppliers, customers, plants and warehouses), and determination of an efficient logistics network configuration is complicated. The optimal configuration will achieve the minimum cost of moving materials and products with all the available resources, through effective strategies addressing issues such as the location and the number of plants and distribution centers, allocation of customer demand, etc.

Optimization models have been successfully applied to these decision problems, and relevant research continues to seek more powerful methods. The logistics network models include a number of formulations ranging from linear models to non-linear ones, and from deterministic models to stochastic ones. Summaries of different studies dealing with the design problem of logistics networks can be found in the work of Vidal and Goetschalckx (1997), Beamon (1998), Erenguc *et al.* (1999), and Pontrandolfo and Okogbaa (1999). Improvements to logistics networks have often resulted in 5% to 15% savings in logistics costs (Kasilingam, 1998).

1.3 Problem Statement

An efficient inbound logistics network is particularly important for the automotive industry because it involves a huge number of auto parts suppliers. Being home to 13 auto manufacturers including General Motors (GM), Ford and Honda, the United States has one of the largest automotive markets in the world. In 2012 these suppliers combined to produce $225.2 billion in industry shipments, making up nearly 4 percent of total U.S. manufacturing.

Efficient design of routes and schedules for moving materials into manufacturing or assembly plants is a central element of inbound logistics operations. Materials must be moved from a relatively large number of suppliers in varying quantities, and different strategies are appropriate for different types of shipments. This dissertation focuses on major categories of suppliers, both of which are served by collection routes that originate at either a plant or a consolidation center, visit several suppliers to pick up material and then return to their origins.

The first supplier category includes suppliers for whom daily pickups may not be required, so the design of a collection system that determines routes and frequency of operations jointly is a central concern. Manufacturers have widely adopted just-in-time (JIT) delivery operations during the last 25 years in an attempt to reduce work-in-process inventory, and this has often resulted in decisions to have every supplier ship material every day, regardless of quantity. This strategy is very efficient at reducing inventory, but may create substantial increases in transportation cost. A focus of this dissertation is on expanding the scope of the system design to include both frequency of pick-up and routing. The frequency of pick-up for these shipments from individual suppliers is based on balancing transportation cost and inventory cost for the materials. The combination of spatial concerns (building good routes) and operating these routes at the best frequency is a crucial issue, and offers opportunities for significant cost reduction.

For suppliers that ship moderately large volumes to a single plant but not enough to fill a truck themselves, one commonly used process is to have plant-based collection routes that stop at multiple suppliers and return to the plant. These routes often operate daily. Such collection routes are widely used in the automotive industry, where they typically serve suppliers that ship between 10% and 70% of a truckload per day to a single assembly plant. Standard practice generally assigns these suppliers to routes so that only one route serves each supplier. However, there are opportunities for cost reduction by allowing some suppliers to be served by multiple routes, with each route picking up a portion of the total shipment. In the vehicle routing literature, this is called "split delivery" routing, although whether the routes are making deliveries or pickups is not critical to the problem definition. One of the problems with the existing solution approaches to the Split Delivery Vehicle Routing Problem (SDVRP) is that there are no controls on the number of routes among which a given supplier can be split, nor on the total number of suppliers that can have split service. A second major concern in the design of collection routes is that variation in production schedules at the destination plant can cause significant day-to-day or week-to-week variability in the load quantities to be picked up at individual suppliers. This dissertation creates a new approach to designing daily collection routes that allow split service with operational controls on the splitting and also incorporate uncertainty in pick up quantities at individual suppliers.

The dissertation builds on elements of traditional vehicle routing as well as broader elements of logistics planning. At the core of the process is a mathematical optimization termed capacitated clustering. The optimization model creates a set of P groups (routes), so that no group exceeds a certain size (vehicle capacity) and all suppliers' shipments are included. The objective of the clustering is to maximize the "effectiveness" of including suppliers within a group. Two different versions of the clustering model are used in this dissertation – one for building clusters that operate at different frequencies for collection of shipments, and the other for creating plant-based runs that operate on a daily basis, which incorporates split service and quantity uncertainty.

1.4 Outline of the Dissertation

The remainder of this dissertation is organized as follows. Chapter 2 presents the work on joint frequency-routing analysis. This work is inspired by the Inventory Routing Problem (IRP) and the Period Vehicle Routing Problem (PVRP), but formulates a new approach to designing inbound material collection route which considers pick-up frequency and spatial design as joint decisions to minimize total logistics (transportation plus inventory) cost and uses a clustering-based solution method. The clustering-based optimization uses an approximation to the actual cost of a routing solution without actual route construction. We show that the problem is analogous to a single-source fixed-charge facility location problem, and near-optimal solutions can be found using an efficient heuristic algorithm. Tests show the effectiveness of this model formulation and a case study demonstrates that substantial total cost savings can be achieved in realistic applications.

Chapter 3 describes how the extended model of the SDVRP is developed for plant-based collection routes. This extended model includes uncertainty in the load quantities at suppliers and controls (either penalties or constraints) designed to improve the “regularity” of service to individual suppliers. These extensions allow inbound logistics managers to design a set of collection routes that is more efficient than the traditional approach of assigning suppliers to single pick-up routes, more controllable than using the standard SDVRP formulation, and more robust under load variations, so that routes do not have to be redesigned as production schedules change. Solutions of large scale problem instances will be very computationally expensive;

therefore an efficient heuristic algorithm is developed. Applications to logistics operations in the automotive industry are used to demonstrate the effectiveness of this extended model and the proposed algorithm, and they are provided in chapter 4.

Chapter 5 contains the conclusions and the directions for continuing work.

CHAPTER 2

COMBINING SERVICE FREQUENCY AND VEHICLE ROUTING FOR MANAGING SUPPLIER SHIPMENTS

2.1 Introduction

Manufacturing or assembly plants often require inbound material movements in varying quantities from a relatively large number of suppliers. The material being shipped from different suppliers may also have quite different value per unit of weight or volume. In addition, some suppliers may provide material to multiple plants operated by the same manufacturer. This leads to complex challenges for organizing the inbound movements to minimize total logistics cost, including both transportation and inventory costs (Blumenfeld *et al*., 1985; Burns *et al*., 1985).

Collection routes operated either from a plant or from a consolidation center to visit several suppliers are commonly used in many different contexts, but generally the decision regarding frequency of pickup for individual suppliers is separated from the construction of collection routes, and often all suppliers are visited at a single common frequency. The focus of this chapter is on considering the frequency of service for individual suppliers jointly with the construction of collection routes operating at different frequencies, to determine a solution that minimizes total logistics cost. Daganzo (1985) and Hall (1985) made early conceptual contributions to this problem, and there is related work on the Period Vehicle Routing Problem (PVRP) and the

Inventory Routing Problem (IRP), although neither the PVRP nor the IRP are focused directly on the problem of interest here. The contribution of this chapter is to construct a formulation based on clustering suppliers in a combined frequency-location domain. This formulation allows very effective solution of large problem instances and creates opportunities for substantial reductions in total logistics cost, as compared to serving all suppliers at a common frequency.

In section 2.2, we provide a small, but realistic, example to illustrate the decisions and tradeoffs involved in the problem and set the stage for the analysis. Section 2.3 discusses previous related research to establish the context for the model described here. Section 2.4 contains the mathematical formulation as a mixed integer linear program. Although small instances of this problem can be solved exactly, heuristic solution methods that are scalable to large problem instances are very important. Section 2.5 describes a heuristic approach for obtaining good solutions quickly, even when there are many suppliers being served. The heuristic solution is based on recognizing that the problem formulation is analogous to a location problem that has been studied by several previous authors. The effectiveness of the problem formulation and the heuristic solution is evaluated in section 2.6, using a set of test problems. Section 2.7 describes a case study, using data from a real system, to illustrate the magnitude of potential cost savings from the joint consideration of service frequency and collection routing. Section 2.8 concludes and discusses directions for further research.

2.2 An Illustrative Situation

To illustrate the problem of interest in a practical situation, we consider a set of 13 suppliers located in Indiana, Ohio and Kentucky that make shipments to a facility in Dayton, Ohio. In the practical situation that motivates this research, the Dayton facility is a consolidation center, and the material being moved is ultimately destined for several manufacturing plants, but our concern is only with the collection operation from the suppliers and delivery to a single location. The model is not dependent on whether that facility is a consolidation center or a plant.

The locations of the suppliers and the Dayton destination facility (small black triangle) are shown in Figure 2-1. These suppliers operate five days per week and have weekly pick-up quantities (shown beside the nodes in Figure 2-1) that range from 235 ft^3 to 8,555 ft^3. The total volume for all 13 suppliers is 27,580 ft^3 per week.

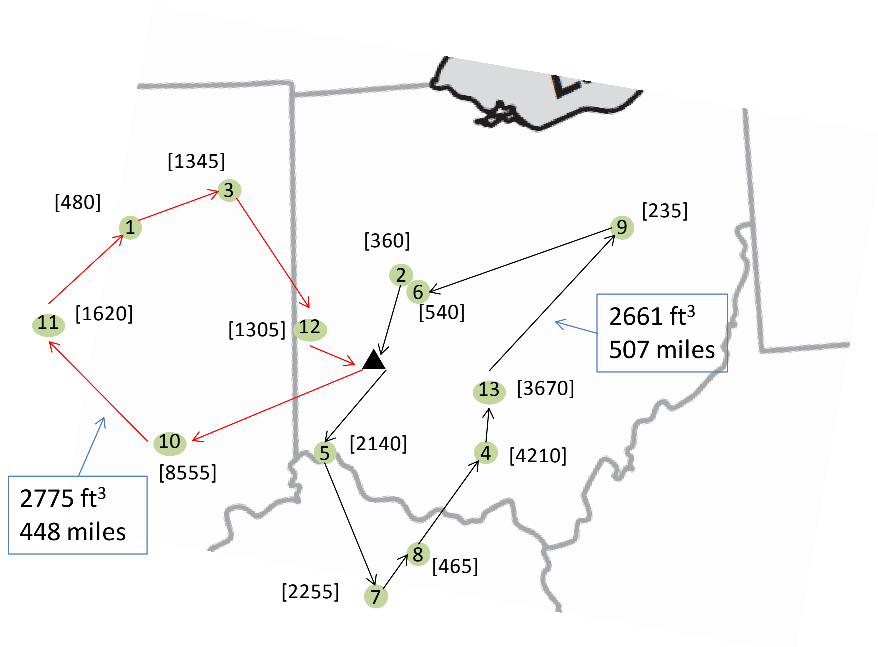


Figure 2-1. 13-supplier example with all-daily service.

The current operation of this system is that all suppliers are visited daily on routes that have their origins and destinations at the Dayton facility. Each route is operated by a truck with 3,000 ft^3 capacity. The strategy of visiting all suppliers every day minimizes on-hand inventory of materials within the system and makes it relatively easy to load uniform daily shipments from the consolidation center to several manufacturing plants, but at a cost of making many visits to suppliers (some of which are for picking up very small quantities of material) and operating a large number of total truck-miles. The all-daily solution includes two routes, as shown in Figure 2-1, totaling 4,775 truck-miles per week. If an average cost of $1.50 per mile is specified for the truck operating cost, the total transportation cost is $7,163 per week.

The average inventory (including both material at the suppliers and material at the destination) for all-daily shipments is the daily quantity moved, 5,436 ft^3 (see Hall, 1985). If this material has an average value of $50/ft^3, and the cost of inventory investment is 15% per year (0.0029 per week), the cost of this inventory is $775 per week, and the total weekly logistics cost is then $7,911.

By considering the spatial design of routes and the frequency of operating those routes as joint decisions, total costs can be reduced. The purpose of the model developed here is to allow explicit tradeoff of vehicle mileage costs and inventory costs by choosing to visit some suppliers less frequently and organizing some routes that operate at those lower frequencies. For example, if we consider three potential pick-up frequencies: daily, twice/week, and weekly, we can construct the solution shown in Figure 2-2, which includes six routes (1 daily, 2 twice/week, and 3 weekly). The total truck-miles operated per week is reduced to 2,837, a 41% reduction. Because some suppliers are visited less frequently, the average amount of inventory in the system increases, and inventory costs increase to $1,967 per week. The total logistics cost of $6,223, however, represents a reduction of 21% from the all-daily strategy.

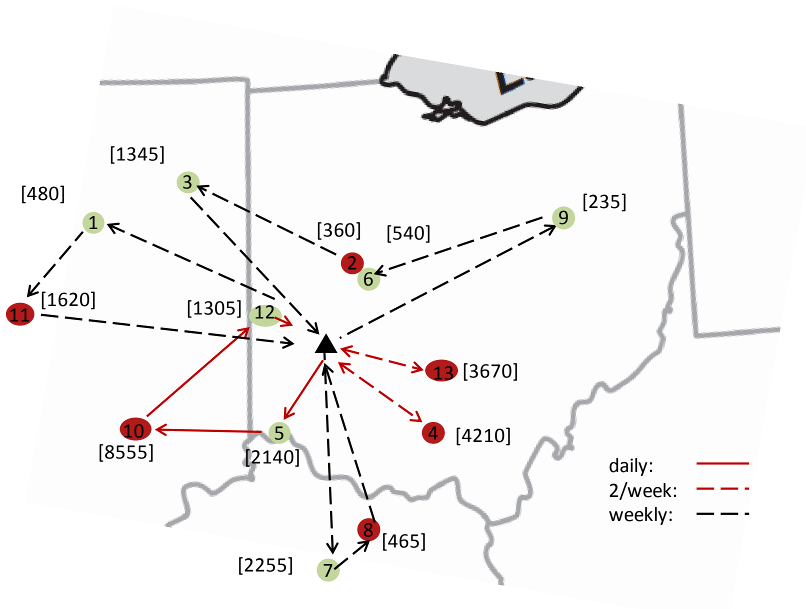


Figure 2-2. 13-supplier example with jointly determined frequency and routing.

As illustrated by this small (but real) example, joint determination of pick-up frequency and routes offers the possibility of very substantial overall cost savings. A model for finding solutions of the character shown in Figure 2-2 can allow these potential cost savings to be realized.

2.3 Previous Related Research

Daganzo (1985) and Hall (1985) both considered an abstract version of the problem, where a set of N suppliers distributed uniformly within an area of size A produce material in varying quantities and with varying value per pound and supply a single location (e.g., assembly plant). Both derived results for frequency of service that minimized total transportation and inventory cost. They used two different assumptions about how suppliers could be separated into frequency classes. Daganzo constructed groups of suppliers based on total value of material shipped per unit time, and then assumed that each supplier class would be served separately at a desired frequency. Hall assumed that suppliers would be served at some multiple of a base vehicle dispatch frequency, with each vehicle dispatch serving a subset of suppliers due for pickup at that time. These efforts are important for their integrated treatment of inventory and transportation costs and for emphasizing the connections between determining appropriate frequencies of service for individual suppliers and determining efficient routes for the transportation of the material to the destination. However, neither approach dealt with the details of actual locations of suppliers or actual route construction.

Russell and Igo (1979) considered a problem they termed “assignment routing” where individual customers must be visited a known number of times per week (e.g., for refuse collection) and the challenge was to assign them to specific days so that the collection of individual day vehicle routing problems had overall minimum transportation cost and can be operated with fixed fleet of available vehicles. Christofides and Beasley (1984) defined a similar problem (the period vehicle routing problem – PVRP) focused on assigning individual customers to specific combinations of days within a period (e.g., Monday-Thursday, Tuesday-Friday, or Wednesday-Saturday). Both of these efforts assumed that the frequency of service for each customer was known a priori, and they focused attention on the individual day routing problems and approximations to tour lengths that could be evaluated easily to assess the effects of changing customer assignments.

Francis and Smilowitz (2006), Francis *et al*. (2006) and Francis *et al.* (2008)

extended the PVRP to make the frequency of service to customers a decision within the model and termed this the service choice version of the problem (PVRP-SC). Customers were assumed to have some given minimum frequency of service, but can be assigned to a route that operated more frequently if such an assignment reduced overall costs. One of the innovations in the PVRP-SC formulation was the inclusion of a benefit term for more frequent service at a given customer. This term may reflect a variety of advantages for increasing service frequency, but certainly one possibility was the reduction in inventory holding costs. A recent review of a wide variety of work on the PVRP was provided by Campbell and Wilson (2014).

The inventory routing problem (IRP) is another similar problem that is generally concerned with supplying a set of customers who use one or more products at varying rates and incorporating those customers into delivery routes at sufficient frequency to avoid stockouts while minimizing transportation costs for the delivery vehicles. In the IRP, the amount of material to be delivered at each customer is a decision within the model. This problem is important for vendor-managed inventories in distribution systems and has been studied by numerous authors. Good summaries of this problem and solution methods were provided by Moin and Salhi (2007), Bertazzi *et al.* (2008) and Bertazzi and Speranza (2012). Coelho *et al.* (2014) provided a comprehensive review of work on the IRP.

Rudiansyah and Tsao (2005) analyzed a problem of restocking vending machines in a vendor-managed inventory system. They created a formulation that contained elements of both the IRP and the PVRP models with consideration of allowable time windows for deliveries. Their approach was based on the PVRP but with the quantities to be delivered considered as decision variables, as in the IRP.

Chuah and Yingling (2005) dealt with a JIT delivery system to an automotive assembly plant that involved finding routes and pickup frequencies from suppliers. Their focus was on suppliers that may be visited multiple times per day and accordingly they placed considerable emphasis on finding routes that visit suppliers within given time windows during the day and adhere to a cap on total inventory. That setting was somewhat similar to the problem considered here, but the focus was on operations within time windows during a day and they did not treat inventory costs directly. That created important differences from the problem illustrated in the previous section.

An important aspect of the PVRP, PVRP-SC and IRP formulations is that they focus quite strongly on the decision to assign a customer to a specific route operated on a specific day during a given period. This places the vehicle routing problems on individual days at the center of the model and the solution process. In the model constructed here, a different approach has been taken that focuses primarily on creating clusters of suppliers (defined both spatially and with respect to pick-up frequency). Construction of actual vehicle routes for specific days is then done as a post-processing step. In the practical applications the authors have seen, collection routes typically have a small number of supplier stops (fewer than 10), so this “cluster first, route second” approach is quite attractive. This has led us to formulate the model as described in detail in section 2.4.

2.4 Problem Formulation

We consider a set of suppliers $i \in I$, and a set of possible frequencies indexed by f. The frequency of service associated with index f is defined as F_f. For the context of interest here (suppliers serving a manufacturing operation), a regular weekly cycle is useful and we can define frequency of service as the number of visits per week. A specific collection route will operate at a selected frequency and include some subset of the suppliers. The combination of frequency and spatial grouping into collection routes focuses attention on the total cost (inventory and transportation) of a solution. A route is built around a combination of a supplier location and a frequency selected (endogenously) as a seed. The model uses the decision variables:

y_{jf} = 1 if supplier j served at frequency index f is selected as a seed point; 0 if not

x_{ijf} = 1 if supplier i is part of a route whose seed is a jf combination; 0 if not.

The parameters of the problem are:

c_{ijf} = marginal cost for serving supplier i on a route with seed point j, operating at frequency index f

c'_{jf} = cost for establishing a route with seed point j, operating at frequency index f

q_i = weekly load to be picked up from supplier i

F_f = frequency (times/week) associated with frequency index f

$V \quad = \quad$ vehicle capacity.

The model is formulated as optimization problem (P2-1), which determines the number of routes to be operated, the frequency for each, and the assignment of suppliers to routes:

$$\text{(P2-1)} \qquad \min \quad \sum_{i,j,f} c_{ijf}\, x_{ijf} + \sum_{j,f} c'_{jf}\, y_{jf} \tag{1}$$

$$\text{s.t.} \qquad \sum_{j,f} x_{ijf} = 1 \quad \textit{for all } i \tag{2}$$

$$\sum_{i} q_i\, x_{ijf} - F_f\, V\, y_{jf} \le 0 \quad \textit{for all } j, f \tag{3}$$

$$x_{ijf} \in (0,1) \quad \textit{for all } i, j, f \tag{4}$$

$$y_{jf} \in (0,1) \quad \textit{for all } j, f \tag{5}$$

Constraint (2) ensures that all suppliers are assigned to some route. Constraint (3) ensures that no assignments are made to routes that do not exist and that the vehicle capacity is respected.

The cost coefficients c_{ijf} in the objective function (1) measure the incremental (marginal) costs of assigning supplier *i* to a seed point specified by location *j* operating at frequency index *f*. Part of this cost is associated with additional distance traveled by the truck operating the route in order to visit supplier *i*. Precise distances depend on the routes constructed, which will in turn depend on the clusters to be formed, so exact determination of the marginal costs is solution-dependent. The purpose of the c_{ijf} values in (1) is to approximate the actual costs in a way that can be specified easily and still lead to near-optimal solutions. If the cost/mile for operating a truck is *g*, and d_{ab} is the distance from point *a* to point *b*, then the incremental transportation cost per week of the assignment x_{ijf} is approximated by $gF_f\left(d_{0i}+\theta\, d_{ij}-d_{0j}\right)$. The parameter θ is an empirical adjustment to allow the approximation to better reflect actual incremental distances. Use of this type of adjustment was first suggested by Yellow (1970). If supplier *i* is itself a seed, the incremental distance is clearly 0.

The second part of the cost coefficients reflects the inventory costs for supplier *i*. If the route operates at frequency F_f, the average shipment size is $\frac{q_i}{F_f}$ and the weekly inventory cost (for average material value *P* and weekly cost of money *R*) is $\frac{RP}{F_f}q_i$. Thus the overall cost coefficients are:

$$c_{ijf} = gF_f\left(d_{0i}+\theta d_{ij}-d_{0j}\right)+\frac{RPq_i}{F_f} \tag{6}$$

The cost coefficients associated with seed selection are based on the round-trip distance from the depot and the frequency of service:

$$c'_{jf} = gF_f\left(2d_{0j}\right) \tag{7}$$

The objective function thus approximates the total truck operating cost plus weekly inventory cost for the solution. The effectiveness of this approximation is evaluated using a set of test problems in section 2.6.

If desired, a simple modification can be made to the c_{ijf} coefficients to incorporate a cost/stop term for pickups at suppliers. If the cost of making a stop at supplier i is e_i and supplier i is assigned to a route that operates with frequency F_f, then a term e_iF_f can be added to c_{ijf} in (6). This makes higher frequency service more costly and encourages more suppliers to be assigned to lower frequency routes.

By denoting a *jf* combination by a single index, *m*, it becomes clear that problem (P2-1) has the same structure as the single-source capacitated facility location problem (SSCFLP). The selection of a seed is analogous to choosing a site in the SSCFLP, and the clustering of suppliers to be served by a route operating at a specific frequency is analogous to determining customers to be served by each facility. This analogy can be exploited because there has been substantial attention given to developing both exact and heuristic solution methods for the SSCFLP (e.g., Hindi and Pienkosz 1999; Holmberg *et al.*, 1999; Holt *et al.*, 1999; Ahuja *et al.*, 2004).

One practical difference between an instance of (P2-1) and a more typical instance of the SSCFLP is that there will normally be many more *jf* combinations than suppliers. In the parlance of location problems, there are many more potential facility sites than there are demand points. This is in contrast to most location applications where the set of potential facility sites is usually much smaller than the set of demand points. This does not affect the structure of the problem, but it does mean that the solution space in the route-frequency application is much larger than in usual location applications.

2.5 Solution of the Optimization Problem

The SSCFLP, to which problem (P2-1) is analogous, is known to be NP-Hard (Ahuja *et al.*, 2004). For small numbers of suppliers and frequency classes, it is possible to use standard integer programming methods to solve the problem. However, as the number of suppliers and/or frequency classes increases, the size of the problem grows very quickly and heuristic solutions are necessary. For example, for 75 suppliers and three frequency classes, the problem has nearly 17,000 integer variables. This implies significant difficulty for direct solution of realistic problems using general integer programming software.

Furthermore, the objective function (eq. 1) is an approximation to the real cost of operating the routes (because we have not explicitly evaluated the sequence of stops on each route), so expending a great deal of computational effort to solve this problem exactly is not likely to be justified. If we can demonstrate that an approximate solution to (P2-1) is capable of creating solutions that achieve near-minimal total cost (after routes operating at different frequencies are constructed), an effective approach to solution for large problem instances is to construct a near-optimal set of seeds and supplier assignments. Then actual routes can be constructed and the total transportation and inventory cost can be computed as a post-processing step. That is the approach followed here. First, we describe the process of finding good approximate solutions to problem (P2-1), using a method developed for the SSCFLP. Once that solution is available, route construction is done using a straightforward traveling salesman algorithm, and total costs are computed. In section 2.6, this approach is evaluated on a battery of test problems.

Very good approximate solutions to the SSCFLP can be obtained using the multi-exchange heuristic developed by Ahuja *et al*. (2004). The algorithm uses a Very Large-Scale Neighborhood (VLSN) technique, alternating between supplier exchanges and route changes (analogous to facility moves). The neighborhood searches defined by supplier exchanges and route changes are embedded in a local improvement algorithm. Restart mechanisms are also included. Our intent here is not to repeat the careful description of the algorithm and its implementation that is contained in Ahuja *et al*. (2004), but to describe briefly the major elements of the solution process and how it translates from the original facility location context to the route-frequency application. The reader interested in more details on the algorithm is referred to the original article.

In the context of problem (P2-1), a supplier neighborhood of a current feasible solution is defined as a set of new solutions obtainable by exchanging suppliers among already chosen frequency-seed combinations. One or more suppliers served by *jf* combination n_1 is moved to another selected combination n_2. Then to maintain load feasibility, another subset of suppliers is moved from *jf* combination n_2 to n_3, etc. If the end of the exchange sequence is a subset of suppliers being moved from some *jf* combination n_p to combination n_1, the exchange is *cyclic*. However, if route-frequency combination n_p can accept the new suppliers it receives without violating its capacity, the exchange sequence may terminate there, and be called a *path* exchange. In general, there will be a very large number of possible exchanges from a given feasible solution, and the algorithm limits the search space to K exchanges, where K is a user-selected parameter. In the computational tests discussed in section 2.6, the value of K used has been set based on problem size, but does not exceed 3 even in large problems. This is sufficient to produce very good solutions. The neighborhood created by supplier exchanges is searched by looking for improving solutions in a dynamically generated graph. The details of generating the graphs and selecting good exchanges to evaluate are described in the original article by Ahuja *et al.* (2004). We will not repeat all these details here, but note that we have selected a breadth-first search rule for seeking good exchanges.

The second type of neighborhood structure used in the algorithm is a route neighborhood. This is a set of solutions obtained either by adding a new route seed-frequency combination, eliminating a current combination, or transferring a current jf combination to another. Changes in the set of route-frequency combinations selected must retain feasibility for the assignment of suppliers, and a local search tries different types of changes in succession, looking for a cost-reducing change. If the change is adding a new route, partial reassignment of suppliers that can be served more cheaply by the new route is accomplished by solving a knapsack problem. If that results in a cost reduction, the change is accepted. If not, a complete reassignment is attempted, using a generalized assignment procedure.

Elimination of existing jf combinations can be effective if the fixed costs associated with a route (c'_{jf}) exceed the benefits of assigning suppliers to that route rather than alternatives. This is evaluated by trying to reassign all suppliers served by the route in question to the other remaining routes. If that does not result in a cost reduction, a further attempt is made to reassign all suppliers to the remaining jf

combinations, using the same generalized assignment procedure used for evaluating route additions.

Transferring one *jf* combination to another may involve a change in frequency, a change in seed location, or both. If the change is in seed location only, the capacity of the route remains the same and the assignment of suppliers is still feasible. Evaluating whether the change reduces costs is quite easy. If the transfer involves a change of frequency, the route capacity changes and a partial or full reassignment of suppliers may be necessary. If it is, this is accomplished via the generalized assignment procedure, and then it is determined whether or not the change reduces costs.

A multistart mechanism is used with multiple initial feasible solutions to avoid having the overall search process be trapped in a local optimum. This mechanism uses the Lagrangean relaxation procedure described by Ahuja *et al*. (2004), based on relaxing the assignment constraints (2) in problem (P2-1). The number of restarts is a user-specified parameter. We have experimented with various numbers of restarts, and found that selecting a value in the range 12-15 produces reliable final results on the test problems we've used.

The implementation of the algorithm used here follows the lead of Ahuja, *et al*. (2004) very closely, except that the interpretation of what they consider to be a "facility move" is quite different in the current application. Implementation of the algorithm has been done in a Windows environment on a desktop computer with a 3.00 GHz processor and 3.00 GB RAM, using C++.

2.6 Computational Experiments

If the problem formulation (P2-1) and the heuristic solution method described in section 2.5 are to have practical usefulness, it is necessary to establish that the objective function defined for the problem is a close approximation to the actual logistics cost of a set of pickup routes in a variety of situations, and that the adoption of the VLSN search algorithm for facility location problems can also identify near-optimal solutions to the frequency-routing problem (P2-1). The purpose of the tests described in this section is to meet those requirements.

A battery of test problems has been created by systematically varying three important problem characteristics:

1) The number of suppliers to be served;
2) The location of the depot (plant or consolidation center) with respect to the suppliers; and
3) The size distribution of the loads to be picked up from the suppliers.

For all the test problems, suppliers are assumed to be randomly distributed within a square area 300 miles on a side. To create specific test cases, we have selected problems with 22, 36, 54, 75 and 100 suppliers from the data sets labeled Augerat, *et al*. A and P, available through the COIN-OR web site (COIN-OR Foundation, 2012).

For each of the five problem sizes, we test three different locations of the depot, one in the upper-left part of the service region, the second near the center, and the third in the lower-right. The location of the depot relative to the set of suppliers affects the overall length and shape of the routes constructed, and we have observed in practice that the depot is not always near the center of the set of supplier locations served.

The third factor in the experiments is the size distribution of supplier material quantities. In practice, it is common to see a roughly exponential distribution, with many small shippers and a few larger ones. Figure 2-3, for example, shows the weekly shipment quantities from 75 suppliers for a manufacturer served by a consolidation center in Ohio. This particular set of suppliers is the case study to be described in more detail in section 2.7.

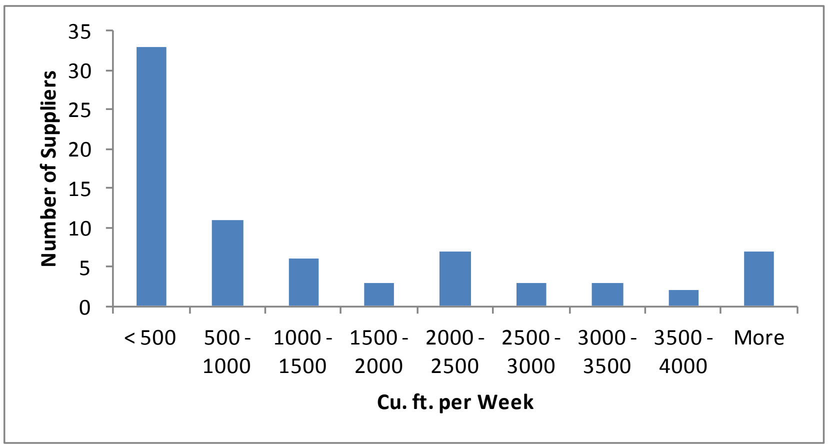


Figure 2-3. Example of supplier weekly load distribution.

In light of this, we have tested two different shifted exponential distributions of supplier quantities, with different averages. Each distribution has a minimum of 200 ft^3 per week. The first has an overall average quantity of 1,500 ft^3 per week, and the second has an average of 750 ft^3 per week. Clearly, the smaller average quantity distribution will result in solutions that use fewer routes and have more stops per route. To test the effect of the distribution of quantities, we have also included a third distribution with individual supplier quantities uniformly distributed between 500 ft^3per week and 2,500 ft^3 per week. This produces an average quantity of 1,500 ft^3 per week, identical to the first exponential distribution, but with fewer small pickup quantities and more large pickups. This is likely to have an effect on route construction in the solutions.

In total, variation of the three factors creates 45 individual test cases representing all the combinations. For each test case, we find a solution using the VLSN heuristic and an optimal solution to problem (P2-1) using LINGO (version 12.0), if possible. Actual routes are created for each solution and total actual costs are computed. The actual cost can be compared with the objective function value from LINGO for the cases where optimal solution of (P2-1) is possible, as well as with the objective function values from the VLSN solutions. These comparisons provide important data about how closely the objective function (eq. 1) approximates actual costs for various solutions, and about how close the VLSN solutions are to the optimal solutions for those cases where optimal solutions can be obtained.

Optimal solutions can be generated using LINGO for all the test cases with 22 and 36 suppliers, and for all but one of the tests with 54 suppliers. For one of the 54-supplier test cases, no optimal solution is available after 200 hours of computation time and the run is aborted. Summary characteristics of the solutions are shown in Table 2-1. For the smallest test cases (22 suppliers), it is clear that solution of problem (P2-1) using commercial software is quite reasonable, with solution times under a minute. However, as the number of suppliers increases, solution times grow very rapidly, and are also highly variable depending on specific problem characteristics. For the 36-supplier test cases, the average solution time is approximately 3,000 seconds (50 mins.), but the times range from less than 1 minute to nearly 4 hours. In general, solution times are smaller when the supplier pickup quantities are smaller (Quantity Distribution 2) and when the depot is near the center of the suppliers (Location 2).

In the 54-supplier test cases, the average solution time for the eight cases that could be solved is 4.6 hours, but ranges from about 5 minutes to more than 14 hours. For test purposes, it is useful to achieve those solutions, but for practical use of the model, the 54-supplier problems are likely to be too large for effective use of commercial MILP software. In general, the practical limit on problem size for using exact solution methods is probably somewhere near the 36-supplier problems. Use of other software packages or faster processors may move the boundary a little, but probably not as high as 50 suppliers.

The results of the objective function evaluations in Table 2-1 provide important confirmation that the function in eq. (1) is a very good approximation of actual transportation and inventory cost. The actual costs of the solutions (after route construction) are within 1% of the objective function value in 14 of the 26 cases, and within 2% in 23 of the 26 cases. The ability of the function in eq. (1) to closely approximate actual costs means that solving problem (P2-1) is an effective way of obtaining solutions that are near-optimal for actual operations, and this implies that if good solutions to (P2-1) can be obtained quickly, these solutions can be very useful in practice.

Table 2-1. Summary of results from LINGO solutions on three smallest problem sets.

Suppliers	Instance	Quantity Distribution	Depot Location	LINGO Obj. Fcn. ($/week)	Actual Cost ($/week)	Difference (%)	Computation Time (s)
	1	1	1	5694	5733	-0.67	11
	2	1	2	5179	5081	1.93	11
	3	1	3	6343	6276	1.08	54
	4	2	1	3692	3764	-1.91	9
22	5	2	2	3292	3311	-0.59	13
	6	2	3	3813	3834	-0.54	13
	7	3	1	5819	5798	0.37	58
	8	3	2	5182	5238	-1.07	7
	9	3	3	6418	6459	-0.64	37
	1	1	1	16130	16100	0.18	14288
	2	1	2	13142	13174	-0.24	358
	3	1	3	15813	15809	0.03	3719
	4	2	1	9685	9742	-0.58	257
36	5	2	2	8492	8456	0.43	97
	6	2	3	9470	9723	-2.60	57
	7	3	1	17022	17159	-0.79	3764
	8	3	2	13376	13376	-0.00	229
	9	3	3	15808	15722	0.55	4244
	1	1	1	15709	15948	-1.50	37560
	2	1	2	13328	13462	-1.00	22800
	3	1	3	16977	17185	-1.21	50460
	4	2	1	9788	9938	-1.51	1319
54	5	2	2	8436	8618	-2.11	314
	6	2	3	9820	9892	-0.73	14409
	7	3	1	16283	17004	-4.24	1342
	8	3	2	13828	14026	-1.41	4620
	9	3	3	NA	NA	NA	NA

Table 2-2 compares the objective function values for the optimal solutions found using LINGO with the solutions found using the VLSN heuristic. Computation times are also listed for comparison. In 11 of the 26 cases for which optimal solutions are available, the VLSN heuristic identifies the same solution, and in 23 of the cases the difference in the objective function values is less than 1%. The largest difference is

2.8% (54 suppliers, instance 7). These excellent heuristic solutions are obtained in a tiny fraction of the computation time required for the optimal solutions. For example, in the 36-supplier test cases, the computation time of the heuristic averages 0.8% of the time required for optimal solution. The heuristic solution times vary with the problem characteristics, but are much less variable than the times required for optimal solutions.

Table 2-2. Comparison of optimal solutions and heuristic solutions.

Suppliers	Instance	Quantity Distribution	Depot Location	LINGO Obj. fcn. ($/week)	LINGO Computation Time (s)	VLSN Obj. fcn. ($/week)	VLSN Computation Time (s)
	1	1	1	5694	11	5745	
	2	1	2	5179	11	5206	
	3	1	3	6343	54	6343	
	4	2	1	3692	9	3703	All times
22	5	2	2	3292	13	3292	< 1 second
	6	2	3	3813	13	3820	
	7	3	1	5819	58	5819	
	8	3	2	5182	7	5190	
	9	3	3	6418	37	6536	
	1	1	1	16130	14288	16130	15
	2	1	2	13142	358	13142	4
	3	1	3	15813	3719	15828	9
	4	2	1	9685	257	9685	2
36	5	2	2	8492	97	8492	3
	6	2	3	9470	57	9470	1
	7	3	1	17022	3764	17274	8
	8	3	2	13376	229	13376	2
	9	3	3	15808	4244	15808	21
	1	1	1	15709	37560	15718	24
	2	1	2	13328	22800	13442	16
	3	1	3	16977	50460	17069	23
	4	2	1	9788	1319	9842	108
54	5	2	2	8436	314	8436	20
	6	2	3	9820	14409	9909	51
	7	3	1	16283	1342	16739	35
	8	3	2	13828	4620	13839	44
	9	3	3	NA	NA	17187	64

For the two largest sets of test cases (75 suppliers and 100 suppliers), no optimal solutions could be obtained within 200 hours of computation time, but the VLSN heuristic reliably obtains solutions in less than 2 minutes. Table 2-3 summarizes the computational results. It is particularly noteworthy that the average computation times for the 54-supplier problems (in Table 2-2), the 75-supplier problems and the 100-supplier problems are 36 seconds, 38 seconds, and 41 seconds, respectively. The computation times increase only slightly for the larger problems, indicating that the heuristic is likely to be useful in a wide variety of practical situations, even if the size of the problem were to increase beyond 100 suppliers.

Table 2-3. Summary of heuristic solutions for larger test cases.

Suppliers	Instance	Quantity Distribution	Depot Location	VLSN Obj. fcn. ($/week)	Computation Time (s)
75	1	1	1	23974	51
	2	1	2	18663	28
	3	1	3	22887	49
	4	2	1	14297	75
	5	2	2	11806	72
	6	2	3	13688	12
	7	3	1	24934	26
	8	3	2	19590	17
	9	3	3	23908	10
100	1	1	1	29202	27
	2	1	2	27816	58
	3	1	3	33643	63
	4	2	1	17436	40
	5	2	2	16918	61
	6	2	3	19240	48
	7	3	1	29496	34
	8	3	2	28419	8
	9	3	3	35427	28

This set of computational tests confirms that the objective function in problem (P2-1) is an accurate approximation to the actual transportation and inventory costs of a solution, and that the VLSN heuristic can produce excellent approximate solutions to problem (P2-1) in modest amounts of computation time. Thus, there is every reason to believe that the approach to combined routing and frequency determination developed here will have important practical applications and usefulness. The following section describes a case study that illustrates the practical value of the approach.

2.7 A Case Study

As a case study in the application of the model and solution process, we use the same destination facility in Dayton, Ohio, used for the small-scale illustration in section 2.2. However, for the case study we use the full set of 75 suppliers that serve this facility (from across Indiana, Ohio, Kentucky and western Pennsylvania), rather than the small subset used in the earlier illustration. The total weekly pickup quantity is 109,920 ft^3, with individual suppliers ranging from 10 ft^3 per week to 8,555 ft^3 per week. The Appendix gives the locations (by Zip Code) and weekly pickup quantities for the suppliers. The average single supplier quantity is 1,466 ft^3 per week, or about one-half of a 3,000 ft^3 truckload. A histogram of the quantities is shown in Figure 2-3, indicating the general pattern of many small suppliers and a few large suppliers. All suppliers operate five days per week and if daily service were provided for all suppliers, the minimum number of routes required is 8.

As in the small example in section 2.2, the potential frequencies considered are daily, twice/week and weekly, the average part value is assumed to be \$50/ft^3, the carrying cost of inventory is 0.288% per week, and trucks cost \$1.50 per mile to operate. No cost is associated with making a stop (i.e., $e_i = 0$ for all suppliers).

If daily service is provided to all suppliers (the current practice at the manufacturer in question), the total distance traveled on the 8 routes is 13,895 miles per week, implying a total transportation cost of $20,843 per week. The weekly inventory cost is $3,166, for a total cost of $24,009 per week.

The experiments in section 2.6 indicate that finding a solution to model (P2-1) using commercial MILP software when there are 75 suppliers is often not computationally practical, but an optimal solution to this particular problem is obtained using LINGO in a little less than an hour of computation time. That solution uses 12 routes (5 operating daily and 7 twice/week). The objective function value is $18,444. By constructing actual routes using the clusters in the LINGO solution, we determine that the total weekly mileage is 8,917, implying total transportation cost of $13,376. The weekly inventory cost is $4,881, and the total cost is $18,257 per week, a 24% reduction from the all-daily solution. The actual cost of the solution after route construction is approximately 1% different from the objective function value for problem (P2-1).

Solving the problem using the VLSN search algorithm also produces a solution with 12 routes (5 operating daily and 7 twice/week). Seven of the selected seed points are different from the optimal solution, but the assignment of suppliers to routes is nearly the same as in the LINGO solution. The reported objective function value from the VLSN algorithm is $18,462, approximately 0.1% above the optimal objective function value, and the actual cost of the solution is $18,302, approximately 0.2% above the actual cost of the optimal solution. Obtaining this solution requires 37 seconds on the same computer used for the LINGO solution. By using the formulation in (P2-1), it is possible to obtain dramatic savings in overall logistics costs relative to the all-daily solution (24% in this case), and the heuristic search algorithm can find a near-optimal solution very quickly.

Figure 2-4 illustrates the daily routes, which are the same in both the optimal and the VLSN solutions. 29 of the 75 suppliers receive daily pickups on these five routes. The supplier nodes shown in red are the seeds chosen in the optimal solution. Although the VLSN algorithm chooses two seeds differently (supplier 56 instead of 23, and supplier 17 instead of 5), the assignment of suppliers to routes is the same in both solutions, so the constructed routes are identical. The average load on these five routes is 2,815 ft^3, 94% of available capacity.

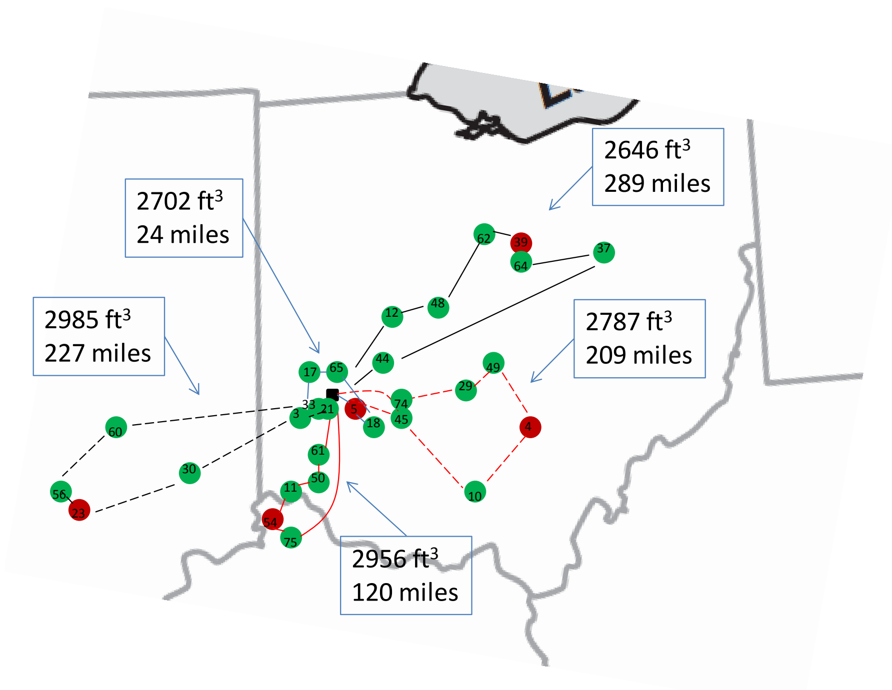


Figure 2-4. Comparison of optimal solutions and heuristic solutions.

Figures 2-5 and 2-6 show the twice/week routes, divided into two subsets for clarity in the maps. Figure 2-5 includes five routes that are identical in both solutions, serving 34 suppliers. In Figure 2-6, part (a) shows the two remaining routes from the optimal solution and part (b) shows the comparable routes from the VLSN algorithm. The differences between the two solutions are confined to the two routes that extend westward from the Dayton facility into Indiana. The VLSN search algorithm chooses different seeds for these two routes and the supplier assignments differ slightly. The total length of the two routes is 15 miles longer in the heuristic solution, and this accounts for the overall cost difference between the heuristic solution and the optimal solution.

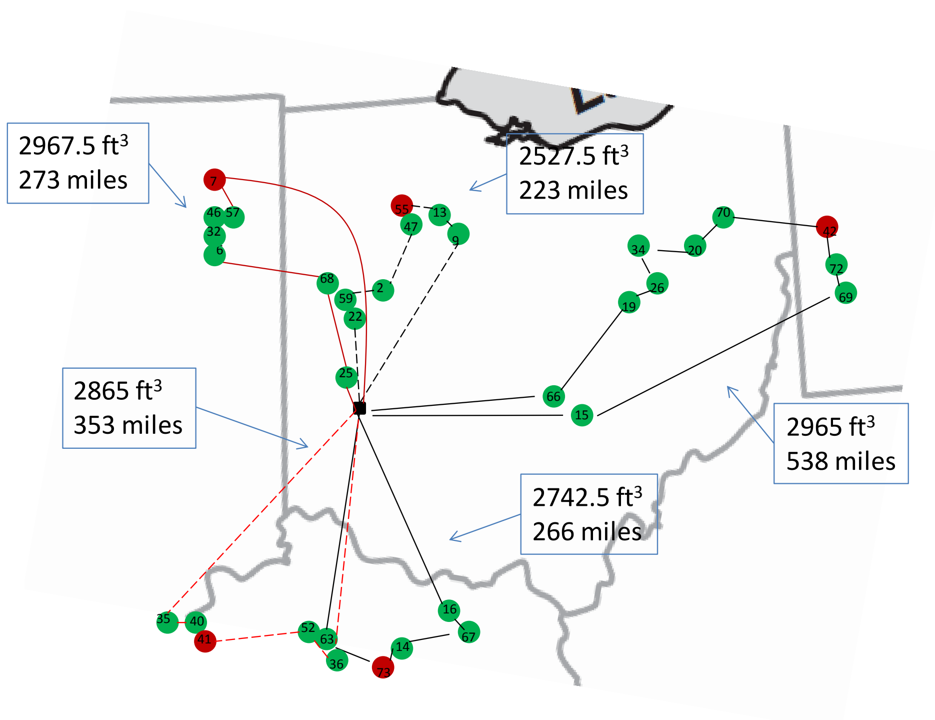
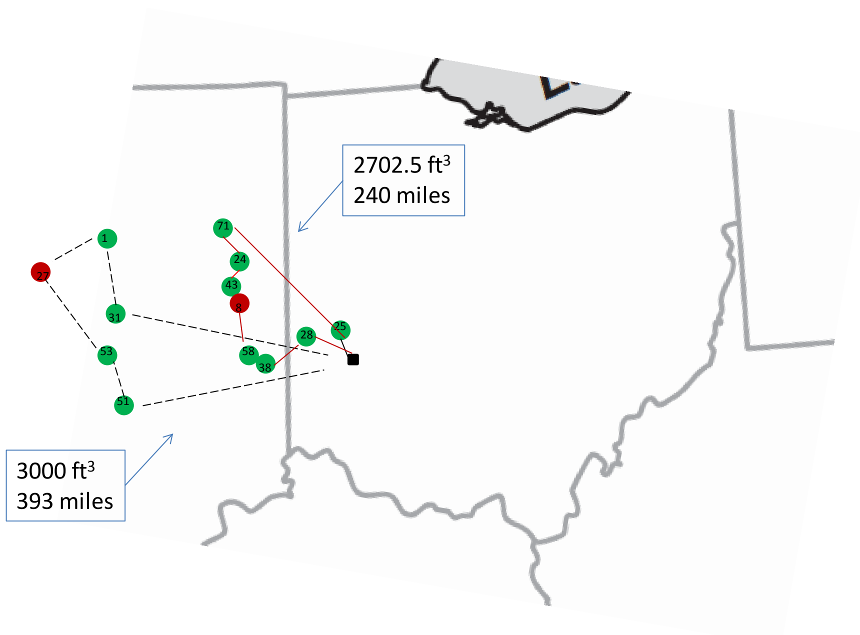
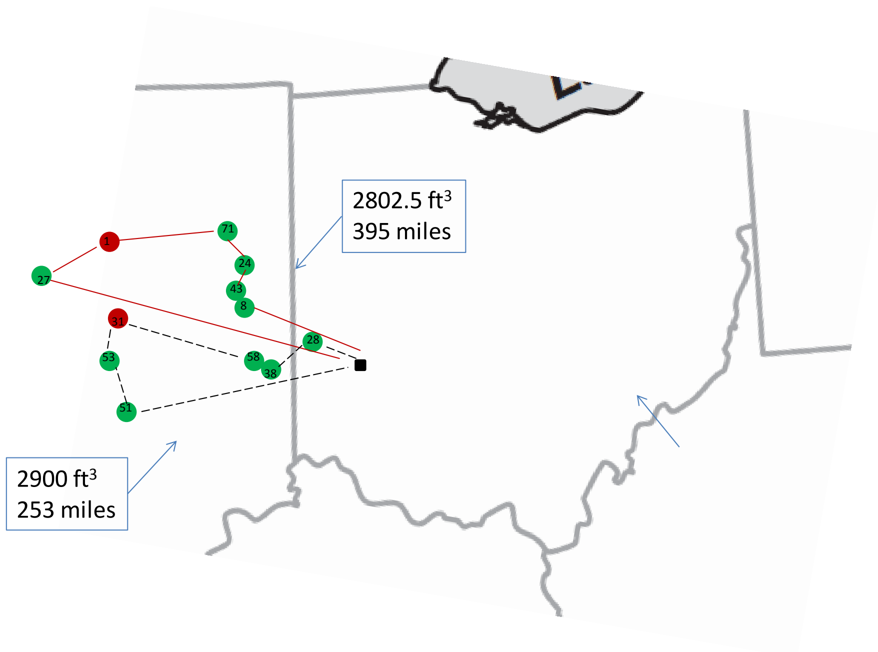


Figure 2-5. Summary of five twice-weekly routes in the case study that are identical in both solutions.

The case study demonstrates that significant savings in total logistics cost can be obtained in a real problem context by finding solutions that combine frequency-of-service decisions with routing decisions. In many supply chain systems, there is a primary focus on inventory reduction through just-in-time delivery operations. However, this case study illustrates that a very substantial transportation cost penalty can be associated with such decisions, and allowing a small amount of additional inventory in the system in order to reduce pickup frequency at some suppliers can yield considerable savings in overall logistics costs. Furthermore, an effective heuristic algorithm is available that can find good frequency-routing solutions very quickly, even in large problem instances.



(a) Routes from the optimal solution.



(b) Routes from the VLSN algorithm solution

.

Figure 2-6. Summary of remaining twice-weekly routes in the case study.

2.8 Conclusions

Collection routes through a consolidation center is a commonly used shipping strategy in organizing inbound material movements, and one important challenge for such complex systems is how to tradeoff transportation and inventory costs in order to minimize total logistics cost. This chapter proposes a clustering-based model that integrates pick-up frequency and spatial grouping. Unlike previous work focusing on the vehicle routing problem in each individual day, the approach views the problem as being primarily a clustering problem, where a route is built around a cluster seed which contains both the spatial information (location) and time information (frequency). Another major contribution of this model is introducing the marginal cost coefficient in the objective function to approximate the actual cost of a solution so that it can be easily computed before actual routes are constructed.

Small instances of the problem can be solved exactly by commercial software; however, the rapid growth in problem size with the increase in the number of suppliers and/or frequency classes and the approximate nature of the objective function suggest that heuristic solutions are necessary. This chapter converts the original problem to the single-source capacitated facility location problem (SSCFLP) by combing the location and frequency into one single index, and explores near-optimal solutions using the Very Large-Scale Neighborhood (VLSN) algorithm developed for SSCFLP. The effectiveness of both the problem formulation and the heuristic solution has been validated by a large battery of test problems. The results of the real case study show that the significant advantage can be achieved by finding solutions that consider both frequency and spatial routing, and demonstrate how this joint determination is important to industry applications.

One direction for further research to enhance this work is to add constraints on specific suppliers that would constrain them to a subset of the possible frequency classes. This may be important in some cases where supplier production schedules might require them to be served at particular frequencies. Another useful direction for enhancement is to incorporate uncertainty in the supplier pickup quantities. In this case, building sets of collection routes that are robust under variations in quantities of material at individual suppliers is desirable, and achieving such robustness may well affect the assignment of suppliers to frequency-based clusters.

CHAPTER 3

DESIGNING SPLIT PICKUP COLLECTION ROUTES WITH UNCERTAIN SUPPLIER QUANTITIES

3.1 Introduction

A common strategy for moving materials into manufacturing or assembly plants is to have plant-based collection routes that stop at multiple suppliers and return to the plant. These routes (sometimes called milkruns) often operate daily. Such collection routes are widely used in the automotive industry, where they typically serve suppliers that ship between 10% and 70% of a truckload per day to a single assembly plant.

A major concern in the design of collection routes is that variation in production schedules at the destination plant can cause significant day-to-day or week-to-week variability in the load quantities to be picked up at individual suppliers. A set of collection routes designed for the average loads may be quite inefficient when loads are below the average and have capacity problems when the loads are above average. Designing routes for some nominal average load factor (e.g., 80-85% of actual truck capacity) is one approach to dealing with this variation, but a design method that explicitly incorporates the variability can produce route designs that are both more efficient and more robust. The new model formulation presented here is intended to address the issue of load variability, including the fact that load variations at individual suppliers are likely to be correlated because they are driven by a common source – variation in production schedules at the destination plant.

Construction of collection routes has many of the characteristics of a classical vehicle routing problem (VRP), but the situation also offers some opportunities that are outside the usual VRP formulation. For example, it may be possible to allow suppliers to be served by more than one route, with each route handling part of the supplier's load. Dror and Trudeau (1989, 1990) introduced this idea with the Split Delivery Vehicle Routing Problem (SDVRP) and demonstrated that allowing splitting of customers has the potential for reducing overall cost of the set of routes. They also described a heuristic algorithm for constructing a split delivery solution. Although the common terminology in the literature refers to customers and deliveries, because of the application of interest here, we will use the terms suppliers and pickups.

Splitting a supplier among several collection routes creates some difficulties at the supplier's facility. To avoid confusion about what parts have been shipped on which route and when they should arrive at the destination plant, separate assignment of specific part numbers for shipment on each route are generally required, and these separated shipments must be staged for loading in different locations. In addition, the scheduled pickup times for each route must usually be coordinated. Although the practical challenges of splitting customer pickups have discouraged its use, the pressure for ever-increasing efficiency in the supply chain promotes consideration of supplier splitting as a means of reducing overall costs.

The model described here is developed by considering the problem as a stochastic capacitated clustering problem where individual suppliers may be part of more than one cluster (route). A subset of suppliers is selected to serve as seeds for eventual clusters, other non-seed suppliers are associated with one or more seeds, and then clusters (representing pickup routes) are formed from associated suppliers for each of several discrete scenarios that represent uncertainty in actual loads from individual suppliers. The concept of *associating* suppliers with seeds (routes) means that the supplier is eligible to be served by that route, but may not be in all demand scenarios. The associations between suppliers and routes create solutions with more regularity for the suppliers, the system manager and the drivers who operate the routes. On any given day, the responsibility for serving an individual supplier rests within a small subset of routes, and drivers need to become familiar with the intricacies of the physical layout at only those suppliers that are associated with the routes they normally operate. This is an implementation of the concept of *customer familiarity* in vehicle routing. Zhong *et al*. (2007) incorporated this idea into a vehicle routing problem for package delivery to improve both service and efficiency, and Smilowitz *et al*. (2013) emphasize both customer familiarity and region familiarity as important elements of workforce management in pickup and delivery operations.

This chapter is organized as follows. Section 3.2 reviews the relevant literature. Section 3.3 describes two versions of the new model formulation that incorporate different mechanisms to control the degree of service splitting at suppliers. Section 3.4 contains the process of defining scenarios to implement the model. Section 3.5 contains an example and includes experiments to compare the solutions from the two model versions. Section 3.6 contains some conclusions and forms the basis for further discussion of solution methods in chapter 4.

3.2 Previous Related Work

The VRP is one of the most studied problems in operations research. A review by Laporte (2009) classifies and summarizes the various solution approaches that have been developed for the VRP since the late 1950's. The SDVRP was first introduced by Dror and Trudeau (1989, 1990) and Archetti and Speranza (2012) provide a recent review of work on the SDVRP. Most of the work on the SDVRP assumes that the quantities to be picked up or delivered at individual vehicle stops are known with certainty. Since that review, additional work on the deterministic SDVRP has been done by Tang *et al*. (2013), Xiong *et al*. (2013), Moghadam *et al*. (2014), Archetti *et al*. (2014), Silva *et al*. (2014) and McNabb *et al*. (2015).

The Vehicle Routing Problem with Stochastic Demand (VRP-SD) is an extension of the VRP where pick-up or delivery quantity at each customer is uncertain. Tillman (1969) did initial work on the VRP-SD and significant additional work began in the 1980's (Stewart and Golden, 1983; Dror and Trudeau, 1986; Bertsimas, 1988; and Dror *et al.,* 1989). The VRP-SD is usually formulated as a stochastic mixed integer program that is solved in two stages. In a first stage, routes are determined such that each supplier is visited exactly once, and in a second stage (when demands are known) a recourse action is taken to address possible route failures, when the quantity assigned to the route exceeds the vehicle capacity. The usual recourse policy when a route failure occurs is that the vehicle detours to the depot to unload and then resumes collections at the first remaining unserved customer in its tour.

Solution approaches for the VRP-SD include both exact methods and heuristic algorithms. The exact methods include Laporte *et al*. (1989, 1992, 1994), Gendreau *et al*. (1995), and Laporte *et al.* (2002). The heuristic algorithms include a variety of adaptations of methods originally designed for the deterministic VRP. Useful examples include Bianchi *et al.* (2004), Ropke and Pisinger (2006), Laporte *et al*. (2010) and Shanmugam *et al.* (2011).

Ak and Erera (2007) proposed a different approach to the VRP-SD, based on a variation in the recourse scheme to be implemented in the event of tour failure. They proposed a paired locally-coordinated (PLC) recourse scheme where some *a priori* vehicle tours are paired to better use available vehicle capacity. In the first stage of the solution, each supplier is assigned to one route. Some routes may be matched together to create a route pair, but no route is included in more than one route pair. In the second stage, if one route in a pair is going to experience a tour failure at some supplier, the unserved suppliers will be added to the end of the partner route. If the partner route has insufficient capacity to serve all the suppliers, the vehicle travels to the depot to unload and then returns to the last visited supplier. A tabu search heuristic is presented to solve the VRP-SD using this alternative recourse policy.

Bouza ëne-Ayari *et al.* (1993) were the first to consider split pickups together with stochastic load quantities. They adapted a savings-based heuristic for the VRP-SD (Dror and Trudeau, 1986) by allowing splitting of the loads at a subset of suppliers into fixed fractions that could be incorporated into different routes. The algorithm has the advantage of being quite simple to implement and represents an important first step in combining consideration of uncertainty in supplier loads with the opportunities offered by splitting the loads.

Lei e*t al.* (2012) suggested adapting the paired-route recourse strategy for the VRP-SD when split deliveries are allowed. Their model makes several important limiting assumptions: (1) on unpaired routes, each supplier is visited by exactly one route (no split service); (2) on paired routes, each supplier can be served by one route or by both routes with pre-planned proportions; and (3) two paired routes share at most one supplier with split service. When route failure occurs at a non-split supplier, the regular recourse policy is applied (i.e., travel to the depot to unload and then return to continue the route at the first unserved supplier). When a failure occurs at a split supplier, each of the two paired routes to which this supplier is assigned serves its predetermined percentage of the demand. An adaptive large-neighborhood search (ALNS) heuristic is developed for solution of their formulation.

The model developed in this thesis takes a different approach to the SDVRP with stochastic demands, formulating it as a stochastic capacitated clustering problem. This formulation includes more flexible recourse strategies than have been employed in previous efforts on the VRP-SD. It also allows incorporation (either through cost penalties or explicit constraints) of elements to provide operational control over the degree of split service created in the solution. This represents a generalization of the SDVRP that can prove very useful in practical applications.

3.3 Model Formulation

We formulate the design of plant-based collection routes as a two-stage stochastic optimization problem. Each route is built around a supplier selected as a seed point, and the choice of the best seed points is part of the solution. Suppliers can be associated with routes, meaning that they are eligible to be served by that route, but may not be in all demand scenarios. A set of discrete scenarios (with associated probabilities) represents the variability in supplier loads, and can reflect correlation among supplier loads, if desired.

The number of seeds (routes) with which a specific supplier is associated plays an important role in governing how split service may be offered. If a supplier is only associated with one route, that route must always pick up the entire load quantity from the supplier and no split service is possible. From the supplier's perspective, service is very regular, and limiting associations reduces operational complexity for the drivers because they need to be familiar with facilities at fewer suppliers. On the other hand, having suppliers associated with multiple seeds creates possibilities for split service that may improve truck capacity utilization and also opens up more opportunities for effective route construction under varying total loads. This important tradeoff is fundamental to the model structure suggested here.

Two versions of the model are considered, differing in the way they incorporate the tradeoff between efficiency and operational regularity reflected in associations. In model (P3-1), associations are assumed to have a cost, and the magnitude of the cost coefficients can be used to encourage or discourage multiple associations for each supplier. In model (P3-2), explicit constraints are imposed on associations and split service to individual suppliers. This represents a more direct approach, rather than the indirect approach through cost penalties, but use of "hard" constraints may in some cases be quite costly. In the following two sub-sections, both versions of the model are laid out in detail. This sets the stage for computational experiments that follow.

3.2.1 Using Cost Penalties

For model (P3-1) the decision variables are defined as follows:

u_j = 1 if supplier *j* is a seed point; 0 if not.

y_{ij} = 1 if supplier *i* is associated with a route whose seed point is *j*; 0 if not.

w_{ij}^s = 1 if supplier *i* is visited by route *j* in scenario *s*; 0 if not.

x_{ij}^s = proportion of pick-up quantity at supplier *i* that is assigned in scenario *s* to a route whose seed point is *j*

The first-stage variables, u_j and y_{ij}, determine seed points for potential routes and associations of suppliers with those routes. Given these first-stage decisions, the second-stage (recourse) variables determine which suppliers are visited by what routes in each demand scenario (w_{ij}^s), and how the required loads are picked up (x_{ij}^s).

The model formulation is shown below. Further discussion of the model parameters follows the model statement.

(P3-1) $$\min \quad \sum_j f_j u_j + \sum_{i,j} g_{ij} y_{ij} + \sum_s \pi_s \sum_{i,j} c_{ij} w_{ij}^s \tag{8}$$

s.t. $$y_{ij} \le u_j \quad \text{for all } ij \tag{9}$$

$$\sum_j x_{ij}^s = 1 \quad \text{for all } i, s \tag{10}$$

$$\sum_i q_i^s x_{ij}^s \leq V \quad \textit{for all } j, s \tag{11}$$

$$x_{ij}^s \leq w_{ij}^s \quad \textit{for all } ij, s \tag{12}$$

$$w_{ij}^s \leq y_{ij} \quad \textit{for all } ij, s \tag{13}$$

$$x_{ij}^s \geq 0 \quad \textit{for all } ij, s \tag{14}$$

$$u_j,\ y_{ij},\ w_{ij}^s \in (0,1) \quad \textit{for all } i,\ j \textit{ and } s \tag{15}$$

The parameters of the problem are defined as follows:

f_j = fixed cost of creating a route whose seed point is *j*

g_{ij} = fixed cost of associating supplier *i* with the route whose seed point is *j*

c_{ij} = marginal cost coefficient for connecting supplier *i* to the route whose seed point is *j*

q_i^s = pick-up quantity at supplier *i* in scenario *s*

V = vehicle capacity

π_s = probability of scenario *s*.

Constraint (9) ensures that suppliers can only be associated with routes (seed points) that have been chosen. For each load scenario, a set of routes is created so that all suppliers' shipments are included (constraint 10) and no route exceeds vehicle capacity (constraint 11). Constraint (12) ensures that material can be picked up from a supplier only if that supplier is visited by the route in question in a specific scenario. Each supplier may be served by one or more routes, but the routes selected must be within the supplier's association set (constraint 13).

The actual objective of route creation is to minimize the total mileage for trucks to make pickups from all suppliers. However, the route mileage depends on the set of suppliers assigned to each route and the sequence of stops constructed. This creates a very complicated objective function that is difficult and computationally expensive to evaluate (as recognized in a wide variety of vehicle routing problems). The coefficients c_{ij} in the objective function of problem (P3-1) represent an approximation to the actual cost of serving supplier i as part of route j, and in this way can be viewed as approximate marginal costs. If a core route runs from node 0 (the plant) to the seed point of route j and back, the increase in total route distance to make an intermediate stop at supplier i is represented as:

$$c_{ij} = d_{0i} + \gamma d_{ij} - d_{0j} \qquad \text{for } i \neq j \tag{16}$$

where $\gamma \geq 1$ is a user-selectable parameter. The inclusion of the parameter γ in eq. (16) is similar to the suggestion by Yellow (1970) for savings-based vehicle routing.

The c_{ij} coefficients do not depend on the clustering solution, so they can be pre-computed from the distance matrix for all ij combinations. In general, there are also other stops on the route in addition to the seed point, so eq. (16) is not necessarily a true measure of the incremental distance associated with adding a stop at supplier i. However, in the context of material collection from suppliers, most routes involve relatively few stops and eq. (16) is a useful approximation.

The objective function coefficients f_j and g_{ij} reflect the costs of selecting seed points and creating associations between seeds and suppliers. We use $f_j = 2d_{0j}$ to establish the round-trip distance to the seed point. Then if point j is a selected seed, $\sum_j f_j u_j + \sum_i c_{ij} w_{ij}^s$ represents quite an accurate approximation to the total route length of route j in scenario s. Empirical testing in chapter 2 (and also in the work done by Dong and Turnquist, 2015) has verified the accuracy of this approximation.

The coefficient g_{ij} represents a cost penalty on the formation of associations. If $g_{ij} = 0$, the optimization has an incentive to associate each supplier with several routes so that the assignments for individual load scenarios will be most efficient for that scenario. However, this may not be the best solution from the perspective of the suppliers and the drivers who must operate the routes. If g_{ij} is made large, the optimization tends to form fewer associations leading to more regular service at each supplier, but the ability to adapt to changing load scenarios is restricted.

3.2.2 Using Explicit Constraints

As an alternative to using cost penalties g_{ij} on creation of associations, it is also possible to formulate the model with explicit constraints on associations and split service to individual suppliers. This requires an additional set of variables:

z_i^s = 1 if supplier i is split between two or more routes in scenario s; 0 if not.

It also requires parameters for the specific constraints:

M = maximum number of routes to which a single supplier can be associated

m = maximum number of routes allowed to visit a single supplier in any specific scenario

N = maximum number of suppliers whose pick-ups can be split across multiple routes in any specific scenario.

The model can then be formulated as problem (P3-2):

(P3-2) $$\min \quad \sum_j f_j u_j + \sum_s \pi_s \sum_{i,j} c_{ij} w_{ij}^s \tag{17}$$

s.t. (9)-(15) above

$$\sum_j y_{ij} \leq M \quad \textit{for all } i \tag{18}$$

$$\sum_j w_{ij}^s \leq m \quad \textit{for all } i, s \tag{19}$$

$$\sum_j w_{ij}^s - (m-1) z_i^s \leq 1 \quad \textit{for all } i, s \tag{20}$$

$$\sum_i z_i^s \leq N \quad \textit{for all } s \tag{21}$$

$$z_i^s \in (0,1) \quad \textit{for all } i \textit{ and } s \tag{22}$$

This version of the model constrains the total number of associations for each supplier (constraint 18), and also ensures that within each scenario a given supplier must be served by no more than m routes (constraint 19). The total number of suppliers that are split among multiple routes in any load scenario is limited to N

(constraint 21). Constraint (20) forces $z_i = 1$ if supplier i is assigned to more than one route.

Problem (P3-2) is clearly larger (in both number of variables and number of constraints) than problem (P3-1), but it offers more direct control over both associations and split service. An objective of computational tests in section 3.5 is to evaluate both model versions, but first we need to focus attention on how scenarios are created for either version of the model.

3.4 Scenario Creation

The purpose of scenario generation in the stochastic model is to create a relatively small set of discrete sets of supplier pick-up quantities, each with a probability of occurrence, so that the range of possible overall daily loads is well-covered, the correlations among individual supplier quantities are represented, and the relative likelihoods of different loading patterns are reflected in the set. In the context of inbound logistics planning, the reflection of correlations among supplier loads is particularly important because the quantities at different suppliers are being driven by a production schedule at the destination plant. Two suppliers that produce parts or material that is used in all production units will have positively correlated pick-up quantities because their volumes are both related to total production volume at the plant. Conversely, suppliers that produce parts for product options may have negatively correlated volumes because production units will have some options and not others. In an automotive assembly application, for example, suppliers of headlights and brake discs are likely to have positively correlated volumes because these are parts used on all production units. However, suppliers of base-model sound systems and in-vehicle navigation systems are likely to have negatively correlated volumes because these options are normally found on different trim levels and a production mix that increases demand for one is likely to decrease demand for the other.

Latin Hypercube Sampling (LHS) is an effective way of generating sample demands that cover the range of possible loads and reflect the marginal distributions of individual suppliers' loads. The original article describing LHS is McKay *et al*. (1979), and this concept has received widespread use in design of experiments for simulation models as well as in creation of sample scenario sets for stochastic programming. To reflect the correlation among individual supplier pick-up quantities, we can modify the LHS set to induce correlation without affecting the marginal distributions, using a procedure described by Iman and Conover (1982). This process leads to an overall strategy that can be called LHS with Correlation, or LHSC.

Consider a set of n suppliers, each of which has an uncertain load quantity, q_i, $i = 1, 2, \ldots, n$, for any time period in which collection service is to be provided. Denote the cumulative distribution function for the random variable q_i as $F_i(q)$. Suppose we want to generate a sample of S realizations of each q_i, indexed by $s = 1, 2, \ldots, S$. Divide the range of q_i into S contiguous intervals of equal probability and generate a sample q_{is} randomly (i.e., uniformly) within each interval for every $i = 1, 2, \ldots, n$, and $s = 1, 2, \ldots, S$.

The LHS sample is constructed by first combining the S samples of q_1 randomly and without replacement with the samples of q_2, also selected randomly and without replacement, to form the S ordered pairs $\left[q_{1s}, q_{2s}\right]$ for $s = 1, 2, \ldots, S$. Then, these pairs are combined randomly and without replacement with the samples of q_3 to create ordered triples $\left[q_{1s}, q_{2s}, q_{3s}\right]$. The process continues through all suppliers until a set of ordered n-vectors $\left[q_{1s}, q_{2s}, q_{3s}, ..., q_{ns}\right]$, $s = 1, 2, \ldots, S$, has been created. This set is an LHS of size S from the n load distributions for the suppliers, sampled consistently with the specified $F_i(q)$. For use in the stochastic optimization, each of these scenarios has equal probability, $1/S$.

Iman and Conover (1982) showed that if a desired rank correlation structure is given for the inputs to a model (in this case, the supplier loads), a restricted pairing procedure produces an LHS with a similar rank correlation structure. The theoretical basis for the procedure is that if X is a row vector of independent samples (i.e., with correlation matrix I), and H is a desired correlation matrix for some transformation of X, then because H is symmetric positive definite, it can be decomposed into $H = LL^T$, where L is a lower triangular matrix, and the transformed vector XL^T has the desired correlation matrix H. In the case of interest here, an element of H denoted h_{ij} , is the desired rank (i.e., Spearman) correlation between the loads at suppliers i and j.

The procedure uses a set of "scores" as the basis for inducing the desired rank correlations. Iman and Conover (1982) suggest using $\Phi^{-1}\left\{\frac{s}{S+1}\right\}$, for $s = 1, 2, \ldots, S$, where Φ^{-1} is the inverse function of the standard Normal distribution. These scores are used to create an S x n matrix, R, where each column is an independent permutation of the S scores. Then each row of R, denoted R_s, has n independent components, where each component assumes one of the score values with equal probability.

If the desired input rank correlation matrix is decomposed into $H = LL^T$, then $R_s L^T$ is a vector that has the desired correlation matrix *H*. Using the entire matrix *R* and computing $R^* = R\,L^T$ produces a matrix whose rows all have the same multivariate distribution as $R_s L^T$, and the rank correlation matrix computed for R^* will approximate *H*. Because *R* is a sample of possible permutations of the scores, its correlation matrix may not be exactly the identity matrix, so the result of the computations may not exactly reflect the desired correlation matrix, but the correlations in the revised LHS will approximate the desired correlations.

If the columns of an input LHS matrix are rearranged so that they each have the same ordering as the corresponding column of R^*, the resulting set of samples has the same sample rank correlation matrix as R^*, and therefore will also approximate *H*. This LHSC process is used here to generate a desired set of sample scenarios for the stochastic optimization.

3.5 An Illustrative Application and Computational Testing

Figure 3-1 illustrates an example with a plant in northeastern Indiana and a group of 14 suppliers – 12 in Ohio two in the very southern edge of Michigan. There are other suppliers to the north and west of the plant, but this group to the east is considered as a subset for collection routing. Figure 3-1 also shows the average quantities (in ft^3, rounded to the nearest hundred) to be picked up from each supplier every day. The trucks are assumed to have a capacity of 3,000 ft^3 and distances are computed using actual highway distances between points.

Consider first a deterministic solution based on the average quantities. If no split pickups are allowed, it is clear that the group of suppliers 1, 8, 9 and 10 present a significant problem. None of them can be grouped together on a route, and none of them can be grouped with the suppliers in the three natural clusters (2-4, 5-7 and 11-14). The result is that the optimal solution is to have seven routes, with four of them being out-and-back routes to a single supplier. This is relatively inefficient, with total mileage of 2,285 miles per day, and an average load on the trucks of only 1,971 ft^3, about two-thirds of capacity.

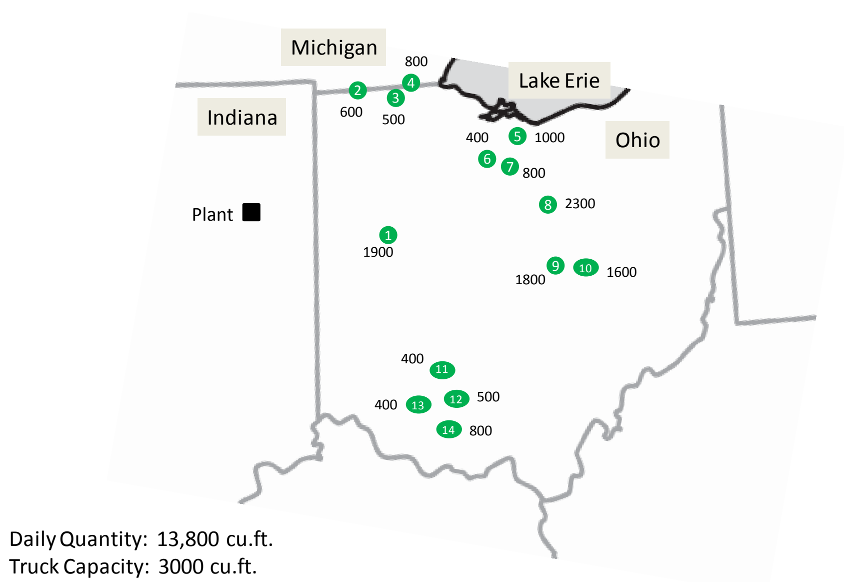


Figure 3-1. Plant, suppliers and average daily quantities (ft^3) for the example.

The deterministic split pickup solution based on average loads with no constraints on routes stopping at a single supplier or the number of suppliers with split service is to operate five routes, with three suppliers (1, 8 and 9) being split between two routes. The resulting route structure is shown in Figure 3-2. The total daily distance is reduced to 1,956 miles, a 15% saving from the no-split solution. In this small example, the potential to constrain splitting at individual suppliers or the number of suppliers with split service may not have great practical significance because the splits created are quite simple. The primary insight offered by the solution shown in Figure 3-2 is that allowing split service in this case creates an opportunity for substantial cost savings.

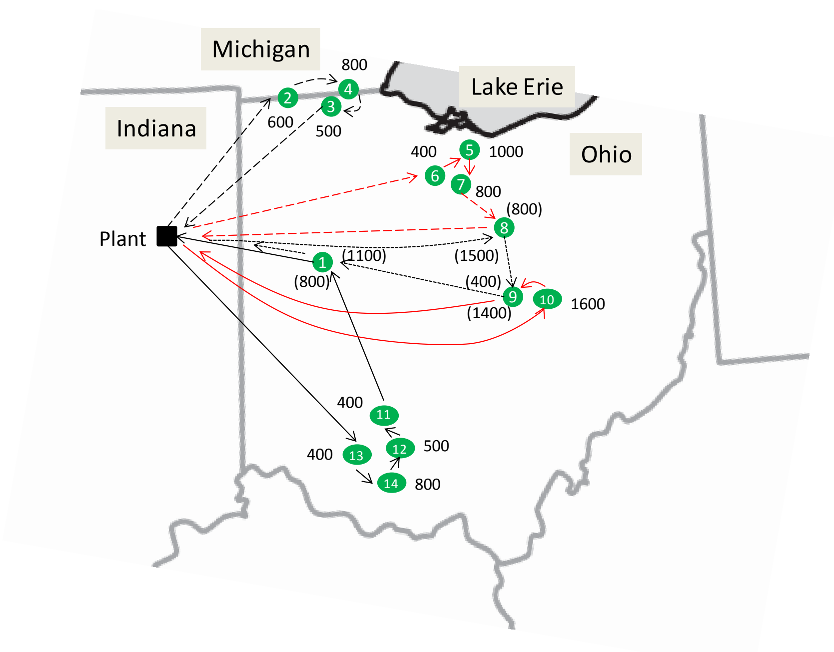


Figure 3-2. Routes for split pickup solution using deterministic average loads.

Neither solution considered thus far includes uncertainty in the load quantities at the suppliers. To introduce load uncertainty into the example, a set of marginal Beta distributions for the loads of individual suppliers is used, together with a correlation matrix to represent the dependence of the load quantities. The parameters of the marginal Beta distributions are shown in Table 3-1. A Beta distribution with parameters $\alpha = \beta = 1$ is a Uniform distribution. If $\alpha = \beta$, the distribution is symmetric, and the variance decreases in both α and β. The suppliers for the example exhibit a variety of shapes in the marginal distributions, but all have relatively large variances.

The correlation matrix among the suppliers is shown in Table 3-2. This matrix shows a mixture of positive and negative correlations, corresponding to complementary and substitute parts being used at the assembly plant.

Table 3-1. Parameters of Beta distributions for daily supplier loads.

Supplier	Lower Limit (cu. ft.)	Upper Limit (cu. ft.)	Alpha	Beta
1	750	3000	1.2	1.1
2	180	1000	1	1
3	325	675	1.4	1.4
4	400	1200	1.2	1.2
5	725	1800	1	3
6	150	650	1.2	1.2
7	275	1300	1.8	1.7
8	1400	3000	1	1
9	450	3000	1.4	1.2
10	450	2600	1.1	1
11	100	700	1.3	1.3
12	180	825	1	1
13	180	600	1.5	1.3
14	225	1325	1	1

Table 3-2. Correlation matrix for daily supplier loads.

Supplier	1	2	3	4	5	6	7	8	9	10	11	12	13	14
1	1													
2	-0.34	1												
3	0.79	-0.21	1											
4	0.89	-0.24	0.98	1										
5	-0.02	-0.33	-0.23	-0.18	1									
6	0.19	-0.2	0.25	0.26	-0.15	1								
7	0.82	-0.3	0.78	0.84	-0.23	0.33	1							
8	0.91	-0.27	0.8	0.9	-0.07	0.18	0.84	1						
9	0.52	-0.33	0.55	0.57	0.05	0.05	0.46	0.55	1					
10	0.97	-0.31	0.77	0.89	-0.02	0.25	0.83	0.92	0.46	1				
11	-0.03	-0.06	-0.03	-0.04	-0.14	0.14	0.19	0	-0.06	-0.02	1			
12	-0.07	0.21	0	-0.03	0.04	-0.11	-0.14	-0.08	-0.35	-0.07	0.05	1		
13	0.07	-0.06	0.06	0.08	-0.23	0.05	0.22	0.12	0.43	0.09	0.01	-0.28	1	
14	0.05	0.1	0	-0.01	0.35	-0.26	-0.25	0.01	-0.08	-0.03	-0.23	0.34	-0.57	1

A sample of 10 scenarios has been created for this example, using the LHSC method described in section 3.4. Each of these scenarios has probability $\pi_s = 0.1$. Table 3-3 summarizes the daily pick-up quantities for the 10 scenarios. The maximum daily pickup quantity in the 10 scenarios is 15,572 ft^3 (scenario 2), so we should expect at least 6 routes in the overall design. A larger set of scenarios could certainly be used and the LHSC method can generate any number desired, but for this example it is useful to consider a small set and be able to examine the nature of the solution quite carefully.

Table 3-3. Daily pick-up quantities (ft^3) for the 14 suppliers considering 10 scenarios.

	Scenario									
Supplier	1	2	3	4	5	6	7	8	9	10
1	2872	2006	1662	1830	2240	2645	1418	2365	1038	1273
2	310	833	456	231	515	604	856	949	753	374
3	359	660	549	451	527	394	413	586	485	616
4	532	1025	465	673	861	584	911	774	978	1156
5	830	1047	894	1594	1088	851	947	726	799	1296
6	602	235	140	375	442	540	346	456	284	497
7	773	969	1083	936	606	465	1256	723	482	810
8	2263	2495	2968	2802	2060	1873	1619	1540	2537	1926
9	1883	2166	1836	1533	632	1142	2380	2767	1357	2561
10	675	1380	2169	1534	922	2316	1731	1008	2566	1984
11	654	297	189	352	546	173	444	259	455	580
12	438	752	791	662	508	612	239	318	269	383
13	568	402	479	534	434	277	221	447	310	348
14	656	1305	774	463	402	871	248	1184	936	1096
Total Load	13415	15572	14455	13970	11783	13347	13029	14102	13249	14900

It is important to note that the routes shown in Figure 3-2 (split-service based on average loads) are infeasible in 5 of these 10 scenarios. In scenarios 1-4 and 10, there is at least one route failure (assigned load exceeds vehicle capacity). The failure in scenario 2 is obvious (the total load, 15,572 ft^3, exceeds the capacity of five routes), but even in a scenario with a total load below the average (scenario 1), there is an infeasibility because there is no way to split suppliers 1, 8 and 9 among the four routes they share and keep all vehicle loads within the 3,000 ft^3 limit. To make this clear, we label the relevant routes from Figure 3-2 as follows:

Route 1: Plant – 6 – 5 – 7 – 8 – Plant

Route 2: Plant – 8 – 9 – 1 – Plant

Route 3: Plant – 10 – 9 – Plant

Route 4: Plant – 13 – 14 – 12 – 11 – 1 – Plant.

The fifth route (Plant – 2 – 4 – 3 – Plant) does not involve any of these three suppliers. Because suppliers 5, 6 and 7 are always served by route 1, their total load (2,205 ft^3 in scenario 1) must be allocated to route 1, and the remaining available capacity for supplier 8 on route 1 is 795 ft^3 Similarly, supplier 10 must be assigned to route 3 (consuming 675 ft^3 and leaving up to 2,325 ft^3 for supplier 9), and suppliers 11-14 must be assigned to route 4, consuming 2,316 ft^3 of space on that route and leaving up to 684 ft^3 of capacity for supplier 1. Denote by v_{ij} the load from supplier i (i = 8, 9, 10) that is carried by route j (j = 1, 2, 3, 4), with $v_{ij} \geq 0$. Then for scenario 1, the following set of conditions must all be satisfied:

$$v_{81} + v_{82} = 2263$$

$$v_{12} + v_{14} = 2872$$

$$v_{92} + v_{93} = 1883$$

$$v_{81} \le 795$$

$$v_{14} \le 684$$

$$v_{93} \le 2325$$

$$v_{12} + v_{82} + v_{92} \le 3000$$

We can assign supplier 9 completely to route 3, reserving the maximum capacity possible for suppliers 1 and 8 on route 2. If we also assign as much as possible of supplier 1 to route 4 and supplier 8 to route 1, the required conditions are reduced to the following:

$$v_{82} = 1468$$

$$v_{12} = 2188$$

$$v_{12} + v_{82} \le 3000$$

This set of conditions clearly has no feasible solution. Similar sets of conditions can be written for scenarios 3, 4 and 10, which also have no feasible solution. Thus, in the face of uncertainty in pickup quantities at individual suppliers, the split pickup solution determined from the average quantities is not effective because it creates numerous route failures.

A more complete incorporation of the uncertainty in the supplier loads is necessary, and the model developed here offers a way to accomplish that. For the implementation in this example, the cost coefficients c_{ij} have been established using the distance matrix, as described in eq. (9), with γ = 1.2. For the first stage, $f_j = 2d_{0j}$ for all j.

We first report results of experiments with problem formulation (P3-1). A baseline is established using $g_{ij} = 0$ for all ij pairs. Then additional experiments are done with g_{ij} values based on the distance between i and j.

For this relatively small problem, solutions are obtained in extensive form using CPLEX, a commercial mixed integer programming package. For this example, the extensive form representation of the problem contains 4,132 variables (2,170 of which are binary) and 4,396 constraints. Solution times for the following experiments using CPLEX version 12.6.1 ranged from 23 seconds to 221 seconds on a personal computer with a 2.0 GHz processor and 8 GB of memory running the 64-bit Windows 7 operating system.

The solution with $g_{ij} = 0$ for all ij pairs selects suppliers 1, 4, 7, 8, 9 and 14 as seeds. Because all $g_{ij} = 0$, every supplier can be associated with all seeds, making them eligible for assignment to any route. However, the actual clusters identified in each scenario (and the constructed routes) do not use all the possible associations.

Figure 3-3 shows the six supplier clusters and route construction for scenario 1. The same set of routes is operated in scenario 5, which has the smallest total pick-up quantity. The pick-up quantities at the suppliers are different in these two scenarios, but no split service is required.

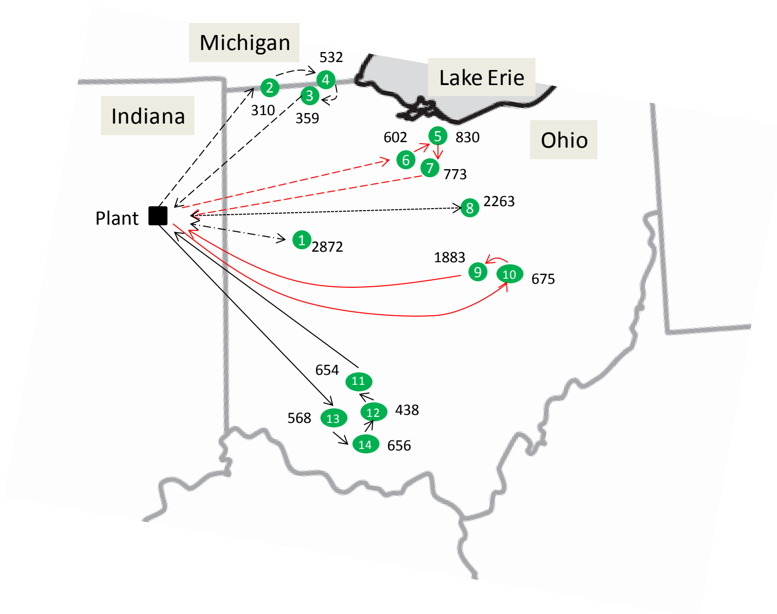


Figure 3-3. Supplier assignments, routes and quantities for scenario 1 when g_{ij} = 0.

Figure 3-4 shows the clusters and routes for scenario 2, which has the largest total pick-up quantity. In this scenario, supplier 9 is split between two routes. Scenario 10 presents another variation, where supplier 9 is split three ways, as shown in Figure 3-5. We'll not present figures for all the route variations across scenarios, but in this example the split service is focused on a single supplier (supplier 9), and that supplier may be split two or three ways, depending on the specific load scenario.

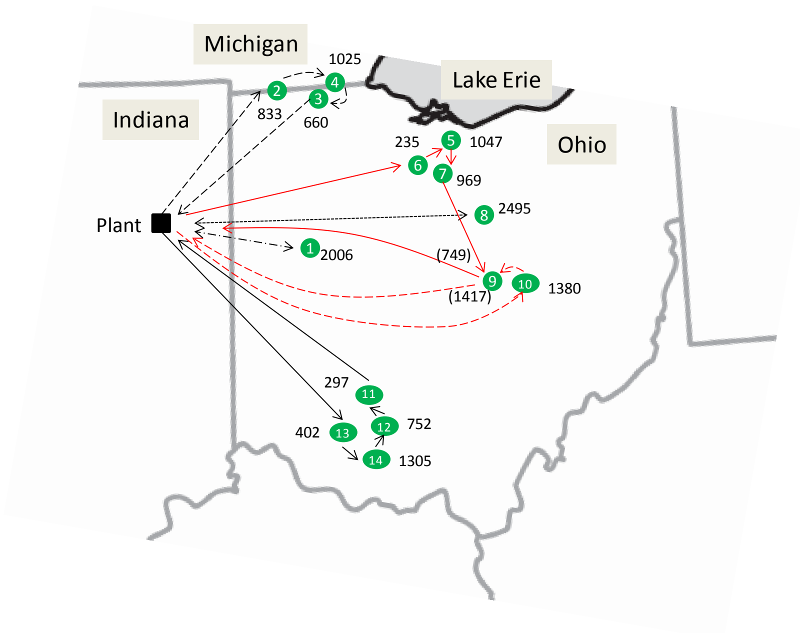


Figure 3-4. Supplier assignments, routes and quantities for scenario 2 when g_{ij} = 0.

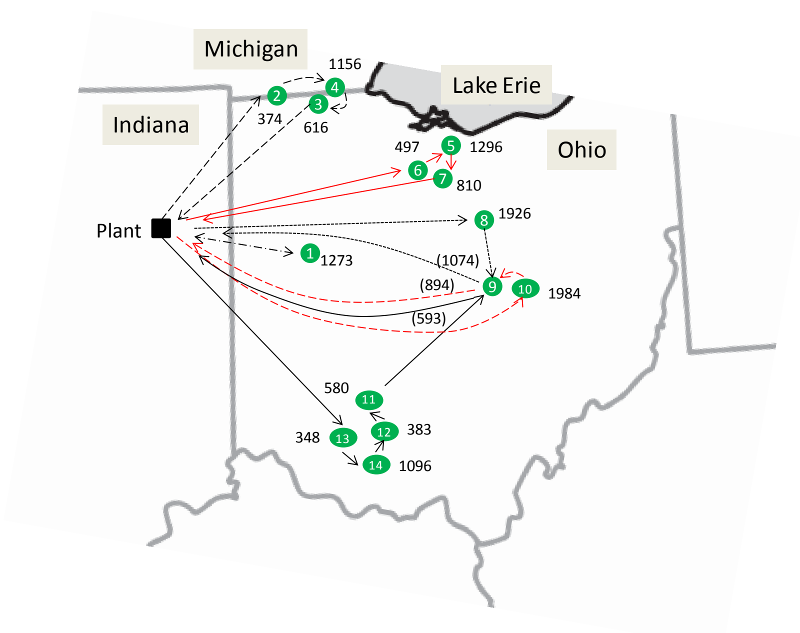


Figure 3-5. Supplier assignments, routes and quantities for scenario 10 when g_{ij} = 0.

The objective function value for the two-stage stochastic program solution is 2,079. The actual expected daily truck mileage across the ten scenarios after explicit route construction for the clusters is 2,007. Thus, the objective function value is within

3.5% of the actual mileage, without having to do explicit route construction as part of the solution process. The accuracy of the approximation used in the objective function has been confirmed in a much larger set of test problems in chapter 2.

To explore the effects of different g_{ij} values, two additional experiments have been run, one using $g_{ij} = 0.1d_{ij}$ and the other using $g_{ij} = d_{ij}$. The first of these experiments creates a small distance-related penalty for an association between a supplier and a seed point, and the second makes that penalty larger. We may expect that in general, the nonzero g_{ij} values will reduce the number of associations and potentially the variation of route construction across scenarios.

When $g_{ij} = 0.1d_{ij}$, the optimal selection of seeds is suppliers 1, 4, 7, 8, 10 and 14. This is a small change (replacing supplier 9 with supplier 10) from the set as selected when $g_{ij} = 0$. The number of associations between suppliers and routes is also reduced. When $g_{ij} = 0.1d_{ij}$, all suppliers have either one or two associations, whereas when $g_{ij} = 0$, all suppliers are associated with all seeds. The primary implication of the reduction in associations is for supplier 9, for which the three-way route split in scenario 10 is eliminated. The revised solution for scenario 10 is shown in Figure 3-6. This involves two suppliers (8 and 9) being split two ways each, instead of a single supplier being split across three routes.

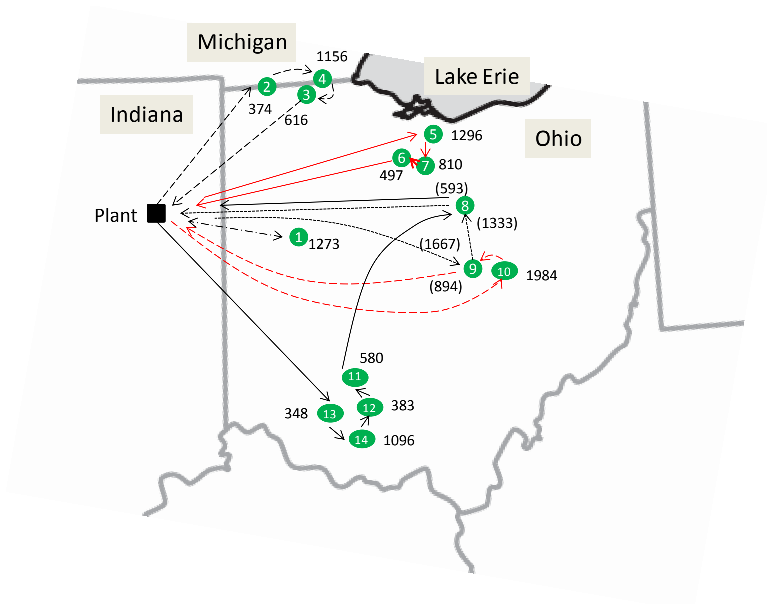


Figure 3-6. Supplier assignments, routes and quantities for scenario 10 when g_{ij} = 0.1 d_{ij}.

The reduction in associations also changes the routing slightly in scenario 8, causing supplier 9 to be split when it isn't in the solutions for $g_{ij} = 0$. The overall expected route length for the solution with $g_{ij} = 0.1d_{ij}$ is 2,009 miles, 2 miles larger than the solution when $g_{ij} = 0$.

The solution obtained when the association costs are increased to $g_{ij} = d_{ij}$ is the same as when $g_{ij} = 0.1d_{ij}$. In this example, at least, placing a small cost on making associations is sufficient to reduce the number of associations to 1 or 2 for each supplier, so increased values of g_{ij} have no further effect on the solution.

Attention is then turned to problem formulation (P3-2). In the baseline solution for (P3-1) with $g_{ij} = 0$, in one scenario a single supplier is split among three pick-up routes. For experiments with problem (P3-2), we set M = 3 and m = 2. That is, each supplier is allowed to be associated with up to three routes, but can actually be served by no more than two of them in any scenario. This forces a different solution from that obtained in the baseline experiment with problem (P3-1). In this small problem, the split service is focused on only one supplier (supplier 9), so there is little concern with the *N* constraint.

Imposition of the constraints creates has no effect on the set of seed points selected. They are the same as in the baseline solution for problem (P3-1). Also, in 7 of the 10 scenarios the routes operated are the same as in the baseline solution. However, in the baseline solution supplier 9 is assigned to four different seeds (7, 8, 9 and 14) across the set of scenarios; in the constrained case, this is not allowed because the maximum number of associations is $M = 3$. In the constrained case, supplier 9 is associated with only three seeds (8, 9 and 14), so the routing solutions in three of the scenarios change slightly. The most substantial change is in scenario 10, where the routing solution must change because the three-way split of supplier 9 shown in Figure 3-5 is not feasible when $m = 2$. For this example problem, the constrained solution for scenario 10 is the same solution shown in Figure 3-6, obtained by setting $g_{ij} = 0.1 d_{ij}$.

The objective function for problem (P3-2) increases very slightly from the baseline (P3-1) solution, but the total expected truck mileage (after routes are constructed) is the same as in the baseline solution. This illustrates that it may be quite possible in practice to find solutions with simpler splitting arrangements (i.e., constraining $m = 2$) without incurring a significant cost penalty.

3.6 Setting the Stage for Further Analysis

The example analysis in section 3.5 illustrates three important points related to the two models developed in this chapter. First, explicit incorporation of uncertainty in individual supplier quantities (and the correlations among suppliers in any given period) allows creation of much more effective and robust solutions for supplier clustering and routing than simply basing a structure of collection routes on average load quantities. Second, either using non-zero g_{ij} coefficients in formulation (P3-1) or using problem (P3-2) can limit the degree of split service provided to individual suppliers. The solutions obtained from the two approaches are likely to be somewhat different, but both can be effective. Third, solutions achieved either with non-zero g_{ij} coefficients in formulation (P3-1) or using problem (P3-2) can limit the degree of splitting at individual suppliers without imposing a substantial increase in total costs. This is important information for logistics managers.

For a small problem like the example in section 3.5, it is straightforward to solve either problem (P3-1) or (P3-2) in extensive form using a commercial MIP solver. However, as either the number of suppliers or the number of scenarios increase, the size of the extensive form problem increases rapidly. This provides an incentive for exploring alternative solution methods using decomposition. The following chapter explores two different decomposition methods frequently used for stochastic programming problems and evaluates their effectiveness using a realistic example problem.

CHAPTER 4

SOLUTION METHODS FOR THE STOCHASTIC PICK-UP PROBLEM

4.1 Decomposition, Optimal Solution, and Heuristics

Problems presented in chapter 3, (P3-1) and (P3-2), are formulated as two-stage stochastic programs. In the first stage, the decision of choosing seed suppliers and associating non-seed suppliers to these seed suppliers has to be made without full information on uncertain load quantities at each supplier. Given values for the first stage variables (u_j and y_{ij}), corrective or recourse actions can be taken in the second stage, to assign suppliers to different routes within other constraints (e.g., capacity constraint).

Stochastic programs are generally quite large because of the proliferation of variables and constraints to represent conditions and actions in several possible future scenarios. As long as the set of potential scenarios considered is finite, it is conceptually possible to convert the stochastic program into an equivalent deterministic program and solve it (in so-called extensive form), but as a matter of practicality, this approach is limited to small problems. For larger-scale applications, a variety of solution approaches have been created based on some form of decomposition. Work on exact and approximation algorithms include those of Kall (1982), Wets (1983), and Laporte *et al.* (1989). Because of the inherent difficulty of solving stochastic programs with integer recourse (as the models proposed in chapter

3), exact algorithms for this class of problems are generally not computationally efficient and methods useful for problems of practical size are heuristics. The focus of this chapter is on heuristic decomposition methods. The two general types of decomposition are stage-based and scenario-based.

The exemplary stage-based decomposition method is the L-shaped method, originally developed by Van Slyke and Wets (1969) for linear programs. Laporte and Louveaux (1993) extended the method to problems with integer variables. The central idea of the L-shaped method is to construct a solution for the first-stage variables, use that solution to condition the solution of the second-stage problem, and then develop additional constraints ("cuts") for the first-stage problem based on dual variables in the second stage. The first stage is then re-solved with the additional cuts and the process repeats. This is analogous to Bender's decomposition in large-scale deterministic linear programs. A duality gap may exist because of integer variables, therefore heuristics will be more efficient for solving stochastic programs with first-stage integer variables (especially for large-scale programs).

Two principal examples of scenario-based decomposition are the progressive hedging approach originated by Rockafellar and Wets (1991) and the dual decomposition approach due to Carøe and Schultz (1999). Scenario-based approaches in two-stage stochastic programs relax the requirement that the first stage decisions must be independent of which second-stage scenario will occur (non-anticipative decisions) and decompose the problem into separate problems for each scenario. They then augment the individual scenario problems with Lagrangian terms to gradually force the first-stage variables of the separate problems into agreement. Progressive hedging and dual decomposition methods accomplish this process in different ways and have a different iterative structure. When the problem contains integer variables, progressive hedging is not guaranteed to converge to an optimal solution, although it has been used successfully as a heuristic in several application areas (e.g., Listes and Dekker (2005), Fan and Liu (2010) and Watson and Woodruff (2011)). Dual decomposition methods can be guaranteed to converge, but often do so quite slowly. Recently, Guo *et al*. (2015) have proposed a way of combining the two methods to achieve the relative computational advantage of progressive hedging with the guaranteed convergence of dual decomposition.

Problems (P3-1) and (P3-2) represent different ways of approaching the design of inbound collection routes under uncertainty, but the formulation (P3-1) appears to be more amenable to decomposition for large-scale problem instances. Thus, in this chapter the focus is only on (P3-1). Section 4.2 describes how the integer L-shaped method (ILSM) is applied. In section 4.3, we adapt the general progressive hedging (PH) method by taking advantage of the problem structure (e.g., all the first stage variables are binary). Section 4.4 presents a realistic case study. In this case study, the Latin Hypercube Sampling with Correlations (LHSC) method is used to generate sample load scenarios and both decomposition methods are tested. This leads to conclusions about which method (ILSM or PH) is likely to be more effective in practical applications.

4.2 Integer L-shaped Method

Let *H(u, y)* be the expected value of the second-stage sub-problem in (P3-1), *H(u,y)* = $\sum_{s} \pi_s \sum_{i,j} c_{ij} w_{ij}^s$, where the values of w_{ij}^s are those that minimize the second-stage cost for scenario *s*, given the values of the first stage variables (u_j and y_{ij}) and the constraints relevant to each scenario. Using this notation, problem (P3-1) can be re-written as (P4-1), where only the first-stage variables appear explicitly.

(P4-1)

$$\min \sum_{j} f_j u_j + \sum_{i,j} g_{ij} y_{ij} + H(u, y) \tag{23}$$

$$\text{s.t.} \quad y_{ij} \le u_j \quad \textit{for all ij} \tag{24}$$

$$u_j, \; y_{ij} \in \{0, 1\} \quad \textit{for all i , j} \tag{25}$$

To evaluate the objective function for a selection of u_j and y_{ij} variables, all the scenario-specific second-stage problems have to be solved to compute *H(u, y)*. The essence of the ILSM is to solve a sequence of simpler problems, building up a piecewise linear lower bound on *H(u, y)*, and eventually converging to a solution of problem (P4-1). The process of building the outer linearization for *H(u, y)* is termed addition of *optimality cuts* to problem (P4-1) in the stochastic programming literature.

In the stage-based decomposition, the first-stage problem is written as:

(P4-2) $$\min \sum_j f_j u_j + \sum_{i,j} g_{ij} y_{ij} + \theta \tag{26}$$

$$\text{s.t.} \quad y_{ij} \le u_j \quad \textit{for all } ij \tag{27}$$

$$\theta \ge e_k + \sum_j a_{kj} u_j + \sum_{i,j} b_{kij} y_{ij} \qquad k = 1, .., K \tag{28}$$

$$u_j, \ y_{ij} \in \{0,1\} \quad \textit{for all } i, j \tag{29}$$

The constraints (28) represent the optimality cuts for the problem. Each is a supporting hyperplane for *H(u, y)*, with the intercept (e_k) and slope coefficients (a_{kj} and b_{kij}) constructed from the solutions of scenario-specific second-stage sub-problems. At any given step in the solution process, *K* is the number of optimality cuts that have been added. As more and more cuts are added, the piecewise linear lower approximation for *H(u, y)* gets better and better, and eventually the solution to (P4-2) yields $\theta = H(u, y)$ and the solution to (P4-1) has been achieved. The key elements in implementation of the ILSM are:

- Ensuring that the second-stage sub-problems have feasible solutions for all selections of *u* and *y*;
- Determining the intercept (e_k) and slope coefficients (a_{kj} and b_{kij}) from the solutions of scenario-specific second-stage sub-problems; and
- Finding an efficient way to solve the second-stage sub-problems because they are solved a large number of times.

To ensure that the second-stage problems are always feasible, we construct an artificial route with infinity capacity in the second stage for all the suppliers, and this route has a corresponding artificial seed supplier. Let *NS* be the number of real suppliers and NS' be the number of all the suppliers. $NS' = NS + 1$, and the extra supplier is the artificial one. We allow all the suppliers to be associated with this dummy route; therefore it assures that every supplier can be assigned to at least one route (i.e., $y_{i,NS'}$ equal to 1 for all the suppliers i) and no infeasible solution would exist. Ideally, when the algorithm finds the optimal solution, this dummy route is not in use and all the suppliers are assigned to actual routes. So the marginal costs of assigning any supplier to this route ($c_{i,NS'}$ for all suppliers i) are set to very large values, which will decrease the desirability of such assignments in the second stage.

Given the seed selection vector (u_j) and association matrix (y_{ij}) from the first stage sub-problem, the scenario-specific sub-problem for the second stage can be written as problem (SP). Note that the u_j variables do not appear explicitly.

(SP) min $$\sum_{i,j} c_{ij} w_{ij}^s \qquad (30)$$

s.t. $$\sum_j x_{ij}^s = 1 \quad \text{for all } i \qquad (31)$$

$$\sum_i q_i^s x_{ij}^s \le V_j \quad \text{for all } j \qquad (32)$$

$$x_{ij}^s \le w_{ij}^s \quad \text{for all } ij \qquad (33)$$

$$w_{ij}^s \le y_{ij} \quad \text{for all } ij \qquad (34)$$

$$x_{ij}^s \ge 0 \quad \text{for all } ij \qquad (35)$$

$$z_i^s,\ w_{ij}^s \in (0,1) \quad \text{for all } i,\ j \qquad (36)$$

Because the artificial route can be loaded with as many materials as possible, $V_{NS'}$ is infinite, while the other route capacities are finite (and in practice, often equal, because a standard truck with a 53-foot trailer is typically used for all routes).

The next two sub-sections discuss the process of constructing the optimality cuts from the solutions of the second-stage subproblems, and section 4.2.3 discusses how those sub-problems can be solved rapidly.

4.2.1 First Type of Optimality Cuts

At every iteration of the algorithm we consider a linear relaxation problem of the scenario *s* subproblem, given the y_{ij} values. The formulation of this relaxed model is shown below.

$$\text{(RP)} \qquad \min \quad \sum_{i,j} c_{ij} w_{ij}^s \tag{37}$$

$$\text{s.t.} \quad \text{(31)-(35) above}$$

$$0 \le z_i^s \le 1; \qquad 0 \le w_{ij}^s \le 1 \quad \textit{for all } i,\, j \tag{38}$$

Let $R_s^*(y)$ be the objective function of (RP) for scenario *s*: $R_s^*(y) = \min \sum_{i,j} c_{ij} w_{ij}^s$;

let π_s be the occurrence probability of the scenarios *s*; and define $\Gamma(y)$ be the expected value of the relaxed second stage problem: $\Gamma(y) = \sum_s \pi_s R_s^*(y)$.

Let λ_{ij}^s be the shadow price for the constraint (34), $w_{ij}^s \le y_{ij}$ *for all ij* , in the linear programming relaxation form, and define Λ_{ij} be the expected value of shadow prices across all the scenarios: $\Lambda_{ij} = \sum_{s} \pi_s \lambda_{ij}^s \quad \textit{for all ij}$.

According to Lemma 3.72 in Louveaux and Schultz (2003), a valid optimality cut at a given $y = y^k$ (k = 1,…, K) is $\theta \geq \Gamma(y^k) - (\Lambda^k)^T (y - y^k)$, where the individual linear equations are indexed by k. This cut can be rewritten as follows:

$$\text{(OCT 1)} \qquad \theta \geq (\Lambda^k)^T y + \left[\Gamma(y^k) - (\Lambda^k)^T y^k \right] \tag{39}$$

where $(\Lambda^k)^T$ is the vector of slope coefficients and $\left(\Gamma(y^k) - (\Lambda^k)^T y^k \right)$ is the intercept.

When the second-stage sub-problems are linear programs (involving only continuous variables), a series of these cuts can approximate *H(u, y)* relatively well. However, problem (SP) involves many binary variables and *H(u, y)* lies above the linear relaxation, $\Gamma(y)$. The looseness of the cuts results in very slow convergence of the solution process. In an effort to speed the process, the intercept term in the implementation has been revised from $\left(\Gamma(y^k) - (\Lambda^k)^T y^k \right)$ to $\left(H(\mathrm{u}^k, y^k) - (\Lambda^k)^T y^k \right)$ to tighten the approximation. This results in cut (OCT 1'):

$$\text{(OCT 1')} \qquad \theta \geq (\Lambda^k)^T y + \left[H(\mathrm{u}^k, y^k) - (\Lambda^k)^T y^k \right] \tag{40}$$

The nature of this change is illustrated generically in Figure 4-1. Although cut (OCT 1') has worked successfully as a heuristic to speed convergence in the current application, these optimality cuts are not guaranteed to be valid because they may cut off some feasible solutions.

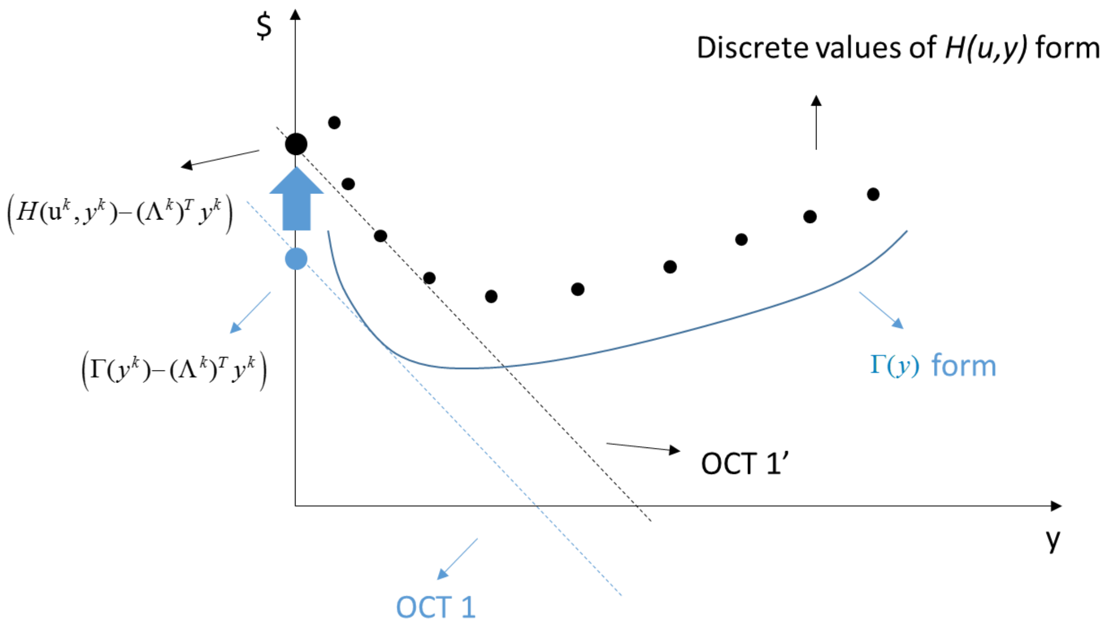


Figure 4-1. Improving the first type of optimality cut by increasing the intercept.

4.2.2 Second Type of Optimality Cuts

Let U^k be the set of variables u_j equal to 1 in the first stage problem of iteration k: $U^k = \left\{ j \mid u_j^k = 1 \right\}$; and let Y^k be the set of variables y_{ij} equal to 1 in the first stage problem of iteration k: $Y^k = \left\{ ij \mid y_j^k = 1 \right\}$. Then a second type of valid optimality cut can be generated as follows (Louveaux and Schultz 2003, Lemma 3.71):

(OCT 2)

$$\theta \geq \sum_{j \in U^v} H(u^k, y^k) u_j + \sum_{ij \in Y^v} H(u^k, y^k) y_{ij} - \sum_{j \notin U^v} H(u^k, y^k) u_j$$
$$- \sum_{ij \notin Y^v} H(u^k, y^k) y_{ij} - H(u^k, y^k)(\left|U^k\right| + \left|Y^k\right| - 1) \tag{41}$$

In other words, an intercept $\beta^k = H(u^k, y^k)(|U^k| + |Y^k| - 1)$ is computed, and slopes of $H\left(u^k, y^k\right)$ are attached to all u_j that are 1 and y_{ij} that are 1, along with slopes of $-H\left(u^k, y^k\right)$ to all u_j and y_{ij} that are zero.

Cuts generated this way are often very loose and not of great value in creating a good approximation of *H(u,y)*. In the application here, these optimality cuts do have some value because they generate supporting hyperplanes that have positive slopes on some u_j and y_{ij} variables. In the optimality cuts of type 1 discussed above, the slopes associated with the y_{ij} variables are constructed from the dual variables in the second-stage sub-problems. In any such sub-problem, increasing a y_{ij} variable can never increase cost, so all the slopes constructed this way are negative. Thus, the type 1 cuts construct an approximation to *H(u,y)* on the "left side" of Figure 4-1, but don't provide a very good approximation on the "right side". The cuts of type 2 can help to fill that in, even though they are relatively loose approximations.

With both types of cuts being constructed, at each iteration of the solution process, two new constraints are added to problem (P4-2).

4.2.3 Solving the Second-Stage Sub-Problems

In the formulation of problem (RP), the distinction between the x_{ij}^s variables and the w_{ij}^s variables ceases to be meaningful, so one set of those variables becomes redundant. If the x_{ij}^s variables are eliminated, the sub-problem for scenario s can be rewritten as follows:

(RP') $$\min \sum_{i,j} c_{ij} w_{ij}^s \tag{42}$$

$$\text{s.t.} \quad \sum_j w_{ij}^s = 1 \quad \textit{for all } i \tag{43}$$

$$\sum_i q_i^s w_{ij}^s \le V_j \quad \textit{for all } j \tag{44}$$

$$0 \le w_{ij}^s \le y_{ij} \quad \textit{for all } ij \tag{45}$$

Then by defining new variables $v_{ij}^s = w_{ij}^s q_i^s$, the problem (RP') can be rewritten to a capacitated transportation problem with "demands" q_i^s and "supplies" V_j. The formulation is shown below:

(TP) $$\min \sum_{i,j} \left(\frac{c_{ij}}{q_i^s} \right) v_{ij}^s \tag{46}$$

$$\text{s.t.} \quad \sum_j v_{ij}^s = q_i^s \quad \textit{for all } i \tag{47}$$

$$\sum_i v_{ij}^s \le V_j \quad \textit{for all } j \tag{48}$$

$$0 \le v_{ij}^s \le q_i^s y_{ij} \quad \textit{for all } ij \tag{49}$$

In this problem, the arc capacities are $q_i^s y_{ij}$, implying either arcs exist and can accommodate full demand q_i^s, or do not exist and limit flow to zero. Dual prices are the net cost changes for introducing or removing arcs. Figure 4-2 illustrates how problem (TP) is converted to a circulation flow network, which can be solved by using the minimum cost flow algorithm.

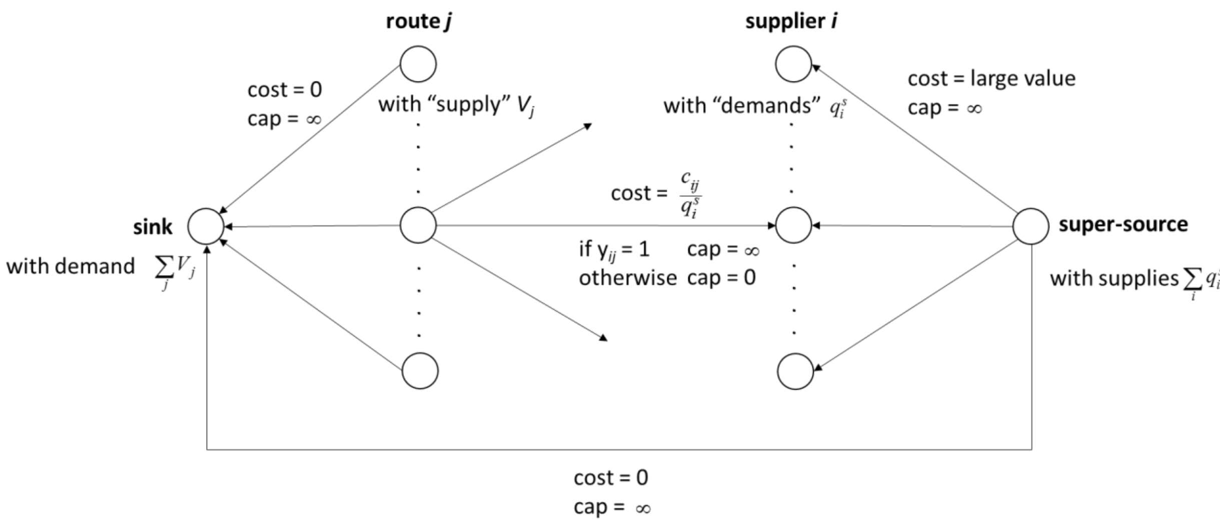


Figure 4-2. Converting to a circulation flow network.

The resulting solution has a cost $R_s^*(y)$ and the dual prices on the arc capacity constraints provide the information for the slopes in the optimality cut. If we let $\pi(n)$ be the node price (dual variable) for node *n* in the network, then the reduced cost for the arc from route *j* to supplier *i* is $\left(\frac{c_{ij}}{q_i^s}\right)+\pi(j)-\pi(i)$. The shadow prices on the arc capacities are non-positive and are given by $\alpha_{ij} = \min\left(0, \frac{c_{ij}}{q_i^s} + \pi(j) - \pi(i)\right)$. To get the dual prices for the cut, we scale α_{ij} by q_i^s, forming $\lambda_{ij}^s = q_i^s \alpha_{ij}$.

The ability to solve the second-stage sub-problems as simple network flow problems is important to the usefulness of the decomposition because these problems must be solved many times through the course of the overall solution process.

4.2.4 General Procedure

The general solution procedure for implementing the ILSM can then be stated as follows. Let z^k be the objective value for the current problem at iteration k and z^* be the estimated optimal objective value.

Step 1: Initialization

$k = 0$

$z^* = \infty$

$\theta = -\infty$

Step 2: Iteration update and decomposition

$k \leftarrow k + 1$

Solve the first stage subproblem (P4-1) for (y^k, u^k, θ^k)

Given (y^k, u^k, θ^k), solve the linear relaxation sub-problems for each scenario to compute the values of λ_{ij}^k for the slope coefficients in cuts of type 1

Compute an estimate of $H(u^k, y^k)$ by using

"integer round up" on the w_{ij}^s variables from the sub-problem

solutions and computing $\hat{H}\left(u^k, y^k\right) = \sum_s \pi_s \sum_{ij} c_{ij} w_{ij}^s$

Step 3: Solution update

Compute the current objective function value:

$$z^k = \sum_j f_j u_j^k + \sum_{ij} g_{ij} y_{ij}^k + \hat{H}\left(u^k, y^k\right)$$

If $z^k < z^*$, then update $z^* = z^k$, and let the current solution be the optimal solution so far.

Step 4: Check for termination

If $\theta^k \geq H(u^k, y^k)$, then stop. Otherwise, impose two optimality cuts (OCT 1') and (OCT 2), and return to Step 2.

4.3 Progressive Hedging Approach

As an alternative to the stage-based decomposition represented by the ILSM, scenario-based decomposition can also be explored and this is done here through creation of a progressive hedging (PH) algorithm. When integer variables are present, PH must be considered as a heuristic, since it is not guaranteed to converge. Potential integration with a dual decomposition method can guarantee convergence to an optimal solution, but dual decomposition often converges quite slowly and our interest is in getting a "good" solution to the problem relatively quickly. Thus, we will consider PH alone as a heuristic.

To use progressive hedging, we decompose the problem by scenario, introducing new variables:

u_j^s = 1 if supplier *j* is a seed point in scenario *s*; 0 if not.

y_{ij}^s = 1 if supplier *i* is associated with a route whose seed point is j in scenario *s*; 0 if not.

The individual scenario models are as follows:

(P4-3) $$\min \quad \sum_j f_j u_j^s + \sum_{i,j} g_{ij} y_{ij}^s + \sum_{i,j} c_{ij} w_{ij}^s \tag{50}$$

s.t. $$y_{ij}^s \le u_j^s \quad \text{for all } ij \tag{51}$$

$$\sum_j x_{ij}^s = 1 \quad \text{for all } i \tag{52}$$

$$\sum_i q_i^s x_{ij}^s \le V \quad \text{for all } j \tag{53}$$

$$x_{ij}^s \le w_{ij}^s \quad \text{for all } ij \tag{54}$$

$$w_{ij}^s \le y_{ij}^s \quad \text{for all } ij \tag{55}$$

$$x_{ij}^s \ge 0 \quad \text{for all } ij \tag{56}$$

$$u_j^s,\ y_{ij}^s,\ w_{ij}^s \in (0,1) \quad \text{for all } i,\ j \tag{57}$$

To make the set of scenario-based solutions implementable, we must satisfy non-anticipative constraints:

$$u_j^s = u_j \quad \forall s \quad \text{and} \qquad y_{ij}^s = y_{ij} \quad \forall s \, .$$

Progressive hedging works by repeatedly solving scenario-specific models and updating the coefficients f_j and g_{ij} so that there is an incentive to select u_j^s and y_{ij}^s values that are common across all scenarios (i.e., satisfy the non-anticipative constraints). The convergence can be aided by adding a valid inequality to each of the scenario-specific problems, and the speed of getting the scenario-specific solutions and updating the coefficients can be aided by using a lower bound approximation for the scenario problems. These adaptations to the general PH concept are discussed in the following two subsections.

4.3.1 Considering Minimum Required Number of Vehicles

One of the issues with convergence in the PH algorithm is that the number of seeds selected in different scenarios is likely to vary as the total amount of material to be picked up varies. Because all material must be picked up in all scenarios, the scenario with largest total load dictates a minimum number of routes that must be operated, and hence a minimum number of seeds that must be selected. Addition of a constraint on the minimum number of seeds selected in each scenario problem is thus a valid constraint and aids the convergence of the algorithm. If the required number of vehicles is at least r, then the following constraint can be added to all individual scenario sub-problems:

$$\sum_j u_j^s \geq r \tag{58}$$

4.3.2 Creating a Lower Bound Problem

Because of the relationships among the variables, $0 \leq x_{ij}^s \leq w_{ij}^s \leq y_{ij}^s \leq u_j^s \leq 1$, and the fact that the coefficients f_j, g_{ij} and c_{ij} are all non-negative, a lower bound on the problem (P4-3) can be constructed by substituting x_{ij}^s for y_{ij}^s and w_{ij}^s and re-writing the problem as (LB):

(LB) $$\min \quad \sum_j f_j u_j^s + \sum_{i,j} \left(g_{ij} + c_{ij} \right) x_{ij}^s \tag{59}$$

$$\text{s.t.} \quad \sum_j x_{ij}^s = 1 \quad \textit{for all } i \tag{60}$$

$$\sum_i q_i^s x_{ij}^s \leq V u_j^s \quad \textit{for all } j \tag{61}$$

$$x_{ij}^s \geq 0 \quad \textit{for all } ij \tag{62}$$

$$\sum_j u_j^s \geq r \tag{63}$$

$$u_j^s \in (0,1) \quad for\ all\ i\ ,\ j \tag{64}$$

Problem (LB) is in the form of a capacitated facility location problem, where a set of facilities of capacity V are to be located at nodes j (the u_j^s variables), and suppliers i are then assigned to specific facilities (the x_{ij}^s variables), with possible splitting of service among facilities. Numerous authors have developed specialized algorithms (either exact solutions or heuristics) for the capacitated facility location problem. Examples include exact solution methods by Van Roy (1986), Leung and Magnanti (1989) and Aardal (1998), and heuristics by Beasley (1993), Gromilund and Ganascia (1997), Lorena and Senne (1999), Barahona and Chudak (2005) and Sun (2012).

Sun (2012) notes that the real advantage of the heuristic algorithms is on large problems (roughly 100 or more locations). For problems smaller than that, a good commercial MIP solver (like CPLEX, for example) can solve the capacitated facility location problems with acceptable speed. For the computational experiments described here, CPLEX has been used to solve the lower bound problems, rather than coding a special-purpose algorithm. However, use of a special-purpose algorithm is an open possibility for extension of this work.

When CPLEX (or other standard MIP solver) is used for the LB problems, adding the constraint on the minimum number of seeds selected (constraint 63) poses no difficulty. However, it can affect the use of special-purpose algorithms designed for the capacitated facility location problem because constraint (63) is generally not included in those problems.

If a solution to problem (LB) is available, a feasible solution to problem (P4-3) can be constructed easily. For *ij* pairs where $x_{ij}^{s}=0$, set $w_{ij}^{s}=0$ and $y_{ij}^{s}=0$. If $x_{ij}^{s}>0$, set $w_{ij}^{s}=1$ and $y_{ij}^{s}=1$. Then evaluate objective function (46) to compute the total cost. This feasible solution creates an upper bound on the objective function value of the optimal solution to problem (P4-3). Because progressive hedging does not require an optimal solution to each individual scenario sub-problem, the feasible solutions created this way can serve as candidate solutions at each iteration. Of course, since these solutions may not be optimal, the progressive hedging algorithm may converge to a sub-optimal overall solution. This issue is evaluated in subsequent computational tests.

4.3.3 General Procedure

The progressive hedging algorithm proceeds through iterations, and the index k is used to indicate the iteration number. This affects the cost coefficients, f_j and g_{ij}, as well as the variables, and in what follows we'll use the notation: u_{j}^{sk}, y_{ij}^{sk}, w_{ij}^{sk} and x_{ij}^{sk} for the variables, and f_{j}^{sk} and g_{ij}^{sk} for the cost coefficients on u and y.

The algorithm uses a set of weights on the variables that are analogous to Lagrange multipliers for the non-anticipative constraints. These weights are defined as γ_{j}^{sk} and λ_{ij}^{sk}, associated with the u_{j}^{sk} and y_{ij}^{sk} variables, respectively. There is also a set of control parameters associated with the first-stage variables, which will be defined as ρ_j^k and ϕ_{ij}^k.

The algorithm can be stated as follows:

Step 1: Initialization

$k = 0$

$\gamma_j^{sk} = 0 \quad \forall s, j$

$\lambda_{ij}^{sk} = 0 \quad \forall i, j, s$

Solve the LB problem defined by (59)-(64) for each *s*, and create a feasible solution from it.

Denote the solutions constructed this way as u_j^{s1}, y_{ij}^{s1}, w_{ij}^{s1} and x_{ij}^{s1}.

Step 2: Iteration update

$k \leftarrow k+1$

For each location *j*, if u_j^{sk} is the same across all scenarios (all 0 or all 1), then set $\rho_j^k = f_j$

If there are differences among scenarios, then set $\rho_j^k = f_j / 2$.

For each *ij* pair, if the values of y_{ij}^{sk} are the same across all scenarios, then set $\phi_{ij}^k = g_{ij}$

If there are differences among scenarios, then set $\phi_{ij}^k = g_{ij} / 2$.

For each location *j*, compute $\bar{u}_j^k = \sum_s \pi_s u_j^{sk}$

For each *ij* pair, compute $\bar{y}_{ij}^{k} = \sum_s \pi_s y_{ij}^{sk}$

Step 3: Weight update

For each location *j* and scenario *s*: $\gamma_j^{sk} = \gamma_j^{s,k-1} + \rho_j^k \left(u_j^{sk} - \bar{u}_j^k \right)$

For each *ij* pair and scenario *s*: $\lambda_{ij}^{sk} = \lambda_{ij}^{s,k-1} + \phi_{ij}^k \left(y_{ij}^{sk} - \bar{y}_{ij}^k \right)$

Step 4: Coefficient update

For each location *j* and scenario *s*: $f_j^{sk} = f_j + \gamma_j^{sk} + \frac{\rho_j^k}{2}\left(1 - 2\bar{u}_j^k\right)$

For each *ij* pair and scenario *s*: $g_{ij}^{sk} = g_{ij} + \lambda_{ij}^{sk} + \frac{\phi_{ij}^k}{2}\left(1 - 2\bar{y}_{ij}^k\right)$

Step 5: Decomposition

For each scenario *s*, use the coefficients f_j^{sk} and g_{ij}^{sk} and solve problem LB for a new lower bound. Construct a feasible solution from it, and denote the solution variables as $u_j^{s,k+1}$, $y_{ij}^{s,k+1}$, $w_{ij}^{s,k+1}$ and $x_{ij}^{s,k+1}$.

Step 6: Check for termination

If all the $u_j^{s,k+1}$ agree across scenarios s, and all the $y_{ij}^{s,k+1}$ agree across scenarios *s*, then stop and report $u_j^{s,k+1}$, $y_{ij}^{s,k+1}$, $w_{ij}^{s,k+1}$ and $x_{ij}^{s,k+1}$ as the optimal solution. If not, return to Step 2.

This algorithm statement is an adaptation of the general progressive hedging method to the particular structure of the inbound material logistics problem with uncertain loads. One of the most important adaptations is the result of knowing that the first-stage variables are all binary. The general progressive hedging method forms the objective for the scenario *s* sub-problem in step 4 as:

$$\min \quad \sum_j f_j u_j^{s,k+1} + \sum_{i,j} g_{ij} y_{ij}^{s,k+1} + \sum_s \pi_s \sum_{i,j} c_{ij} w_{ij}^{s,k+1} + \sum_j \gamma_j^{sk} u_j^{s,k+1} + \sum_{i,j} \lambda_{ij}^{sk} y_{ij}^{s,k+1}$$
$$+ \frac{1}{2}\left[\sum_j \rho_j^k \left(u_j^{s,k+1} - \overline{u}_j^k \right)^2 + \sum_{i,j} \phi_{ij}^k \left(y_{ij}^{s,k+1} - \overline{y}_{ij}^k \right)^2 \right]$$

The terms involving the coefficients γ_j^{sk} and λ_{ij}^{sk} are the Lagrangian terms and the last set of terms creates a penalty for scenario-specific first-stage variables that deviate from the average across scenarios. This penalty increases the rate of convergence of the first-stage variables from the various scenario problems to a common value.

Because the first-stage variables are all binary, the use of the quadratic terms can be avoided. Consider, for example, one of the quadratic terms $\left(u_j^{s,k+1} - \overline{u}_j^k \right)^2$. Because $\overline{u}_j^k = \sum_s \pi_s u_j^{sk}$, the u_j^{sk} were all either 0 or 1, and the π_s are probabilities, the value of $\overline{u}_j^k$ is in the interval between 0 and 1, and is a constant for the solution of the scenario *s* sub-problem to obtain $u_j^{s,k+1}$, $y_{ij}^{s,k+1}$, $w_{ij}^{s,k+1}$ and $x_{ij}^{s,k+1}$.

We know that $\left(u_j^{s,k+1} - \overline{u}_j^k \right)^2 = \left(u_j^{s,k+1} \right)^2 - 2\overline{u}_j^k u_j^{s,k+1} + \left(\overline{u}_j^k \right)^2$, and because $u_j^{s,k+1}$ must be either 0 or 1, $\left(u_j^{s,k+1} \right)^2 = u_j^{s,k+1}$. Thus, the quadratic term can be written as

$\left(u_j^{s,k+1} - \overline{u}_j^{k}\right)^2 = \left(1 - 2\overline{u}_j^{k}\right) u_j^{s,k+1} + \text{constant}$. In the optimization, the constants can be ignored and we can group terms with adjusted coefficients, as shown in step 4 above. These adjusted coefficients are used in the original objective function formulation for solution of the sub-problems.

The termination check in step 6 can also be modified slightly to speed convergence of the algorithm. In computational tests, it has been noted that the convergence of the non-anticipative constraints $u_j^s = u_j \quad \forall s$ occurs fairly quickly, but that the $y_{ij}^s = y_{ij} \quad \forall s$ constraints require many more iterations. That is, the seed locations are chosen fairly quickly, but the set of associations for individual suppliers converges much more slowly. This occurs because not all associations are used in every scenario, so the individual scenario solutions tend to have too few $y_{ij}^s = 1$.

Because progressive hedging is being used as an heuristic solution method, an effective strategy for reaching a solution is to terminate the algorithm when "most" of the scenarios agree on a set of associations: $y_{ij}^s = 1$. These can then be reported as the final values: $y_{ij} = 1$. The potential cost of this termination criterion is that a few too many associations may be reported in the final solution. This is also evaluated in subsequent computational experiments.

4.4 *A Case Study for Computational Testing*

In chapter 3, a model is formulated for designing inbound material collection routes to manufacturing facilities that incorporates uncertainty in the pickup quantities at individual suppliers, the ability to use split pickups (assigning parts of a suppliers load to different routes), and operational controls on the degree of splitting that is allowed. An example analysis in chapter 3 focuses on a small-scale problem with 14 suppliers and 10 different load scenarios. For that small problem, it is reasonable to solve the two-stage stochastic optimization problem in extensive form using a commercial mixed integer programming solver (CPLEX).

The purpose of this section is to describe a larger case study in the use of the model, demonstrating its value in a practical situation. The plant for the case study is an automotive assembly plant in Kansas City, Kansas, from which daily collection routes operate to gather inbound material from 44 suppliers located in 16 U.S. states. Over the period during which supplier shipment data were collected, the average total daily quantity picked up was 31,361 ft^3 or about 11 truckloads. The routes operated originate at the plant, typically visit 2-7 suppliers and then return to the plant. Over the period observed (33 weeks of production), the total daily quantities for pickup at these suppliers ranged from 17,182 ft^3 to 44,102 ft^3. This wide variation in total quantity makes it quite difficult to design a stable set of collection routes.

In addition to describing a larger scale application of the model formulation, this section also discusses how the proposed heuristic methods in the previous sections work. An attempt to solve the case study formulation in extensive form with CPLEX proved unsuccessful. For example, in a 10-scenario example, after approximately 9 hours of computation on a Windows-based personal computer with 2.00 GHz processor and 8.00 GB memory, a feasible solution had been obtained, but the gap between the objective value for that solution and the lower bound on the objective was still 3.4% and the run aborted due to insufficient memory to store the branching tree. While it would be possible to increase the computer hardware capability brought to bear on the problem, the long computation times experienced to date with the small set of scenarios indicate that solving the problem in extensive form is unlikely to be a successful strategy for actual applications. Thus, it is important to use the decomposition approaches to solving the stochastic optimization.

4.4.1 The Case Study Setting

Figure 4-3 illustrates the plant location and the locations of the 44 suppliers. Many of the suppliers are located in Michigan, Indiana and Ohio, the core of the automotive industry, but in total the suppliers cover a very wide geographic area. The large distances involved in making collections from these suppliers and transporting material to the plant imply that an efficient route structure is important monetarily. The size of the average pickup quantity varies quite widely across suppliers, from 127 ft^3 per day to 1,859 ft^3 per day. The trucks used for the collection routes have a nominal capacity of 3,000 ft^3, so on average the individual supplier loads range from about 4% of a truckload to about 62% of a truckload.

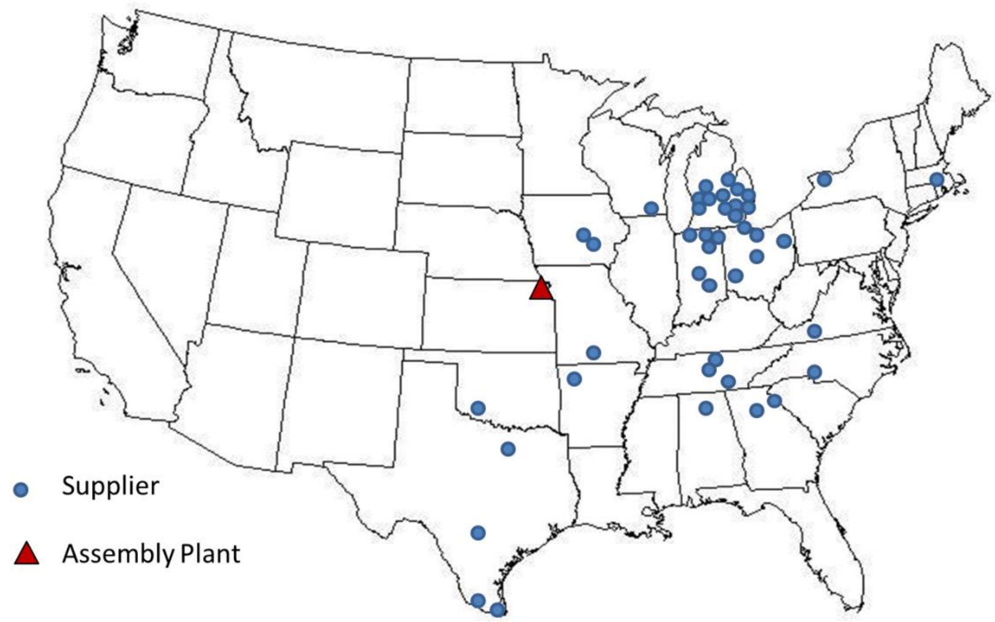
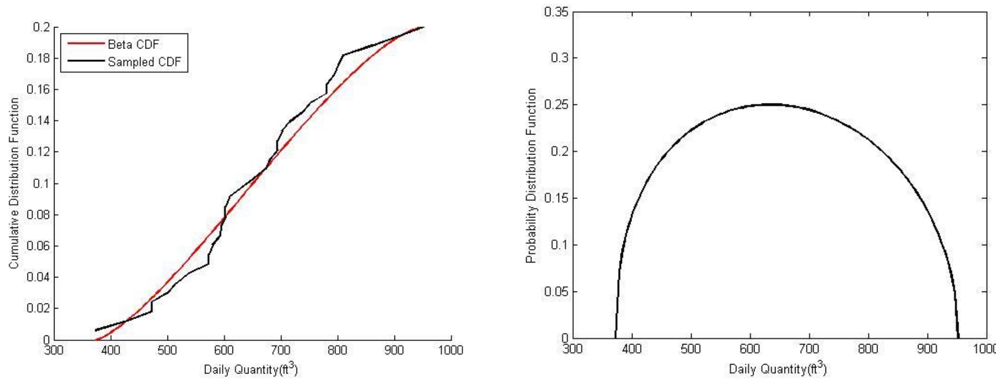


Figure 4-3. Locations of plant and suppliers.

The plant operates five days per week with a production schedule that is determined on a weekly basis. Orders for materials are placed with suppliers as a weekly quantity, with the total to be divided into five equal parts for daily pickup. Thus, within a week, the pickup quantities at individual suppliers are known and constant, but between weeks there can be substantial variation. Data on weekly quantities picked up from all suppliers over a 33-week period were collected and analyzed to characterize the week-to-week variability.

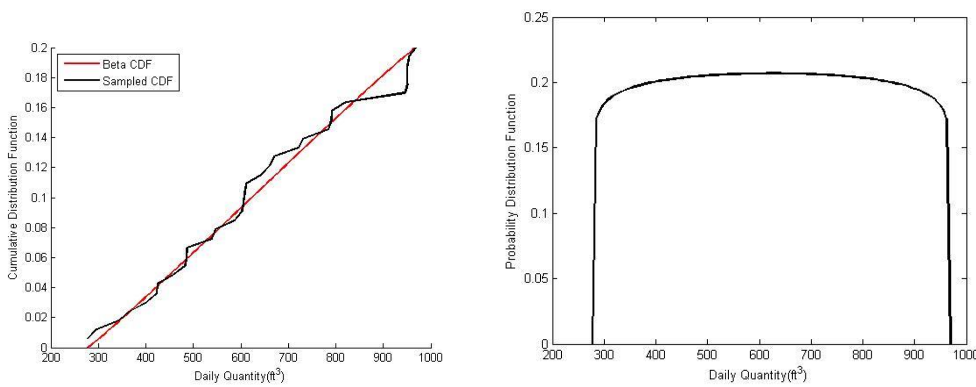
Because the week-to-week variation in quantities to be picked up is a result of plant production schedules and not unknown exogenous influences, there is a legitimate question about whether the variability in this context should be treated as uncertainty or simply as predictable variations over time.

For each supplier, a Beta distribution was estimated to fit the observed set of quantities. The quantities were scaled to daily values, rather than weekly, because that is how they are used in the routing analysis. Figures 4-4 to 4-7 show four examples of the fitted vs. sample cumulative distribution functions (part a) and the associated estimate of the probability density function (part b). These four examples illustrate that individual suppliers have quite different distributions of quantity. The flexibility of the Beta distribution, however, accommodates these widely varying individual load distributions. None of the 44 estimated Beta distributions could be rejected at the 5% level using a Kolmogorov-Smirnov test of goodness-of-fit, so the Beta family appears to be a good choice for the representation of supplier loads. Table 4-1 summarizes the parameters (min, max, α and β) for all of the distributions. Table 4-1 also shows the resulting values of mean and standard deviation for the estimated distributions, and the calculation of the coefficient of variation. Most of the suppliers (35 of 44) have coefficients of variation between 0.19 and 0.34, and five of the suppliers have a coefficient of variation of 0.5 or above. This indicates quite large week-to-week variability in the quantities to be picked up from individual suppliers.



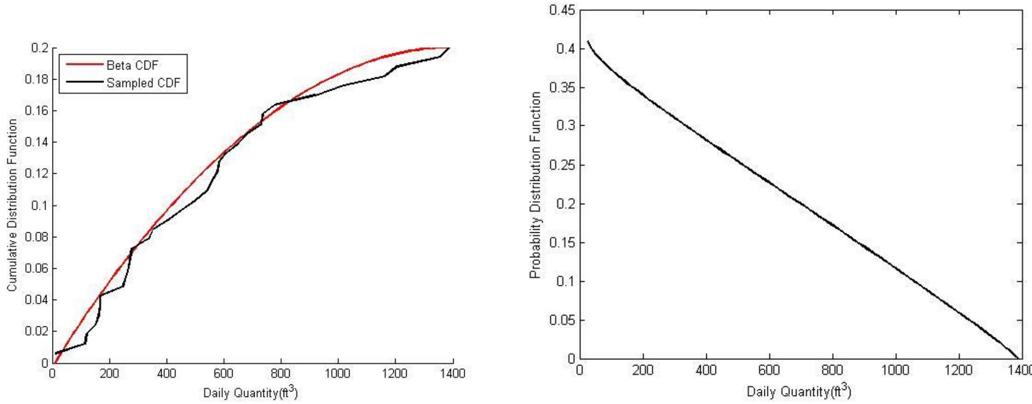
(a) (b)

Figure 4-4. Sample CDF for daily loads from supplier 2, with associated fitted Beta distribution.



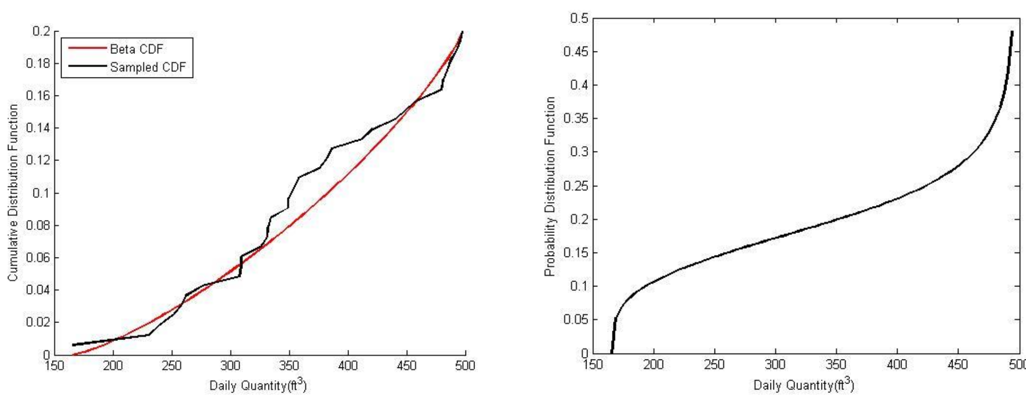
(a) (b)

Figure 4-5. Sample CDF for daily loads from supplier 3, with associated fitted Beta distribution.



(a) (b)

Figure 4-6. Sample CDF for daily loads from supplier 9, with associated fitted Beta distribution.



(a) (b)

Figure 4-7. Sample CDF for daily loads from supplier 25, with associated fitted Beta distribution.

Table 4-1. Summary of parameters for estimated Beta distributions of supplier daily shipments.

Supplier	Zip Code	State	Min Daily Shipment (ft^3)	Max Daily Shipment (ft^3)	α Parameter	β Parameter	Average Daily Shipment (ft^3)	Standard Deviation (ft^3)	Coefficient of Variation
1	02038	MA	260	750	0.953	0.780	529	147	0.28
2	14020	NY	370	960	1.404	1.492	656	149	0.23
3	24382	VA	270	1000	1.057	1.057	635	207	0.33
4	28273	NC	230	530	1.180	0.933	398	84	0.21
5	30024	GA	300	700	1.191	1.222	497	108	0.22
6	30643	GA	190	430	1.537	1.618	307	59	0.19
7	35824	AL	120	320	1.598	1.618	219	49	0.22
8	37066	TN	750	1600	0.770	0.719	1190	269	0.23
9	37214	TN	10	1580	0.978	1.926	539	376	0.70
10	37398	TN	100	1520	0.762	1.407	599	381	0.64
11	43015	OH	370	830	1.166	1.037	613	128	0.21
12	43528	OH	400	1150	0.922	0.852	790	225	0.28
13	44144	OH	430	1860	0.838	1.237	1008	400	0.40
14	44316	OH	270	2290	0.720	0.692	1300	650	0.50
15	44883	OH	690	1420	0.860	0.660	1103	228	0.21
16	45420	OH	590	1280	1.321	1.181	954	184	0.19
17	46176	IN	430	1610	0.880	0.744	1069	363	0.34
18	46222	IN	770	2800	0.822	0.719	1853	635	0.34
19	46526	IN	120	410	0.840	0.649	284	91	0.32
20	46530	IN	620	1750	0.807	0.717	1218	355	0.29
21	46705	IN	110	250	1.504	1.485	180	35	0.19
22	46706	IN	290	990	0.860	0.719	671	217	0.32
23	48026	MI	130	490	0.982	0.938	314	105	0.33
24	48036	MI	1000	2240	1.149	0.898	1696	353	0.21
25	48047	MI	160	500	1.292	0.812	369	94	0.25
26	48066	MI	120	470	1.142	0.989	308	99	0.32
27	48089	MI	190	440	1.334	1.214	321	66	0.21
28	48174	MI	0	700	1.667	2.027	316	161	0.51
29	48706	MI	110	460	1.887	1.984	281	79	0.28
30	49250	MI	400	930	1.137	0.963	687	150	0.22
31	49307	MI	310	800	0.931	0.853	566	147	0.26
32	49417	MI	270	2500	0.815	0.625	1532	708	0.46
33	49507	MI	170	610	0.868	0.760	405	135	0.33
34	49512	MI	370	1210	1.122	0.951	825	239	0.29
35	52347	IA	240	550	1.387	1.241	404	81	0.20
36	52361	IA	90	220	1.824	1.441	163	31	0.19
37	53115	WI	340	830	0.956	0.907	591	145	0.25
38	65714	MO	110	420	0.998	0.914	272	91	0.33
39	72701	AR	22	280	0.814	1.191	127	73	0.57
40	73542	OK	670	2380	0.737	0.593	1618	557	0.34
41	75051	TX	380	1550	0.904	0.872	976	351	0.36
42	78154	TX	220	460	1.009	0.901	347	70	0.20
43	78503	TX	1080	2600	0.959	0.912	1859	448	0.24
44	78521	TX	540	1360	0.710	0.584	990	269	0.27

It is also important to recognize that the quantities from individual suppliers do not vary independently because the variation is driven by changes in the production schedule at the assembly plant. Two suppliers that produce parts or material that is used in all production units will have positively correlated pick-up quantities because their volumes are both related to total production volume at the plant. For example, suppliers of headlights and brake discs are likely to have positively correlated volumes because these are parts used on all production units. Conversely, suppliers that produce parts for product options may have negatively correlated volumes because production units will have some options and not others. For example, suppliers of base-model sound systems and in-vehicle navigation systems are likely to have negatively correlated volumes because these options are normally found on different trim levels and a production mix that increases demand for one is likely to decrease demand for the other.

For the suppliers in this case, the correlation matrix of load quantities is shown in Table 4-2. Most of the correlations are positive and many of them are quite high ($\geq$ 0.7). A few pairs (e.g., suppliers 3 and 25, 12 and 20) are perfectly correlated, indicating parts that are used on all vehicles. Supplier 39 is the one most notably negatively correlated with several other suppliers.

Table 4-2. Correlation matrix of load quantities.

Supplier#	1	2	3	4	5	6	7	8	9	10	11
1	1.000	0.822	0.961	0.876	0.880	0.662	0.498	0.955	0.572	0.837	0.899
2	0.822	1.000	0.670	0.943	0.943	0.815	0.777	0.912	0.401	0.581	0.938
3	0.961	0.670	1.000	0.745	0.750	0.530	0.359	0.872	0.564	0.903	0.775
4	0.876	0.943	0.745	1.000	1.000	0.837	0.765	0.973	0.444	0.606	0.995
5	0.880	0.943	0.750	1.000	1.000	0.833	0.758	0.975	0.441	0.611	0.996
6	0.662	0.815	0.530	0.837	0.833	1.000	0.898	0.780	0.323	0.489	0.807
7	0.498	0.777	0.359	0.765	0.758	0.898	1.000	0.669	0.253	0.336	0.713
8	0.955	0.912	0.872	0.973	0.975	0.780	0.669	1.000	0.506	0.748	0.984
9	0.572	0.401	0.564	0.444	0.441	0.323	0.253	0.506	1.000	0.540	0.456
10	0.837	0.581	0.903	0.606	0.611	0.489	0.336	0.748	0.540	1.000	0.640
11	0.899	0.938	0.775	0.995	0.996	0.807	0.713	0.984	0.456	0.640	1.000
12	0.985	0.755	0.990	0.829	0.833	0.614	0.457	0.932	0.565	0.882	0.854
13	0.712	0.467	0.691	0.544	0.551	0.284	0.098	0.625	0.233	0.438	0.581
14	0.398	0.262	0.462	0.349	0.344	0.244	0.204	0.392	0.198	0.322	0.338
15	0.889	0.944	0.763	0.996	0.997	0.830	0.753	0.981	0.456	0.633	0.997
16	0.578	0.659	0.487	0.598	0.591	0.632	0.466	0.580	0.388	0.383	0.582
17	0.961	0.668	0.999	0.745	0.750	0.525	0.353	0.873	0.564	0.901	0.776
18	0.962	0.674	0.997	0.758	0.762	0.534	0.364	0.879	0.564	0.889	0.786
19	0.967	0.693	0.997	0.761	0.766	0.546	0.372	0.884	0.557	0.903	0.791
20	0.984	0.749	0.992	0.822	0.827	0.606	0.445	0.927	0.570	0.883	0.848
21	0.897	0.934	0.789	0.974	0.975	0.799	0.713	0.970	0.427	0.670	0.977
22	0.957	0.660	0.997	0.732	0.737	0.521	0.341	0.863	0.555	0.889	0.763
23	0.960	0.666	1.000	0.741	0.746	0.523	0.349	0.870	0.566	0.902	0.772
24	0.912	0.915	0.819	0.965	0.968	0.802	0.716	0.983	0.467	0.712	0.976
25	0.954	0.907	0.845	0.942	0.944	0.751	0.609	0.968	0.544	0.722	0.958
26	0.966	0.681	0.997	0.764	0.768	0.540	0.369	0.884	0.573	0.895	0.793
27	0.917	0.919	0.820	0.951	0.955	0.777	0.676	0.975	0.470	0.717	0.967
28	0.307	0.377	0.187	0.432	0.427	0.434	0.455	0.372	0.220	0.045	0.409
29	0.860	0.573	0.885	0.687	0.690	0.401	0.259	0.783	0.493	0.617	0.706
30	0.936	0.916	0.847	0.955	0.958	0.778	0.670	0.985	0.474	0.739	0.970
31	0.989	0.793	0.977	0.867	0.871	0.660	0.505	0.955	0.558	0.865	0.889
32	0.946	0.638	0.958	0.688	0.693	0.479	0.266	0.815	0.574	0.843	0.723
33	0.968	0.685	0.998	0.763	0.767	0.541	0.367	0.884	0.563	0.895	0.792
34	0.765	0.609	0.758	0.666	0.668	0.533	0.447	0.733	0.461	0.625	0.669
35	0.913	0.928	0.811	0.974	0.977	0.809	0.719	0.986	0.481	0.703	0.984
36	0.899	0.932	0.792	0.959	0.963	0.773	0.685	0.972	0.436	0.687	0.974
37	0.984	0.842	0.944	0.917	0.920	0.690	0.559	0.980	0.545	0.816	0.935
38	0.961	0.670	0.999	0.745	0.750	0.524	0.354	0.873	0.570	0.898	0.775
39	-0.109	0.063	-0.140	0.113	0.106	0.049	0.335	0.029	0.140	-0.160	0.079
40	0.961	0.670	1.000	0.745	0.750	0.528	0.354	0.873	0.566	0.902	0.776
41	0.955	0.658	0.997	0.731	0.735	0.500	0.335	0.861	0.565	0.912	0.763
42	0.866	0.941	0.733	0.993	0.994	0.843	0.757	0.968	0.404	0.610	0.992
43	0.982	0.849	0.945	0.918	0.920	0.736	0.606	0.980	0.540	0.833	0.931
44	0.991	0.805	0.973	0.876	0.880	0.663	0.514	0.961	0.558	0.858	0.898

(a)

Supplier#	12	13	14	15	16	17	18	19	20	21	22
1	0.985	0.712	0.398	0.889	0.578	0.961	0.962	0.967	0.984	0.897	0.957
2	0.755	0.467	0.262	0.944	0.659	0.668	0.674	0.693	0.749	0.934	0.660
3	0.990	0.691	0.462	0.763	0.487	0.999	0.997	0.997	0.992	0.789	0.997
4	0.829	0.544	0.349	0.996	0.598	0.745	0.758	0.761	0.822	0.974	0.732
5	0.833	0.551	0.344	0.997	0.591	0.750	0.762	0.766	0.827	0.975	0.737
6	0.614	0.284	0.244	0.830	0.632	0.525	0.534	0.546	0.606	0.799	0.521
7	0.457	0.098	0.204	0.753	0.466	0.353	0.364	0.372	0.445	0.713	0.341
8	0.932	0.625	0.392	0.981	0.580	0.873	0.879	0.884	0.927	0.970	0.863
9	0.565	0.233	0.198	0.456	0.388	0.564	0.564	0.557	0.570	0.427	0.555
10	0.882	0.438	0.322	0.633	0.383	0.901	0.889	0.903	0.883	0.670	0.889
11	0.854	0.581	0.338	0.997	0.582	0.776	0.786	0.791	0.848	0.977	0.763
12	1.000	0.694	0.452	0.845	0.527	0.990	0.991	0.992	1.000	0.862	0.986
13	0.694	1.000	-0.017	0.560	0.317	0.695	0.704	0.696	0.694	0.561	0.720
14	0.452	-0.017	1.000	0.328	0.370	0.462	0.465	0.453	0.454	0.360	0.450
15	0.845	0.560	0.328	1.000	0.575	0.763	0.773	0.778	0.838	0.974	0.751
16	0.527	0.317	0.370	0.575	1.000	0.486	0.496	0.511	0.526	0.581	0.500
17	0.990	0.695	0.462	0.763	0.486	1.000	0.998	0.998	0.992	0.791	0.997
18	0.991	0.704	0.465	0.773	0.496	0.998	1.000	0.997	0.992	0.800	0.995
19	0.992	0.696	0.453	0.778	0.511	0.998	0.997	1.000	0.993	0.809	0.996
20	1.000	0.694	0.454	0.838	0.526	0.992	0.992	0.993	1.000	0.856	0.987
21	0.862	0.561	0.360	0.974	0.581	0.791	0.800	0.809	0.856	1.000	0.778
22	0.986	0.720	0.450	0.751	0.500	0.997	0.995	0.996	0.987	0.778	1.000
23	0.989	0.695	0.459	0.759	0.488	1.000	0.998	0.998	0.991	0.787	0.998
24	0.887	0.567	0.355	0.980	0.544	0.818	0.821	0.828	0.880	0.959	0.808
25	0.903	0.673	0.279	0.951	0.624	0.845	0.849	0.859	0.898	0.934	0.841
26	0.992	0.700	0.454	0.779	0.499	0.998	0.999	0.996	0.993	0.805	0.994
27	0.884	0.574	0.346	0.968	0.544	0.819	0.815	0.830	0.879	0.952	0.809
28	0.240	0.061	0.327	0.425	0.229	0.182	0.177	0.187	0.235	0.342	0.172
29	0.881	0.726	0.633	0.690	0.483	0.887	0.895	0.878	0.883	0.705	0.893
30	0.907	0.587	0.381	0.970	0.553	0.846	0.843	0.857	0.902	0.957	0.836
31	0.996	0.684	0.449	0.880	0.553	0.977	0.979	0.981	0.995	0.894	0.971
32	0.943	0.745	0.357	0.705	0.515	0.957	0.954	0.956	0.945	0.735	0.961
33	0.993	0.702	0.461	0.779	0.499	0.999	0.998	0.998	0.994	0.807	0.996
34	0.769	0.349	0.578	0.676	0.397	0.755	0.748	0.752	0.770	0.695	0.740
35	0.881	0.564	0.342	0.987	0.562	0.810	0.812	0.821	0.875	0.964	0.798
36	0.863	0.554	0.344	0.975	0.532	0.791	0.789	0.804	0.857	0.960	0.779
37	0.980	0.673	0.440	0.926	0.558	0.945	0.950	0.951	0.977	0.934	0.936
38	0.990	0.694	0.466	0.763	0.491	1.000	0.998	0.997	0.992	0.791	0.997
39	-0.090	-0.280	0.073	0.104	-0.148	-0.139	-0.127	-0.152	-0.097	0.087	-0.165
40	0.990	0.693	0.464	0.763	0.494	1.000	0.998	0.998	0.992	0.790	0.997
41	0.985	0.685	0.448	0.748	0.473	0.997	0.995	0.995	0.986	0.778	0.993
42	0.818	0.545	0.312	0.994	0.567	0.732	0.744	0.749	0.811	0.970	0.720
43	0.980	0.649	0.443	0.927	0.578	0.944	0.949	0.951	0.977	0.931	0.937
44	0.996	0.689	0.441	0.890	0.549	0.974	0.976	0.978	0.994	0.900	0.968

(b)

Supplier#	23	24	25	26	27	28	29	30	31	32	33
1	0.960	0.912	0.954	0.966	0.917	0.307	0.860	0.936	0.989	0.946	0.968
2	0.666	0.915	0.907	0.681	0.919	0.377	0.573	0.916	0.793	0.638	0.685
3	1.000	0.819	0.845	0.997	0.820	0.187	0.885	0.847	0.977	0.958	0.998
4	0.741	0.965	0.942	0.764	0.951	0.432	0.687	0.955	0.867	0.688	0.763
5	0.746	0.968	0.944	0.768	0.955	0.427	0.690	0.958	0.871	0.693	0.767
6	0.523	0.802	0.751	0.540	0.777	0.434	0.401	0.778	0.660	0.479	0.541
7	0.349	0.716	0.609	0.369	0.676	0.455	0.259	0.670	0.505	0.266	0.367
8	0.870	0.983	0.968	0.884	0.975	0.372	0.783	0.985	0.955	0.815	0.884
9	0.566	0.467	0.544	0.573	0.470	0.220	0.493	0.474	0.558	0.574	0.563
10	0.902	0.712	0.722	0.895	0.717	0.045	0.617	0.739	0.865	0.843	0.895
11	0.772	0.976	0.958	0.793	0.967	0.409	0.706	0.970	0.889	0.723	0.792
12	0.989	0.887	0.903	0.992	0.884	0.240	0.881	0.907	0.996	0.943	0.993
13	0.695	0.567	0.673	0.700	0.574	0.061	0.726	0.587	0.684	0.745	0.702
14	0.459	0.355	0.279	0.454	0.346	0.327	0.633	0.381	0.449	0.357	0.461
15	0.759	0.980	0.951	0.779	0.968	0.425	0.690	0.970	0.880	0.705	0.779
16	0.488	0.544	0.624	0.499	0.544	0.229	0.483	0.553	0.553	0.515	0.499
17	1.000	0.818	0.845	0.998	0.819	0.182	0.887	0.846	0.977	0.957	0.999
18	0.998	0.821	0.849	0.999	0.815	0.177	0.895	0.843	0.979	0.954	0.998
19	0.998	0.828	0.859	0.996	0.830	0.187	0.878	0.857	0.981	0.956	0.998
20	0.991	0.880	0.898	0.993	0.879	0.235	0.883	0.902	0.995	0.945	0.994
21	0.787	0.959	0.934	0.805	0.952	0.342	0.705	0.957	0.894	0.735	0.807
22	0.998	0.808	0.841	0.994	0.809	0.172	0.893	0.836	0.971	0.961	0.996
23	1.000	0.816	0.843	0.997	0.817	0.176	0.886	0.844	0.976	0.959	0.999
24	0.816	1.000	0.939	0.828	0.982	0.346	0.719	0.986	0.912	0.750	0.828
25	0.843	0.939	1.000	0.856	0.951	0.407	0.745	0.957	0.925	0.848	0.859
26	0.997	0.828	0.856	1.000	0.821	0.179	0.887	0.849	0.981	0.956	0.998
27	0.817	0.982	0.951	0.821	1.000	0.407	0.716	0.997	0.909	0.765	0.829
28	0.176	0.346	0.407	0.179	0.407	1.000	0.282	0.402	0.277	0.190	0.194
29	0.886	0.719	0.745	0.887	0.716	0.282	1.000	0.747	0.869	0.855	0.889
30	0.844	0.986	0.957	0.849	0.997	0.402	0.747	1.000	0.930	0.791	0.856
31	0.976	0.912	0.925	0.981	0.909	0.277	0.869	0.930	1.000	0.927	0.982
32	0.959	0.750	0.848	0.956	0.765	0.190	0.855	0.791	0.927	1.000	0.959
33	0.999	0.828	0.859	0.998	0.829	0.194	0.889	0.856	0.982	0.959	1.000
34	0.756	0.722	0.664	0.750	0.714	0.384	0.737	0.737	0.769	0.720	0.751
35	0.807	0.997	0.951	0.820	0.986	0.373	0.711	0.989	0.909	0.748	0.821
36	0.789	0.982	0.943	0.796	0.995	0.407	0.695	0.992	0.891	0.733	0.803
37	0.943	0.945	0.948	0.953	0.940	0.316	0.856	0.956	0.989	0.888	0.952
38	1.000	0.819	0.845	0.997	0.820	0.182	0.891	0.847	0.977	0.958	0.999
39	-0.144	0.105	-0.052	-0.123	0.038	0.030	-0.071	0.027	-0.076	-0.240	-0.143
40	1.000	0.818	0.845	0.998	0.819	0.178	0.887	0.846	0.977	0.957	0.999
41	0.997	0.803	0.837	0.995	0.806	0.166	0.873	0.834	0.971	0.957	0.995
42	0.729	0.969	0.936	0.751	0.953	0.399	0.657	0.956	0.857	0.673	0.750
43	0.943	0.948	0.945	0.952	0.940	0.314	0.839	0.958	0.991	0.887	0.952
44	0.972	0.922	0.931	0.978	0.918	0.275	0.869	0.938	0.999	0.924	0.979

(c)

Supplier#	34	35	36	37	38	39	40	41	42	43	44
1	0.765	0.913	0.899	0.984	0.961	-0.109	0.961	0.955	0.866	0.982	0.991
2	0.609	0.928	0.932	0.842	0.670	0.063	0.670	0.658	0.941	0.849	0.805
3	0.758	0.811	0.792	0.944	0.999	-0.140	1.000	0.997	0.733	0.945	0.973
4	0.666	0.974	0.959	0.917	0.745	0.113	0.745	0.731	0.993	0.918	0.876
5	0.668	0.977	0.963	0.920	0.750	0.106	0.750	0.735	0.994	0.920	0.880
6	0.533	0.809	0.773	0.690	0.524	0.049	0.528	0.500	0.843	0.736	0.663
7	0.447	0.719	0.685	0.559	0.354	0.335	0.354	0.335	0.757	0.606	0.514
8	0.733	0.986	0.972	0.980	0.873	0.029	0.873	0.861	0.968	0.980	0.961
9	0.461	0.481	0.436	0.545	0.570	0.140	0.566	0.565	0.404	0.540	0.558
10	0.625	0.703	0.687	0.816	0.898	-0.160	0.902	0.912	0.610	0.833	0.858
11	0.669	0.984	0.974	0.935	0.775	0.079	0.776	0.763	0.992	0.931	0.898
12	0.769	0.881	0.863	0.980	0.990	-0.090	0.990	0.985	0.818	0.980	0.996
13	0.349	0.564	0.554	0.673	0.694	-0.280	0.693	0.685	0.545	0.649	0.689
14	0.578	0.342	0.344	0.440	0.466	0.073	0.464	0.448	0.312	0.443	0.441
15	0.676	0.987	0.975	0.926	0.763	0.104	0.763	0.748	0.994	0.927	0.890
16	0.397	0.562	0.532	0.558	0.491	-0.148	0.494	0.473	0.567	0.578	0.549
17	0.755	0.810	0.791	0.945	1.000	-0.139	1.000	0.997	0.732	0.944	0.974
18	0.748	0.812	0.789	0.950	0.998	-0.127	0.998	0.995	0.744	0.949	0.976
19	0.752	0.821	0.804	0.951	0.997	-0.152	0.998	0.995	0.749	0.951	0.978
20	0.770	0.875	0.857	0.977	0.992	-0.097	0.992	0.986	0.811	0.977	0.994
21	0.695	0.964	0.960	0.934	0.791	0.087	0.790	0.778	0.970	0.931	0.900
22	0.740	0.798	0.779	0.936	0.997	-0.165	0.997	0.993	0.720	0.937	0.968
23	0.756	0.807	0.789	0.943	1.000	-0.144	1.000	0.997	0.729	0.943	0.972
24	0.722	0.997	0.982	0.945	0.819	0.105	0.818	0.803	0.969	0.948	0.922
25	0.664	0.951	0.943	0.948	0.845	-0.052	0.845	0.837	0.936	0.945	0.931
26	0.750	0.820	0.796	0.953	0.997	-0.123	0.998	0.995	0.751	0.952	0.978
27	0.714	0.986	0.995	0.940	0.820	0.038	0.819	0.806	0.953	0.940	0.918
28	0.384	0.373	0.407	0.316	0.182	0.030	0.178	0.166	0.399	0.314	0.275
29	0.737	0.711	0.695	0.856	0.891	-0.071	0.887	0.873	0.657	0.839	0.869
30	0.737	0.989	0.992	0.956	0.847	0.027	0.846	0.834	0.956	0.958	0.938
31	0.769	0.909	0.891	0.989	0.977	-0.076	0.977	0.971	0.857	0.991	0.999
32	0.720	0.748	0.733	0.888	0.958	-0.240	0.957	0.957	0.673	0.887	0.924
33	0.751	0.821	0.803	0.952	0.999	-0.143	0.999	0.995	0.750	0.952	0.979
34	1.000	0.718	0.705	0.760	0.761	0.056	0.754	0.741	0.639	0.764	0.769
35	0.718	1.000	0.987	0.944	0.811	0.095	0.810	0.795	0.977	0.947	0.918
36	0.705	0.987	1.000	0.928	0.793	0.063	0.791	0.777	0.962	0.926	0.901
37	0.760	0.944	0.928	1.000	0.944	-0.015	0.945	0.937	0.906	0.993	0.993
38	0.761	0.811	0.793	0.944	1.000	-0.135	0.999	0.996	0.732	0.944	0.973
39	0.056	0.095	0.063	-0.015	-0.135	1.000	-0.143	-0.138	0.076	-0.017	-0.064
40	0.754	0.810	0.791	0.945	0.999	-0.143	1.000	0.997	0.733	0.945	0.974
41	0.741	0.795	0.777	0.937	0.996	-0.138	0.997	1.000	0.718	0.936	0.966
42	0.639	0.977	0.962	0.906	0.732	0.076	0.733	0.718	1.000	0.908	0.867
43	0.764	0.947	0.926	0.993	0.944	-0.017	0.945	0.936	0.908	1.000	0.993
44	0.769	0.918	0.901	0.993	0.973	-0.064	0.974	0.966	0.867	0.993	1.000

(d)

Positive correlations are very important for route planning because the variance in total load quantity on constructed routes will be larger than the sum of the variances of the individual supplier loads. However, nearly all treatments of vehicle routing problems with stochastic demand assume independence among the loads at individual stops. The approach taken for the model used here incorporates the correlations into the route construction. This is done by defining a collection of discrete load scenarios with associated probabilities and using those scenarios as the basis for a stochastic optimization of the collection routes. The scenarios generated for the stochastic model should be a relatively small set, but defined so that the range of possible overall daily loads is well-covered, the correlations among individual supplier quantities are represented, and the relative likelihoods of different loading patterns are reflected in the set.

The method of Latin Hypercube Sampling with Correlations (LHSC) described in chapter 3 is an effective way of generating sample load scenarios. It can be used to generate any desired number of scenarios, given inputs of the marginal load distributions for individual suppliers (as summarized in Table 4-1 above) and the rank correlation between pairs of suppliers (as summarized in Table 4-2).

4.4.2 Computational Experiments

The computational experiments discussed here focus on versions of the case study problem using 10 and 25 scenarios and using penalty coefficients on associations of $g_{ij} = 0.1\ d_{ij}$ for all ij pairs. The 10-scenario instance of the problem has about 40,000 variables (about 21,000 of which are binary) and about 60,000 constraints. As mentioned earlier, an attempt to solve the 10-scenario version of the problem using CPLEX failed, with the run aborting after approximately 9 hours of computation. However, at that point a feasible solution was available with a gap of 3.5% between the solution and the lower bound. That solution is used here as a benchmark against which to compare the solutions from the heuristic decomposition algorithms.

Table 4-3 summarizes the daily pick-up quantities for the 10 scenarios, each of which has probability $\pi_s = 0.1$. The maximum daily pickup quantity in the 10 scenarios is 34,527 ft^3 (scenario 10), so at least 12 routes are required in the overall solution.

Table 4-3. Daily pick-up quantities (ft^3) for the 44 suppliers considering 10 scenarios.

	Scenario									
Supplier	1	2	3	4	5	6	7	8	9	10
1	739	615	702	644	312	512	368	557	405	468
2	486	610	817	750	428	801	890	527	576	662
3	515	906	426	841	386	701	755	316	984	599
4	259	470	313	446	351	428	289	486	522	393
5	303	412	429	694	553	461	572	646	497	385
6	329	217	237	323	365	296	259	396	346	284
7	213	152	245	277	179	220	202	253	155	312
8	1403	1022	1237	841	801	1107	1538	1379	989	1595
9	466	322	33	802	1148	249	655	491	170	946
10	610	243	491	212	138	409	1081	787	1420	940
11	736	664	497	425	438	655	530	809	761	577
12	646	502	930	466	563	803	989	1087	716	1124
13	1260	630	739	1058	456	1699	550	1113	1526	929
14	623	284	2281	1040	1993	1309	1671	507	1529	2141
15	1381	1321	1096	1011	712	853	1398	914	1262	1201
16	892	1061	696	1190	1209	977	828	1136	594	931
17	1331	1185	1559	1429	1493	524	628	933	1062	715
18	2695	2344	1313	2045	973	1565	925	2763	1648	2133
19	123	218	232	369	175	405	383	343	272	310
20	1691	800	1332	1710	699	946	1097	1125	1505	1395
21	144	216	205	170	198	118	230	169	188	154
22	876	925	353	666	801	718	594	462	990	404
23	329	385	187	418	288	444	226	273	158	471
24	1787	2098	1992	1347	1693	1845	1548	1139	1187	2212
25	492	360	346	259	449	405	206	468	285	438
26	313	378	351	127	170	231	304	463	258	418
27	245	355	316	431	275	411	293	384	323	212
28	189	593	391	453	336	112	296	473	47	221
29	169	211	201	303	351	311	418	246	265	360
30	866	707	473	544	914	611	763	643	823	464
31	509	360	345	749	769	699	593	627	544	420
32	1016	1349	1927	2402	1177	454	678	2439	1621	2262
33	577	601	239	495	289	383	458	211	324	536
34	1129	631	1191	717	944	1033	454	550	864	783
35	427	356	274	304	404	434	337	468	543	501
36	187	216	169	154	183	98	138	197	128	161
37	620	714	823	757	395	496	450	367	648	559
38	219	349	266	139	282	409	208	385	167	316
39	27	155	135	111	206	240	171	46	60	74
40	2195	814	1912	1676	1189	673	2380	1423	1541	2220
41	876	1327	417	840	1377	1498	1146	506	1052	712
42	231	389	322	452	369	434	263	343	405	292
43	1789	2374	1399	1238	2226	2073	2571	1967	1704	1127
44	714	923	618	1315	1350	1229	546	1071	897	1170
Total Load	32627	30764	30457	32640	30007	30279	30879	31888	31961	34527

In the best feasible solution from CPLEX, the selection of seeds is 2, 4, 10, 15, 18, 21, 26, 29, 37, 38, 41 and 43, and nearly all suppliers are associated with 1-3 seeds. Supplier 18 (located in Indianapolis, IN) has four associations. The overall daily expected route length for the solution across the 10 scenarios is 19,900 miles. Table 4-4 shows the actual daily truck mileage of each route (seed) in each scenario.

Table 4-4. Daily truck mileage for the 12 routes across the 10 scenarios using CPLEX.

	Scenario									
Seed	1	2	3	4	5	6	7	8	9	10
2	2898	2987	2786	2898	2883	2786	2766	2987	2786	2786
4	2158	2145	2349	2117	2104	2145	2158	2145	2145	2158
10	1598	1431	1431	1618	1431	1431	1411	1431	1431	1411
15	1597	1475	1524	1642	1616	1646	1616	1501	1597	1646
18	1036	1036	1036	1245	1036	1289	1036	1036	1036	1474
21	1509	1402	1402	1402	1402	1524	1402	1402	1305	1599
26	1614	1598	1598	1614	1598	1598	1614	1598	1681	1598
29	1613	1736	1769	1628	1595	1754	1595	1628	1736	1754
37	1321	1338	1338	1338	1338	1043	1338	1338	1338	1338
38	370	370	370	370	1088	563	1088	370	370	1335
41	1282	2236	1282	1749	2166	2192	2166	2236	1749	1238
43	2232	2212	2212	2212	2212	2212	2212	2212	2212	2212
Total	19228	19966	19097	19833	20469	20183	20402	19884	19386	20549

These routes are very long, reflecting the geography over which the collection system operates. The shortest routes (370 miles) can be operated within one day, but many of the routes would require 3-5 days, even with teams of drivers. This is not critical for comparison of the solutions from the various algorithms, but it is worth noting that implementation of these solutions is complex, with multiple vehicles operating on each route simultaneously. The scale of this problem also implies that the collection process is expensive, and efforts to improve its efficiency are likely to be important.

The ILSM Algorithm

The ILSM converges very slowly and the computer with 2.00 MHz processor and 8.00 GB of memory runs out of memory after about 10 hours. The primary difficulty is that a large number of optimality cuts are necessary to create an effective lower approximation for $H(u,y)$ and this large number of added constraints increases the solution time and memory requirements for the first-stage problem. The large number of required cuts is a result of the large number of y_{ij} variables.

Although the ILSM algorithm initially appeared attractive because the second-stage sub-problems could be reduced to network flow problems and solved rapidly, the difficulties with the first-stage solutions have proved crippling, and the ILSM algorithm has been abandoned. On the other hand, the PH algorithm has proven much more effective and further experiments focus on its use.

The Progressive Hedging Algorithm

As discussed in section 4.3.3, the convergence of constraints $y_{ij}^s = y_{ij} \quad \forall s$ is observed to occur fairly slowly compared to the $u_j^s = u_j \quad \forall s$ constraints. For the experiments discussed here, the modification to the termination check in step 6 is implemented so that when the convergence of $u_j^s = u_j \quad \forall s$ has been reached, the algorithm terminates when "most" of the associations to seeds have converged. For a given *ij* pair, if $y_{ij}^s = 1$ for most, but not all, scenarios, the cost to the objective function of the overall problem for making $y_{ij}^s = 1$ in the remaining scenarios is some fraction of

g_{ij} (corresponding to the fraction of scenarios where currently $y_{ij}^{s} = 0$). If the value of g_{ij} is small, or the fraction of scenarios to be "flipped" is small, nudging the solution toward convergence by setting $y_{ij}^{s} = 1 \ \ \forall s$ has small cost and may speed the solution process considerably. For the experiments discussed here, this threshold on g_{ij} is considered to be the average value of all the penalty coefficients used in the model, defined as $\overline{g_{ij}}$. If supplier i is associated with seed j in some (but not all) scenarios, and $g_{ij} \le \overline{g_{ij}}$, the termination check assumes that this ij pair has converged.

The PH solution time for the 10-scenario problem is approximately 75 seconds. Although its objective function value (23,288) is about 5% above the value for the best CPLEX feasible solution (22,103), the total expected truck mileage, 20,244, is within 1.5% of the overall expected travelling distance from CPLEX. The daily traveling distance for each route under each scenario is shown in Table 4-5.

Table 4-5. Daily truck mileage for the 12 routes across the 10 scenarios using PH.

	Scenario									
Seed	1	2	3	4	5	6	7	8	9	10
2	2786	2987	2786	2786	2987	2987	2786	2998	2786	2786
4	2146	2146	2349	2146	2147	2147	2477	2146	2367	2396
10	1618	1618	1618	1618	1431	1431	1431	1618	1618	1431
15	1665	1475	1665	1616	1616	1616	1616	1524	1597	1646
21	1524	1408	1677	1366	1408	1408	1366	1366	1366	1455
25	1719	1719	1719	1719	1719	1719	1719	1719	1719	1719
28	1384	1532	1425	1425	1425	1425	1425	1425	1863	1425
32	1391	1391	1391	1496	1496	1496	1496	1391	1391	1391
36	1282	572	1345	1427	572	572	572	1427	1427	1427
38	370	370	370	370	1158	1158	1088	370	370	1397
41	1282	2236	1282	2236	2192	2192	2236	2236	2236	1282
44	2232	2212	2212	2212	2212	2212	2212	2212	2212	2232
Total	19399	19666	19839	20417	20363	20363	20424	20432	20952	20587

Since the center of the models developed in chapter 3 is capacitated clustering, particular attention should be paid to how clusters are formed by difference solution methods. The PH selects suppliers 2, 4, 10, 15, 21, 25, 28, 31, 36, 38, 41 and 44 as seeds. Figure 4-8 illustrates where the seeds suppliers are located for both the CPLEX and PH solutions. The supplier nodes in yellow are the common seeds by these two methods, green ones are chosen by CPLEX only, and purple ones are chosen by the PH only.

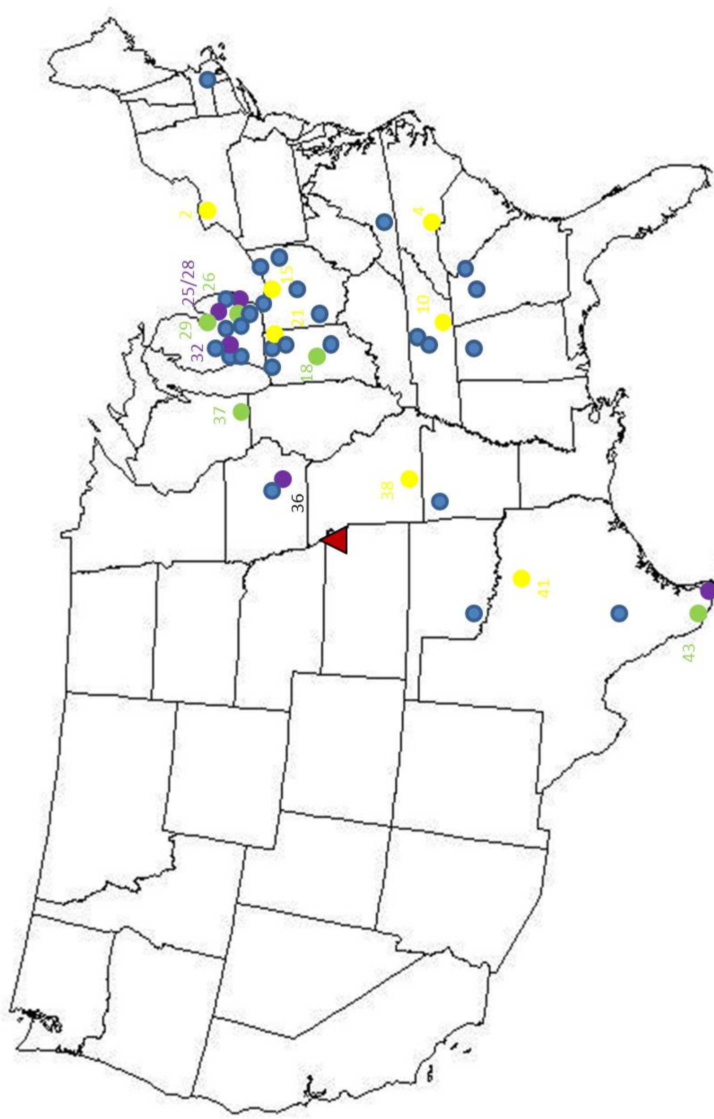


Figure 4-8. Selected seed suppliers by CPLEX and PH.

From Figure 4-8 we can see that CPLEX and PH select 7 common seed suppliers (out of 12), which are suppliers 2, 4, 10, 15, 21, 38, and 41. For the other five routes, they pick different seed suppliers but within the same cluster. For example, for the route serving suppliers in south Texas, CPLEX chooses supplier 43 as the seed while the PH algorithm chooses supplier 44. Such a difference in seed selection does not substantially affect the route construction, and it explains why the total expected route lengths are close. In general, both methods generate three routes serving suppliers south and west of the plant (in Texas, Oklahoma, Arkansas and Missouri), two routes serving suppliers in the southeastern area (Tennessee, Alabama, Georgia, North Carolina and Virginia), and seven routes for the dense clusters of suppliers in the Mideast (Michigan, Indiana, Ohio, etc.). Considering its speed of computing, the PH is an effective strategy for solving this problem.

. Figure 4-9 presents the empirical cumulative distribution functions (CDFs) of the total truck mileage in each scenario for CPLEX solution and PH solution. These two CDFs are similar in shape, with the PH one shifting to right by 344 ft^3 on average. It is consistent with the fact that CPLEX and PH generate similar route patterns and the overall expected travelling distance of the PH solution is about 1.5% above the distance of CPLEX solution.

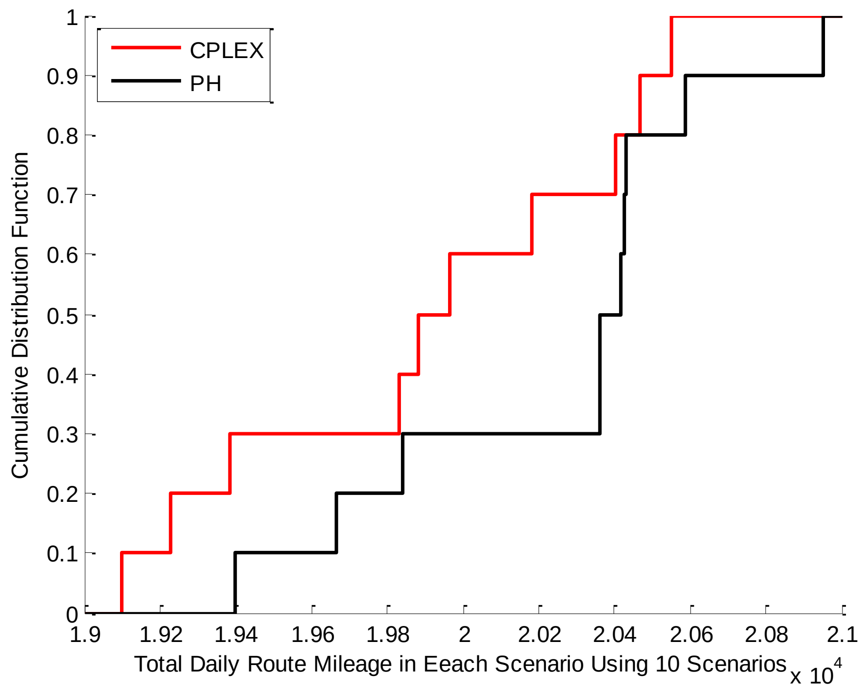


Figure 4-9. Empirical CDFs for total truck mileage for the 12 routes across the 10 scenarios using CPLEX and PH.

Table 4-6 lists split deliveries for the CPLEX and PH solutions. It is noted that the PH solution has more split services. For the CPLEX solution, in each scenario there are 2-6 split suppliers. Scenario 5 has the fewest split suppliers (only suppliers 24 and 43) because of the smallest total pick-up quantity, and there are six split suppliers in scenarios 4 and 10 due to the large total loads. Most split suppliers are served by two routes, and which suppliers are split depends on their daily pick-up quantities in each scenario. The three-way splits occur on supplier 18 in scenarios 1 and 4, supplier 24 in scenario 2, and supplier 12 in scenario 10. It is reasonable for these suppliers to be split across multiple routes because suppliers 12 and 24 are located in/near Michigan where many suppliers are clustered, and supplier 18 is located in central Indiana, which is near several routes. For the PH solution, 6-8 suppliers are visited by multiple routes in each scenario, and more suppliers are split three ways. Supplier 18 is split across four routes in scenarios 4 and 6 as a result of its central location in the map. The association matrix (y_{ij}) also demonstrates that there may be many splits because about one third of suppliers are associated with multiple routes, and supplier 18 is associated with 6 routes.

Table 4-6. Split deliveries for the CPLEX and PH solutions using 10 scenarios.

scenario	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers
1	18	10,15,18	24	2,26	30	21,26	41	41,43						
2	16	4,15	18	15,18	24	2,26,29	43	41,43						
3	12	15,29	14	2,4	24	26,29	32	29,37						
4	9	4,10	12	15,29	18	10,15,18	24	2,26	30	21,26	32	29,37	42	41,43
5	24	2,26	43	41,43										
6	12	15,29	13	2,15	20	18,21	24	26,29	43	41,43				
7	16	4,15	30	21,26	43	41,43								
8	12	15,29	18	15,18	24	2,26	32	29,37	43	41,43				
9	13	2,15	15	15,26	24	26,29	42	41,43						
10	12	15,21,29	14	2,15	16	4,18	18	18,38	24	26,29	32	29,37		

(a) CPLEX solution

scenario	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers
1	8	4,10	13	2,15	16	4,15	18	10,36	20	21,36	24	25,28	30	21,28		
2	8	4,10	16	4,15	18	10,21	24	2,25,28	30	21,28	34	28,32	43	41,44		
3	12	15,28	14	2,4	16	4,15	18	10,21	24	21,25,28	32	32,36	34	21,32		
4	13	2,15	16	4,15	18	10,15,21,36	24	25,28	31	25,32	32	32.36	43	41,44		
5	14	2,15	16	4,15	18	15,21	24	2,25	30	21,28	31	25,32	40	38,42	43	41,44
6	13	2,15	14	2,4	18	10,15,21,28	24	25,28	31	25,32	39	38,41	43	41,44		
7	8	4,10	14	2,15	16	4,15	18	15,21	24	25,28	31	25,32	40	38,41	43	41,44
8	8	4,10	12	15,28	16	4,15	18	10,21,36	24	25,28	32	2,32,36	43	41,44		
9	13	2,15	14	2,4,28	18	10,21,36	24	25,28	32	32,36	34	28,32	43	41,44		
10	8	4,10,38	12	15,21,28	14	2,4,15	18	36,38	24	25,28	32	32,36	41	41,44		

(b) PH solution

10 scenarios vs. 25 scenarios

Extending the problem size from 10 scenarios to 25 scenarios creates an opportunity to test the PH algorithm in an instance of even greater complexity, and to compare the resulting collection of routes to see the effects of including more scenarios in the stochastic analysis. Table 4-7 presents the daily pick-up quantity at every supplier across the scenarios, each of which has equal probability $\pi_s = 0.04$. The largest total daily load is 34,401 ft^3 in scenario 16, so the minimum number of routes required for all the scenarios is 12. There is no solution from CPLEX available for the 25-scenario instance of the problem, so comparisons are made between the two solutions from the PH algorithm.

Table 4-7. Daily pick-up quantities (ft^3) for the 44 suppliers considering 25 scenarios.

	Scenario											
Supplier	1	2	3	4	5	6	7	8	9	10	11	12
1	317	694	631	343	538	573	500	361	521	280	381	407
2	694	462	521	754	951	536	665	853	550	636	739	432
3	596	347	317	845	693	483	291	457	544	387	754	449
4	529	503	337	404	441	423	357	315	294	481	392	518
5	534	661	324	644	401	345	474	491	518	320	373	481
6	251	294	345	323	372	222	315	245	206	327	309	340
7	209	275	197	252	121	297	174	246	138	214	233	218
8	1333	761	1519	1115	1010	1407	902	1025	1274	770	1159	963
9	141	16.2	538	735	1258	673	784	1106	290	948	465	434
10	519	104	1169	1246	294	201	1392	255	689	177	434	157
11	583	570	764	520	787	603	798	741	686	825	374	478
12	981	807	1098	414	747	446	1034	612	1065	1144	757	998
13	1170	579	872	1330	1008	1042	1351	1098	978	646	447	1610
14	345	1530	1852	1194	2243	2289	1618	407	1795	1696	1093	651
15	820	1118	944	765	929	1188	1419	844	1340	692	1303	1148
16	935	607	916	1088	948	892	834	1045	1140	1169	654	1217
17	1105	565	1435	1591	754	444	1497	980	1045	1469	1167	1603
18	2546	2783	2294	850	2452	928	2178	2321	1142	1183	1865	2682
19	240	394	167	370	183	349	315	257	143	404	282	397
20	1229	1141	1651	806	1092	669	1460	1041	1719	932	1571	636
21	182	177	137	210	196	190	155	230	202	161	225	232
22	419	377	365	489	978	733	605	682	556	581	880	833
23	293	338	411	279	210	393	259	295	240	409	202	145
24	1444	1005	1959	1148	1514	1863	1274	2221	2035	1462	1355	1691
25	210	287	343	362	414	475	455	437	397	407	265	165
26	227	207	326	405	390	286	310	190	355	160	143	240
27	432	276	349	305	360	384	289	327	296	416	318	369
28	336	572	536	339	427	393	310	229	133	200	673	485
29	272	327	314	358	191	195	310	340	211	381	156	146
30	450	670	558	647	787	584	918	708	735	776	540	503
31	401	799	418	764	448	357	523	739	690	718	773	568
32	2292	483	2192	2020	1621	744	821	2106	1057	2496	289	2271
33	309	468	205	558	485	538	346	577	285	173	336	245
34	587	994	884	649	1171	772	1099	1147	957	558	910	420
35	497	407	316	482	297	366	271	538	357	454	443	488
36	192	127	130	171	197	181	161	143	149	157	139	167
37	455	531	596	629	422	383	464	569	357	552	687	409
38	305	189	152	379	149	223	291	119	280	216	362	259
39	255	25	168	272	238	91	155	76	226	101	48	180
40	1958	2370	680	964	817	1827	1615	2217	2047	1355	2104	2023
41	1315	844	1487	1257	1155	1456	538	584	1052	1403	624	452
42	443	427	434	448	456	267	309	380	335	240	289	304
43	1084	2225	2007	1627	2294	1662	1525	1302	2048	2513	1801	2423
44	540	1248	904	968	849	1258	1169	999	1294	1360	706	944
Total Load	29975	29584.2	33762	31319	33288	29631	32530	31855	32371	31949	29020	31781

(a)

							Scenario						
Supplier	13	14	15	16	17	18	19	20	21	22	23	24	25
1	717	649	452	595	421	746	302	559	734	674	709	475	654
2	622	782	504	482	804	702	410	600	839	907	573	653	763
3	592	874	405	767	971	509	806	945	888	674	707	935	633
4	511	308	374	456	348	474	376	452	240	265	269	491	413
5	569	557	404	438	614	590	599	429	506	454	357	693	631
6	389	299	211	353	380	267	403	288	237	280	262	361	408
7	301	199	156	183	256	273	191	227	261	168	164	239	287
8	1545	1222	1197	1365	860	796	1591	1069	879	1453	1504	1446	1587
9	1011	890	614	1322	255	139	231	94	581	406	69	198	330
10	636	352	787	921	1130	400	892	1030	571	245	457	751	133
11	436	483	624	709	638	661	750	451	505	400	656	544	728
12	695	944	537	864	669	490	480	817	903	908	1121	624	581
13	1667	1530	762	456	493	1270	1781	618	711	915	784	1485	554
14	994	1311	485	2091	549	2070	740	1442	2004	822	302	2177	918
15	1405	1241	980	1279	1380	877	1061	1400	1363	1081	1035	1225	738
16	975	1277	730	1125	783	756	814	1021	692	1113	999	852	1212
17	1336	620	635	838	1209	1534	712	1305	1263	862	480	925	1398
18	2064	2733	1987	1012	2615	2473	1542	1699	1798	1379	803	1283	1565
19	267	385	408	376	202	351	123	303	334	141	190	292	225
20	1180	1293	886	1748	1690	1546	1400	1501	772	946	732	1637	1316
21	124	220	214	142	174	147	156	245	199	120	129	187	167
22	918	538	658	953	323	807	300	792	764	455	983	869	927
23	351	173	310	374	230	469	441	486	144	184	458	430	367
24	1162	2081	1635	2171	1823	1737	1562	1325	2135	1905	2006	2206	1755
25	454	249	465	377	481	313	326	365	492	431	299	220	497
26	455	417	470	294	201	373	339	132	275	441	369	423	261
27	236	400	258	227	419	337	322	246	215	370	394	272	199
28	60	164	364	448	207	293	96	268	413	524	257	15	152
29	217	286	228	278	406	180	239	298	257	346	419	249	382
30	804	603	860	411	838	693	888	485	630	904	750	868	527
31	314	636	610	457	541	489	675	735	585	369	498	346	656
32	1244	2433	905	1163	1403	1701	414	583	2466	1834	1499	1900	2360
33	603	522	434	192	264	443	377	391	238	402	591	499	594
34	674	534	1053	1122	461	750	718	1200	806	1017	849	935	479
35	469	516	336	282	327	442	377	395	385	254	526	422	426
36	166	184	189	175	207	200	151	99	111	217	136	116	204
37	802	512	758	777	483	798	648	739	619	676	722	363	826
38	202	416	125	321	375	394	317	341	255	240	168	405	353
39	55	132	37	151	61	140	111	40	114	29	207	190	84
40	763	1027	2361	1685	870	1158	2238	1372	1510	1795	2317	1246	2273
41	462	1132	776	1096	1387	856	949	1216	989	687	738	1548	418
42	357	248	260	419	281	354	384	321	410	329	368	225	393
43	1188	1409	1218	2153	2587	1374	2117	1515	1852	2353	1751	2445	1940
44	1056	656	782	1353	1324	648	1194	1339	1143	1083	579	751	589
Total Load	31048	33437	28444	34401	31940	32020	30543	31178	33088	31058	29186	34416	31903

(b)

The running time of the PH heuristic for the 25-scenario instance is 996 seconds (about16.5 minutes), as compared with 75 seconds for the 10-scenario version. Thus, the computational burden of additional scenarios is substantial, but it is still practical to solve the 25-scenario version of this realistic problem using the PH algorithm. The total expected route length across the 25 scenarios is 21,160 miles, as compared with 20,244 in the 10-scenario version of the problem.

Seeds selected by the PH are suppliers 2, 4,10,17,21, 24, 29, 32,35,39,41, and 44. Figure 4-10 illustrates the location of seed suppliers for 10-scenario and 25-scenario cases by the PH method. The green nodes are seeds selected only for 10-scenario example, the purple nodes are seeds selected only for 25-scenairo example, and common seed suppliers are yellow nodes (7 of 12). For the other five routes, it is noted suppliers (15, 17), suppliers (24, 25), suppliers (28, 29), suppliers (35, 36) and suppliers (39, 39) are pairs from different clusters, and the PH method picks one or another from these pairs as seed suppliers for 10-scenario and 25-scenario tests.

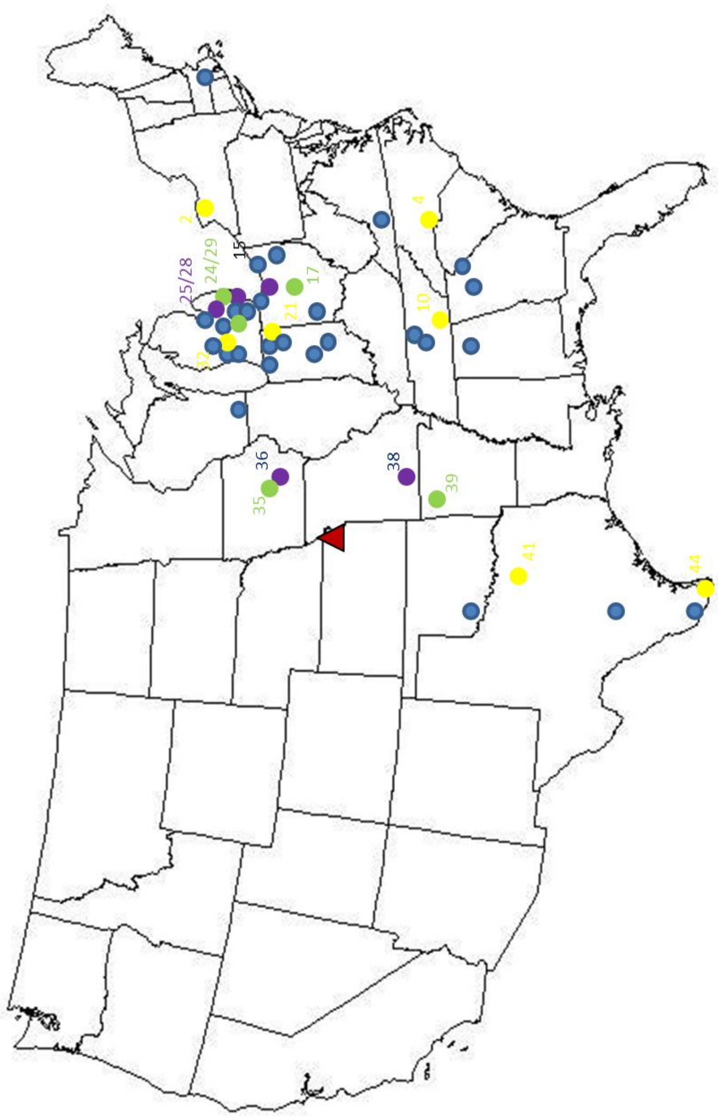


Figure 4-10. Selected seed suppliers by PH using 10 scenarios and 25 scenarios.

Table 4-8 shows the split deliveries for the PH solution using 25 scenarios. The number of split suppliers across the scenarios ranges from 4 suppliers (in scenario 15) to 9 suppliers (in scenarios 5, 12 and 16). In most of scenarios, there are 6-8 split suppliers, as there were in the 10-scenario version of the problem. Table 4-8 also illustrates that although most of suppliers are split two ways, there are more three-way, four-way, or even five-way splits. Besides the supplier 18 as discussed above, suppliers 14, 24 and 32 are served by more than three routes in at least one scenario. These suppliers are located either in the center of this distribution map, or in/near the dense clusters in Michigan and northern Ohio. Table 4-7 (b) and Table 4-8 demonstrate that the PH algorithm is likely to produce solutions with split service and that the degree of splitting (both in the number of split suppliers and the number of routes across which suppliers are split), increases slightly as the set of load scenarios increases.

Table 4-8. Split deliveries for the PH solution using 25 scenarios.

scenario	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers	split supplier	seed suppliers
1	14	2,4	15	2,24	17	17,21	18	17,35	24	24,29	32	29,32,35	40	39,41				
2	14	2,4	16	4,29	17	17,21	18	17,24	20	21,32	24	24,29	40	39,41	43	41,44		
3	8	10,39	14	4,21	15	2,24	18	17,35,39	20	21,35	24	24,29,35	32	32,35	43	41,44		
4	8	10,17	14	2,4,21	16	4,17,29	24	24,29	32	29,32,35	43	41,44						
5	11	2,17	14	4,17	16	4,10	18	35,39	20	17,21	24	24,29	30	21,29	32	29,32,35	43	40,44
6	14	4,10,21,24,29	15	2,24	20	21,32	40	39,41	43	41,44								
7	8	10,17	13	4,21,32	14	2,17	15	17,24	20	17,21,35	24	24,29	44	41,44				
8	13	2,4	16	4,10	18	10,17	24	21,24,29,35	32	17,32,35								
9	14	21,29	15	2,24	16	4,17	18	10,17	20	17,21	24	24,29,32,35	40	39,42	43	41,44		
10	14	4,10,21	15	2,21	24	24,29	30	21,29	32	17,21,32,35	40	39,41	43	41,44				
11	14	2,4	18	17,24	20	2,21	24	2,29	30	21,24								
12	13	2,10	14	2,4	15	21,24	16	4,21	18	10,17,35	24	24,29	32	29,32	40	39,41	43	41,44
13	8	10,39	13	2,24,29,35	16	4,17	18	17,35	20	21,35	32	29,32						
14	13	2,35	14	2,10,17	16	4,17	18	17,39	20	21,32	24	24,29,35	32	32,35				
15	24	2,24,29	30	2,21	32	29,32	40	40,41										
16	14	17,29	15	2,24	16	4,17	18	10,17	20	2135	24	24,29,35	32	32,35	40	39,41	43	41,44
17	14	2,4	15	2,24	16	4,10,21	18	10,17,39	20	21,32	24	24,29	40	39,41				
18	13	2,4	14	4,10,17	18	17,35,39	20	21,35	24	24,29	32	29,32,35						
19	13	2,21,32	14	2,4	24	24,29	30	21,29	40	39,41	43	41,44						
20	14	2,4	15	2,4	16	4,17,21	18	17,35	20	10,21,32	24	24,29	44	41,44				
21	14	2,4,10,17	15	21,24	18	17,35	24	24,29,35	32	33,35	43	41,44						
22	15	21,24	16	4,10	24	2,17,21,29,35	32	21,32	40	39,41	43	41,44						
23	15	2,24	24	17,21,24	30	21,29	32	29,32	40	39,41								
24	13	2,24	14	17,21,24,29	16	4,10	18	17,39	20	10,21	32	29,32,35	40	39,41	43	41,44		
25	14	2,4	16	4,10,17	18	17,35	24	24,29	30	21,29	32	29,32,35						

Compared to the PH solution using 10 scenarios, the 25-scenario case has more split pickups and longer daily route lengths. Figure 4-11 demonstrates the empirical CDFs for the total travelling distance in each scenario for the 10-scenairo and 25-scenario cases. Although the expected route length in the 25-scenario run is only about 4.5% higher than in the 10-scenario run, the 25-scenario experiment has a somewhat different distribution of outcomes, with more than half of the scenarios having total travel distance between 21,000 miles and 23,000 miles, a range not seen at all in the 10-scenario experiment.

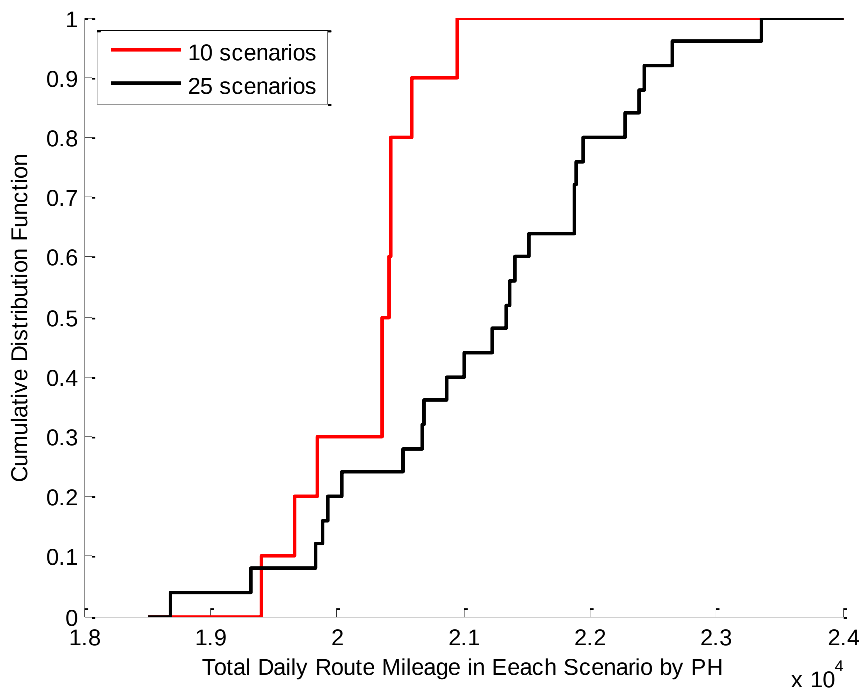


Figure 4-11. Empirical CDFs for total truck mileage for the 12 routes across the 10 scenarios and 25 scenarios using PH.

If the scenarios are created carefully, reflecting both the marginal distributions of load quantities at individual suppliers and the correlations among loads driven by the nature of production schedules at the assembly plant, using a relatively small number of scenarios (e.g., 10) can produce solutions that reflect the average total travel distance reasonably well, but the distribution of values across the scenarios may be somewhat biased. Using a larger number of scenarios (e.g., 25) is likely to produce a better reflection of the distribution of total distances (or costs). However, there is a significant penalty in computation time, and for many route design exercises the smaller number of scenarios may be sufficient.

4.5 Conclusions

This chapter follows up chapter 3 to suggest two heuristic algorithms for solving the stochastic collection route design problem (P3-1) for instances of realistic size. The two methods explored are the integer L-shaped method (ILSM) and progressive hedging (PH). These are two commonly used decomposition approaches for stochastic programming problems, but operate from different perspectives. The ILSM is a stage-based decomposition method. For the problem of interest here, the ILSM has proven to be ineffective. The process of adding optimality cuts to the first-stage problem to create an approximation for the second-stage cost operates too slowly and creates too many additional constraints in the first-stage problem. ILSM does not appear to be a useful strategy for the problem of interest here.

However, PH has been shown to work quite effectively. The decomposition of PH is scenario-based. It decomposes the problem into individual problems for each scenario, updates coefficients of first-stage variables based on the scenario-specific solutions, and resolves each individual scenario sub-problem. This process repeats until the non-anticipative constraints are satisfied (i.e., the solutions of the first-stage variables converge across all the scenarios). In a case study from the automotive industry involving 44 suppliers, the PH algorithm achieved a good solution in approximately 75 seconds of computation. As a comparison, CPLEX required nearly 9 hours on the same computer to achieve a similar solution.

The use of the PH algorithm together with a careful strategy of constructing a relatively small set of representative scenarios (e.g., using the Latin Hypercube Sampling with Correlations (LHSC) method) is an effective tool for designing collection routes that are both efficient and robust under volume variations at suppliers. It does not guarantee optimal solutions, but achieves good solutions quickly for problem instances of realistic size, making it useful in practical situations.

CHAPTER 5

CONCLUSIONS AND FUTURE RESEARCH

5. 1 Contributions

Spending in the U.S. logistics industry totaled $1.45 trillion in 2014, and represented 8.3 percent of annual gross domestic product (GDP). Improving the efficiency of inbound logistics networks has economic significance. The goal of this research is efficient design of routes and schedules for moving materials into manufacturing or assembly plants, focusing on two subsets of suppliers: 1) those that ship modest quantities of material and for whom the frequency-of-service decision should be integrated with the design of routes to collect the material; and 2) shippers that are served by daily collection routes, but for whom there may be considerable variation in daily pick-up quantity.

Separate optimization models and solution methods have been developed for these two cases. The models have been tested using data from the automotive industry and both offer substantial opportunities for improving the effectiveness of inbound material collection operations.

The following are the main research contributions of this dissertation:

- **A clustering approach** is proposed for designing route and schedules. Unlike previous work focusing on the vehicle routing problem in each individual day, the approach views the problem as being primarily a clustering problem, which groups suppliers to different seeds first and constructs routes later. This mathematical optimization is termed capacitated clustering.

- **The marginal cost coefficient** is introduced in the objective function to approximate the actual cost of a solution so that it can be easily computed before actual routes are constructed. This approximation has proven quite accurate for the typical routes in the automotive industry, which stop at a relatively small number of suppliers.
- **Pick-up frequency and spatial clustering are integrated** in the developed model to minimize total logistics (transportation plus inventory) cost for suppliers with small quantities of materials, where daily pickups may not be required.
- **Showing the equivalence** between the frequency-spatial clustering problem and the single-source capacitated facility location problem (SSCFLP), allowing near-optimal solutions to be achieved efficiently using the Very Large-Scale Neighborhood (VLSN) algorithm developed for SSCFLP.
- **Substantial total cost savings can be achieved in realistic applications** by using the clustering-based optimization problem that considers operating frequency and spatial grouping jointly, as demonstrated in a case study.
- **Formulating the design of plant-based collection routes with uncertain loads as a two-stage stochastic optimization problem** for suppliers that ship larger volumes daily. This formulation includes opportunities for split pick-ups at suppliers as a way of partially buffering the uncertainty in the load

quantities and also incorporates controls (either penalties or constraints) designed to improve the "regularity" of service to individual suppliers.

- **Developing an effective decomposition method** that shows progressive hedging (PH) for achieving good (although not necessarily optimal) solutions for the stochastic optimization.

- **Showing how a previously developed method, Latin Hypercube Sampling with Correlations (LHSC), can be used** as an effective tool for constructing a relatively small set of representative scenarios of supplier pick-up quantities to represent the correlations among individual supplier quantities and reflect the relative likelihoods of different loading patterns.

- **The effectiveness of model formulations and proposed heuristic algorithms** has been demonstrated by computational experiments and real-world applications.

5. 2 Future Research

This dissertation analyses two main types of suppliers that ship different volumes of materials to manufacturing or assembly plants. Therefore the work consists of two parts, focusing on each supplier type. In the future, we can extend the work from one type to another. In other words, the model that combines service frequency and vehicle routing can be enhanced by incorporating uncertainty in the supplier pickup quantities, and routes designed for uncertain loads can be operated with different frequency to make the clustering of suppliers frequency-based as well as spatial. It is worth noting that the models developed her consider one manufacturing plant or consolidation center at a time. When optimizing the whole system with thousands of suppliers, the first step is to sort these suppliers to their associated plants or consolidation centers following production schedules.

One useful direction for further research on integrating pick up service and spatial clustering is to add constraints on specific suppliers that would constrain them to a subset of the possible frequency classes. This extension is important because in practice suppliers may be served at particular frequencies as required by production schedules.

For current work about stochastic pick up quantities, although the designs are more robust to accommodate variation in production schedules, the number of routes in use are the same across all the scenarios, even for the scenarios with small total daily loads, where routes might be redundant. Allowing the flexibility of operating different number of routes in each scenario will improve the efficiency of designing routes. Another type of research that could be done is to combine progressive hedging with dual decomposition methods. The dual decomposition methods can be guaranteed to converge but often do so quite slowly. However, by using progressive hedging to generate a good starting solution for the dual decomposition, this combination may generate better quality solutions for the stochastic pick-up problem within reasonable computation times.

APPENDIX

Supplier locations and quantities for case study.

Supplier Index	ZIP Code	Quantity (cu. ft. / week)		Supplier Index	ZIP Code	Quantity (cu. ft. / week)
1	46947	480		39	44903	3105
2	43311	360		40	40214	465
3	45342	180		41	40258	1395
4	43138	225		42	16001	1365
5	45404	565		43	47348	295
6	46806	1345		44	45502	75
7	46723	475		45	45177	4115
8	47348	310		46	46803	840
9	43351	495		47	45840	3255
10	45690	4210		48	43015	3250
11	45204	2140		49	43102	1230
12	43319	540		50	45040	4190
13	43316	10		51	46131	680
14	40353	2255		52	40601	20
15	43764	10		53	46222	2870
16	41041	465		54	41042	7815
17	45377	6480		55	45840	290
18	45420	270		56	47201	3950
19	44690	235		57	46825	1325
20	44706	180		58	47374	290
21	45439	585		59	45365	480
22	45365	165		60	46176	1535
23	47201	8555		61	45066	205
24	47359	95		62	44827	2145
25	45373	60		63	40511	10
26	44659	955		64	44904	2045
27	47987	1620		65	45420	4160
28	45346	1305		66	43025	65
29	43113	3670		67	41041	2160
30	47240	120		68	45833	1635
31	46221	350		69	15219	960
32	46725	255		70	44460	990
33	45439	2035		71	46750	2805
34	44633	300		72	15116	870
35	47112	2875		73	40391	595
36	40511	975		74	45385	485
37	44633	2070		75	41042	430
38	47374	305				

REFERENCE

Aardal, K. (1998). “Capacitated Facility Location: Separation Algorithm and Computational Experience,” *Mathematical Programming*, 81, 149-175.

Ahuja, R.K., J.B. Orlin, S. Pallottino, M.P. Scaparra and M.G. Scutella. (2004). “A Multi-Exchange Heuristic for the Single-Source Capacitated Facility Location Problem,” *Management Science*, 50(6), 749-760.

Ak, A. and A. L. Erera. (2007). “A Paired-Vehicle Recourse Strategy for the Vehicle Routing Problem with Stochastic Demands,” *Transportation Science*, 41, 222–237.

Archetti, C., N. Bianchessi and M.G. Speranza. (2014). “Branch-and-Cut Algorithms for the Split Delivery Vehicle Routing Problem,” *European Journal of Operational Research*, 238(3), 685-698.

Archetti, C. and M.G. Speranza. (2012). “Vehicle Routing Problems with Split Deliveries,” *International Transactions in Operational Research*, 19(1-2), 3-22.

Ballou, R.H. (2003). *Business Logistics Management.* Prentice Hall, Englewood Cliffs, New Jersey.

Barahona, F. and F. A. Chudak. (2005). “Near-optimal Solutions to Large Scale Facility Location Problems,” *Discrete Optimization*, 2(1), 35-50.

Beamon, B.M. (1998). “Supply chain design and analysis: models and methods,” *International Journal of Production Economics*, 55, 281–294.

Beasley, J.E. (1993). “Lagrangian Heuristics for Location Problems,” *European Journal of Operational Research*, 65, 383-399.

Bertazzi, L., M. Savelsbergh and M.G. Speranza. (2008). “Inventory Routing,” in Golden, B., S. Raghavan and E. Wasil (eds), *The Vehicle Routing Problem: Latest Advances and New Challenges*, Springer, Berlin, 49-72.

Bertazzi, L. and M.G. Speranza (2012). “Inventory Routing Problems: An Introduction,” *European Journal of Transportation and Logistics*, 1, 307-326.

Bertsimas, D.J. (1988). *Probabilistic Combinatorial Optimization Problems*, Ph.D. Thesis, Report No. 193, Operations Research Center, Massachusetts Institute of Technology, Cambridge, MA.

Bianchi, L., M. Birattari, M. Chiarandini, M. Manfrin, M. Mastrolilli, L. Paquete, O. Rossi-Doria, and T. Schiavinotto. (2004). “Metaheuristics for the Vehicle Routing Problem with Stochastic Demands, ” in X.Yao, E.K.Burke, J.A.Lozano, J.Smith, J.J.Merelo-Guervos, J.A.Bullinaria, J.E.Rowe, P.Tino, A.Kaban, and H.Schwefel (eds.), *Parallel Problem Solving from Nature - PPSN VIII*, Springer, Birmingham.

Blumenfeld, D.E., L.D. Burns and J.D. Diltz. (1985). “Analyzing trade-offs between transportation, inventory and production costs on freight networks,” *Transportation Research Part B: Methodological*, 19(5), 361-380.

Bouza'iene-Ayari, B., M. Dror, and G. Laporte. (1993). “Vehicle Routing with Stochastic Demands and Split Deliveries,” *Foundations of Computing and Decision Sciences,*18, 63-69.

Burns, L.D., R.W. Hall, D.E. Blumenfeld and C.F. Daganzo. (1985). “Distribution Strategies that Minimize Transportation and Inventory Costs,” *Operations Research*, 33(3), 469-490.

Campbell, A.M. and J.H. Wilson. (2014). “Forty Years of Periodic Vehicle Routing,” *Networks*, 63(1), 2-15.

Carøe, C.C. and R. Schultz. (1999). “Dual Decomposition in Stochastic Integer Programming,” *Operations Research Letters*, 24(1-2), 37-45.

Chuah, K.H. and J.C. Yingling. (2005). “Routing for a Just-in-Time Supply Pickup and Delivery System,” *Transportation Science*, 39(3), 328-339.

Christofides, N. and J.E. Beasley. (1984). “The Period Routing Problem,” *Networks*, 14, 237-256.

Coelho, L.C., J.-F. Cordeau and G. Laporte. (2014). “Thirty Years of Inventory Routing,” *Transportation Science*, 48(1), 1-19.

COIN-OR Foundation. (2012). *Computational Infrastructure for Operations Research*, www.coin-or.org/SYMPHONY/branchandcut/VRP/data/index.htm.old, accessed October 15, 2014.

Council of Supply Chain Management Professionals. (2015). 26th Annual State of Logistics Report. Washington D.C.

Daganzo, C.F. (1985). “Supplying a Single Location from Heterogeneous Sources,” *Transportation Research Part B: Methodological*, 19(5), 409-419.

Daganzo, C.F. and G.F. Newell. (1986). “Handling operations and the lot size trade-off,” *Transportation Research Part B: Methodological*, 27(3), 167-183.

Dong, Z. and M.A. Turnquist. (2015). “Combining Service Frequency and Vehicle Routing for Managing Supplier Shipments,” *Transportation Research, Part E: Logistics and Transportation Review*, 79, 231-243.

Dror, M., G. Laporte, and P. Trudeau. (1989). “Vehicle Routing with Stochastic Demands: Properties and Solution Frameworks,” *Transportation Science*, 23, 166-176.

Dror, M. and P. Trudeau. (1986). “Stochastic Vehicle Routing with Modified Savings Algorithm,” *European Journal of Operational Research*, 23, 228-235.

Dror, M. and P. Trudeau. (1989). “Savings by Split Delivery Routing,” *Transportation Science*, 23(2), 141-145.

Dror, M. and P. Trudeau. (1990). “Split Delivery Routing,” *Naval Research Logistics*, 37(3), 383-402.

Erenguc, S. S., N. C. Simpson., and A. J. Vakharia. (1999). “Integrated production/distribution planning in supply chains: an invited review,” *European Journal of Operational Research*, 115, 219–236.

Fan, Y. and C. Liu. (2010). “Solving Stochastic Transportation Network Protection Problems Using the Progressive Hedging-based Method,” *Networks and Spatial Economics*, 10(2), 193-208.

Francis, P. M. and K.R. Smilowitz. (2006).”Modeling Techniques for Periodic Vehicle Routing Problems,” *Transportation Research Part B: Methodological*, 40(10), 872-884.

Francis, P. M., K. R. Smilowitz and M. Tzur. (2006). “The Period Vehicle Routing Problem with Service Choice,” *Transportation Science*, 40(4), 439-454.

Francis, P. M., K. R. Smilowitz, and M. Tzur. (2008). “The Period Vehicle Routing Problem and its Extensions,” in Golden, B., S. Raghavan and E. Wasil (eds), *The Vehicle Routing Problem: Latest Advances and New Challenges*, Springer, Berlin, 73-102.

Gendreau, M., G. Laporte, and R. Seguin. (1995). “An Exact Algorithm for the Vehicle Routing Problem with Stochastic Demands and Customers,” *Transportation Science*, 29(2), 143- 155.

Grolimund, S. and J.G. Ganascia. (1997). “Driving Tabu Search with Case-based Reasoning,” *European Journal of Operational Research*, 103, 326-338.

Guo, G., G. Hackebeil, S.M. Ryan, J.-P. Watson and D.L. Woodruff. (2015). “Integration of Progressive Hedging and Dual Decomposition in Stochastic Integer Programs,” *Operations Research Letters*, 43(3), 311-316.

Hall, R.W. (1985). "Determining Vehicle Dispatch Frequency when Shipping Frequency Differs among Suppliers," *Transportation Research Part B: Methodological*, 19(5), 421-431.

Hindi, H. and K. Pienkosz. (1999). "Efficient Solution of Large Scale, Single-Source, Capacitated Plant Location Problems," *Journal of the Operational Research Society*, 50, 268-274.

Holmberg, K., M. Ronnqvist and D. Yuan. (1999). "An Exact Algorithm for the Capacitated Facility Location Problems with Single Sourcing," European *Journal of Operational Research*, 113(3), 544-559.

Holt, J., M. Ronnqvist and S. Tragantalerngsak. (1999). "A Repeated Matching Heuristic for the Single-Source Capacitated Facility Location Problem," *European Journal of Operational Research*, 116(1), 51-68.

Horowitz, A.D. and C.F. Daganzo. (1986). "A graphical method for optimizing a continuous review inventory systems," *Production and Inventory Management*, 27(4), 3-45.

Iman, R.L. and W.J. Conover. (1982). "A Distribution-free Approach to Inducing Rank Correlation among Input Variables," *Communications in Statistics: Simulation and Computing B*, 11(3), 311-334.

Kasilingam, R. G. (1998). *Logistics and transportation: Design and planning*. Dordrecht, Kluwer Academic.

Kall, P. (1982). "Stochastic programming", *European Journal of Operational Research*, 10, 125-130.

Laporte, G. (2009). “Fifty Years of Vehicle Routing, ” *Transportation Science*, 43(4), 408-416.

Laporte, G. and F.V. Louveaux. (1993). “The Integer L-Shaped Method for Stochastic Integer Programs with Complete Recourse,” *Operations Research Letters*, 13(3), 133-142.

Laporte, G., F.V. Louveaux and L. Van Hamme. (2002). “An Integer L-Shaped Algorithm for the Capacitated Vehicle Routing Problem with Stochastic Demands,” *Operations Research*, 50, 415-423.

Laporte, G., F.V. Louveaux, and H. Mercure. (1989). “Models and Exact Solutions for a Class of Stochastic Location-Routing Problems,” *European Journal of Operational Research*, 39, 71-78.

Laporte, G., F.V. Louveaux, and H. Mercure. (1992). “The vehicle routing problem with stochastic travel times,” *Transportation Science*, 26, 161-170.

Laporte, G., F.V. Louveaux, and H. Mercure. (1994). “A Priori Optimization of the Probabilistic Traveling Salesman Problem,” *Operations Research*, 42, 543-549.

Laporte, G., R. Musmanno, and F. Vocaturo. (2010). “An Adaptive Large Neighbourhood Search Heuristic for the Capacitated Arc-Routing Problem with Stochastic Demands,” *Transportation Science*, 44(1), 125–135.

Lei, H., G. Laporte, and B. Guo. (2012). “The Vehicle Routing Problem with Stochastic Demands and Split Deliveries,” *Information Systems and Operational Research*, 50(2), 59-71.

Leung, J.M.Y. and T.L. Magnanti. (1989). “Valid Inequalities and Facets of the Capacitated Plant Location Problem,” *Mathematical Programming*, 44, 271-291.

Lewis, H. T., W. Culiton, and J. D. Steele. (1956). *The role of air freight in physical distribution.* Division of Research, Graduate School of Business Administration, Harvard University, Boston, Massachusetts.

Listes, O. and R. Dekker. (2005). “A Scenario Aggregation Based Approach for Determining a Robust Airline Fleet Composition,” *Transportation Science*, 39(3), 367-382.

Louveaux, F.V. and R. Schultz. (2003). “Stochastic Integer Programming,” Chapter 4 in A. Ruszczynski and A. Shapiro (eds.), *Stochastic Programming*, Handbooks in

Operations Research and Management Science, Vol. 10, Elsevier Science, Amsterdam.

Lorena, L.A.N. and E.L.F. Senne. (1999). “Improving Traditional Subgradient Scheme for Lagrangean Relaxation: An Application to Location Problems,” *International Journal of Mathematical Algorithms*, 1, 133-151.

McKay, M.D., R.J. Beckman and W.J. Conover. (1979). “A Comparison of Three Methods for Sampling Values of Input Variables in the Analysis of Output from a Computer Code,” *Technometrics*, 21, 239-245.

McNabb, M.E., J.D. Weir, R.R. Hill and S.N. Hall (2015). “Testing Local Search Move Operators on the Vehicle Routing Problem with Split Deliveries and Time Windows,” *Computers and Operations Research*, 56, 93-109.

Moghadam, S.S., S.M.T.F. Ghomi and B. Karimi. (2014). “Vehicle Routing Scheduling Problem with Cross Docking and Split Deliveries,” *Computers and Chemical Engineering*, 69, 98-107.

Moin, N.H. and S. Salhi. (2007). “Inventory Routing Problems: A Logistical Overview,” *Journal of the Operational Research Society*, 58(9), 1185-1194.

Pontrandolfo, P. and O. G. Okogbaa. (1999). “Global manufacturing: a review and a framework for planning in a global corporation,” *International Journal of Production*

Economics, 37(1), 1–19.

Rockafellar, R.T. and R.J.-B. Wets. (1991). "Scenarios and Policy Aggregation in Optimization under Uncertainty," *Operations Research*, 16(1), 119-147.

Ropke, S. and D. Pisinger. (2006). "An Adaptive Large Neighborhood Search Heuristic for the Pickup and Delivery Problem with Time Windows," *Transportation Science*, 40, 455–472.

Rudiansyah, A. and D. Tsao. (2005). "An Integrated Model of the Periodic Delivery Problems for Vending-Machine Supply Chains," *Journal of Food Engineering*, 70, 421-434.

Russell, R. and W. Igo (1979). "An Assignment Routing Problem," *Networks*, 9, 1-17.

Shanmugam, G., P. Ganesan and P.T. Vanathi. (2011). "Meta Heuristic Algorithms for Vehicle Routing Problem with Stochastic Demands," *Journal of Computer Science,* 7(4), 533-542.

Silva, M.M., A. Subramanian and L.S. Ochi. (2014). "An Iterated Local Search Heuristic for the Split Delivery Vehicle Routing Problem," *Computers and Operations Research*, 53, 234-249.

Smilowitz, K., M. Nowak and T. Jiang. (2013). “Workforce Management in Periodic Delivery Operations,” *Transportation Science*, 47(2), 214-230.

Stewart, W.R., Jr. and B.L. Golden. (1983). “Stochastic Vehicle Routing: A Comprehensive Approach,” *European Journal of Operational Research*, 14, 371-385.

Sun, M. (2012). “A Tabu Search Heuristic Procedure for the Capacitated Facility Location Problem,” *Journal of Heuristics*, 18, 91-118.

Tang, J., Y. Ma, J. Guan and C. Yan. (2013). “A Max-Min Ant System for the Split Delivery Weighted Vehicle Routing Problem,” *Expert Systems with Applications*, 40(18), 7468-7477.

Tillman, F. (1969). “The Multiple Terminal Delivery Problem with Probabilistic Demands,” *Transportation Science*, 3, 192-204.

Van Roy, T.J. (1986). “A Cross Decomposition Algorithm for Capacitated Facility Location,” *Operations Research*, 34(1), 145-163.

Van Slyke, R. and R.J.-B. Wets. (1969). “L-Shaped Linear Programs with Applications to Optimal Control and Stochastic Programming,” *Society for Industrial and Applied Mathematics Journal of Applied Mathematics*, 17(4), 638-663.

Vidal, C. J. and M. Goetschalckx. (1997). “Strategic production-distribution models: A critical review with emphasis on global supply chain models,” *European Journal of Operational Research*, 98, 1–18.

Watson, J.-P. and D.L. Woodruff. (2011). “Progressive Hedging Innovations for a Class of Stochastic Mixed-Integer Resource Allocation Problems,” *Computational Management Science*, 8(4), 355-370.

Wets, R. (1983). "Stochastic programming: Solution techniques and approximation schemes", in: A. Bachem, M. GrStschel and B. Korte (eds.), *Mathematical Programming: State-of-the-Art*, Springer, Berlin, 566-603.

Xiong, Y., D. Gulczynski, D. Kleitman, B. Golden and E. Wasil. (2013). “A Worst-Case analysis for the Split Delivery Vehicle Routing Problem with Minimum Delivery Amounts,” *Optimization Letters*, 7(7), 1597-1609.

Yellow, P.C. (1970). “A Computational Modification to the Savings Method of Vehicle Scheduling,” *Operational Research Quarterly*, 21(2), 281-283.

Zhong, H., R.W. Hall and M. Dessouky. (2007). “Territory Planning and Vehicle Dispatching with Driver Learning,” *Transportation Science*, 41(1), 74-89.