Quantitative Models for Managing Multi-Node Store Replenishment Logistics in the Fast-Food Supply Chain

by

Alejandro Vigo Camargo

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Industrial and Operations Engineering)
in the University of Michigan
2023

Doctoral Committee:

Professor Yavuz A. Bozer, Chair
Professor Amy M. Cohn
Professor Mark S. Daskin
Professor Izak Duenyas

Alejandro Vigo Camargo

avigo@umich.edu

ORCID iD: 0000-0003-4589-7985

Dedication

To my parents, Gina Camargo and Gilberto Vigo.

I would not be who I am without your unconditional love and support.

Acknowledgements

I have been incredibly lucky to be surrounded by the best group of people and support system these last couple of years. I could write an entire book full of my gratitude for them, and it would still not be enough. However, I cannot pass the opportunity to thank them, so I'll try to be concise.

To my sister, Vivianette Vigo, thank you for being my best friend. I honestly don't know what life would be without you (mainly because you're my older sister), and I hope I never have to find out. To some of my favorite people in the world – Josue Juan, Daymi Muñiz, and Lou Maldonado – thank you for always taking the trip, picking up the phone, or just sending a text. Having you in my corner for more than 10 years has meant the world to me. To the friends Michigan has gifted me for life– Jocelyn Modelski, Parker Koch, and Marissa Torres, thank you for making Michigan feel like home. Even as we have moved away, from Michigan and each other, I'm so grateful for your friendship and encouragement. I also would have not gotten through the cold Michigan winters without the warm embrace of the queer community in Ann Arbor. I never expected to find such welcoming queer spaces that would allow me to build friendships I will treasure forever. I'm particularly grateful to Ryan Ross, Josh Lott, Alex Thum, Oliver Lockhart, Mike Thompson-Brusstar, Bryan Terrazas, and Pat Zabawa, for welcoming me into their lives and allowing me to be myself always, I hope you know you'll never get rid of me.

I would like to acknowledge all the extraordinary public school teachers in Puerto Rico that in some shape or form guided me throughout my career. I am also intensely grateful to the faculty in the Industrial Engineering department at the University of Puerto Rico - Mayaguez for being my first guides into this field. I want to give special thanks to my undergraduate research advisor, Dr. Hector Carlo, for introducing me to the world of research and logistics, for encouraging me to purse a PhD, and being a source encouragement over the years.

I want to thank my colleagues at the Industrial and Operations Engineering department for inspiring me with their cool research and showing me the path forward. Special thanks to the 2015 IOE PhD cohort, for getting me through that grueling transition to graduate school and their constant support. To my advisor, Prof. Yavuz A. Bozer, thank you for your guidance and feedback along the way. To my dissertation committee, thank you for sharing your time with me and your valuable input on my work. I want to thank the supply chain group at our industry collaborator in the fast-food sector, who provided data and input that was invaluable in shaping the scope and focus of this study. I'm also thankful to the IOE department and to the Rackham Graduate School for the financial support throughout my PhD.

I want to end by saying how deeply grateful I am for my family. Thank you all for always being there and believing in me. To my parents, I will never find the right words to say thank you for everything. Mom, thank you for always pushing me to be better and work harder. I'm sorry for making you worry so much when I don't pick up the phone, but know that your love always grounds me. Dad, thank you for being an incredible father and always looking out for what's best for me. I may not always take your advice, but I hope you know your voice always guides me. I truly am the luckiest to have you both in my life.

Table of Contents

List of Tables

List of Figures

List of Appendices

Abstract

Fast-food restaurants represent a significant sector of the retail industry, with consumers spending over $310 billion in a single year, that has shown consistent yearly average growth of 6% over the last 10 years. However, there is a surprising scarcity of studies focused on the replenishment logistics of this sector. Considering how replenishment logistics are critical to any fast-food company success and how costly the logistical challenges in this area can be, it makes this a particularly interesting area of study. The work discussed in this dissertation addresses this gap by developing quantitative models and decision tools focused on the replenishment logistics in the fast-food industry.

We introduce the store replenishment problem (SRP), which is concerned with minimizing the logistics costs associated with replenishing stores in a network over a fixed time horizon. The objective function of the SRP was defined in collaboration with a supply chain group at a well-known fast-food chain to include four critical cost components associated with the replenishment logistics: transportation, labor, fleet size, and route-time overage costs. We formulate the SRP as a mixed-integer program using a set-partitioning approach. The formulation uses a set of pre-generated potential routes as input and then concurrently determines fleet size, delivery routes, and chooses which routes are going to be completed by a single-driver or a team of two. Using real-world data, we show that the proposed heuristic outperforms the current industry baseline and that the multi-component objective function obtains superior solutions to those obtained with one-dimensional objectives, such as minimizing only the travel distance or the fleet size.

The second model presented focuses on a generalization of the SRP, where building a flexible delivery schedule is included as a decision variable (SRP-FS). This flexible schedule determines both the timing and the quantity of the deliveries at each store while observing the storage capacity of the stores. The objective of the SRP-FS is also to minimize replenishment costs and uses the same multi-component cost structure as the SRP. We developed a two-step simulated annealing metaheuristic, that incorporates an adjusted Savings Algorithm to solve the vehicle routing component. A series of self-generated test problems and real-world data from our industry collaborator are used to evaluate the performance of the heuristic. The results show that the proposed metaheuristic is capable of finding good solutions in reasonable times and that significant cost reductions can be obtained by introducing a flexible delivery schedule.

The last model discussed in this dissertation is concerned with the impact of the store network composition on the SRP. Due to the nature of the fast-food industry, store locations are not uniformly spread through regions and can result in areas with high density of stores areas with relatively isolated stores in remote locations. This mixed composition in the store network can present a logistical challenge for decision-makers when planning store replenishment routes. We propose a quantitative model that exploits the clustered nature of the store network into the solution approach. Using a clustering heuristic, we are able to simplify the decision space of the problem and formulate the SRP as a bin-packing problem to assign clusters to routes. Our computational results show that the proposed heuristic outperforms the original SRP method in almost every test instance, particularly in instances based on real-world data from our industry collaborator.

Chapter 1

Introduction

The replenishment function, in its most general form, presents a costly logistic challenge for most supply chains (Akkerman *et al.* 2010). This is particularly critical in the retail industry, where an effective store replenishment process is needed to ensure the success of companies by allowing costumer access to their products. Fast-food restaurants represent a significant sector in the retail industry, with the USDA (2018) reporting that consumers spent over $310 billion in 2017 alone, with an average consistent growth of 6% year to year. While food supply chains have been extensively studied in the literature (Lemma *et al*. 2014), there is a surprising scarcity of studies focused on replenishment logistics in this large sector of the industry. This dissertation is intended to address the above gap by developing quantitative models and decision tools to manage the replenishment logistics in the fast-food industry.

Fast-food store networks are often composed of a large number of stores spread out across multiple regions, and in some cases multiple countries. Popular fast-food chains such as those under the Yum! Brands umbrella have thousands of locations worldwide. In the US alone, KFC operates over 4,000 stores, while Taco Bell reports 6,611 locations (*Yum! Brands*, no date). Due to the size of the above networks, and the need to maintain inventory regularly and consistently available at all the stores, many of the fast-food companies operate (or have contracts with) regional distribution centers (DCs) out of which all the stores in a given region are replenished on a regular basis. While there are different business models under which fast-food companies manage their replenishment logistics, a common approach is for the replenishment logistics to be coordinated centrally by their supply chain group. The work discussed in this dissertation is based on such an approach and is concerned with the challenges involved in managing the replenishment logistics for a given region with a known DC and network of stores.

A summary of each chapter and its focus are presented as follows. In Chapter 2 we formally define the Store Replenishment Problem (SRP), which is the primary focus of this dissertation. The SRP objective is to minimize the overall replenishment cost by finding an ideal set of routes to satisfy the (known) store demands over a (known) planning horizon, given a fixed delivery schedule, considering both the fleet size and single-driver versus team routes (two drivers). The objective function of the SRP was developed jointly with our industry collaborator (a well-known and large fast-food chain) and formulated to include four key cost components that are critical to the overall replenishment costs: cost due to distance traveled, labor cost, fleet size costs, and the additional cost incurred by those routes that may exceed their time limits.

The multi-component nature of the cost function is one of the key contributions of this chapter. To the best of our knowledge, the SRP is the first problem to incorporate single vs team-driver routes, fleet sizing, labor costs, and route-time overage costs, all concurrently, in making the routing decisions. Through the computational results we present, we are able to show how the above SRP objective function outperforms one-dimensional objective functions (such as minimizing only the travel distance) that have frequently been used in routing problems. A second contribution of this chapter are the insights obtained from our results, using real-world data from our industry collaborator, and sensitivity analysis that can be critical for logistics managers. We developed a solution approach for the SRP that generates a set of potential routes which are then used by a mixed-integer formulation of the SRP as input to find a solution that minimizes overall costs. While being a heuristic approach, it allows for real-world problems to be solved in practical times with commercial solvers like CPLEX.

Chapter 3 is concerned with extending the SRP to include the delivery schedule itself as a decision variable. That is, the SRP with flexible schedule (SRP-FS) generalizes the SRP by relaxing the assumption that the delivery schedule is given and fixed, while still considering the other decision elements (fleet size, routing, single/team routes) and using the same multi-dimensional objective function. While introducing a flexible schedule can have potential cost benefits to the SRP, it significantly increases the complexity of the problem. To gain insight into the problem we first studied a simplified version of the model, where only direct deliveries (that is, out-and-back trips) are allowed for each store. The near-optimal heuristic approach we present for the direct-delivery version is then used to build an initial solution for the more complex milkrun version of the problem (where each truck visits two or more stores). To address the increased complexity of the SRP-FS, we developed a two-step simulated annealing metaheuristic that in each iteration first builds a delivery schedule and then evaluates its corresponding routes until a good solution is identified. The introduction of the SRP-FS and the proposed solution method is the main contribution of Chapter 3. To the best of our knowledge, the SRP-FS is the first study to incorporate this particular set of decision variables and cost components simultaneously. Using real-world data, we are also able to assess the cost benefits of introducing a flexible schedule when compared to the current industry baseline.

In Chapter 4, we study the impact of the location of the stores in the network and exploit store-clustering in solving the SRP. In the fast-food sector, the decision of where to locate the stores is often based on customer demand, local trends and competition, and the number and location of existing stores in the area. Due to the variability in these factors, store locations are often not uniformly spread through the region, resulting in areas with high density of stores, while other areas contain only a small number of stores (or sometimes individual stores) in relatively remote locations. Such a mixed composition of store locations in the network presents a logistical challenge when planning replenishment routes to supply the stores. The key contribution of Chapter 4 is the development of a solution approach that takes advantage of the clustered nature of the store locations in order to simplify the SRP and find better solutions efficiently. Using a clustering-based heuristic, we group stores in the service region to reduce the decision space of the problem. We then formulate the SRP as a bin-packing problem with the objective of minimizing replenishment costs by assigning the above clusters into bins that represent delivery routes. These assignments are then used by a multi-step routing heuristic to build the final routes. The computational results show that the proposed clustering-based heuristic outperforms the original SRP method introduced in Chapter 2 with respect to both solution quality and runtime.

Each chapter discussed above was structured as a stand-alone research article to be submitted for review and potential publication. Thus, each Chapter was prepared with its own introduction, literature review, and results/discussions. The final chapter of the dissertation summarizes the overall work that was performed, and it includes a discussion of the results and possible future research directions.

References

Akkerman, R., Farahani, P. and Grunow, M. (2010) *Quality, safety and sustainability in food distribution: A review of quantitative operations management approaches and challenges*, *OR Spectrum*. doi: 10.1007/s00291-010-0223-2.

Lemma, Y., Kitaw, D. and Gatew, G. (2014) 'Loss in Perishable Food Supply Chain: An Optimization Approach Literature Review', *International Journal of Scientific & Engineering Research*, 5(5).

USDA (2018) *Expenditure for Food at Home and Away from Home, USDA*. Available at: https://www.ers.usda.gov/topics/food-markets-prices/food-service-industry/market-segments/.

Yum! Brands (no date) *Yum.com*. Available at: https://www.yum.com/wps/portal/yumbrands/Yumbrands (Accessed: 24 April 2020).

Chapter 2

Development and Application of a Cost-Driven Decision Model for Store Replenishment Logistics in the Fast-Food Sector[1]

1. Introduction

Although consumer spending at fast-food restaurants reached $310 billion in 2017 (USDA, 2018), there is a scarcity of academic research on the supply chain logistics of this sector. A significant portion of food supply related literature is focused on production, with applications in the agriculture industry, and on reducing food waste; see, for example, (Lemma, *et al.* 2014). While production and waste reduction are relevant, the highly competitive fast-food sector relies on the timely and efficient replenishment of their stores, which is fertile ground for academic research.

This study is focused on the fast-food sector, where customer loyalty, driven by price and quality of the brand, is key for success (Shokri, *et al.* 2014). The store networks for fast-food companies often span large regions or countries and may involve thousands of locations. For example, as of 2018, Domino's Pizza reports over 5,400 stores in the US, with plans to expand to about 8,000 locations (Klein, 2018). Similarly, Yum! Brands operates over 43,000 locations globally, including 4,062 KFC stores, 7,447 Pizza Hut stores, and 6,611 Taco Bell stores in the US alone (*Yum! Brands*, nd). Given the large number of stores, their geographical spread, and the need to replenish them frequently, the store network is often divided into regions, with a distribution center (DC) serving each region. Usually, the store replenishment operations are

[1] This Chapter was accepted for publication as: Vigo Camargo, A. and Bozer, Y. A. (*in press*) 'Development and application of a cost-driven decision model for store replenishment logistics in the fast-food sector.' *International Journal of Logistics Systems and Management*

coordinated through a company's central planning group, using their own DCs and trucks/drivers. Even when such operations are outsourced, using a central DC for each region is fairly common.

Similar settings are also found in the grocery-food sector. However, grocery stores carry a much larger number and variety of stock keeping units (SKUs), and in many cases (such as wine, soft drinks, and dairy products) they use a *hybrid model* where some goods are delivered through a DC while others are delivered by the vendors through a direct store delivery (DSD) model, which represents about 25% of their sales (GMA, 2008). Some of the major grocery-food suppliers/producers are moving away from the DSD model, opting instead to ship their goods to their customers' DCs (Staff, 2019). While such a trend brings the grocery-food sector closer to our study, given the above differences in scale and variety, and the fact that DSD is still common for groceries, we limit our attention to the fast-food sector.

Store replenishment is a primary source of cost (Akkerman *et al.* 2010). The travel costs between the DC and the stores is a significant contributor to the overall cost, which is perhaps why a majority of the publications reviewed in a survey of routing papers by Braekers *et al.* (2016) have distance-dependent objective functions. However, other factors, such as the labor cost associated with delivering inventory, the fleet cost, and whether routes are driven by a single driver or a team of two, can also play a major role in determining the total replenishment cost. Balancing the above costs while making resource allocation decisions and meeting multiple time/distance constraints is often a complex challenge, which perhaps persuaded some to use simpler, one-dimensional objectives. While distance minimization alone may be suitable for some routing problems, for the fast-food store replenishment problem (SRP), using a one-dimensional objective neglects key trade-offs that may lead to inferior solutions. In this paper we address the above challenge by developing a cost-based model that incorporates distance-based costs as well as cost of delivery labor and other cost elements into the objective function to minimize the overall cost of replenishment logistics. Using real-world data from our industry collaborator, we show how such an approach leads to better solutions and provides key managerial insights to improve competitiveness.

2.1 Objective and Contribution

We present a cost-driven model to minimize the logistics costs associated with the SRP, which we define as determining the number of trucks, the delivery routes, and which routes will be single- or two-driver (team) routes to support a given delivery plan (i.e., set of orders for each store in a region) over a fixed time horizon, with a known DC location and known distances. In addition to fleet-sizing and vehicle routing, the model considers single vs team routes and the unloading labor. The objective is to minimize the total replenishment cost to satisfy store orders while accounting for route-time and truck-capacity constraints.

Given a delivery plan, the SRP is structurally similar to a vehicle routing problem (VRP), since the main concern is the construction of low-cost routes to deliver the necessary inventory to the stores. However, the SRP is also concerned with minimizing the fleet cost and the labor cost associated with unloading inventory. Additionally, the SRP considers the added cost incurred by routes that extend beyond their time limits (overage routes). Unlike traditional vehicle routing models, the proposed model considers single-driver and team routes, which are a common practice (Goel *et al.*, 2019). Although the same truck might be used for a single- or team-route, the limits on route length and usable truck capacity are different since team routes can complete longer routes and deliver larger quantities. Accordingly, single- and team-routes have a specific cost structure and constraint set associated with each. Capturing the interplay between the above costs is an essential component of the SRP.

We formulate the SRP as a mixed-integer programming (MIP) model, with an objective that incorporates multiple cost elements. Due to complex routing decisions and constraints, a set of pre-generated potential routes are used as input. To generate the routes, we utilize a simple clustering heuristic. The proposed model, including the objective function, was developed with extensive input from the supply chain team of a national fast-food company who was our collaborator for the study.

The main contributions of the paper are twofold. First, we developed a cost-based objective that incorporates multiple, relevant cost elements associated with fast-food store replenishment logistics. To the best of our knowledge, this is the first store replenishment study that incorporates single/team-driver routes, route overage costs, and fleet sizing decisions concurrently. We show that our cost function outperforms one-dimensional objective functions, frequently adopted in routing problems. Second, through our computational results, using real-world data and a series of sensitivity analyses, we obtain insights that can be critical for management in the fast-food sector. Furthermore, the proposed approach allows large-scale problems to be solved with a standard solver such as CPLEX.

The remainder of the paper is organized as follows. In Section 2, we review the relevant literature. The MIP formulation and the cost function are presented in Section 3. The route-generating clustering heuristic is described in Section 4. In Section 5, we discuss the computational results and present benchmark comparisons. The sensitivity analyses we performed and the results are presented in Section 6. Lastly, Section 7 presents our conclusions and possible future research directions.

2. Literature Review

Most problems that involve routing are modelled as a Vehicle Routing Problem (VRP), which is an extensively-researched, NP-hard problem (Lenstra and Kan, 1981), with books and survey articles devoted to the topic; see, for example Sharda *et al.* (2008), Laporte (2009), Toth and Vigo (2014), and Braekers *et al.* (2016). Food distribution has been one of the many applications for the VRP. For example, Tarantilis and Kiranoudis (2001) focused on perishable foods and developed a metaheuristic to minimize the distance-based cost for milk distribution. In a related paper, Tarantilis and Kiranoudis (2002) developed an algorithm to minimize the total distance traveled in a meat distribution application.

As VRP research expanded, more variants emerged. A recent "concise review" of emerging VRP variants cites over 300 papers (Vidal *et al.*, 2019). Two extensions of the VRP that are relevant to the SRP are the Period(ic) VRP (PVRP) and the Multi-Period VRP (MVRP). The PVRP focuses on selecting a delivery schedule from a predefined set of schedules for each node in a network, and building routes to support the selected delivery schedule (Francis *et al.*, 2008). In the classical PVRP, the customers must be visited with certain frequencies during the time horizon. While both the SRP and PVRP include routing decisions over multiple periods, as explained below, there are specific differences between them.

The PVRP was introduced by Beltrami and Bodin (1974), and its first formulation was presented in Christofides and Beasley (1984) as an integer program (IP). Building upon the IP formulation of the PVRP, Cordeau *et al.* (1997) proposed a Tabu search heuristic to solve a mixed-integer formulation of the PVRP. Similarly, in Chao *et al.* (1995), which was later expanded by Gulczynski *et al.* (2011), the authors propose a two-step heuristic that first solves an IP to develop an initial delivery schedule, and then uses an improvement heuristic to find an ideal routing and schedule. Additionally, other heuristics developed originally for the VRP were extended to the PVRP by Alegre *et al.* (2007) and Hemmelmayr *et al.* (2009).

Francis *et al.* (2006) presented an approximating method to solve the PVRP with "service choice," which allows the model to choose the frequency of visit for each customer, and change the structure of the delivery schedule accordingly. However, as is the case with most of the PVRP literature, the delivery frequencies are based on predetermined patterns (such as daily deliveries, a delivery every other day, etc.), narrowing the possibilities. Unlike the PVRP, the SRP assumes that the delivery schedule over the time horizon has been fixed and it does not necessarily follow a predetermined pattern.

Virtually all the PVRP papers assume the number of trucks is given, and they use it to limit the number of routes completed in each period. In contrast, the SRP treats the number of trucks as a decision variable. One exception is Gaudioso and Paletta (1992), who propose a heuristic to solve the PRVP with the main objective of minimizing the number of trucks required. Their model is similar to the SRP in that the number of trucks is a decision variable and the maximum travel time for each route is limited. However, the SRP does not focus only on the number of trucks; it is concerned with minimizing the overall cost, which includes travel and labor cost in addition to the truck cost. Another key difference is that, although we also limit the maximum route-time, the SRP allows a route to exceed the limit by imposing an overage cost.

The MVRP, on the other hand, which was formally defined by Archetti *et al.* (2015), shares many similarities with the PVRP, the main one being its goal to build routes for a set of customers over multiple time periods. While the MVRP also selects a delivery schedule, it assumes the customers have delivery due dates, and in contrast to the PVRP, there is no fixed frequency for the deliveries. Both Archetti *et al.* (2015) and Larrain *et al.* (2019) formulated the MVRP with the objective of minimizing the transportation cost, considering inventory holding and a late-delivery penalty. In López-Santana *et al.* (2018), the authors present a three-phase approach for the MVRP to minimize the total travel time. Wen *et al.* (2010) studied a dynamic extension of the MVRP, with customer demands updated each day. The authors present a three-phase heuristic to solve a multi-objective formulation of the problem that tries to minimize travel distance and customer wait time, while balancing work. Although the MVRP has a broader decision scope than the PVRP, the differences we described between the SRP and PVRP apply to the MVRP as well, particularly the fixed schedule, the number of trucks being a decision variable, and our multi-component cost function.

To the best of our knowledge, none of the published PVRP or MVRP studies explicitly consider single- vs team-driving routes. Our formulation explicitly considers them, which bears some similarity to the Heterogeneous VRP (HVRP). Introduced by Golden *et al.* (1984), the HVRP focuses on route optimization with a fleet of vehicles of varying parameters, such as vehicle capacity. While the SRP assumes a homogeneous fleet of vehicles, the cost, the (effective) capacity, and the length limit of each route depends on whether it is a single-driver or team route. This is comparable to choosing between two types of vehicles in a non-homogeneous fleet. However, most HVRP publications do not consider multi-period time horizons, and they only focus on fleet dimensioning. Further details on the HVRP are presented in Koç *et al.* (2016).

An advantage of our formulation of the SRP lies in the objective function, which incorporates multiple cost elements that account for miles traveled, labor, truck cost, excess route-time costs, and single- vs team-routes (see Section 3). Instead of being one-dimensional, our objective function captures the inherent trade-offs between the above cost elements, which not only leads to superior solutions but also reveals managerial insights as we show later in the paper. The proposed SRP heuristic serves as at useful tool for planning and decision-making in the fast-food supply chain and similar settings.

3. Formulating the SRP

We present our assumptions, the cost model, and the resulting MIP model.

3.1 Modelling Assumptions

The time horizon is modelled as a finite set of discrete time periods. For our application, each day represents a time period. Store demand is represented by the orders a store places (at most one order per store per day), which are fulfilled by the DC. Based on the store orders, a delivery plan is developed for the region; it specifies the delivery quantity and the time period for each delivery at each store over the time horizon. Given the delivery plan, the following assumptions are made for the MIP model:

(1) Each store must receive the planned delivery quantity in the specified time period; no early or late deliveries are allowed.

(2) The stores are responsible for managing their own inventories and they each own the store inventory. Therefore, inventory carrying cost at the stores is not considered. We assume sufficient stock to support the delivery plan is available at the DC. Expiration dates and shelf life are considered by the DC in filling the store orders.

(3) Routes are classified into single-driver and team routes (two drivers).

(4) The trucks are physically identical. However, the usable (or effective) capacity of a truck depends on the type of route (single-driver or team).

(5) A truck can complete at most one route per time period.

(6) The trucks and drivers are available at the beginning of each time period.

(7) A store is visited by at most one truck per time period (no split deliveries).

(8) Potential routes are predetermined using a route-generation heuristic.

(9) Store demand (measured in pounds) is for a single, aggregate SKU.

As the first assumption suggests, the SRP is concerned with minimizing cost by choosing the appropriate routes and fleet size needed to support the delivery plan. This assumption is aligned with the business model of our industry collaborator, where each store has a predetermined delivery plan. Furthermore, although the stores are charged a nominal delivery fee, all the transportation and delivery costs are borne by the company, which gives them a strong incentive to minimize the cost of supporting the delivery plan. While the delivery plan itself is also of interest, the development/optimization of the delivery plan is beyond the scope of our paper.

The second assumption helps us focus on supporting the delivery plan instead of inventory management at the DC. Per our industry collaborator, stock shortages at the DC are rare, and if it occurs, they have systems in place to minimize the impact on the stores. Unless there is a major supply disruption, shortages at the DC are not common and they are rectified quickly.

The third assumption stems from an industry practice of using multiple drivers on a route to allow longer routes and larger delivery quantities, without violating federal hours-of-service regulations (MacMillan, 2018). Assumption 4 addresses single-driver vs team routes, as each type of route has specific parameters. Assumptions 5, 6, and 7 are made primarily to simplify the model.

The eighth assumption, i.e., using a set partitioning approach for our formulation, is advantageous when capturing complex cost functions and intra-route constraints (Toth and Vigo, 2014). The last assumption, i.e., using an aggregate SKU, allows us to limit the total weight on a truck without tracking each SKU individually. Our industry collaborator also expresses their delivery data in total pounds using an aggregate SKU.

3.2 The Cost Model

The cost model and the MIP formulation (section 3.3) are general-purpose, and they capture the inherent trade-offs in the SRP. While the primary computational results in our study are based on specific parameter values furnished by our industry collaborator, whose input was instrumental in the development and validation of the model, the cost model and the MIP formulation have broad applicability, meaning they can be used in other SRP applications as long as the appropriate parameter values are specified.

Using a one-dimensional objective, such as just minimizing distances, is common in the literature. However, a one-dimensional objective is at best a surrogate objective that captures only a portion of the problem, and it may fall short of adequately capturing key trade-offs inherent to the SRP. For example, in some cases, one may be able to reduce overall cost by using a team route instead of a single-driver route, even if the total distance traveled increases.

Working with our industry collaborator, we defined the costs associated with the SRP based on (1) the miles traveled, (2) the labor associated with unloading inventory delivered (in pounds), and (3) the number of trucks needed. Another important aspect we gathered from our industry collaborator and incorporated into the model is that, although US federal regulations limit the amount of time a driver can work continuously to 14 hours (FMCSA, 2017), some routes require more time due to the size of the delivery region. In such cases, the driver is given an allowance to rest overnight, and the route becomes a two-day route. Team routes may also be used in such cases, since two drivers can alternate their working periods. Even with team routes, however, some routes might be long enough to qualify as a two-day route. Since each time period is modelled as a day, a fourth cost component, namely, an “overage cost,” was introduced (for both single and team routes) to capture the additional cost of those routes that exceed the one-day time limit. The above four cost components are used in the objective function of the MIP model, which is described in the next section.

3.3 MIP Formulation

The MIP model for the SRP uses the following parameters:

N = set of stores in the region ($i = 1, \dots, |N|$)

R = set of routes ($r = 1, \dots, |R|$)

T = planning horizon ($t = 1, \dots, |T|$)

m_r = distance in miles of route r

f_{ir} = indicator parameter; equals 1 if store i is visited in route r, 0 otherwise.

d_i^t = demand in pounds (lbs) of store i in time period t

A distinctive aspect of the proposed formulation is that both single and team routes are considered. Single-driver routes have more restrictive limits on truck capacity and route time length than team routes as well as different cost parameters. To capture the above differences, additional parameters are defined as follows:

U_V^1 (U_V^2) = upper limit of truck capacity, in pounds, for single (team) routes

U_T^1 (U_T^2) = upper limit of route time length, in minutes, for single (team) routes

l_r = 1 if route r is a team route; 0 if a single route

B_{lb} = unloading time, in minutes/pound unloaded

B_n = stop time, in minutes/stop (equal for each stop)

B_m = travel time, in minutes/mile traveled

C_V = daily cost/truck required

C_m^1 (C_m^2) = cost/mile traveled in a single (team) route

C_{lb}^1 (C_{lb}^2) = cost/pound delivered in a single (team) route

C_U = fixed overage cost for each route that exceeds the route-time upper limit

The upper limits on truck capacity are determined based on the space and weight limitations of the truck as well as what managers consider to be a reasonable amount of inventory a driver can unload on a route. The upper limits on route times, on the other hand, are defined subject to federal hours of service regulation depending on single vs team routes. Route times are calculated by considering the driving time, a fixed stopping time/store, and the unloading time/store.

The decision variables are defined as follows:

a_r^t = 1 if route r is used in time period t; 0 otherwise

q_{ir}^t = number of pounds delivered to store i on route r in time period t

x_r^t = 1 if length of route r in mins in period t goes over the upper limit; 0 otherwise

p_r^t = amount of time in minutes that route r in period t exceeds the upper limit

K = fleet size

The objective function is based on the four cost elements described in Section 3.2. The first term reflects the distance cost for each single or team route used:

$$\sum_{\mathrm{t}} \sum_{r} [C_m^2\, l_r + C_m^1 (1 - l_r)]\, m_r\, a_r^t \tag{1}$$

Similarly, the second term reflects the labor cost incurred due to unloading inventory, based on single or team route:

$$\sum_{t} \sum_{r} \sum_{i} [C_{lb}^2\, l_r + C_{lb}^1 (1 - l_r)]\, q_{ir}^t \tag{2}$$

The third term calculates the overage cost for each route that exceeds the route time limit:

$$\sum_{t} \sum_{r} C_U\, x_r^t \tag{3}$$

Lastly, the final term in the objective function reflects the cost of providing K trucks over the time horizon:

$$C_V\, T\, K \tag{4}$$

The sum of the four components constitutes the objective function shown below. Given a delivery plan, and a set of potential routes, the proposed formulation determines the number of trucks required and the routes to be used in each period to minimize the total cost over the time horizon. The MIP formulation is presented as follows:

$$\min \sum_{t} \sum_{r} [C_m^2\, l_r + C_m^1 (1 - l_r)]\, m_r\, a_r^t + \sum_{t} \sum_{r} \sum_{i} [C_{lb}^2\, l_r + C_{lb}^1 (1 - l_r)]\, q_{ir}^t + \sum_{t} \sum_{r} C_U\, x_r^t + C_V\, T\, K \tag{5}$$

$$s.t. \ \sum_{r} q_{ir}^t = d_i^t \qquad \forall\ i \in N, t \in T \tag{6}$$

$$\sum_{i} q_{ir}^t \leq a_r^t\, [\, U_V^2\, l_r + U_V^1\, (1 - l_r)] \qquad \forall\ r \in R, t \in T \tag{7}$$

$$\sum_{r} a_r^t\, f_{ir} \leq 1 \qquad \forall\, i \in N, t \in T \tag{8}$$

$$\sum_{r} a_r^t \leq K \qquad \forall\ t \in T \tag{9}$$

$$\sum_{i} (B_{lb}\, q_{ir}^t) + \sum_{i \in \{i \mid d_i^t > 0\}} (B_n\, \frac{q_{ir}^t}{d_i^t}) + B_m\, m_r\, a_r^t \leq [U_T^2\, l_r + U_T^1 (1 - l_r)] + p_r^t \qquad \forall\, r \in R, t \in T \tag{10}$$

$$p_r^t \leq [U_T^2\, l_r + U_T^1 (1 - l_r)]\, x_r^t \qquad \forall\, r \in R, t \in T \tag{11}$$

$$x_r^t \leq 1 \qquad \forall\, r \in R, t \in T \tag{12}$$

$$K \geq 0 \qquad \text{integer} \tag{13}$$

$$a_r^t \in \{0,1\} \qquad \forall\, r \in R, t \in T \tag{14}$$

$q_{ir}^t \geq 0 \qquad \forall\, i \in N, r \in R, t \in T \qquad (15)$

$x_r^t \in \{0, 1\} \qquad \forall\, r \in R, t \in T \qquad (16)$

$p_r^t \geq 0 \qquad \forall\, r \in R, t \in T \qquad (17)$

Constraints (6) ensure that store demand is met in each period, while constraints (7) enforce the truck capacity. Constraint set (8) prevents split deliveries. Constraint set (9) ensures that the number of routes used in any period does not exceed the number of trucks. Set (10) calculates the estimated route time and determines if it exceeds the time limit. The left-hand side of the constraint contains three terms that calculate the amount of time needed for unloading the inventory, the fixed time associated with each stop on the route, and the drive time, respectively. On the right-hand side, the variable p_r^t captures the excess route time that goes over the upper limit. Set (11) identifies the number of routes that exceed the time limit to assign the appropriate overage cost in the objective function. While constraints (12) ensure that the route time for overage routes is bounded. Constraints (13) – (17) define the variables and the non-negativity constraints.

4. Clustering Heuristic for Route Generation

One challenge associated with the set partitioning-based approach we adopted is generating potential routes *a priori.* For example, with just 50 stores, if every route with up to 7 stores is feasible, it would yield over 118 million routes. To address this challenge, we use a simple clustering heuristic to pre-generate a manageable set of potential routes. It is adapted from the nearest-neighbor-based clustering and routing heuristic (nCAR) presented by Sarkar *et al.* (2018). Our goal is not to develop an optimum-seeking algorithm for the SRP, but rather to use a simple heuristic to obtain solutions in reasonable times that allow us to investigate the impact of the various cost elements in the objective function on the solution structure and at the same time glean managerial insights made possible by a cost model with multiple elements. Given the groundwork we laid with the proposed model, subsequent studies can explore other solution techniques, such as column generation/pricing schemes, to obtain better solutions.

The proposed clustering heuristic considers each period sequentially, and it allows for more than one set of parameters to be considered when building the routes, which lets the heuristic accept different parameters between single and team routes when determining the composition and size of the routes. Usable truck capacity is a constraining factor when determining which store will be added next to a cluster, while route time limits are used to constrain the cluster size.

For each period, given a set of stores that must be visited, the heuristic creates multiple clusters of increasing size (measured by the number of stores). For each cluster, a route is generated by finding the shortest path through the stores in the cluster, beginning and ending at the DC. The parameters for truck capacity and route time for both single and team routes are used as input, along with a distance matrix, store demand for each period, and the maximum number of stops allowed on a route, say, Z.

Given the above parameters, the following heuristic is employed, where z serves as an iteration counter, and s reflects the current cluster size. The steps are executed for each time period t, considering first the parameters for single-driver routes and then team routes:

1. Let N_t represent the (sub)set of stores that must be visited in period t.

2. Set $z = 1$. Select store $i \in N_t$ as the seed store of cluster i_z and set $s = 1$. Initialize the cluster demand to the demand of store i, i.e., set $d_{i_z} = d_i^t$, and the cluster time length to the driving time from the DC to store i plus the time associated with visiting store i, which is equal to the unloading time, $d_i^t B_{lb}$, and the fixed time per stop, B_n.

3. If $s = z$, or the cluster time length exceeds the maximum route length in time (U_T^1, U_T^2), go to Step 5. Otherwise, go to Step 4.

4. Determine the next store k to add to the cluster as follows:

   4a. For each store already in the cluster, identify its nearest neighbor in distance. (Ties are broken by the smallest store number.) Verify that adding the nearest neighbor to the cluster would not violate the truck capacity. If it does, then consider the next closest neighbor. Continue in this fashion until a feasible neighbor is found, or there are no neighbors that can be added. The above is repeated for each store already in the cluster, leading to a set of potential stores.

   4b. If the set of potential stores from Step 4a is empty, then set $z = Z$ and go to Step 7. Otherwise continue to Step 4c.

   4c. Given the set of potential stores identified in Step 4a, evaluate the increase in travel distance by inserting each potential store, one at a time, before and after its nearest neighbor in the cluster. Choose the position that yields the smallest increase. Record the increase in distance for each potential store.

   4d. Set the store that leads to the smallest increase in distance as store k and add it to the cluster. Update the cluster size (i.e., set $s = s + 1$), the cluster demand, (i.e., set $d_{i_z} = d_{i_z} + d_k^t$), and the cluster time length to include store k. Go to Step 3.

5. Compare the new cluster with previously saved clusters to avoid duplicates.

6. Determine the shortest path to visit all the stores in the cluster from the DC. Save the shortest path as a route and add it to the set of routes R, identifying it as a single or team route.

7. If $z = Z$, then remove store i from N_t and go to Step 2. Otherwise, set $z = z + 1$, and go to Step 3.

The shortest path in Step 6 was obtained by solving a Traveling Salesman Problem (TSP) using Google's OR-Tools Python library (Google, 2019) which uses a greedy descent heuristic to find a solution. Although the routes are generated by time period, the model is not restricted to use any route for a specific time period; rather, it considers the entire pool of routes when deciding which routes to use. The above heuristic allows some flexibility in the order in which the routes are built since the final set of routes does not change with the order in which the seed stores are selected, or whether single or team routes are generated first.

5. Computational Results

In this section we discuss two sets of results. For the first set, we solved a series of test instances of varying sizes with randomly generated data to evaluate the size limitations of the SRP heuristic. We also compare the results obtained from the SRP heuristic with baseline results obtained from alternative, one-dimensional objective functions. For the second set, using real-world data provided by our industry collaborator, we developed multiple problem instances to discuss the solutions obtained from the SRP heuristic and to compare them to the industry baseline as well as the other one-dimensional baselines. The SRP heuristic was programmed with 64-bit Python 3.7.4, while the MIP model was solved with CPLEX 12.7 on a Windows computer with a 64-bit 2.50 GHz Intel Core i7 and 8 GB of RAM.

5.1 Randomly Generated Test Instances

We randomly generated a series of test instances of varying sizes to evaluate how the proposed heuristic performs with different data scenarios as shown in Table 2-3. The two main factors we varied on each instance is the problem size, defined by the number of stores, and the delivery frequency, defined by how many times during the time horizon each store needs replenishment.

The largest problem size tested was 350 stores which is a representative average upper limit on the number of stores in a region for our industry collaborator. Store locations were randomly scattered following a uniform distribution within a 600 × 600 grid, with the DC placed at the center (plots of the locations on each instance are included in the Appendix). The distance matrix was calculated for each instance assuming Euclidean distances. The frequency of delivery for each store was randomly generated following a normal distribution, with the average changing according to the level shown in Table 2-1. Once a frequency was determined for each store, a uniform distribution was used to randomly assign each demand to a period within the time horizon. The demand for each store was also generated using a normal distribution based on the parameters shown in Table 2-3. Data from our industry collaborators were used to define the values for all other parameters in the model (truck capacity, unit costs, etc.), these values are discussed in Section 5.2

Table 2-1. Parameters for the Random Test Instances

Variable	Level	Parameter Value
Problem Size	3 levels	$N = 10$
		$N = 100$
		$N = 350$
Order Frequency	Low	$\mu = 1; \sigma = 0.60$
	Average	$\mu = 2; \sigma = 0.60$
	High	$\mu = 4; \sigma = 0.60$
Store Demand		$\mu = 2030; \sigma = 760.15$

For each combination 5 random test instances were generated following the procedure described above, for a total of 45 test instances. Using the SRP heuristic, we were able to solve all 45 instances within a time limit of 3 hours. Table 2-2, where low and high demand frequency are denoted by “L” and “H,” respectively, shows the average total runtime for each combination as well as the portion of the total time dedicated to the CPLEX solver under the MIP Runtime columns.

Table 2-2. Average runtime (in secs) per test instance

Instances	Total Runtime		MIP Runtime	
	Avg	**Std Dev**	**Avg**	**Std Dev**
N10-L	2.05	0.81	0.25	0.09
N10	3.71	1.13	0.56	0.18
N10-H	3.33	0.92	0.34	0.15
N100-L	34.67	3.40	0.95	0.15
N100	60.40	7.77	1.92	0.69
N100-H	76.99	7.48	3.18	0.93
N350-L	1,913.68	198.05	17.51	2.80
N350	5,072.02	724.94	56.72	1.05
N350-H	10,109.55	1,395.58	139.26	9.63

As expected, the total runtime increases with the number of stores in the network. Within instances of the same size, the total runtime increases as the order frequency increases. The majority of the overall runtime is dedicated to the clustering heuristic and building the MIP model. These two processes account for 94% of the total runtime on average across all the test instances. On the other hand, the average solver runtime across all the instances was under a minute, which indicates that, with the proposed approach, we are able to find solutions for various scenarios in a practical time. A key factor to consider is that, as the number of stores and the order frequency increase, the clustering heuristic generates a substantially larger number of routes, which leads to an increase in the size of the MIP since some variables and constraints are indexed by the number of routes and stores. This explains the significant increase in runtime for the instances with 350 stores which generate approximately 10,000 routes each.

To evaluate the performance of our proposed multi-dimensional objective function, we solved the same set of instances with alternative/one-dimensional objectives for comparison as follows: 1) Minimizing only the truck cost (i.e., the fleet size), and 2) Minimizing only the cost of miles traveled (i.e., using only a distance-based objective as most studies do). For both of the alternative objectives, once a solution was found by the model, we calculated the total cost of the solution using our proposed cost model to have a direct comparison. In Table 2-3, we summarize the average percentage difference between our proposed multi-dimensional objective function (MIP Soln) and the two alternative one-dimensional ones using the same set of instances.

Table 2-3. Comparison of Alternative Objective Functions

Instances	MIP Soln	Min Dist	% Gap	Min Truck	% Gap
N10-L	$9,153.02	$9,413.92	2.7%	$9,329.00	1.8%
N10	$13,009.37	$13,205.86	1.6%	$14,941.01	15.0%
N10-H	$19,321.88	$19,457.41	0.7%	$21,266.34	9.9%
N100-L	$39,917.25	$39,932.94	0.0%	$47,861.54	19.8%
N100	$53,877.55	$54,367.47	0.9%	$65,996.03	22.5%
N100-H	$86,304.88	$86,713.60	0.5%	$99,367.30	15.1%
N350-L	$136,899.52	$138,081.19	0.9%	$161,656.75	19.4%
N350	$166,462.99	$167,316.36	0.5%	$195,828.80	17.6%
N350-H	$238,665.22	$239,848.54	0.5%	$273,002.85	15.5%

These result show that on average the multi-dimensional objective function outperforms both of the commonly used one-dimensional objectives across all the instances. Overall the solutions obtained when only truck cost is considered are 15% higher than the ones obtained with our objective function on average. The solutions obtained with the distance-based objective have a significantly smaller gap, being overall about 1% over our solutions. However, the differences in Table 2-3 ultimately depend on the specific cost parameters used and it does not diminish the theoretical significance of incorporating multiple, relevant cost components in the objective function. For example, if the labor cost increases relative to fuel cost, the differences in Table 2-3 would likely be more pronounced.

5.2 Industry Data

Two problem instances were developed using data from our industry collaborator. Each instance is based on orders received from 181 stores over 6 days. The two instances are representative of the operations of our collaborator, where plans are made on a weekly basis with no Sunday deliveries. Both instances involve the same 181 stores but use different delivery plans. Figure 2-1 shows the location of the stores with the map details omitted to maintain data confidentiality.

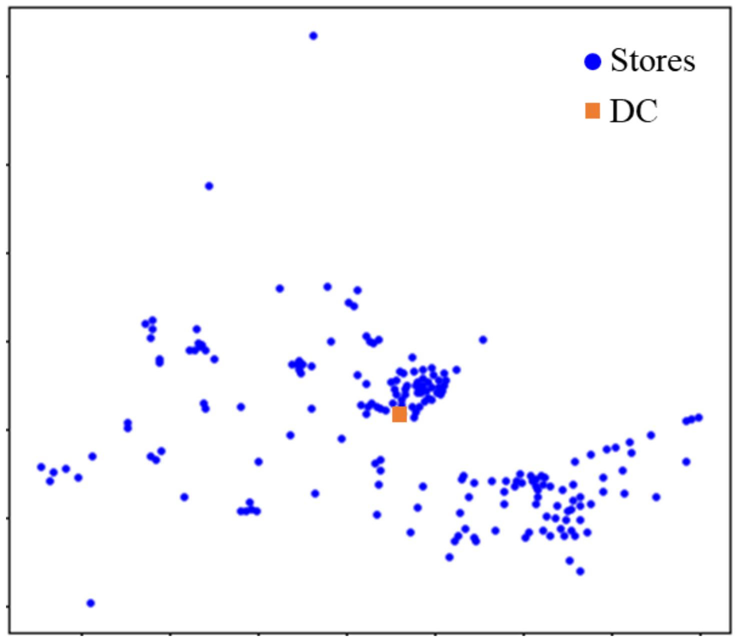


Figure 2-1. Map of Store and DC Location

Store demand and a distance matrix are the main inputs to our approach. Actual road distances for the distance matrix were obtained from the Google Maps API (Google, 2020). Additional input data includes the truck capacity and the maximum route time for both single and team routes, the average minutes/mile traveled, and the appropriate values for the cost parameters. Since the parameter values and cost estimates used for the test instances represent actual values used by our industry collaborator, they are not shown for confidentiality reasons. However, we can report their relative relationships. For example, the truck capacity for team routes is 35% higher than that of single-driver routes, as two drivers are allowed to handle higher quantities for delivery. Similarly, the maximum route-time is 30% higher for team routes. However, compared to single routes, team routes are 20% higher in cost/mile traveled, and 35% higher in cost/pound delivered. The overage cost for exceeding the route time limit was set as the daily truck cost, to account for the incremental usage of the trucks on such routes.

The results obtained from the SRP heuristic are compared to the industry baseline cost, which was calculated employing the data obtained from our collaborator and their current routes, using the same cost model presented in Section 3. Our industry collaborator uses a commercial routing software to construct their routes and update them manually as new stores are opened and to accommodate weekly changes in demand.

5.3 Discussion of Industry Data Results

The "MIP solution" we get from solving the model with CPLEX is exact. However, since the MIP model considers only the routes generated by the clustering heuristic, the MIP solution is a heuristic solution. The overall runtime for the SRP heuristic is 11.37 (14.8) mins for instance 1 (2), with the MIP solver taking only 6-9 secs of this time. The bulk of the time is used to generate the routes (3,582 and 3,294 routes for instance 1 and 2, respectively) and build the MIP model. In addition to the industry baseline, we solved the same two instances with alternative/one-dimensional objectives for comparison as follows: 1) Minimizing only the truck cost (i.e., the fleet size), 2) Minimizing only the cost of miles traveled (i.e., using only a distance-based objective as most studies do), and 3) Using the same cost function but allowing only single-driver routes.

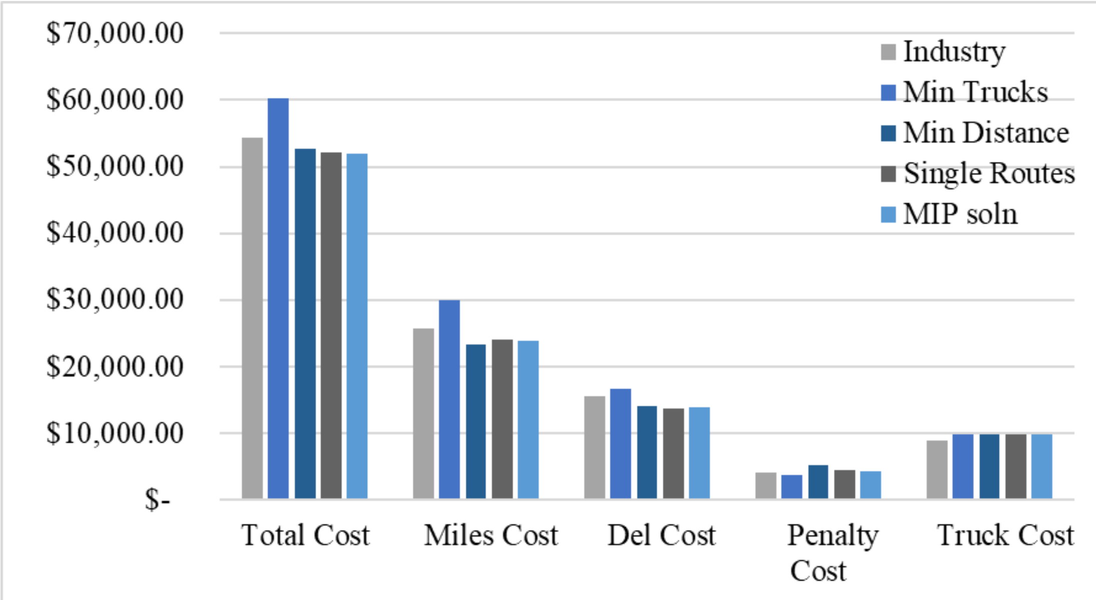


Figure 2-2. Cost Comparison, Instance 1

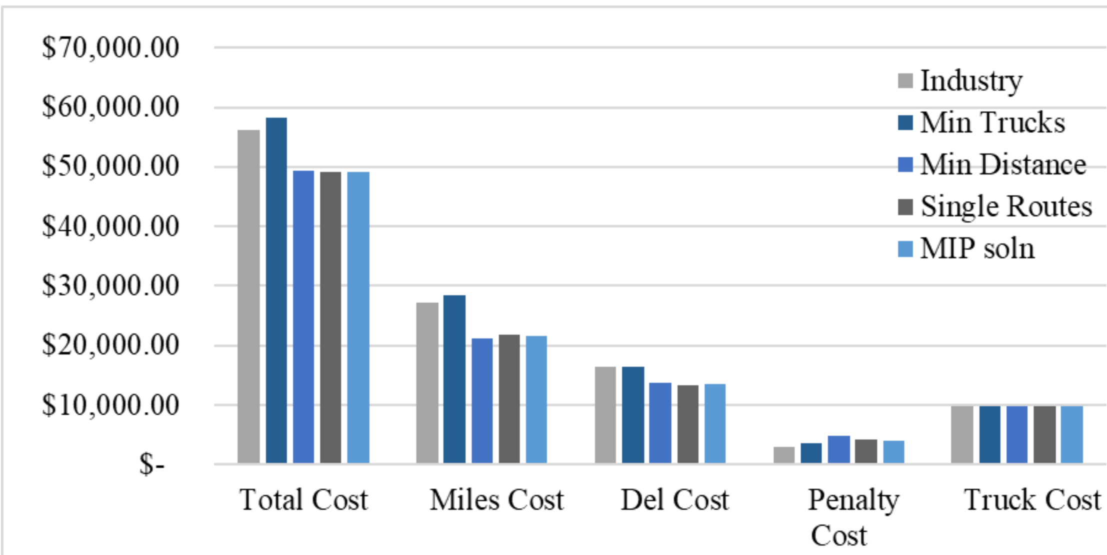


Figure 2-3. Cost Comparison, Instance 2

Figure 2-2 (2-3) shows the cost of the MIP solution versus the alternative baselines for instance 1 (2). For both instances, the MIP solution outperforms all the baselines. Relative to the industry baseline, our solution yields a reduction of about 4% and 13% in total cost for each instance. This is a significant cost reduction for vehicle routing problems where, considering the large transportation volumes typically associated with these problems, even a 2% reduction is considered substantial (Hasle and Kloster, 2007). Looking at the breakdown by cost component, the graphs show that the MIP solution does not incur the lowest cost by component. For example, the distance-only objective solution has a miles traveled cost that is 3% lower than that of the MIP solution, similarly the truck cost objective has a solution with an overage cost 11% lower than the MIP solution for instance 1 (4% for instance 2). However, when we consider all the cost dimensions together, the MIP solution identifies trade-offs that result in a lower overall cost. Table 2-4 presents a detailed comparison of the MIP solution and the industry baseline.

Table 2-4. Baseline versus MIP Results

	Baseline - 1	MIP Soln - 1	Baseline - 2	MIP Soln - 2
Total Cost	$54,283.48	$51,946.13	$56,284.60	$49,399.04
Miles Cost	$25,661.31	$23,936.69	$27,191.51	$22,129.20
Del Cost	$15,585.85	$13,936.15	$16,353.05	$13,789.10
Overage Cost	$4,147.92	$4,296.06	$2,962.80	$3,703.50
Truck Cost	$8,888.40	$9,777.24	$9,777.24	$9,777.24
Total Runtime (min)	-	16.53	-	6.55
MIP Runtime (min)	-	0.18	-	0.11
Miles	19,834.29	19,091.66	20,066.70	18,184.66
Trucks	10	11	11	11
Visits	408	408	395	395
Miles/Visit	48.61	46.79	50.80	46.04
Delivery Trips	55	60	58	56
Routes Over Limit	28	29	20	25
% Over	50.9%	48.33%	34.5%	45%
No. of Team Routes	22	2	38	4

The percentage differences observed in these comparisons may change based on the parameter values used and specific problems tested (recall that the proposed model can be implemented with any set of user-provided data). Nevertheless, the above results support our main message that the SRP problem is a multi-dimensional problem with labor costs (driving as well as delivery) and equipment (truck) costs, and therefore, using traditional one-dimensional objectives may lead to inferior solutions. By considering multiple cost elements concurrently, our model is able to quantify trade-offs and ultimately identify superior solutions, which is important not only for reducing costs but also for providing managerial insights. Furthermore, some of the trade-offs may not be intuitive or obvious at first glance, such as achieving a lower total cost with longer routes or more delivery trips, which makes the proposed model an interesting and effective heuristic. The breakdown shown for the MIP solutions and the industry baseline in Table 2-4 highlights some of the key trade-offs which provides valuable managerial insights.

The MIP solution for the first instance shows a slight increase in the fleet size and the number of overage routes. However, the overall cost is reduced since 90% fewer team routes are deployed. Similarly, for the second instance, the heuristic solution reduces the number of team routes by 95%, which, combined with the reduced total miles traveled, yields a 13% reduction in total cost. Although our industry collaborator relies heavily on team routes, our solutions suggest that extending single routes by incurring overage costs is, in general, better than extending them through team driving (one would also need to consider some of the non-tangible trade-offs between single and team driving before making such changes).

Further insights can be gained by examining the truck capacity utilization and the route lengths (in time). When we compare the truck contents on each route with the route times, it becomes clear that the time limit is the tighter constraint. The graphs in Figure 2-4 (Figure 2-5) show the truck content in pounds for each truck used in the solution and the route time for each trip in the first instance under the MIP solution (industry baseline). The guidelines in the graphs show the upper limits for the truck capacity and the route times. The guideline for the truck capacity on team routes is not shown due to the scale chosen to best depict the data.

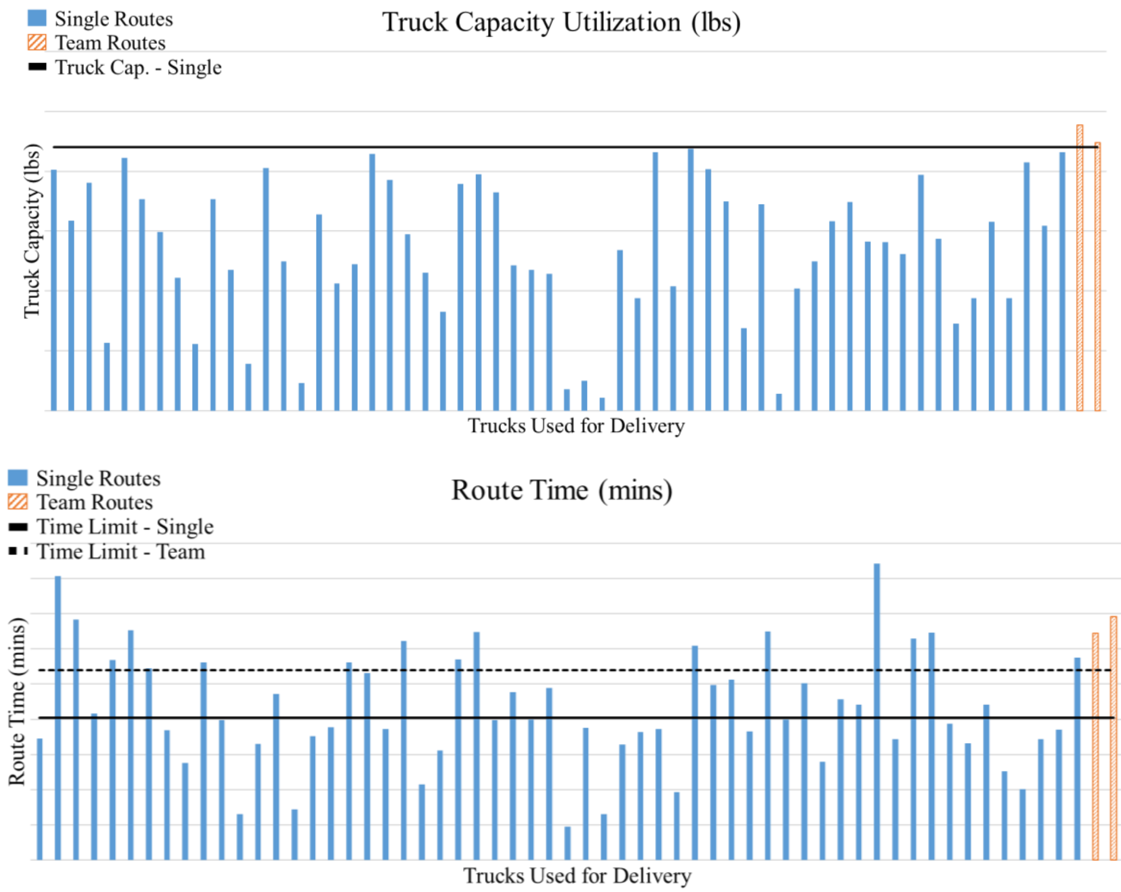


Figure 2-4. Truck Contents and Route Length for the MIP solution for Instance 1.

The graphs for the truck capacity utilization show that, for instance 1, only 5 routes (10%) in the baseline use more than 90% of the truck capacity, while 15% of the routes in the heuristic solution are over the 90% threshold. Instance 2 shows similar results with 5% and 18% of the routes over the 90% threshold for the baseline and MIP solution, respectively (graphs for instance 2 are shown in the Appendix). However, in both the baseline and the MIP solution for the two instances, the *average* truck capacity utilization ranges between 61%-65%, with the capacity utilization of individual trucks ranging from 5% to 99%. Such a wide range reflects the challenging nature of the SRP. The routes with low truck utilization are single-store routes, which are also known as *out-and-back* deliveries. For the baseline, with the exception of one, the routes with the lowest truck capacity utilization have a route time close to the upper limit, which indicates that the store visited was far from the DC. In contrast, in the SRP heuristic solution, single-store routes do not exceed 50% of the upper time limit. This suggests that, although these stores are at a reasonable distance from the DC, the SRP heuristic replenishes them individually, while more distant stores are part of multi-stop routes. Our results imply that further research on the merits of mixing out-and-back deliveries with multi-stop routes would be well-justified.

For route times, the baseline (MIP solution) shows that about 51% (48%) of the routes are over the one-day limit for the first instance, and 35% (45%) for the second instance. These results indicate that, as we stated earlier, there is a tendency to accept overage costs instead of increasing the number of routes or changing other elements in the solution. This result can be influenced by the overage cost; we study how it interacts with the solution structure in the sensitivity analysis discussed in Section 6.

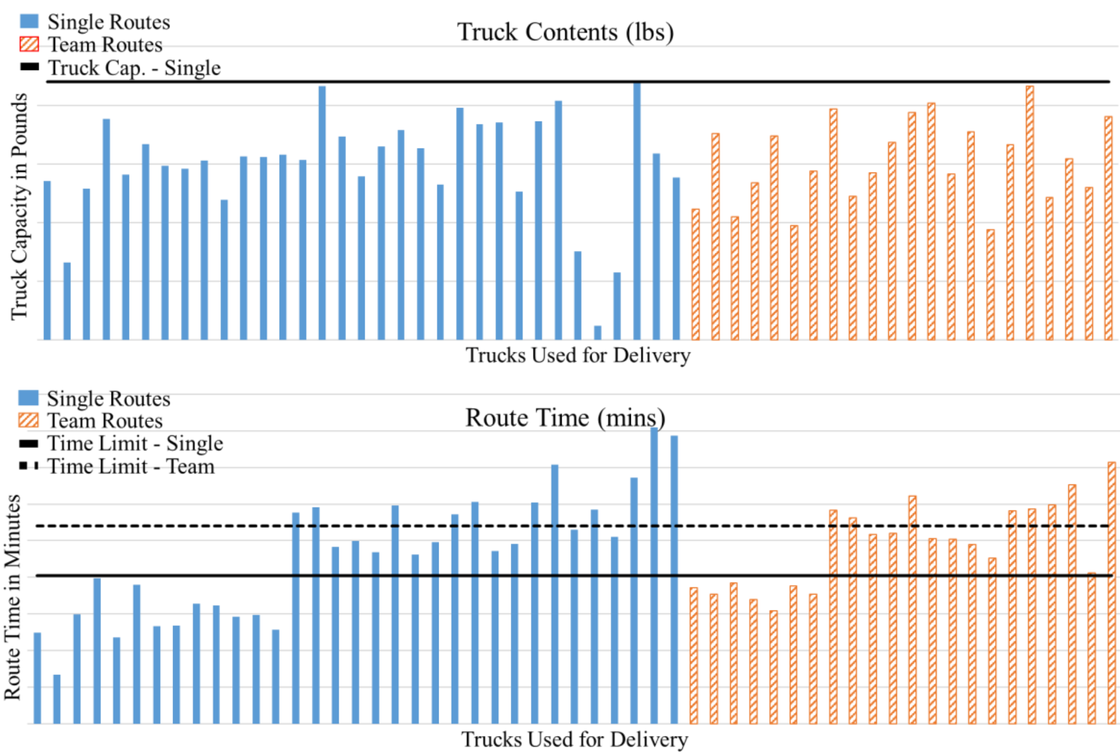


Figure 2-5. Truck Contents and Route Length for the Baseline for Instance 1.

5.4 Split Deliveries

The above results disallow split deliveries (SDs), which is consistent with the operations of our industry collaborator. SDs are often disallowed in the literature as well, although the SDVRP has received more attention in recent years; see, for example, Archetti and Speranza (2008) for a review of SDVRP and its comparison with the traditional VRP. Even though SDs are not used by our collaborator, there are no systemic rules against SDs, and exploring its impact seems justified due to the low average truck capacity utilization (61%-65%). With SDs, a higher truck utilization can be achieved, which in turn may reduce the number of routes and total distance for a lower overall cost.

Removing constraint (4) in Section 3, and using the same two problem instances, we obtain the results shown in Figures 2-6 and 2-7, which indicate that, compared to the baseline, SDs reduce the total cost by 7% and 15% for the first and second instance, respectively. Although the optimality gap is very small (average of 0.55%), we note that the SD results are based on incumbent solutions obtained from the MIP model after it reached an upper limit of 3 hours without being able to confirm their optimality. Table 2-5 summarizes the performance of the SRP heuristic with and without SDs.

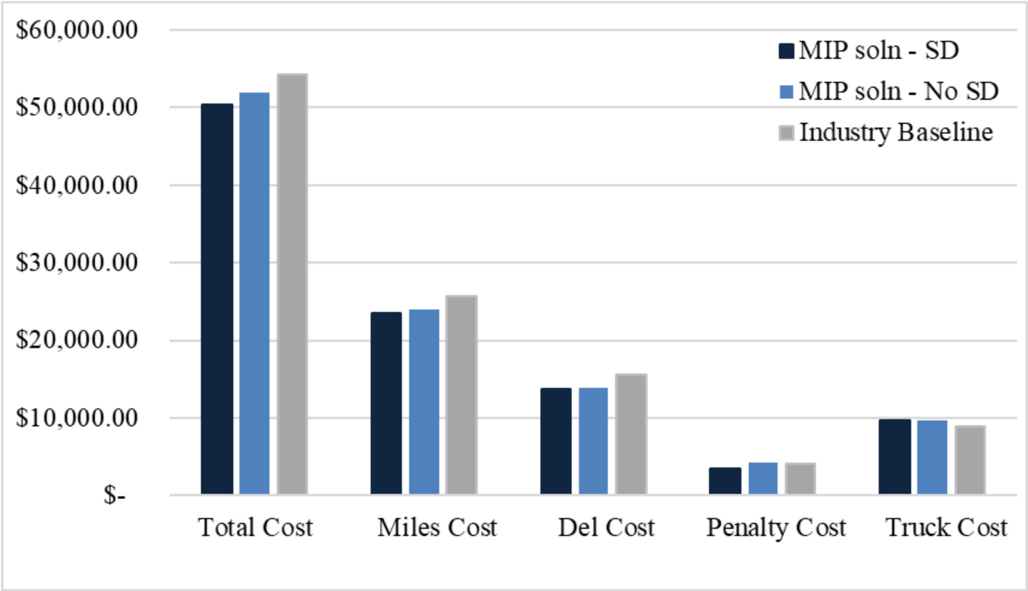


Figure 2-6. Cost Impact of Split Deliveries, Instance 1

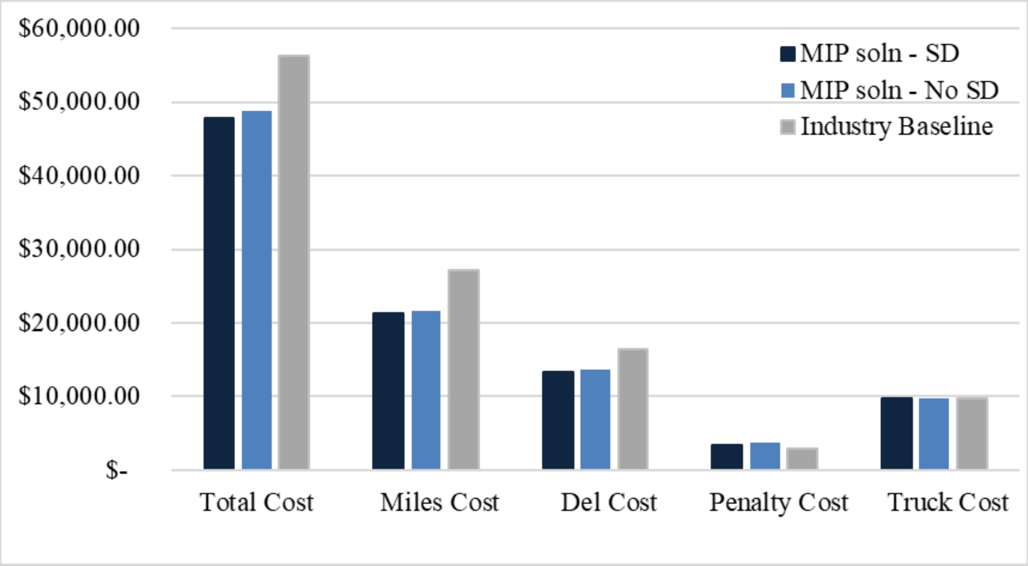


Figure 2-7. Cost Impact of Split Deliveries, Instance 2

Table 2-5. Performance of the SRP Heuristic, Split Deliveries.

	MIP Soln No SD Instance 1	MIP Soln SD Instance 1	MIP Soln No SD Instance 2	MIP Soln SD Instance 2
Total Cost Reduction	4.3%	7.1%	13.4%	15.2%
Optimality Gap	0%	0.59%	0%	0.50%
Total Runtime (min)	16.53	204.84	10.65	193.90
MIP Runtime (min)	0.18	180.73	0.15	181.12

For both instances, allowing SDs results in a further reduction of about 3% in total cost. This can motivate further study on the impact of incorporating SDs into store replenishment in the fast-food sector. Factors such as the model runtime and the additional disruption to store operations caused by SDs should also be considered.

6. Sensitivity Analyses

In this section we perform sensitivity analyses to better understand how each cost component impacts the solution. The analysis was performed by re-solving the first industry data instance with the SRP heuristic (with no SDs) while changing individual cost parameter values from zero to three times their original value, and keeping the other parameters fixed. As expected, when a cost parameter is changed, the total cost changes; however, we show the *magnitude* of the change relative to the change in the parameter value as well as possible changes in the *structure* of the solution. While one may also vary the cost parameters two or more at a time, our approach is needed as a starting point before trying more complicated changes.

Figure 2-8 includes four graphs that show how each cost component and the overall cost change as a particular cost parameter was increased. Since the distance cost consistently represents the largest portion of the total cost, it is not surprising that increasing the cost/mile traveled in both single and team routes has the biggest impact in the total cost. While the lines for the other cost components do not change in the graph, changes in the unit distance cost affects the total miles traveled every time it is increased. More specifically, when the unit costs are increased from 0 to $3x$, the total miles traveled decreases by 14%. This change also led to more than doubling the number of overage routes, and a reduction from 21% to 9% in the proportion of team routes used.

On the other hand, the overage cost appears to have the most impact in the *structure* of the solution, while causing the least increase in total cost. As the overage cost increases, the other components of the model adjust to keep the total cost from increasing significantly. As the overage cost increases from 0 to $3x$, the number of overage routes decreases by almost 60%, which leads to a small increase (3%) in the total miles traveled, and a significant change in single versus team routes, with the proportion of team routes going from 0% to 32%. This shows that as overage cost increases, team routes become advantageous since they allow larger delivery quantities and longer route times. Changes to the overage cost is the only scenario where we observed a clear and significant change in the solution structure and the percent contribution of the individual cost components. Although the change in the solution structure is significant, remarkably, the change in the total cost is relatively small as shown in the graph.

Changes in the unit delivery cost caused an anticipated reduction in the total number of team routes used. However, the biggest reduction happens when the unit costs are increased from 0 to $1x$, with the proportion of team routes decreasing by 80%. Further increases in this cost only lead to slight changes in the solution structure. The line graph shows how the total cost increases with the unit delivery costs, while the other cost components show almost no change. Similarly, examining the impact of changes in the truck cost, we can conclude that, within the range of values we examined, the solution structure is not sensitive to changes in the truck cost. As this cost parameter increased, the total number of trucks used did not change, and as the graph shows, it did not cause any significant changes in other cost components.

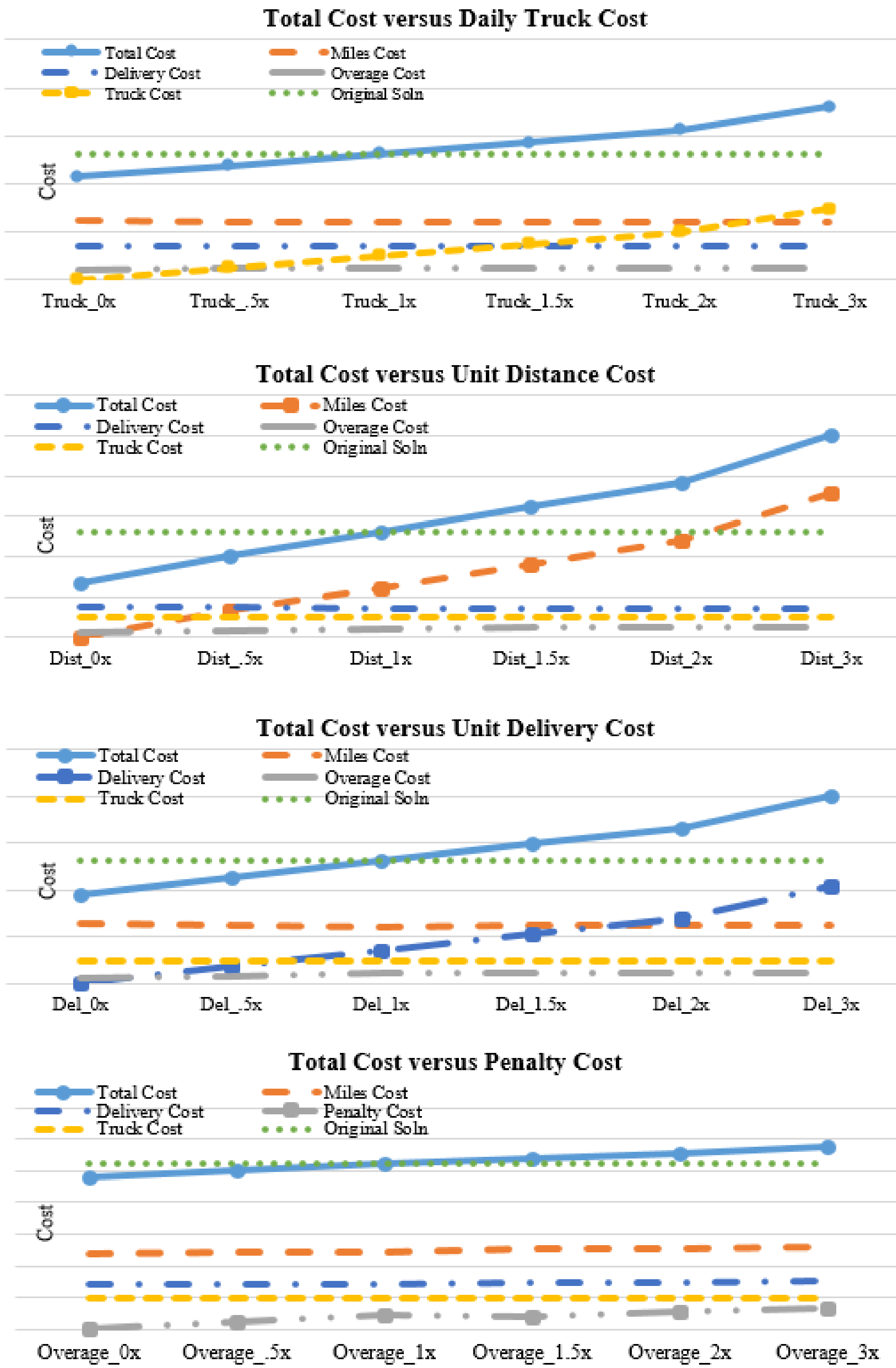


Figure 2-8. Sensitivity Analysis Graphs

7. Conclusions

We introduced a MIP formulation for the SRP and developed a heuristic to minimize the overall logistics costs associated with replenishing store in a specified region. The proposed approach derives its strength from its simplicity and effectiveness. It is simple in that we are able to obtain results in a reasonable time on a modest computer, using a commercial solver like CPLEX. It is effective in that using real-world data, our model identifies better solutions with a lower total cost than the baseline cost established from our industry collaborator's operations.

The two central contributions of this work are the cost-driven multi-dimensional objective function we developed, and the key managerial insights we obtain from the solutions. The proposed SRP heuristic uses a cost model (developed jointly with our industry collaborator) which captures trade-offs that are inherent to the SRP, and it allows us to capture distance-based and labor-based costs as well as equipment and time overage costs. Our numeric results based on industry data show that using traditional one-dimensional objectives, such as minimizing only the total distance or the fleet size, may lead to inferior solutions. Furthermore, our cost model considers single-driver and team routes. To our knowledge, this study may be the only one to consider team routes, overage cost, and fleet-sizing decisions simultaneously. Additionally, the general structure of the cost model, and the user-defined parameter values, allows for the proposed SRP heuristic to be generalized and applied to other distribution logistics problems with multiple nodes.

The results from this study also provide key managerial insights for decision-makers in the fast-food industry. Our results show that total cost can be reduced with extended single-driver routes as opposed to using a larger number of team routes. Through a series of sensitivity analyses, we established that among the cost components, the distance cost and the overage cost for extended routes show the highest impact on the overall cost and the solution structure, respectively. This can motivate further study into how the overage cost is defined and what other factors may influence extending the routes beyond their current one-day limit.

Although the results from the SRP heuristic showed a reduction in total cost, there are potential areas of study that can further improve the solutions, as shown for example with the SD results. Other future directions to study include a non-homogeneous fleet of trucks (which would allow for lighter or shorter routes, while simplifying urban-area deliveries) and incorporating multiple groups of SKUs.

References

Akkerman, R., Farahani, P. and Grunow, M. (2010) *Quality, safety and sustainability in food distribution: A review of quantitative operations management approaches and challenges*, *OR Spectrum*. doi: 10.1007/s00291-010-0223-2.

Alegre, J., Laguna, M. and Pacheco, J. (2007) 'Optimizing the periodic pick-up of raw materials for a manufacturer of auto parts', *European Journal of Operational Research*, 179(3), pp. 736–746. doi: 10.1016/j.ejor.2005.03.063.

Archetti, C., Jabali, O. and Speranza, M.G. (2015) 'Multi-period Vehicle Routing Problem with Due dates', *Computers and Operations Research*. Elsevier, 61, pp. 122–134. doi: 10.1016/j.cor.2015.03.014.

Archetti, C. and Speranza, M.G. (2008) 'The Split Delivery Vehicle Routing Problem: A Survey', in Golden, B., Raghavan, S., and Wasil, E. (eds) *The Vehicle Routing Problem: Latest Advances and New Challenges*. Boston, MA: Springer, pp. 103–122. doi: https://doi-org.proxy.lib.umich.edu/10.1007/978-0-387-77778-8_5.

Beltrami, E.J. and Bodin, L.D. (1974) 'Networks and vehicle routing for municipal waste collection', *Networks*, 4(1), pp. 65–94. doi: 10.1002/net.3230040106.

Braekers, K., Ramaekers, K. and Van Nieuwenhuyse, I. (2016) 'The vehicle routing problem: State of the art classification and review', *Computers and Industrial Engineering*. Elsevier Ltd, 99, pp. 300–313. doi: 10.1016/j.cie.2015.12.007.

Chao, I.-M, Golden, B.L. and Wasil, E. (1995) 'An improved heuristic for the period vehicle routing problem', *Networks*, 26(1), pp. 25–44. doi: 10.1002/net.3230260104.

Christofides, N. and Beasley, J.E. (1984) 'The period routing problem', *Networks*, 14(2), pp. 237–256. doi: 10.1002/net.3230140205.

Cordeau, J.F., Gendreau, M. and Laporte, G. (1997) ‘A tabu search heuristic for periodic and multi-depot vehicle routing problems’, *Networks*, 30, pp. 105–119. doi: 10.1002/(SICI)1097-0037(199709)30:2%3C105::AID-NET5%3E3.0.CO;2-G.

FMCSA (2017) *Summary of Hours of Service Regulations, Federal Moto Carrier Safety Administration*. Available at: https://www.fmcsa.dot.gov/regulations/hours-service/summary-hours-service-regulations.

Francis, P., Smilowitz, K. and Tzur, M. (2006) ‘The Period Vehicle Routing Problem with Service Choice’, *Transportation Science*, 40(4), pp. 439–454. doi: 10.1287/trsc.1050.0140.

Francis, P., Smilowitz, K. and Tzur, M. (2008) ‘The Period Vehicle Routing Problem and its Extensions’, in Golden, B., Raghavan, S., and Wasil, E. (eds) *The vehicle routing problem: latest advances and new challenges*. New York, pp. 73–102.

Gaudioso, M. and Paletta, G. (1992) ‘A Heuristic for the Periodic Vehicle Routing Problem’, *Transportation Science*, 26(2), pp. 86–92. doi: 10.1287/trsc.26.2.86.

GMA (2008) *Powering Growth Through Direct Store Delivery*. Available at: https://www.yumpu.com/en/document/read/6749704/direct-store-delivery-grocery-manufacturers-association.

Goel, A., Vidal, T. and Kok, A.L. (2019) *To team up or not – Single versus team driving in European road freight transport*. Rio de Janeiro, Brasil. Available at: https://w1.cirrelt.ca/~vidalt/papers/Team-vs-Single.pdf.

Golden, B. *et al.* (1984) ‘The fleet size and mix vehicle routing problem with time windows’, *Computers & Operations Research*, 11(1), pp. 49–66. Available at: https://doi-org.proxy.lib.umich.edu/10.1016/0305-0548(84)90007-8.

Google (2019) *Google OR-Tools, Google*. Available at: https://developers.google.com/optimization.

Google (2020) *Google Maps Platform, Google*. Available at: https://cloud.google.com/maps-platform/.

Gulczynski, D., Golden, B. and Wasil, E. (2011) ‘The period vehicle routing problem: New heuristics and real-world variants’, *Transportation Research Part E: Logistics and Transportation Review*. Elsevier Ltd, 47(5), pp. 648–668. doi: 10.1016/j.tre.2011.02.002.

Hasle, Geir and Kloster, O. (2007) ‘Industrial Vehicle Routing’, in Hasle, G., Lie, K., and Quak, E. (eds) *Geometric Modelling, Numerical Simulation, and Optimization*. Berlin, Heidelberg: Springer Berlin Heidelberg, pp. 397–435. doi: https://doi-org.proxy.lib.umich.edu/10.1007/978-3-540-68783-2_12.

Hemmelmayr, V. C., Doerner, K. F. and Hartl, R. F. (2009) ‘A variable neighborhood search heuristic for periodic routing problems’, *European Journal of Operational Research*. Elsevier B.V., 195(3), pp. 791–802. doi: 10.1016/j.ejor.2007.08.048.

Klein, D. (2018) *Red-Hot Domino's and the Road to 30,000 Restaurants*, *QSR*. Available at: https://www.qsrmagazine.com/pizza/red-hot-dominos-and-road-30000-restaurants.

Koç, Ç. *et al.* (2016) 'Thirty years of heterogeneous vehicle routing', *European Journal of Operational Research*, 249(1), pp. 1–21. doi: 10.1016/j.ejor.2015.07.020.

Laporte, G. (2009) 'Fifty years of vehicle routing', *Transportation Science*, 43(4), pp. 408–416. doi: 10.1287/trsc.1090.0301.

Larrain, H. *et al.* (2019) 'Exact solution methods for the multi-period vehicle routing problem with due dates', *Computers and Operations Research*. Elsevier Ltd, 110, pp. 148–158. doi: 10.1016/j.cor.2019.05.026.

Lemma, Y., Kitaw, D. and Gatew, G. (2014) 'Loss in Perishable Food Supply Chain: An Optimization Approach Literature Review', *International Journal of Scientific & Engineering Research*, 5(5).

Lenstra, J. K. and Kan, A.H.G.R. (1981) 'Complexity of vehicle routing and scheduling problems', *Networks*, 11(2), pp. 221–227. doi: 10.1002/net.3230110211.

López-Santana, E., Rodríguez-Vásquez, W.C. and Méndez-Giraldo, G. (2018) 'A hybrid expert system, clustering and ant colony optimization approach for scheduling and routing problem in courier services', *International Journal of Industrial Engineering Computations,* 9(3), pp. 369–396. doi: 10.5267/j.ijiec.2017.8.001.

MacMillan, C. (2018) *What You Need to Know About Team Truck Driving Jobs*, *Smart Trucking*. Available at: https://www.smart-trucking.com/team-truck-driving/ (Accessed: 25 April 2020).

Sarkar, C., Paul, H.S. and Pal, A. (2018) 'A Scalable Multi-Robot Task Allocation Algorithm', in *Proceedings - IEEE International Conference on Robotics and Automation*. IEEE, pp. 5022–5027. doi: 10.1109/ICRA.2018.8460886.

Sharda, R. *et al.* (2008) 'The Vehicle Routing Problem: Latest Advances and New Challenges'. Boston, MA: Springer Science+Business Media, LLC (Operations Research/Computer Science Interfaces,1387-666X ;43). Available at: http://dx.doi.org/10.1007/978-0-387-77778-8.

Shokri, A., Oglethorpe, D. and Nabhani, F. (2014) 'Evaluating sustainability in the UK fast food supply chain: Review of dimensions, awareness and practice', *Journal of Manufacturing Technology Management*, 25(8), pp. 1224–1244. doi: 10.1108/JMTM-04-2013-0031.

Staff, E. (2019) 'Nestle Joins Kellogg, Hostess and Others before It to Exit DSD to Reduce Logistics Costs', *Supply Chain Digest*, June. Available at: http://www.scdigest.com/ontarget/19-06-04-1.php?cid=15536.

Tarantilis, C.D. and Kiranoudis, C.T. (2001) 'A meta-heuristic algorithm for the efficient distribution of perishable foods', *Journal of Food Engineering*, 50(1), pp. 1–9.

Tarantilis, C.D. and Kiranoudis, C.T. (2002) 'Distribution of fresh meat', *Journal of Food*

Engineering, 51(1), pp. 85–91.

Toth, P. and Vigo, D. (eds) (2014) *Vehicle Routing: Problems, Methods, and Applications*. Philadelphia: Society for Industrial and Applied Mathematics : Mathematical Optimization Society (MOS-SIAM series on optimization). Available at: http://epubs.siam.org/doi/book/10.1137/1.9781611973594.

USDA (2018) *Expenditure for Food at Home and Away from Home, USDA*. Available at: https://www.ers.usda.gov/topics/food-markets-prices/food-service-industry/market-segments/.

Vidal, T., Laporte, G. and Matl, P. (2019) ‘A concise guide to existing and emerging vehicle routing problem variants’, *European Journal of Operational Research*. doi: 10.1016/j.ejor.2019.10.010.

Wen, M. *et al.* (2010) ‘The dynamic multi-period vehicle routing problem’, *Computers and Operations Research*, 37(9), pp. 1615–1623. doi: 10.1016/j.cor.2009.12.002.

Yum! Brands (no date) *Yum.com*. Available at: https://www.yum.com/wps/portal/yumbrands/Yumbrands (Accessed: 24 April 2020).

Chapter 3

The Store Replenishment Problem with Flexible Delivery Schedule and Limited Store Capacities[2]

1. Introduction

A critical success factor for a retail company is product availability. A sector where this is particularly true is the fast-food sector. Despite the challenges posed by the Covid-19 pandemic, many fast-food companies were able to successfully continue their operations and provide their customers with the products and services they expect from a fast-food outlet.

At the center of the operational success of many fast-food companies, one often finds a well-organized and fairly sophisticated supply chain group managing the distribution and replenishment logistics for their network of stores. The highly competitive nature of the fast-food sector has been a key motivator for the above groups to find ways to reduce cost and improve replenishment logistics, which led to the Store Replenishment Problem (SRP) described by Vigo Camargo & Bozer (2022). The objective of the SRP is to determine the least overall cost associated with satisfying the specified store demands in a fast-food network with a central distribution center (DC), considering the delivery routes, resources, and labor necessary to meet the above goal. The SRP assumes that the orders placed by the stores constitute a fixed delivery schedule; i.e., if a store has placed, say, two orders during the planning horizon, one for period t_1

[2] This Chapter was submitted for publication and is under review as: Vigo Camargo, A. and Bozer, Y. A. (2022) 'The Store Replenishment Problem with Flexible Delivery Schedule and Limited Store Capacities.'

and one for period t_2, then the two orders are scheduled for delivery on those two periods, in the quantities that were ordered, with no late or early deliveries allowed. However, since each vehicle visit to a store contributes to the cost, allowing flexibility in the form of early deliveries presents a strong opportunity to further reduce the overall replenishment cost by utilizing the delivery trips more efficiently and potentially reducing the number of store visits required.

As discussed by Jaigirdar *et al*., (2022), decision-making for a food distribution network is a challenging endeavor, which can be further complicated by the significant costs associated with distribution logistics. According to Akkerman *et al*., (2010), store replenishment is a primary source of cost in food distribution. Since small percentage decreases in cost can lead to substantial savings in routing problems (Hasle & Kloster, 2007) the potential to identify cost reductions through a flexible delivery schedule is what motivated our study.

To understand the impact of allowing early deliveries of food inventory, we study the SRP with a flexible schedule (i.e., SRP-FS), which is an extension and generalization of the SRP. The SRP-FS still takes the store orders as input but builds a delivery schedule, over a given multi-period planning horizon, that allows early deliveries while ensuring that store orders are met on time (i.e., no late deliveries). Allowing early deliveries potentially reduces the overall cost but it also requires that we consider the storage capacity of each store. Once inventory is delivered at a store, it is owned by the franchise location rather than the company (Vigo Camargo & Bozer, 2022). Therefore, we do not account for the inventory holding cost in the SRP-FS model, but we still include the inventory balance equations since we need to keep track of inventory at each store to ensure that a store’s storage capacity is not exceeded.

Furthermore, to keep track of the inventory level at each store and estimate the remaining capacity, we also have to consider how the store inventory is consumed over time, which is a function of the end-customer demand. Obtaining up-to-date data on inventory consumption at each store is possible with a modern point-of-sale system but using such data to construct the delivery schedule, one period at a time, would amount to dynamic delivery routing and inventory allocation at the DC every night, which is beyond the scope of our study and also beyond the goals of our industry collaborator. Instead, we consider alternative ways to model inventory consumption at each store within the scope of our study and the data available to us. Although the final results obtained from the model depend on the inventory consumption pattern, we stress that the proposed model and the solution procedure can be used with alternative inventory consumption patterns if needed.

To tackle the SRP-FS, we developed a two-step metaheuristic based on Simulated Annealing to first build a candidate delivery schedule, and then use the Adjusted Team Savings Algorithm, as a second step, to build and evaluate the delivery routes for the candidate delivery schedule. To develop insights, we first studied a simplified version of the model, assuming that only direct (out-and-back) deliveries are allowed at each store. For the direct delivery approach, we present a heuristic policy that yields a near-optimal solution. The heuristic policy is then extended to the multi-stop version of the model and it is used to build the initial feasible solution for the proposed metaheuristic. Data obtained from our industry collaborator, i.e., the supply chain group of a leading and well-known national fast-food company, were used to test and evaluate the performance of the proposed heuristic.

Various problems studied in the literature share similarities with the SRP-FS. Problems such as the Petrol Station Replenishment Problem (Brown & Graves, 1981), and the more general Periodic Vehicle Routing Problem (Francis *et al.*, 2008), have similar structures and objectives. However, there are also key differences between the above problems and the SRP-FS, such as the objective function formulation and the way the delivery schedule is treated (these differences are discussed further in Section 2). The objective function of the SRP-FS incorporates multiple cost elements based on the distance cost, the truck cost, the labor cost, and the additional cost associated with route-time extensions. Furthermore, the SRP-FS considers both single-driver and two-driver routes (team routes), which enhances the model since team routes offer extended route-time limits and they are common in the industry (Goel *et al.*, 2019). While each type of route can be completed using the same trucks, depending on whether it is a single- or team-route, key parameters such as the maximum usable truck capacity, the route time limit, and the other costs associated with the route (such as the distance and labor cost) change accordingly.

To summarize, the main contribution of this paper is the introduction of an extension to the SRP where we study the potential reduction in the overall cost of replenishing stores in a network through a flexible delivery schedule. We develop a two-step simulated annealing metaheuristic that iteratively builds a delivery schedule and its corresponding routes to find a good solution to the problem. To the best of our knowledge, our study is the first one that incorporates multiple cost components into the objective function of the SRP, considers both team and single-driver routes, and allows a flexible delivery schedule simultaneously. Additionally, we present a near-optimal heuristic for the direct shipment version of the SRP-FS.

In the following section we present past studies in the literature that are relevant to the SRP-FS. We then introduce a formal mathematical formulation of the SRP-FS in Section 3. In Section 4 we discuss the direct delivery case of the problem and present a solution method. Section 5 focuses on the heuristic approach developed to solve the SRP-FS, and the computational results are presented in Section 6. Last, in Section 7 we present our conclusions and potential directions for future research.

2. Literature Review

Determining the truck routes is one of the decisions made in the SRP-FS, making the problem part of a group of problems that has been studied extensively in the literature. Laporte (2009), Toth & Vigo (2014), and Mor & Speranza (2022), among others, describe progress made in vehicle routing problems (VRPs) over many years, including some of its well-known extensions. When considering the SRP-FS, we can compare it to similar problems that have been examined in the past. One such problem of particular interest is the Flexible Periodic VRP (or FPVRP), which is an extension of the traditional Periodic VRP (PVRP), first discussed by Archetti *et al.* (2017). At its core, the PVRP and its extensions involve decisions analogous those made by the SRP-FS, as it also selects a delivery schedule along with its corresponding routing (Francis *et al*., 2008). However, the PVRP assumes that costumers follow a particular demand frequency pattern over the time horizon and makes the delivery schedule selection based on a predetermine set of potential visit frequencies. This is a common feature of PVRP extensions, for example the PVRP with service choice, studied by Francis *et al*. (2006), where the service choice involves letting the model choose how often each customer is to be visited, is still limited to pick from predetermined frequency patterns over the planning horizon. The FPVRP differs from other PVRP extensions in that its flexibility is based on deciding when to visit each customer and how much inventory to deliver on each visit. Several studies in the literature focused on the FPVRP since its introduction. Archetti *et al.* (2018) build on the original MIP formulation of the problem and present a two-phase metaheuristic to solve the FPVRP. More recently, Huerta-muñoz *et al.* (2022) present a MILP formulation for the FPVRP with heterogeneous vehicles, and they develop a kernel search-based metaheuristic to solve it.

In general, FPVRP studies assign a total demand to each customer that must be satisfied over the planning horizon. Based on this total demand, the FPVRP is concerned with determining how much inventory to deliver on each visit. In contrast, in the SRP-FS, store orders (in specific quantitates) must be delivered by specific time periods within the planning horizon, and if early deliveries are made, the stores carry inventory forward as long as store demand is satisfied by the specified time period. This feature makes the SRP-FS similar to another extension of the PVRP, known as the PVRP with Time Windows (PVRPTW). According to Cordeau *et al.* (2001), in the PVRPTW, each customer must be visited a specified number of times within a pre-determined time interval. The authors present a Tabu search heuristic to solve the problem. However, their work only considers a particular subset of potential schedules on which the customers can be visited throughout the planning horizon. Rothenbächer (2019) presents an exact algorithm to solve the PVRPTW based on a Branch-and-Price-and-Cut scheme, where the potential schedules to visit the customers are flexible in that they are allowed to overlap and consider different frequencies. However, this is still not equivalent to the flexibility inherent in the SRP-FS since the SRP-FS is not constrained by a predetermined set of potential visit schedules; the only restriction in scheduling the deliveries is to ensure that the store orders are met by specific periods, without exceeding the store capacity.

The Petrol Station Replenishment Problem (PSRP) also deals with the replenishment and routing logistics associated with delivering inventory to a number of customers in a given region. Originally formulated by Brown & Graves (1981), the PSRP was motivated by the distribution operations of petroleum companies in North America. A key difference between the PSRP and the SRP-FS is that the former considers unique delivery vehicles that contain compartments that can hold different petroleum products, and each compartment must be assigned to a specific customer on each route. Since its introduction, the PSRP was studied by Cornillier *et al.* (2008a), who present an exact algorithm to solve the single-day PSRP, and then by Cornillier *et al.*, (2008b), who present a heuristic to solve the multi-period PSRP. The multi-period PSRP was also studied by Boers *et al.* (2020) who formulate the problem as an MILP and then present a decomposition heuristic to solve it. Al-Hinai & Triki (2020) recently presented a two-level evolutionary algorithm to solve the Periodic PSRP, where periodicity is introduced in the delivery schedule as well as the service choice of the customers. The above extensions of the PSRP share structural elements with the traditional PVRP in that potential delivery schedules are predefined. Again, this differentiates the SRP-FS from the other problems in that it is not constrained by predefined delivery schedules. Furthermore, in contrast to the PSRP, the SRP-FS is not limited by partitions in the delivery trucks that have to be assigned to individual stores and have more flexibility in the amount to deliver at each store.

Another well-known and related routing problem is the Inventory Routing Problem (IRP), which is primarily concerned with building routes to minimize the transportation and inventory holding costs while determining the frequency and quantity for each delivery. The IRP has been studied extensively in the literature since its introduction by Bell *et al*. (1983). Surveys on studies concerned with the IRP can be found in Andersson *et al*. (2010), Moin & Salhi (2007), and Coelho *et al*. (2014). While the IRP shares a similar structure with the SRP-FS in that the delivery quantities, the schedule, and the routes are decision variables, they differ in the factors that impact these decisions. The IRP is traditionally focused on considering the inventory holding costs at the stores, which the SRP-FS does not consider due to the nature of the application. Furthermore, while both problems consider the inventory consumption rate of each customer, the IRP uses this rate as the demand for each customer in order to determine the delivery schedule. In contrast, the SRP-FS has to ensure that the specific order quantities are delivered by specific periods, while still considering how the stores consume inventory and account for the limited store capacity.

Unlike the other problems discussed in this section, the SRP-FS incorporate multiple cost components associated with travel, labor, fleet size, and route-time in its objective function, while considering the differences between team and single-driver routes and allowing a flexible delivery schedule. The above flexibility in the schedule is not restricted to predefined frequencies, as is the case in the PVRPTW and PSRP. Additionally, the SRP-FS ensures that stores receive what they order by the specified period, as opposed to determining the delivery quantities based on the consumption rate for each period, as in the IRP, or on a total demand, as is defined by the FVRP. These differences distinguish the SPR-FS from previous problems studied in the literature.

3. Formulating the SRP-FS

3.1 Modelling Assumptions

The SRP-FS is modelled using a finite planning horizon with discrete time periods. For our particular application, and for our industry collaborator, a time period represents a day. The SRP-FS is concerned with building a delivery plan, which includes the delivery schedule (that is, when each store is visited), the delivery quantities, and the routes for the trucks, based on specified store orders that are to be replenished through the DC. To formulate the SRP-FS as an MIP, the following assumptions are made:

(1) Early deliveries are allowed as long as each store receives the inventory quantity they

ordered by the specified periods; no late deliveries are allowed.

(2) Inventory is consumed at each store at a known and predetermined rate. Alternative inventory consumption rates can be considered by the proposed model.

(3) Once it is delivered, each store (franchisee) owns the inventory; therefore, the inventory carrying cost at the stores is not part of the objective function.

(4) Stores have finite capacity for storing inventory. None of the store orders may exceed the store's capacity.

(5) A route can be completed by a single driver or a team of two drivers (team routes).

(6) The usable capacity of a truck and the route-time limit depends on the type of route (single-driver or team).

(7) A truck can only complete one route per period.

(8) All the trucks and drivers are available at the start of each period.

(9) In a given period, a store may receive at most one delivery. No split deliveries are allowed.

(10) Store orders are measured in pounds and are represented as a single aggregate SKU.

The first assumption reflects the introduction of a flexible schedule. Provided the quantities ordered by each store have been delivered by the specified periods, a flexible schedule allows early deliveries in order to explore possible cost savings. The second assumption is also due to a flexible schedule. Since deliveries can be made earlier, the model needs to account for how inventory is consumed at each store in order to keep track of available space at each store for subsequent deliveries. As stated earlier, although real-time inventory usage data may be available, using such data and changing the model to a nightly dynamic decision-making model is beyond the scope of our study. Instead, we use pre-defined inventory consumption rates for each store based on the inventory quantities ordered by the store.

The remaining assumptions (3-10) are consistent with the SRP (Vigo Camargo & Bozer, 2022). The third assumption follows the business model commonly used in the fast-food sector, where inventory at the stores is owned by the franchisee rather than the company. Team routes (assumptions 5 and 6) is a common practice in the trucking industry; it allows larger delivery quantities (hence, larger usable truck capacity) and longer routes, while respecting federal regulations limiting hours-of-service for the drivers. Assumptions 7 through 10 are introduced mainly to simplify the model while still capturing general practices found in fast-food supply operations.

3.2 MIP Formulation

The mathematical formulation of the SRP-FS is an extension of the formulation presented in Vigo Camargo & Bozer (2022). For continuity, similar notation is used as defined below:

N = set of locations in the region ($i = 0, \ldots, |N|$), where 0 denotes the DC
T = planning horizon ($t = 1, \ldots, |T|$)
K = set of trucks ($k = 1, \ldots, u$), where u is an upper bound for the total trucks used
L = set of indexes to identify single-driver (1) and team routes (2)
d_i^t = demand in pounds (lbs) for store i in time period t
U_V^l = upper limit on truck capacity, in pounds, for route type l
U_T^l = upper limit of route time length, in minutes, for route type l
m_{ij} = distance in miles from store i to j
V = maximum number of stops allowed in a route
B_{lb} = unloading rate, in minutes per pound unloaded
B_n = time per stop, in minutes per stop on a route
B_m = travel time rate, in minutes per mile traveled on a route
C_V = daily cost per truck required

C_m^l = cost/mile traveled in a route for type route l

C_{lb}^l = cost/pound delivered in a route for type route l

C_U = fixed overage cost for each route that exceeds the route time upper limit

To account for possible early deliveries, the model keeps track of available inventory at each store throughout the planning horizon. Also, the limited storage capacity at each store must be accounted for to avoid unrealistically large one-time deliveries; a concern voiced by our industry collaborator in light of limited storage space at each store.

To model the above concerns, the following additional parameters are introduced to extend the SRP:

I_i^0 = starting inventory in pounds at store i

U_S = store capacity, upper limit in pounds

r_i^t = inventory consumption rate at store i in period t

The objective function is to minimize the overall cost associated with replenishing the stores from a given DC, considering a flexible schedule and the space limitations imposed by both the trucks and the stores. The decision variables are defined as follows:

x_{ijk}^{tl} = 1 if truck k, completing a route type l, travels from store i to j in time period t

q_{ik}^{tl} = number of pounds delivered to store i on truck k, on a route type l, in time period t

a_k^{tl} = 1 if truck k, completing a route type l, is used on period t

I_i^t = inventory in pounds at store i at the end of time period t

z_r^t = 1 if the time length of route r goes over the upper limit in minutes on period t

H = number of trucks required over the time horizon

The mathematical formulation of the SRP-FS is presented as follows:

$$\min \sum_{t,k,l} \sum_{i,j} C_m^l \, m_{ij} \, x_{ijk}^{tl} + \sum_{t,l} \sum_{k} \sum_{i} C_{lb}^l \, q_{ik}^{tl} + \sum_{t} \sum_{r} C_U \, z_k^{tl} + C_V \, T \, H \tag{1}$$

$$s.t. \quad I_i^{t-1} + \sum_{k,l} q_{ik}^{tl} - r_i^t = I_i^t \qquad \forall\ i \in N\backslash\{0\}, t \in T \qquad (2)$$

$$\sum_{k,l} \sum_{o \le t} q_{ik}^{ol} \ge \sum_{o \le t} d_i^o \qquad \forall\ i \in N\backslash\{0\}, t \in T \qquad (3)$$

$$\sum_{k,l} q_{ik}^{tl} + I_i^{t-1} \le U_S \qquad \forall\ i \in N\backslash\{0\}, t \in T \qquad (4)$$

$$\sum_{i>0} q_{ik}^{tl} \le a_k^{tl}\, U_V^l \qquad \forall\ k \in K, t \in T, l \in L \qquad (5)$$

$$\sum_{i,k,l} x_{ijk}^{tl} \le 1 \qquad \forall\, j \in N\backslash\{0\}, t \in T \qquad (6)$$

$$\sum_{i,j} x_{ijk}^{tl} \le a_k^{tl}\, V \qquad \forall\, k \in K, t \in T, l \in L \qquad (7)$$

$$\sum_{i} x_{ijk}^{tl}\ U_S \ge q_{jk}^{tl} \qquad \forall\, j \in N\backslash\{0\}, k \in K, t \in T, l \in L \qquad (8)$$

$$\sum_{k,l} a_k^{tl} \le H \qquad \forall\ t \in T \qquad (9)$$

$$\sum_{i} \left(B_{lb}\, q_{ik}^{tl}\right) + \sum_{i>0,j} B_n\, x_{ijk}^{tl} + \sum_{i,j} B_m\, m_{ij}\, x_{ijk}^{tl} \le U_T^l\,(1 + z_k^{tl}) \qquad \forall\, k \in K, t \in T, l \in L \qquad (10)$$

$$z_k^{tl} \le 1 \qquad \forall\, k \in K, t \in T, l \in L \qquad (11)$$

$$\sum_{j} x_{0jk}^{tl} = a_k^{tl} \qquad \forall\, k \in K, t \in T, l \in L \qquad (12)$$

$$\sum_{i} x_{ijk}^{tl} - \sum_{o} x_{jok}^{tl} = 0 \qquad \forall\, k \in K, t \in T, j \in N, l \in L \qquad (13)$$

$$a_k^{tl} \in \{0,1\} \qquad \forall\, k \in K, t \in T, l \in L \qquad (14)$$

$$\sum_{(i,j)\,\in S,k} x_{ijk}^{tl} \le |S| - 1 \qquad \forall\, S \subset N, t \in T, l \in L \qquad (15)$$

$$u \ge H \ge 0 \qquad \text{integer} \qquad (16)$$

$$x_{ijk}^{tl} \in \{0,1\} \qquad \forall\, i,j \in N, \mathrm{k} \in \mathrm{K}, t \in T, l \in L \qquad (17)$$

$$q_{ik}^{tl} \ge 0 \qquad \forall\, i \in N, k \in K, t \in T, l \in L \qquad (18)$$

$$I_i^t \ge\ 0 \qquad \forall\ i \in N, t \in T \qquad (19)$$

$$z_k^{tl} \geq 0 \qquad \forall\, k \in K, t \in T, l \in L \qquad \textbf{(20)}$$

The objective function in equation (1) represents the four cost components associated with SRP-FS; namely, the mileage cost due to the distance traveled, the delivery cost based on the quantity of inventory delivered (by the pound), the overage cost of exceeding a route time limit, and the fleet cost (for providing H trucks throughout the planning horizon). Constraints (2) represent the basic inventory balance constraints, and they incorporate the rate of inventory consumption at each store. Constraints (3) are used to ensure that demand is met by the period expected for every store. Constraints (4) impose the store capacity, while constraints (5) ensure that the truck capacity is not exceeded on either single-driver or team routes. Constraints (6) disallow split deliveries, while constraints (7) impose the maximum number of stops allowed in a route. Constraints (8) ensure that the delivery and the visit variables are linked. Constraints (9) ensure that the number of trucks used in any time period does not exceed the fleet size. The route time is calculated with constraints (10), including those routes which may incur an overage cost. Constraints (11) ensure that the route overage is bounded. Constraints (12) link the number of trucks used with every departure from the DC. Constraints (13) represent the routing balance constraints, while constraints (14) ensure that each truck is used at most once per period. Set (15) represents the subtour elimination constraints, while constraint sets (16) – (20) define the variables and represent the non-negativity constraints.

The above extension, combined with the routing component and subtour elimination constraints, significantly increases the complexity of the model compared to the original SRP with a fixed schedule. To gain insight into the solution structure of the SRP-FS, we first studied the direct delivery version of the problem, also known as out-and-back deliveries. In the following section we discuss the direct delivery approach and a procedure to find a near-optimum solution.

4. SRP-FS with Direct Deliveries

With direct deliveries from the DC, where a truck makes only an out-and-back trip to a store, SRP-FS is simplified significantly. Since the distance cost per mile traveled does not change over the planning horizon, and all the store orders have to be delivered, the solution to the direct-delivery version of the SRP-FS focuses on the total number of deliveries made and the number of trucks needed, where the latter is dictated by the period with the maximum number of trucks used. Whether each route is completed by a single driver or a team, and whether or not a route exceeds its time limit, also impacts the final cost but these factors can be easily adjusted after a delivery schedule is developed.

Except for the number of trucks needed, each store can be treated independently. Therefore, we first focus on minimizing the total number of deliveries. Considering each store individually and trying to minimize the number of deliveries resembles the Capacitated Lot Sizing Problem with Inventory Bounds (CLSP-IB). The traditional Lot Sizing Problem (LSP) is typically used in a production setting where there is demand for a product to be manufactured over a known planning horizon. The objective of the LSP is to determine in which periods to manufacture the product and how much to manufacture in order to meet demand, while minimizing the fixed and variable costs (Brahimi, *et al*. (2017)). The fixed cost is based on the setup cost for production, while the variable cost is based on the inventory holding cost, and in some cases the unit production cost. The CLSP-IB is an extension of the traditional LSP in that it limits the quantity that can be manufactured in each period, and it limits how much inventory can be stored between periods.

Under the direct delivery assumption, the truck capacity of the SRP-FS is analogous to the manufacturing capacity in the CLSP-IB. Similarly, the store capacity can be expressed as the inventory bounds in the CLSP-IB. In terms of the objective function, both problems incur a fixed cost (i.e., the out-and-back distance cost in the SRP-FS and the production setup cost in the CLSP-IB) as well as a variable cost directly associated with the number of items delivered (SRP-FS) or manufactured (CLSP-IB). Since the fixed cost in each problem remains the same through the planning horizon (cost per mile traveled and the setup cost), the main decision is focused on minimizing the number of deliveries (SRP-FS) or production runs (CLSP-IB). Hence, the two problems can be solved using similar solution methods.

The LSP has been studied extensively in the literature, including the CLSP-IB extension (see Gharaei, *et al*. (2021)). In Florian & Klein (1971), the authors present an algorithm to obtain an optimal solution to the CLSP-IB with concave costs and a stationary production capacity. As an extension to this work, Akbalik *et al*. (2015) present two algorithms for the CLSP-IB for the case with concave costs and the case with Wagner-Whitin (WW) costs, which establishes that the cost of producing and holding inventory for a specific period is greater than or equal to the cost of just producing in the next period. For concave costs, the authors adapt the algorithm in Florian & Klein (1971) to incorporate inventory bounds and they formulate the problem as a Shortest Path Problem. For WW costs, they define recursive formulas for sub-plans (smaller intervals of time) within the given planning horizon and they obtain an optimal solution through dynamic programming.

The above algorithms use the concept of "inventory periods," defined as the time between a period where inventory is at full capacity and the next period when inventory is completely depleted. By assuming that the last period in the planning horizon is also the end point of an inventory period (i.e., there is no inventory left at the end of the planning horizon), the above algorithms are able to identify one period within the sub-plans to produce a fraction of the total capacity, and then assign the other k necessary periods of production at full capacity in order to meet the exact demand. However, assuming that no inventory is left at the end of the planning horizon would not be appropriate in some cases, particularly in applications like the fast-food industry where inventory is rarely depleted completely, and any remaining inventory would normally be carried over to the next planning horizon. The decision policy described below is shown to yield an optimal solution for such applications, including the direct delivery version of the SRP-FS. The policy establishes that whenever the inventory level at a store at the start of a period is not enough to meet the expected consumption for that period, a delivery should be scheduled with enough inventory to not exceed the minimum of three items: 1) the truck capacity (based on a single-driver), 2) the available capacity at the store, and 3) the sum of the store orders in subsequent time periods. The third item is included to ensure that we do not deliver more inventory than what was ordered by the store over the planning horizon.

To formalize the proposed policy and prove its optimality, let $x_i^t \in X$ be a binary variable to indicate whether or not a delivery is made in period $t \in T$ at store i, while q_i^t is defined as the quantity of inventory delivered to store i in period t. The truck capacity and the store capacity are denoted by U_V and U_S, respectively. Inventory available at the start of period t is defined by the variable I_i^t. Furthermore, d_i^t (integer) denotes the store order in period t. We assume that $d_i^t \leq U_S$ for all t, and that each store consumes inventory at a known rate of r_i^t. A feasible solution to the above problem shows a schedule of periods in which a delivery is made as well as the amount delivered on each visit, while satisfying the capacity constraints and the store orders. The following definition leads to a mathematical formulation of the problem and a solution:

Definition: A set of positive integer values, $X = \{x_i^1, x_i^2, \ldots, x_i^T\}$, is a solution to the direct delivery SRP-FS if the following constraints are satisfied:

$$x_i^t\, U_V \geq q_i^t \quad \forall\, i, t \tag{21}$$

$$\sum^{t} q_i^o \geq \sum^{\mathrm{t}} d_i^o \;\; \forall\, i, t \tag{22}$$

$$U_S \geq q_i^t + I_i^t \geq r_i^t \;\; \forall\, t \tag{23}$$

$$x_i^t \geq 0,\;\; q_i^t \geq 0,\;\; I_i^t \geq 0 \;\; \forall\, i, t \tag{24}$$

where $I_i^{t-1} = q_i^{t-1} + I_i^{t-1} - r_i^{t-1}$ for all $t > 1$. With the above definition, a formal proposition for the optimal policy is established as follows.

Proposition: Suppose a set of integer positive values, X^*, represents a schedule of deliveries to a store such that a delivery is scheduled whenever the inventory available at the start of the period is less than the expected consumption for that period, and the quantity delivered is the difference between the current inventory level and the maximum inventory allowed at the store. That is,

$$x_i^t = \begin{cases} 1 & \text{if } q_i^t > 0 \\ 0 & \text{otherwise} \end{cases} \tag{25}$$

where,

$$q_i^t = \begin{cases} U_S - I_i^t & \text{if } \;\; I_i^t < d_i^t \text{ or } \sum^{t-1} q_i^o < \sum^{t} d_i^o \\ 0 & \text{otherwise} \end{cases} \tag{26}$$

Then X^* is an optimal solution to the direct delivery SRP-FS. A proof of the optimality of the above proposition for an individual store is presented in the Appendix.

The proposed policy provides us with a delivery plan that minimizes the total number of visits needed to satisfy the store orders. If a store is considered individually, minimizing the number of visits to the store yields the optimal solution. However, when all the stores are considered together, we have to account for the number of trucks used and its impact on the total cost of the solution. To address the fleet cost, we developed the following heuristic that takes an initial delivery schedule, built from the proposed policy, and minimizes the number of trucks needed over the planning horizon:

1. For each period *t*, calculate the average number of trucks needed up to that period:

$$V_t = \frac{\sum_i \sum_{o \le t} x_i^o}{t}$$

2. Determine a lower bound on the number of trucks needed (L_V) using the maximum value amongst the average calculated in step 1; that is, set $L_V = \max_t \{ V_t \}$.
3. Starting from the first period, identify the deliveries in the initial delivery schedule that meet the following criteria as “fixed”:
    a. It is scheduled on the first period
    b. It is scheduled directly after a “fixed” delivery
    c. The available space at the store on the previous period is not enough to cover the necessary inventory to meet the store order and/or expected consumption
4. If any period has a larger number of “fixed” deliveries than the calculated lower bound (L_V), update the lower bound to the number of “fixed” deliveries.
5. Build a final delivery schedule:
    a. All “fixed” deliveries will remain on the period they are on.
    b. Starting from the second period, assign any delivery from the initial schedule that is not “fixed” to the earliest period available that meets the following criteria:
        i. The number of deliveries on the available period is less than the lower bound L_V

ii. The available space at the store is greater than or equal to the amount of inventory delivered in the initial schedule

iii. If no previous period is feasible, then keep the delivery on the same period

6. Every time a delivery is scheduled, update the available space at the store.

We tested the proposed heuristic using three test instances generated with random data. Each instance consisted of 350 stores and 10 time periods. Each store was assigned a storage capacity of 3,000 lbs, while store orders per period for each store were randomly generated using a uniform distribution between 300 and 2,700 lbs in multiples of 10. The initial inventory level at each store was also randomly generated using the same method as the store orders. For this particular test, we assumed the inventory consumption rate is equal to the store order, i.e., inventory is consumed in the same period the store order is expected. We varied the truck capacity between instances to create a different delivery frequency profile for each instance, ranging across high, medium, and low. A higher truck capacity reduces the need for frequent deliveries, while a lower truck capacity increases the delivery frequency for each store. We compared the results obtained from the heuristic with the optimum solutions obtained from CPLEX using a modified version of the MIP model presented in the previous section.

Table 3-1. Summary of Results Using the Direct Delivery Heuristic

350 Stores; T = 10	**Delivery Intensity**								
Solution	**High**			**Medium**			**Low**		
	No. Visits	No. Trucks	Time (s)	No. Visits	No. Trucks	Time (s)	No. Visits	No. Trucks	Time (s)
CPLEX	2319	233	92.12	1797	180	1076.3	969	168	1197.1
Heuristic	2319	(260) 260	1.68	1797	(193) 180	1.57	969	(185) 168	0.98

% of Diff	0.00%	10.38%		0.00%	0.00%		0.00%	0.00%	

Table 3-1 summarizes the total number of visits (or deliveries) for each instance as well as the number of trucks needed. For the results obtained with the heuristic, the Table shows the final number of trucks found by the heuristic, as well as the initial number of trucks found when the initial delivery schedule is developed in parenthesis. The heuristic identifies the optimal solution for both instances within seconds when the delivery frequency is low (i.e., 20% of the periods have a delivery scheduled) or medium (i.e., 40% of the periods have a delivery scheduled). When the delivery frequency is high (i.e., deliveries for each store are, on average, scheduled on 70% of the periods), the heuristic finds a solution that is 10% over the optimum. Although a high delivery frequency is unlikely to be encountered in the fast-food industry, the above results show that the proposed heuristic is a practical tool to develop an initial delivery plan in a very short amount of time. In the following section we discuss the solution approach we developed for the milk-run version of the SRP-FS (i.e., each truck is allowed to visit two or more stores on one route). By adjusting how the truck lower bound is calculated in the Out-and-Back policy discussed in this section, we were able to use it to build an initial solution for the milk-run version of the problem. The adjustment involves determining the truck lower bound by comparing the average number of trucks required based on store visits (considering maximum number of stops allowed in a route) and the average number of trucks required based on inventory volume (considering single-driver truck capacity), and choosing the larger between them. After an initial schedule is built, we used the traditional Savings Algorithm (Clarke & Wright, 1964) to build corresponding routes and generate an initial feasible solution to the SRP-FS.

5. A Two-Step Simulated Annealing Heuristic for the SRP-FS

Turning our attention back to the milk-run version of the SRP-FS, it becomes evident that its solution structure, similar to the direct delivery version, can also be broken down into two main components: 1. Building a delivery schedule, and 2. Forming the deliveries into routes. The heuristic developed for the direct delivery version of the problem was also composed of two steps that were aligned with the two components of the solution structure, i.e., minimizing the number of store visits, and minimizing the number of trucks used. We therefore decided to use the same two-step structure for tackling the milk-run version of the SRP-FS, where each step is aligned with the two components of the solution structure, i.e., the first step focuses on building a feasible delivery schedule, while the second focuses on constructing the delivery routes.

The overall structure of the solution procedure we developed for the SRP-FS is based on Simulated Annealing (SA). First discussed by Kirkpatrick *et al*. (1983), SA heuristics have been used in the literature to solve a wide range of optimization problems. One of the main advantages of SA for complex combinatorial problems is that it can avoid being trapped by local minima by accepting non-improving solutions in the search neighbourhood (Eglese, 1990). Studies that employ SA have been shown to obtain good solutions to problems in vehicle routing, scheduling, and lot-sizing; see, for example, the surveys by Koulamas *et al*. (1994), and Vidal *et al*. (2013), among others, which motivated us to use it for the SRP-FS.

The first step in the proposed heuristic uses SA to find a feasible delivery schedule and the corresponding delivery quantities. However, in order to evaluate the cost of the solution, the routing component associated with the delivery schedule needs to be built. The second step of the heuristic addresses the construction of the routes. The routing component is nested within the iterative SA structure and uses an adjusted savings algorithm to build routes based on the delivery schedule created in the first step, and to determine which routes should be single-driver or team routes. The routes from the second step are used to calculate the overall cost of the solution, which allows the SA algorithm to either accept or reject the trial solution before proceeding to the next iteration. In the following sections we present and discuss the details of the above two algorithms.

5.1 Simulated Annealing Heuristic for Developing the Delivery Schedule

The general structure used for our SA heuristic is based on the annealing scheme used by Bozer & Carlo (2008) and discussed in Tompkins *et al*. (2010). The following notation is used for the SA heuristic:

$s^0/s/s'$ = initial/ current/ candidate solution

s^* = current solution found by the heuristic with the lowest cost

α = temperature reduction factor

T = set of temperatures t_i that follow the cooling schedule for the annealing process, where $t_i = t_0\ \alpha^i$ for all $i > 1$

t_0 = initial temperature

τ = parameter used to calculate an initial temperature

e = fixed epoch length

$f_j(s)$ = objective value of the jth accepted candidate solution, s, in an epoch

$\bar{f}_e$ = mean objective value of the accepted solutions in an epoch, where $\bar{f}_e = \sum_{j=1}^{e} f_j(s)\ /e$

$\bar{f}'_e$ = overall mean objective value of all the accepted candidate solutions during previous epochs to the current one for a given temperature

ϵ_i = threshold value to determine system equilibrium at temperature i

M = maximum number of epochs to evaluate over all temperature

I = counter to record the temperature index where the current best solution, s^*, was found

N = maximum number of successive temperature changes allowed without an improvement on the best solution, s^*

The proposed SA heuristic proceeds as follows and is depicted in Figure 3-1.

Step 1. Determine the total cost of the initial solution, $f(s^0)$. Use this value to set an initial temperature ($t_0 = f(s^0)/\tau$) and initialize $t_1 = \alpha t_0$, $I = 1$, and $i = 1$. Set the initial solution as the current solution for the SA, $s = s^0$.

Step 2a. Randomly select a period (p) and a store (n) that is visited on the selected period. Create a list of all other periods where it would be feasible to shift part or the entire delivery to store n, using the following rules:

- If the potential period is before period p, then check if there is space available at the store.
- If the potential period is after period p, then first check if part of the delivery can be delayed to the future period, or if it is required to meet store demand before the potential period. If part or all of the delivery can be delayed, then check if there is available space at the store on the future period.

Step 2b. From the list of potential periods, select the alternative period (p') with the lowest average truck utilization. If no deliveries are scheduled on the potential periods, then chose one randomly.

Step 2c. Determine the quantity (y') to be shifted to period p'.

- If $p' < p$, the quantity to shift will be the minimum between the space available at the store on the selected period, and the original delivery quantity.

- If $p' > p$, the quantity to shift will be the minimum between the delivery quantity that can be delayed and the space available at the store on the selected period.

Update the delivery schedule with the new delivery quantities, $y_n^p = y_n^p - y'$ and $y_n^{p'} = y_n^{p'} + y'$.

Step 2d. Using the updated delivery schedule, build a routing plan using the Adjusted Team Savings Algorithm for the altered periods *p* and *p'* (see Section 5.2). Set the delivery schedule and resulting routes as the new candidate solution s'.

Step 2e. Calculate the change in total cost $\Delta f = f(s) - f(s')$. If $\Delta f > 0$, then s' is accepted, set $s = s'$ and go to Step 2f; otherwise go to Step 2e.

Step 2f. Using a uniform distribution $U(0,1)$, sample a random variable x. If $x \leq \exp\left(\frac{\Delta f}{t_i}\right)$, then candidate solution s' is accepted ($s = s'$), go to Step 2g otherwise repeat the perturbation process from Step 2a.

Step 2g. Update the current objective value $f(s) = f(s')$. If $f(s) < f(s^*)$, then set the current solution as the best solution found ($s^* = s,\ f(s^*) = f(s)$) and record the temperature index $I = i$. If the number of candidate solutions accepted equals the epoch length e, go to Step 3; otherwise, go to Step 2a to repeat the perturbation process.

Step 3. If equilibrium has not been reached at temperature t_i, that is, if the percentage difference between the mean objective value of the accepted solution on the current epoch and the overall mean objective value of the accepted solutions on previous epochs is over a specified threshold $\left(\frac{|\bar{f}_e - \bar{f}_e'|}{\bar{f}_e'} \geq \epsilon_i\right)$, restart the counter of accepted solutions and repeat the perturbation process from Step 2a. If equilibrium has been reached, then update the temperature by setting $i = i + 1$ and $t_i = t_0\, \alpha^i$. If the maximum number of successive temperature changes without improvements has been reached $(i - I \geq N)$, **STOP**; otherwise go to Step 4.

Step 4. If the total number of epochs evaluated is less than the maximum (M), repeat the perturbation process from Step 2a; otherwise **STOP.**

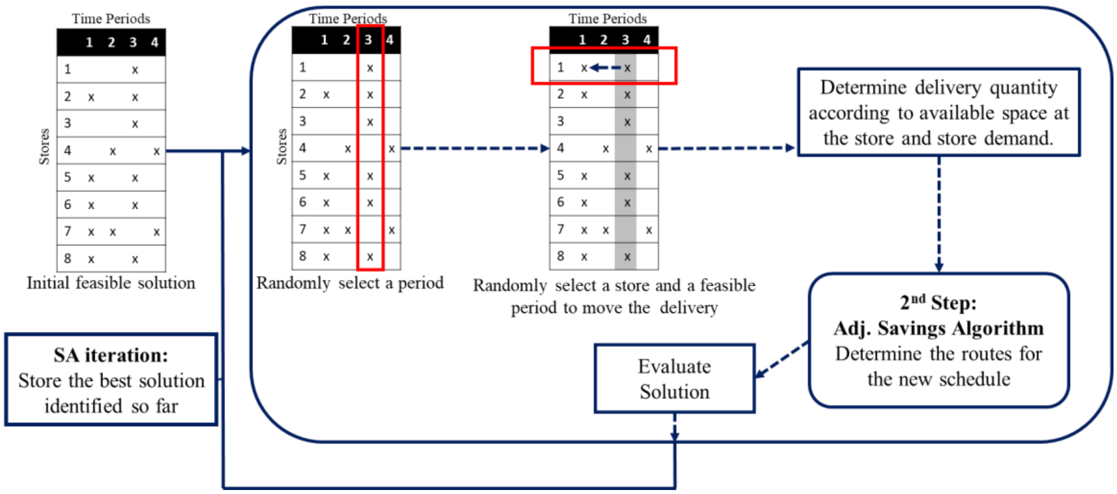


Figure 3-1. SA Heuristic Flowchart

One of the key steps in the above SA heuristic is how we perturb the current solution to obtain a candidate solution as shown in Step 2b. We evaluated a variety of alternative perturbation schemes using a test instance and compared it with choosing an alternative randomly. We found that selecting the period with the lowest truck utilization yielded the best solutions.

5.2 Adjusted Team Savings Algorithm for the Delivery Routes

Given a candidate delivery schedule generated by the SA heuristic, we need to build the delivery routes to evaluate the candidate schedule and determine the total cost of the solution. The Adjusted Team Savings algorithm (TS) was developed and integrated into the SA algorithm for this purpose. Similar to the classical savings algorithm developed by Clarke & Wright (1964), the TS starts from a set of routes (see below) and uses the potential savings that could be generated by changing a route from single-driver to a team route and potentially including more stores into the route. The details of the TS are presented below and summarized in Figure 3-2.

Step 1. Using the traditional Savings Algorithm, build single-driver routes to visit all the stores scheduled for delivery in period t. Calculate the cost of each baseline route (C_r^B). These single-driver routes serve as a baseline for the savings calculation.

Step 2. For each baseline route r, calculate its Team Savings S_r:

Step 2a. Let route r have the parameters of a Team route and continue adding stores to the route until truck capacity or maximum number of stops are met.

Step 2b. Calculate the new cost of route t using team route cost parameters (C_r^T). If additional stores were added to route r, then for comparison, calculate the cost of visiting these additional stores in a separate route (C_r^A).

Step 2c. Determine the Team Savings for each route r, $S_r = C_r^B + C_r^A - C_r^T$

Step 3. Select the route with the highest Team Savings S_r. If $S_r > 0$ go to Step 4; otherwise STOP and return the saved routes so far to the SA heuristic.

Step 4. Save the selected route as a final solution route. Rebuild baseline routes for all stores that are not included in the saved routes and go to Step 2.

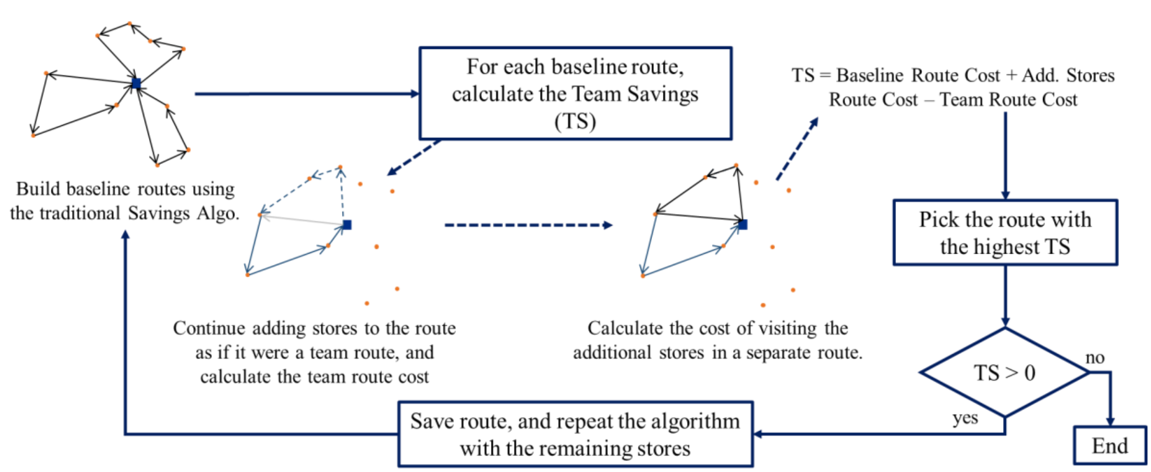


Figure 3-2. Adjusted Team Savings Algorithm Flowchart

6. Computational Results

The two-step SA heuristic (Section 5) was coded using Python 3.7.4 and all instances were run on a Windows computer with a 64-bit 2.50 GHz Intel Core i7 processor and 8GB of RAM. Two main sets of instances were used to obtain computational results: 1) Test set with randomly generated data; 2) Industry data set provided by our industry collaborator.

The set of randomly generated data was originally created and used for the SRP in Vigo Camargo & Bozer (2022). A total of 45 instances are contained in this set with sizes ranging from 10 to 350 stores ($N = 10, 100, 350$), with a time horizon of a week. The store locations were randomly generated in a 600 x 600 grid with the DC located at the center. Store demand was also randomly generated following a normal distribution $N(\mu = 2{,}030,\ \sigma = 760.15)$, based on the average value for store demand in the industry data instances. Each instance is characterized by low, medium and high frequency store orders, which was randomly generated for each store using a normal distribution with parameters $\mu = 1, 2, 4$ orders in the planning horizon, respectively, and a standard deviation of 0.60 for all three frequencies. A subset of these randomly generated instances was used as a test set to determine the parameter values to use for the SA heuristic, which we discuss in the following section.

6.1 SA Parameters

An important factor in the performance of a SA heuristic is the values assigned to the annealing parameters. Using guidelines discussed in Meller & Bozer (1996), we set the initial temperature as a fraction of the initial solution objective value using the parameter τ. Using this approach to set the initial temperature determines the initial threshold acceptance, i.e., setting $\tau = 40$ (with $\alpha = 1$) gives a solution that is 5% over the initial solution an initial acceptance probability of

0.135. We also defined a cooling schedule that uses α as the base cooling rate. The other two key parameters that can impact the performance of the SA heuristic are the epoch size (e) and the maximum number of temperature changes without an improvement in the solution (N). To determine the value of the above four parameters, we completed a preliminary experiment using three instances from the set of randomly generated instances of varying size.

Table 3-2. Alternative Values for SA Parameters

Variable		Parameter Value
α	Base cooling rate	*0.7, 0.8, 0.9*
τ	Initial temperature parameter	*20, 40, 80*
e	Epoch size	*15, 30, 60*
N	Max. number of temperature changes without improvements	*5, 10*

Table 3-2 summarizes potential values assigned to each parameter. From these values we built 54 parameter combinations and used them to solve each instance three times for a total of 648 runs. We compared the average solution obtained from each parameter combination with the lowest value found on each corresponding instance size by calculating the percentage difference between the objective values. Figure 3-3 shows the average percentage difference obtained from each parameter combination across all three instance sizes. From these preliminary results, we selected the combination of parameter settings that yielded the lowest average percentage difference. While the graph shows various parameter sets that are close to the lowest solution line, two individual sets obtained the lowest percentage difference, highlighted in the graph with an increased marker size. Looking at the breakdown of the results from these two parameter sets, we identified that one set performed better in smaller instances, while the other set performed better in larger instances. We selected the parameter set that performed better in larger instances (marked with a red outline in the graph), as these would be more representative of industry data. The values for this set correspond to $\alpha = 0.8, \tau = 40, e = 15, N = 10$.

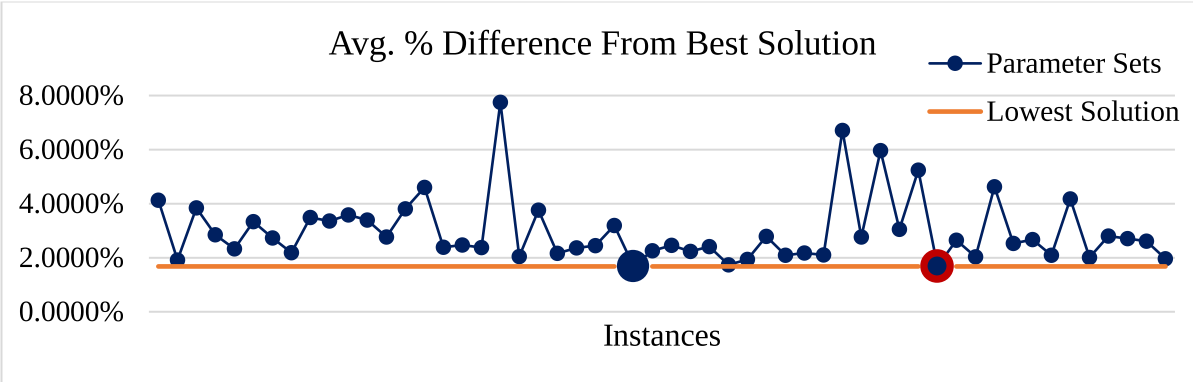


Figure 3-3. SA Parameter Testing

6.2 SA Heuristic Performance

To evaluate the performance of the proposed heuristic, we used the set of instances with randomly generated data. We solved all 45 instances with the SA heuristic using the parameters identified in the previous section. As a comparison, we also solved a version of the SRP-FS where we assume infinite truck capacity (ITC), for both types of routes, to serve as a benchmark. Furthermore, we used the formulation presented in Section 3 to solve the problem directly using CPLEX optimization software. However, due to the complexity and size of the problem, CPLEX was unable to solve the instances with 100 and 350 stores. For the instances with 10 stores, CPLEX reported incumbent solutions and their optimality gap after running for an hour. Due to the pseudo-random nature of the SA heuristic, we ran each instance three times and summarized the results in Table 3-3. We first discuss the results for all the instances and compare them with the ITC benchmark, and then we discuss the results obtained from CPLEX, which are summarized in Table 3-4.

Table 3-3. SA Heuristic Performance

	SRP-FS			ITC			% Diff
Instance	**Best Sol**	**Avg Sol**	**Avg Time**	**Best Sol**	**Avg Sol**	**Avg Time**	**ITC**
n10-1	9,923.58	10,070.75	1,405.06	9,119.95	9,388.12	164.92	8.81%
n10-1-H	17,061.81	17,497.49	1,219.67	16,470.11	16,640.07	908.21	3.59%

n10-1-L	7,100.71	7,200.85	242.80	6,185.35	6,857.21	225.53	14.80%
n10-2	9,832.52	9,919.68	455.65	8,975.78	9,060.80	336.34	9.55%
n10-2-H	15,930.02	16,363.90	1,362.62	14,108.38	14,720.53	1,482.85	12.91%
n10-2-L	6,864.76	6,864.76	350.26	6,236.26	6,621.39	236.93	10.08%
n10-3	8,748.35	8,949.07	389.65	8,083.03	8,083.03	313.99	8.23%
n10-3-H	14,124.00	14,786.81	891.00	12,260.53	12,699.46	711.48	15.20%
n10-3-L	7,057.15	7,272.33	643.51	6,209.77	6,450.89	143.11	13.65%
n10-4	8,512.20	8,524.86	2,737.07	7,572.14	7,826.01	525.50	12.41%
n10-4-H	14,224.39	14,944.91	2,000.22	12,697.45	12,854.64	346.64	12.03%
n10-4-L	6,192.65	6,463.67	434.82	6,010.07	6,018.22	165.75	3.04%
n10-5	9,424.10	9,605.88	1,127.08	9,127.44	9,138.64	111.57	3.25%
n10-5-H	17,599.53	18,005.99	486.86	16,250.24	16,451.59	587.11	8.30%
n10-5-L	6,455.45	6,585.78	94.56	6,283.86	6,283.86	122.27	2.73%
						Avg	**9.24%**
n100-1	52,246.38	52,818.75	3,613.79	49,715.72	50,925.75	3,048.01	5.09%
n100-1-H	92,901.38	93,928.69	3,661.91	90,103.41	90,311.98	3,650.85	3.11%
n100-1-L	38,316.24	38,582.24	3,611.22	38,214.97	39,531.74	3,070.20	0.27%
n100-2	57,782.79	58,254.54	3,620.76	56,327.53	56,866.26	3,458.82	2.58%
n100-2-H	92,558.58	93,405.59	3,653.33	88,476.45	89,469.17	3,646.78	4.61%
n100-2-L	34,120.59	35,155.33	3,610.42	33,792.43	35,207.49	5,408.93	0.97%
n100-3	57,043.01	57,702.15	3,622.06	55,978.29	56,578.24	3,535.86	1.90%
n100-3-H	93,857.39	94,255.18	3,658.80	90,244.60	91,271.36	3,658.51	4.00%
n100-3-L	37,168.58	38,263.42	3,237.63	35,870.97	37,086.38	3,433.32	3.62%
n100-4	52,595.99	53,112.09	3,620.32	50,760.93	51,346.73	3,419.54	3.62%
n100-4-H	94,312.42	95,552.99	3,657.86	91,503.37	92,323.32	3,652.98	3.07%
n100-4-L	34,670.03	35,592.53	3,558.49	34,442.83	35,614.49	3,610.44	0.66%
n100-5	50,443.29	50,853.20	3,618.88	49,577.22	50,146.83	3,620.71	1.75%
n100-5-H	87,295.91	87,647.42	3,660.96	85,777.74	86,094.94	3,656.69	1.77%
n100-5-L	33,948.29	34,481.43	3,610.85	33,608.59	34,186.48	3,612.05	1.01%
						Avg	**2.53%**
n350-1	161,766.78	162,362.88	3,826.98	143,800.75	144,261.18	3,779.37	12.49%
n350-1-H	266,950.06	267,463.82	4,506.34	252,334.26	254,189.63	4,319.87	5.79%
n350-1-L	111,016.14	111,290.11	3,677.32	109,124.05	109,822.33	3,689.26	1.73%
n350-2	159,444.87	159,726.92	3,792.49	142,444.05	143,925.80	3,832.99	11.94%
n350-2-H	260,978.22	261,373.96	4,361.89	252,341.19	252,810.27	4,251.96	3.42%
n350-2-L	102,113.67	102,430.09	3,673.67	95,045.71	95,705.67	3,663.05	7.44%
n350-3	154,599.79	155,164.67	3,779.09	141,562.99	143,649.03	3,777.37	9.21%
n350-3-H	263,099.04	263,326.03	4,267.45	252,834.28	253,090.52	4,146.99	4.06%
n350-3-L	105,354.96	106,137.34	3,685.49	97,496.93	97,969.08	3,671.27	8.06%
n350-4	159,284.80	160,307.88	3,796.63	137,672.59	139,674.75	3,810.48	15.70%
n350-4-H	261,794.04	263,179.84	4,491.39	248,542.02	249,466.64	4,274.37	5.33%
n350-4-L	102,920.99	103,565.69	3,681.87	94,779.12	95,369.98	3,670.75	8.59%
n350-5	155,915.76	156,192.11	3,801.30	139,295.52	140,339.97	3,748.63	11.93%
n350-5-H	261,047.01	261,299.51	4,361.10	247,083.48	247,734.52	4,291.79	5.65%
n350-5-L	100,508.43	101,310.92	3,692.40	95,797.74	96,691.68	3,667.99	4.92%

Avg	**7.75%**
Total Avg	**6.51%**

Table 3-3 shows the best and average solution found with the SA heuristic for each instance as well as the average runtime in seconds. Under the ITC block, we show the same information for the benchmark runs as a comparison. The last column shows the percentage difference between the best solution found by the SA heuristic and the benchmark solution based on ITC.

Overall, the SA heuristic solutions have a gap of 6.51% with the benchmark solutions. The percentage difference between the SA solutions and the ITC solutions range from 0.27% to 15.7%, with the smaller instances yielding solutions with a larger average gap than the other two group of instances. Solving the instances with 10 stores using the MIP formulation and CPLEX, we obtain the results shown in Table 3-4 where the incumbent solutions, along with the optimality gaps reported by the solver, were obtained with a maximum solver time set at one hour. CPLEX not able to solve larger instances of 100 and 350 stores due to the size of the model.

Table 3-4. MIP Solutions for n10 Instance

Instance	SRP-FS Best Sol	MIP Best Sol	Opt Gap	% Diff
n10-1	9,923.58	8,595.91	9.53%	15.45%
n10-1-H	17,061.81	16,309.28	13.49%	4.61%
n10-1-L	7,100.71	6,622.51	6.00%	7.22%
n10-2	9,832.52	9,158.19	16.93%	7.36%
n10-2-H	15,930.02	15,098.00	26.13%	5.51%
n10-2-L	6,864.76	6,846.35	13.50%	0.27%
n10-3	8,748.35	8,385.17	11.23%	4.33%
n10-3-H	14,124.00	13,161.83	15.44%	7.31%
n10-3-L	7,057.15	6,855.96	16.19%	2.93%
n10-4	8,512.20	7,991.03	8.72%	6.52%
n10-4-H	14,224.39	13,401.95	14.86%	6.14%

n10-4-L	6,192.65	6,041.41	2.87%	2.50%
n10-5	9,424.10	9,079.45	25.01%	3.80%
n10-5-H	17,599.53	17,080.49	31.68%	3.04%
n10-5-L	6,455.45	6,424.42	21.46%	0.48%
			Avg	**5.17%**

With an average of 5.17%, and a median value of 4.61%, the percentage difference between the SA heuristic solution and the best solution found by solving the MIP model directly seem comparable to the difference we observed between the SA and ITC solutions for larger instances. These results support the ability of the SA heuristic to find quality solutions for a variety of instances, with different demand profiles and sizes, in a reasonable time. In the following section we discuss the results obtained using industry data and how they compare to the current industry benchmark and previous results from the traditional SRP.

6.3 Industry Data Results

There are two instances of industry data, and each instance consists of store orders placed by a network of 181 stores over one week of operations. These stores are all served by the same DC in the given region. To determine the distance between each store and the DC, we used the Google Maps API (Google, 2020) to calculate actual road distances. As with the randomly generated instances, the usable truck capacity for team routes was set to be 35% higher than single-driver routes, while team routes were given 30% more route time before incurring in overage costs. The cost parameters were obtained directly from our industry collaborator and are not shown due to confidentiality. In relative terms, the cost/mile traveled for team routes is 20% higher than single-driver routes, and the cost/pound delivered is 35% higher for team routes. Overage cost for routes that exceed the maximum set time reflects the daily cost of having a truck available.

We compared the SA solutions obtained for the SRP-FS with the solutions obtained for the traditional SRP and the solution that represents the current industry baseline. The solutions to the traditional SRP are taken from Vigo Camargo & Bozer (2022), who use a clustering heuristic to solve the SRP with no early deliveries considered. Figures 3-4 and 3-5 show the comparison and cost breakdown between the above three solutions for the two instances of industry data. The graphs in these figures show how the SRP-FS solution reduces the total cost relative to the industry baseline by 15% and 20% for the two instances. Furthermore, when compared with the solution obtained for the traditional SRP, the SA heuristic for the SRP-FS obtains a solution with an 11% and 10% cost reduction for the two instances, which represents a significant sum in terms of absolute logistics costs.

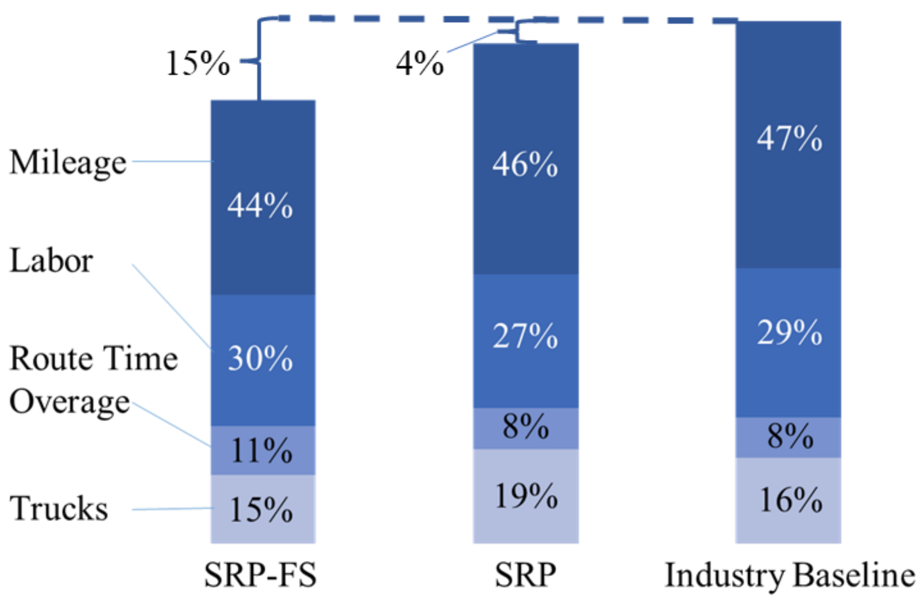


Figure 3-4. Industry Data Comparison – Instance 1

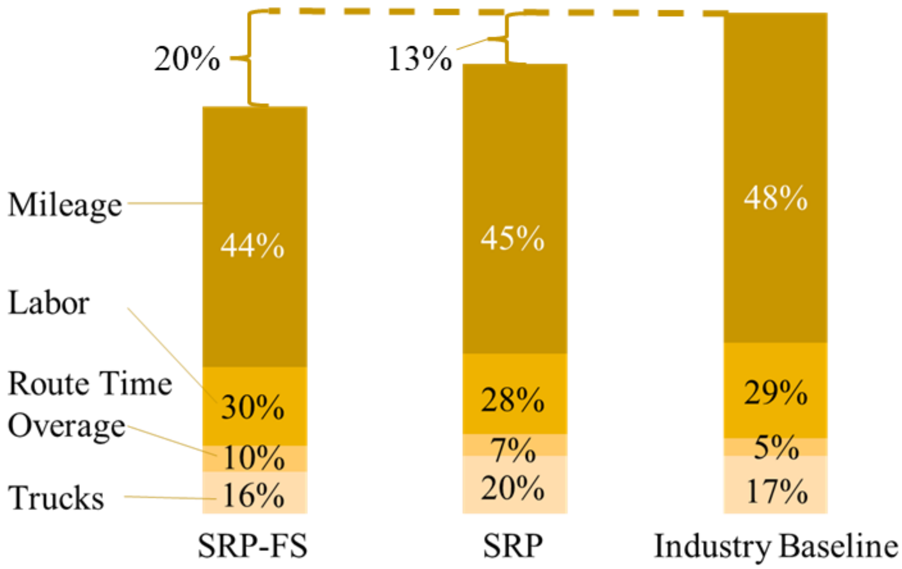


Figure 3-5. Industry Data Comparison – Instance 2

The significant reduction in total cost for both instances reflects the benefit of allowing early deliveries. The solution for the SRP-FS has 13% less miles traveled for instance 1, and 17% for instance 2. Likewise, it uses approximately 20% fewer trucks in both instances, and does not employ any team routes at all, in contrast to over 20 team routes used in the industry baseline in both instances. While no team routes are used, the SRP-FS solution does increase the number of routes that exceed the maximum route time, with a 25% increase in instance 1, and a 55% increase in instance 2. When comparing the solutions from the SRP-FS with those of the traditional SRP, we see a similar trend with reductions in the number of trucks used (27% on both instances), and in the total miles traveled (10% and 13% for the two instances). The SRP solution employs far less team routes than the industry baseline, but it is still outperformed by the SRP-FS, which does not use any team routes. In contrast, the SRP-FS solutions have 20% more routes incurring overage costs than the SRP solutions. Even with the above differences, the SPR-FS manages to find a solution with 9%-11% lower costs than the traditional SRP, for both instances. These reductions in the total cost can be attributed to the flexibility introduced by allowing early deliveries. When comparing the solutions for the SRP-FS with the fixed schedule solution, we found that 3% of the deliveries scheduled for instance 1 were made in an earlier period, while for the second instance, 15% of the deliveries were made earlier.

The reductions observed in the number of team routes used goes along with an increase in the number of longer routes. While longer routes incur an overage cost, by not using team routes, the model is able to identify savings in both labor and mileage costs. These longer routes allow the SRP-FS to make better use of the truck capacity throughout the planning horizon. Table 3-5 shows a summary of the truck utilization from all three solution sets for both instances. We observe that, under the SRP-FS, the truck utilization ranges between 27%-73% for the two instances, while both the SRP and industry baseline have a significantly larger range of 5%-99% for both instances. The flexibility inherent in the SRP-FS reduces the need to use trucks to make single, small deliveries and to employ team routes. With team routes being a common practice in the trucking industry, the results shown here can motivate further study into how team routes are employed and the impact they may have beyond their economic impact.

Table 3-5. Truck Utilization Summary

	Instance 1			Instance 2		
	SRP-FS	SRP	Industry Baseline	SRP-FS	SRP	Industry Baseline
Min	33%	5.0%	5.4%	27%	5.3%	7.3%
Max	73%	99.3%	100%	73%	99.4%	99.0%
Average	64%	61.8%	61.1%	64%	64.6%	52.7%

Allowing early deliveries comes, on the other hand, with the constraint imposed by the limited store capacity. For our study we set the capacity of all the stores to be 1.5 times the largest store order in the network, to follow the assumption that no store order is larger than the store capacity. Based on this parameter, the SRP-FS solution shows that the average store capacity utilization ranges from 5%-34% for instance 1, and 5%-54% for instance 2, with an average of 12% and 15%, respectively. While this result indicates that the truck capacity is the limiting factor for our data, adjusting these values allows the model to be used to assess the potential savings associated with increasing the storage capacity at the stores. For example, stores located far from the DC may benefit from larger storage capacities.

Last, the cost reduction obtained by allowing early deliveries is particularly interesting because it can be a significant adjustment to the more common approach in the fast-food industry of using a fixed delivery schedule each week, where only the store orders are updated from one week to the next. The potential of further reducing their overall costs by 10% can be a significant motivator to allow early deliveries as long as the store capacity is not exceeded.

7. Conclusions

In this paper, we extend the traditional SRP and introduce the SRP-FS, where the delivery schedule as well as the amount of inventory to deliver on each visit are part of the decision variables. The SRP-FS has the objective of minimizing the overall replenishment costs associated with the distance traveled, the labor costs, route-time overage costs, and the truck cost, while ensuring that store orders are delivered by the specified period, and neither the store capacity nor the truck capacity are exceeded. To the best of our knowledge, this is the first study in the literature to incorporate the above cost components into the decision-making process while considering single-driver and team routes and at the same time allowing a flexible delivery schedule. As part of the contributions of this paper, we presented a MIP formulation of the SRP-FS with subtour elimination constraints. However, due to the size and complexity of the model, a commercial optimization software (CPLEX) was unable to solve even small instances of 10 stores to optimality. To tackle larger and realistic instances, we developed a two-step SA heuristic that incorporates an adjusted savings algorithm for the routing component of the problem. Another contribution of the paper is the study of the direct delivery version of the SRP-FS, where a truck is allowed to make only out-and-back deliveries to each store. For the direct delivery version of the problem, we developed a heuristic policy that yields a near-optimal solution.

Test instances with 10, 100 and 350 stores, and randomly generated store demand were used to evaluate the performance of the proposed SA heuristic. Comparing the SA solutions with those solutions obtained with an infinite truck capacity as a baseline, we observed that the SA solutions have a gap of about 7% on average across all instances. While the MIP was unable to solve the smaller instances to optimality, we were able to obtain incumbent solutions for instances with 10 stores and compare them to the solutions obtained from the SA heuristic. The comparison showed an average gap of about 5% between the SA heuristic solution and the MIP incumbent solution (with an average optimality gap of about 15.5%). While the analysis of these results is limited by the fact that we could not obtain optimal solutions to the test instances, we were able to show that the ITC benchmark used for comparison has a comparable gap between the SA heuristic results and those obtained from the MIP model and thus could serve as a lower bound for the problem.

The quality of the solutions obtained with the SA heuristic is further supported by the results obtained using two instances of industry data obtained from our collaborator in the fast-food industry. We were able to use the SA heuristic to obtain a solution for both instances and compare them with the current industry baseline as well as with the traditional SRP solution with a fixed delivery schedule. In this comparison, the SRP-FS outperforms both the industry baseline and the traditional SRP. The cost reductions are reflected in the reductions we obtain in the number of trucks required, the number of store visits, the distance traveled, and the number of team-routes used. The results show that the proposed SA heuristic is capable of managing data instances of realistic size while obtaining good solutions that provide key managerial insights concerning early deliveries and store capacities.

Several managerial insights are gained from the study. First, the results highlight the significant impact that a flexible delivery schedule has on the overall cost. Our results show that allowing early deliveries can lead to reductions in most cost elements while still meeting store orders on the specified periods. However, it is common in industry to use predetermined delivery days for each store (perhaps due to its simplicity), and changing to a flexible delivery schedule would require operational changes. Such changes would require input from and participation by different stakeholders across the supply chain, particularly from franchise owners. Further analysis of such changes is beyond the scope of our study. Second, the results may also impact how management utilizes available resources, such as the delivery trucks and the fleet size, through the planning horizon, and how work is assigned to drivers by balancing the use of single-driver and team routes. Our results suggest that current industry applications may be employing an excess of team routes, and that a reduction in such routes can help decrease the overall cost. Further study into the use of team vs single-driver routes and their impact in the fast-food supply chain seems well-justified to develop additional insights.

The results in this study are based on the assumption that the planning horizon of the SRP-FS is given and fixed. Future areas of research can focus on using data-analytic tools to incorporate a look-ahead component for the planning horizon to potentially improve the decisions made by the model. Incorporating future demand data from stores can impact the quantities that are delivered to stores as future consumption is taken into account. The results of the study are also based on limited space available to store inventory at each store. Available space at each store is often at a premium and adding more space may or may not be feasible. However, insufficient store space would force frequent deliveries, which would drive up the cost especially for stores that are not close to the DC. Future research may focus on how store capacity impacts the solution structure and replenishment decisions. Such research would also have interesting managerial implications as it can better inform space requirements imposed by management on future stores based on their locations and demand.

Another direction of future research we are currently exploring is focused on the impact the store locations (relative to the DC and relative to each other) has on the decisions made by the SRP-FS. Factors like clusters of stores, the size of the clusters, and their distance to the DC, can be exploited to further reduce cost and make better replenishment decisions. In industries like the fast-food industry, the decision of where to open new stores is often driven by projected customer demand, existing store locations, and market competitors. This can lead to areas in a network with a high density of stores due to high customer demand, or remote locations at a significant distance from the DC. Understanding how such factors influence the solution structure and incorporating them into the decision-making process can have significant benefits and important managerial implications for reducing the replenishments costs as well as assessing the locations of current and future stores.

References

Akbalik, A., Penz, B., & Rapine, C. (2015). Capacitated lot sizing problems with inventory bounds. *Annals of Operations Research*, *229*(1), 1–18. https://doi.org/10.1007/s10479-015-1816-6

Al-Hinai, N., & Triki, C. (2020). A two-level evolutionary algorithm for solving the petrol station replenishment problem with periodicity constraints and service choice. *Annals of Operations Research*, *286*, 325–350. https://doi.org/https://doi.org/10.1007/s10479-018-3117-3

Andersson, H., Hoff, A., Christiansen, M., Hasle, G., & Løkketangen, A. (2010). Industrial aspects and literature survey: Combined inventory management and routing. *Computers and Operations Research*, *37*(9), 1515–1536. https://doi.org/10.1016/j.cor.2009.11.009

Archetti, C., Fernández, E., & Huerta-muñoz, D. L. (2017). The Flexible Periodic Vehicle Routing Problem. *Computers and Operations Research*, *85*, 58–70. https://doi.org/https://doi.org/10.1016/j.cor.2017.03.008

Archetti, C., Fernández, E., & Huerta-muñoz, D.L. (2018). A two-phase solution algorithm for the Flexible Periodic Vehicle Routing Problem. *Computers and Operations Research*, *99*, 27–37. https://doi.org/10.1016/j.cor.2018.05.021

Bell, W.J., Dalberto, L.M., Fisher, M.L., Greenfield, A.J., Jaikumar, R., Kedia, P., … Prutzman, P. J. (1983). Improving the Distribution of Industrial Gases with an On-Line Computerized Routing and Scheduling Optimizer. *Interfaces*, *13*(6), 4–23. https://doi.org/10.1287/inte.13.6.4

Boers, L., Atasoy, B., Correia, G., & Negenborn, R.R. (2020). The Multi-period Petrol Station Replenishment Problem: Formulation and Solution Methods. In E. Lalla-Ruiz, M. Mes, & S. Voß (Eds.) (pp. 600–615). Cham: Springer International Publishing.

Bozer, Y.A., & Carlo, H.J. (2008). Optimizing inbound and outbound door assignments in less-than-truckload crossdocks. *IIE Transactions*, *40*(11), 1007–1018. https://doi.org/10.1080/07408170802167688

Brahimi, N., Absi, N., Dauzère-Pérès, S., & Nordli, A. (2017). Single-item dynamic lot-sizing problems: An updated survey. *European Journal of Operational Research*, *263*(3), 838–863. https://doi.org/10.1016/j.ejor.2017.05.008

Brown, G.G., & Graves, G.W. (1981). Real-Time Dispatch of Petroleum Tank Trucks. *Management Science*, *27*(May 2022), 19–32. https://doi.org/https://doi.org/10.1287/mnsc.27.1.19

Clarke, G., & Wright, J.W. (1964). Scheduling of Vehicles from a Central Depot to a Number of Delivery Points. *Operations Research*, *12*(4), 568–581. https://doi.org/10.1007/s13398-014-0173-7.2

Coelho, L.C., Cordeau, J., & Laporte, G. (2014). Thirty Years of Inventory Routing. *Transportation Science*, *48*(1), 1–19. https://doi.org/https://doi.org/10.1287/trsc.2013.0472

Cordeau, J., Laporte, G., & Mercier, A. (2001). A unified tabu search heuristic for vehicle routing problems with time windows. *Journal of Operational Research Society*, *52*, 928–936.

Cornillier, F., Boctor, F.F., Laporte, G., & Renaud, J. (2008a). A heuristic for the multi-period petrol station replenishment problem. *European Journal of Operational Research*, *191*, 295–305. https://doi.org/10.1016/j.ejor.2007.08.016

Cornillier, F., Boctor, F.F., Laporte, G., & Renaud, J. (2008b). An exact algorithm for the petrol station replenishment problem. *Journal of the Operational Research Society*, *59*, 607–615. https://doi.org/10.1057/palgrave.jors.2602374

Eglese, R.W. (1990). Simulated annealing: A tool for operational research. *European Journal of Operational Research*, *46*(3), 271–281. https://doi.org/10.1016/0377-2217(90)90001-R

Florian, M., & Klein, M. (1971). Deterministic Production Planning with Concave Costs and Capacity Constraints. *Management Science*, *18*(1), 12–20. https://doi.org/https://doi.org/10.1287/mnsc.18.1.12

Francis, P., Smilowitz, K. and Tzur, M. (2006) 'The Period Vehicle Routing Problem with Service Choice', *Transportation Science*, 40(4), pp. 439–454. doi: 10.1287/trsc.1050.0140.

Francis, P., Smilowitz, K., & Tzur, M. (2008). The Period Vehicle Routing Problem and its Extensions. In B. Golden, S. Raghavan, & E. Wasil (Eds.), *The vehicle routing problem: latest advances and new challenges* (pp. 73–102). New York.

Goel, A., Vidal, T., & Kok, A.L. (2019). *To team up or not – Single versus team driving in European road freight transport*. Rio de Janeiro, Brasil. Retrieved from https://w1.cirrelt.ca/~vidalt/papers/Team-vs-Single.pdf

Google. (2020). Google Maps Platform. Retrieved from https://cloud.google.com/maps-platform/

Huerta-Muñoz, D.L., Archetti, C., Fernández, E., & Perea, F. (2022). The Heterogeneous Flexible Periodic Vehicle Routing Problem : Mathematical formulations and solution algorithms. *Computers and Operations Research*, *141*(105662). https://doi.org/https://doi.org/10.1016/j.cor.2021.105662

Kirkpatrick, A.S., Gelatt, C.D., Vecchi, M.P., Science, S., Series, N., & May, N. (1983). Optimization by Simulated Annealing. *Science*, *220*(4598), 671–680. https://doi.org/https://doi.org/10.1126/science.220.4598.671

Koulamas, C., Antony, S., & Jaen, R. (1994). A survey of simulated annealing applications to operations research problems. *Omega*, *22*(1), 41–56. https://doi.org/10.1016/0305-0483(94)90006-X

Laporte, G. (2009). Fifty years of vehicle routing. *Transportation Science*, *43*(4), 408–416. https://doi.org/10.1287/trsc.1090.0301

Meller, R.D., & Bozer, Y.A. (1996). A new simulated annealing algorithm for the facility layout problem. *International Journal of Production Research*, *34*(6), 1675–1692. https://doi.org/10.1080/00207549608904990

Moin, N. H., & Salhi, S. (2007). Inventory routing problems: a logistical overview. *Journal of the Operational Research Society*, *58*(9), 1185–1195. https://doi.org/10.1057/palgrave.jors.2602264

Mor, A., & Speranza, M.G. (2022). Vehicle routing problems over time : a survey. *Annals of Operations Research*, 1–21. https://doi.org/https://doi.org/10.1007/s10479-021-04488-0

Rothenbächer, A. (2019). Branch-and-Price-and-Cut for the Periodic Vehicle Routing Problem with Flexible Schedule Structures. *Transportation Science*, *53*(3), 850–866. https://doi.org/https://doi.org/10.1287/trsc.2018.0855

Tompkins, J.A., White, J.A., Bozer, Y.A., & Tanchoco, J.M.A. (2010). *Facilities Planning*. John Wiley & Sons.

Toth, P., & Vigo, D. (Eds.). (2014). *Vehicle Routing: Problems, Methods, and Applications*. Philadelphia: Society for Industrial and Applied Mathematics : Mathematical Optimization Society. Retrieved from http://epubs.siam.org/doi/book/10.1137/1.9781611973594

Vidal, T., Crainic, T.G., Gendreau, M., & Prins, C. (2013). Heuristics for multi-attribute vehicle routing problems: A survey and synthesis. *European Journal of Operational Research*, *231*(1), 1–21. https://doi.org/10.1016/j.ejor.2013.02.053

Vigo Camargo, A., & Bozer, Y.A. (2022). *Development and Application of a Cost-Driven Decision Model for Store Replenishment Logistics in the Fast-Food Sector*. [Manuscript submitted for publication].

Chapter 4

Impact of Network Geography for the Store Replenishment Problem in the Fast-Food Sector[3]

1. Introduction

The logistics associated with replenishing stores in a network is a complex problem with significant cost implications that can be found in a multitude of industries. The Store Replenishment Problem (SRP) was heavily motivated by the logistical challenges encountered in the fast food industry (Vigo Camargo and Bozer, 2022). Decisions related to the number and locations of the fast-food stores are often based on factors such as customer demand, franchising agreements, local trends and competition, as well as the location of existing stores. Given the non-uniform distribution of population centers across a geographic area, stores are often located in an uneven manner, with some areas representing a high density of stores, while other stores are located in fairly remote locations, away from other stores and the distribution center (DC) from which all the stores are replenished. Such a mixed composition of store locations in the network presents both a challenge and an opportunity in terms of logistics planning and developing ideal routes to meet store demand while minimizing overall logistics costs.

In this paper we focus on the store network composition and its impact on solving the SRP. Due to the nature of the fast-food industry, clusters of stores are often found in and around high-population centers. The key question we address is whether or not one can develop better and simpler solution procedures to the SRP, including the routing decisions, by taking the above

[3] This Chapter will be submitted for review and possible publication as: Bozer, Y. A. and Vigo Camargo, A. (2023) 'Impact of Network Geography on the Store Replenishment Problem in the Fast-Food Sector.'

clusters into account and what such a procedure would look like. While clustering heuristics are not novel, we show that store clusters can indeed be exploited for tackling the SRP. The solution procedure we propose incorporates a bin-packing formulation and a routing component with a clustering heuristic to identify solutions that outperform previous results and with significantly shorter runtimes.

The SRP was first defined by Vigo Camargo & Bozer (2022) as a replenishment problem that incorporates routing and fleet-sizing decisions as well as selecting between single-driver versus team routes (two drivers), while minimizing a multi-component cost function that considers distance, delivery labor, fleet size, and route time overage costs. In this paper we introduce an alternative solution procedure for the SRP based on the same assumptions as in the above paper of a given and fixed delivery schedule in a known store network. We use a clustering heuristic to group the stores into clusters and thus reduce the decision space for the problem. The above clusters are then used as input for a bin packing model, which assigns the clusters into bins that represent trucks/routes to be used in minimizing the four cost components of the original objective function. Lastly, we use a multi-step routing heuristic to build the final delivery routes based on the assignments obtained from the bin packing solution.

Data obtained from the supply chain group of an (inter)national, well-known fast-food company, who served as our industry collaborator for this study, were used to evaluate the performance of the proposed heuristic. We compare the results of the new clustering-based approach above with the baseline cost obtained from our industry collaborator as well as with the solutions presented in Vigo Camargo & Bozer (2022) and show that the new heuristic outperforms both.

The remainder of the paper is organized as follows. In Section 2 we discuss some of the relevant literature. We introduce our solution procedure and discuss its key components in Section 3. Section 4 presents the computational results obtained using the proposed heuristic. Lastly, we discuss our conclusions and areas of future research in Section 5.

2. Literature Review

The objective of this study is to introduce a solution method for the SRP that incorporates the clusters that occur in the store network in order to facilitate the decision-making process. The SRP was first defined in Vigo Camargo & Bozer (2022), which includes a literature review focused on the structure of the problem. For this paper, we will discuss relevant literature to the solution method we are proposing and how it relates to the SRP.

Routing problems have been extensively studied in the literature over the years. Surveys like those by Sharda *et al.* (2008) and Mor and Speranza (2022) discuss a variety of studies focused on routing problems and the different techniques used to solve them. Some of these studies have employed a type of methodology that can be described as 'cluster first-route second' (Bodin, 1975) where the set of nodes that need to be visited is divided into smaller subsets and then smaller routing problems are solved.

In Hiquebran *et al.* (1993), the authors use a nearest neighbor algorithm to group nodes in a network and in conjunction with a simplified simulated annealing present a solution method to the traditional vehicle routing problem (VRP). Motivated by the waste collection logistics, Kim *et al.* (2006) presented the waste collection VRP, which differs from the traditional VRP in that trucks are allowed to make stops at disposal stations when they are full and complete additional routes in the same period. The authors used a clustering-based heuristic, which was based on the K-means algorithm, and an insertion algorithm to build the routes to solve the waste collection

VRP. Similarly, Dondo and Cerdá (2007) presented a multi-step heuristic approach to solve the multi-depot VRP with a heterogeneous fleet, where clusters are used to simplify and solve a direct MILP formulation of the problem. Shortly after, Ghoseiri and Ghannadpour (2010) used an urgency-based clustering heuristic to assign trains to depots and then used a hybrid genetic algorithm to find a solution for the multi-depot VRP with a homogenous fleet. In Qi *et al.* (2012), the authors use spatial and temporal information to cluster nodes in the network and use a genetic algorithm to solve the VRP with time windows (VRPTW). More recently, Fachini and Armentano (2020) present an exact algorithm for the heterogeneous fixed fleet VRPTW that uses Benders decomposition to divide the problem into a generalized assignment problem and TSP subproblems.

Bramel and Simchi-Levi (1995) introduced a location based heuristic to solve the capacitated VRP (CVRP), which approximates the problem to a capacitated concentrator location problem to connect subsets of stores and turn them into routes. Shin and Han (2011) present a three-step heuristic for the CVRP using a distance-only objective function. Their cluster first approach uses a centroid-based clustering heuristic that uses a strategy of starting new clusters using seed nodes that are the furthest from the other clusters built. After employing a cluster-improvement process, they solve traveling salesman problems (TSP) on each cluster to build the final routes. To solve the CVRP, Kao and Chen (2013) used combinatorial particle swarm optimization to cluster customers and then simulated annealing to sequence the corresponding routes.

While the proposed heuristic in this paper employs a similar cluster first-route second approach as those discussed above, our approach is motivated by the replenishment logistics in the fast-food sector and incorporates elements that differ from those in the other studies discussed here. The structure of the SRP, with is multi-dimensional cost objective function, and the incorporation of fleet sizing and single versus team routes, with the routing decisions, make it stand apart from the other VRP-focused papers. Cömert *et al.* (2017) uses a cluster first-route second approach to solve a routing problem motivated by a supermarket supply chain, but the problem is formulated as a VRPTW with the objective of minimizing total time. Furthermore, our approach includes formulating the SRP as a bin packing problem and directly solving the model to obtain route assignments, which to the best of our knowledge, was not an approach used on other cluster first-route second papers.

The bin packing problem (BPP) has been vastly studied in the literature over the years (Munien & Ezugwu, (2021), including in vehicle routing applications. However, in most of these applications the BPP is used to model the loading of the trucks, rather than the assignment of stores to routes, as we propose in our solution method. Hamdi-Dhaoui *et al.* (2014) presented a variant of the VRP that incorporates the loading of the truck as part of the decisions with the objective of both minimizing travel cost and balancing the loads across the fleet. The loading component is defined as a two-dimensional BPP and a genetic algorithm is used to find a solution. Heßler *et al.* (2022) disregards the routing component and focuses on the problem of loading trucks constrained by both weight and volume in a direct delivery setting. With the objective of minimizing the fleet size, the authors present a heuristic solution method to solve the problem.

Another variant of the BPP considers the interaction between items assigned to the same bin. This is relevant to our study, since when multiple clusters are assigned to a route, the additional distance between the two clusters needs to be considered by the model. Li *et al.* (2014) discusses the two-dimensional BPP with conflict penalties, where items may incur in additional cost when assigned to the same bin. The authors define the problem with the objective of minimizing the total number of bins used and the conflict penalties incurred, and present an algorithm as a solution method. A one-dimensional version of this problem was also introduced by Khanafer *et al.* (2012) where column-generation and tabu search were used to find a solution.

In these studies, the interaction between items in a bin are considered as conflict that need to be avoided. In our approach, the interaction between clusters can be inevitable and thus it was necessary that the additional distance between clusters was considered rather than actively discouraged. Our formulation of the SRP as a BPP includes a constraint that captures the additional distance traveled between pairs of clusters assigned to the same bin. This distance is then included as part of the cost component associated with distance traveled in the objective function. Additionally, by using a clustering heuristic to reduce the number of nodes in the network, our formulation of the BPP can be solved directly using a commercial solver like CPLEX for even large real-world instance of 300 stores in practical time. In the following section we further discuss the various components of our proposed solution method.

3. Solution Method

As described in Section 1, the core idea in our study is to create clusters of stores to simplify the decision-making process for the SRP. By creating such clusters, we can formulate the SRP as a bin-packing problem, where the bins represent the routes to be completed by a single truck. The goal of the bin-packing formulation is to assign the clusters to routes in order to supply the necessary inventory to each store in the cluster while trying to minimize the logistics costs associated with the distance traveled, delivery labor, truck fleet size, and route time overages. The proposed solution procedure is composed of three main components. The first component groups the stores in the network into clusters. The resulting clusters are used as input for the bin-packing formulation of the SRP, which represents the second component. Lastly the solution of the bin-packing problem is used to create the necessary routes. In the following sections we discuss the above components in more detail.

3.1 The Clustering Heuristic

To create and adjust the initial set of clusters for our model, we adapted elements from the well-known K-means clustering heuristic (MacQueen, 1967). The K-means clustering heuristic initially builds K clusters, whose centroids are then adjusted and each store is (re)assigned to the nearest centroid. Our clustering heuristic uses a similar iterative approach and incorporates components that more directly fit the needs of the SRP. The steps of the clustering heuristic are described below:

Step 1. Identify the stores that are visited in period t and randomly select a store to serve as the centroid for a cluster.

Step 2. Build the cluster by assigning stores, one at a time, nearest to the selected centroid that meet the following criteria:

- The store is not already assigned to another cluster.
- Number of stores assigned to the cluster is less than the maximum allowed (M).
- Distance from the store to the centroid is less than the maximum distance allowed (R).
- Estimated time to visit all the stores in the cluster does not exceed the maximum route time allowed (U_T).

Step 3. If no more stores can be added to the current cluster, randomly select another unassigned store as the centroid for a new cluster and repeat Step 2. If all the stores have been assigned to a cluster, go to Step 4.

Step 4. Calculate new centroids for all the clusters based on the locations of the assigned stores. For each cluster, assign the store nearest to the calculated centroid as the cluster's new centroid.

Step 5. If the new centroids are the same as the previous ones, or the distance between the new and previous centroids is less than a user-specified threshold (D), then finalize the clusters and go to Step 6. Otherwise, restart the store assignment process using the new centroids and following the criteria in Step 2.

Step 6. For each cluster, calculate the average distance between the stores assigned to the cluster and the cluster centroid.

Step 7. For each cluster i, find the nearest store to the centroid of cluster i that is assigned to another cluster j. Verify if it is feasible to reassign the store from cluster j to i using the following criteria:

- The number of stores in cluster i is not over the maximum allowed (M).
- Distance between the store and centroid i does not exceed the maximum allowed (R).
- The additional estimated time to visit the stores in cluster i does not exceed maximum route time (U_T).
- The sum of the average cluster distance (between the stores and the centroid) is reduced by the exchange.

If the exchange is feasible, reassign the store and repeat Step 7 for the next cluster until all clusters have been checked; otherwise go to Step 8.

Step 8. Find the furthest store from the centroid in cluster i and verify if a two-store exchange between another cluster j is feasible using the same criteria outlined in Step 7 for both clusters. If the exchange is feasible, reassign the stores and repeat Step 7 for the next cluster until all the clusters have been checked.

Step 9. After all clusters have been checked, calculate the new centroids for each cluster whose stores were exchanged. Assign the store nearest to the calculated centroid as the cluster's new centroid. STOP if any of the following criteria are met:

- The centroids did not change.
- The distance between the new and previous centroids is less than D.
- N iterations of the heuristic have been completed.

Otherwise, start a new iteration from Step 2 using the new centroids.

The clusters obtained from the above heuristic are used as input for the bin packing model described next.

3.2 The Bin Packing Problem Formulation

The bin packing problem (BPP), which has been studied extensively in the past, is concerned with partitioning a given set of items with particular weights into the minimum number of (identical) bins without exceeding the capacity of the bins (Munien & Ezugwu, 2021). The BPP has been used successfully in a variety of applications ranging from scheduling, resource allocation, and logistics in areas such as production and health care; for the details, please see the survey by Munien & Ezugwu (2021). We formulate the SRP as a BPP to assign clusters of stores to routes to be completed by a single truck. Under this approach, the clusters and the store demands serve as the weighted items, while the routes, that are constrained by capacity and maximum route time, represent the bins. In order to build this formulation, we made the following assumptions:

1. The delivery schedule is given and fixed over the planning horizon.
2. Each store (or franchisee) owns and manages their own inventory.

3. Cluster demand includes store demand for each store assigned to the cluster.
4. A “bin” represents a route to be completed by single truck employing a single driver or a team.
5. Usable truck capacity and maximum route-time is defined by the type of route it completes (team or single-driver).
6. Trucks can only complete one route per period.
7. All the trucks and drivers are available at the start of each period.
8. No cluster of stores will exceed truck capacity or the maximum number of stops allowed on a route.
9. A cluster can only be assigned to one route in each period. No split deliveries to are allowed.
10. Store orders and truck capacity are measured in pounds. Demand is represented as a single aggregate SKU.

The above assumptions are consistent with the SRP formulation introduced by Vigo Camargo & Bozer (2022). The key differences are introduced in assumptions 3, 4 and 8, which consider the clusters and the bin packing approach used in this study. Assumption 3 helps define the input data the model uses by defining how store demand is translated to the clusters. The fourth assumption defines how the “bins” are translated to the SRP in the form of routes to be completed. Lastly, with assumption 8, we ensure the feasibility of the problem and maintain that, at minimum, a cluster can always be assigned to a route.

We use the following notation to build the bin packing formulation of the SRP:

N, R, T, L	Sets for clusters, routes, time period, and route type.
U_V^l	Usable truck capacity for route type l.

U_T^l Max. route time before overage cost is incurred for route type *l*.

C_V Cost per time period of having a truck available for use.

C_m^l Cost per mile traveled on a route type *l*.

C_{lb}^l Cost per pound of inventory delivered on a route type *l*.

C_U Overage cost incurred by a route that exceeds the max. route time.

B_m Time per mile traveled.

B_{lb} Time per pound delivered.

B_n Time per stop in a route.

V Maximum number of stops allowed in a route.

m_i Average distance in cluster *i*.

d_i Store demand of the stores in cluster *i*.

n_i Number of stores in cluster *i*.

m_i^{DC} Distance from centroid of cluster *i* to the DC.

m_{ij}^C Distance between cluster *i* and cluster *j*.

The decision variables for the model are defined as follows:

x_{ir}^l 1 if cluster *i* is assigned to truck *r* of route type *l*; 0 otherwise.

y_r^l 1 if truck *r* with route type *l* is used.

z_r^l 1 if truck *r* with route type *l* exceeds maximum route time; 0 otherwise.

P_r Maximum distance between clusters assigned to *r*.

K Fleet size.

The mathematical formulation is presented below:

$$min\, S_V\, T\, K + \sum_{i,r,l} x_{ir}^l\, m_i\, n_i\, C_m^l + \sum_{r,l} P_r \frac{C_m^l}{2} + \sum_{i,r,l} x_{ir}^l\, d_i\, S_{lb} + \sum_{r,l} z_r^l\, C_U \tag{1}$$

$$s.t.\ \sum_{r,l} x_{ir}^l\, U_V \geq n_i \qquad \forall\, i \in N \tag{2}$$

$$\sum_i x_{ir}^l\, d_i \leq U_T^l\, y_r^l \qquad \forall\, r \in R,\ l \in L \tag{3}$$

$$\sum_i x_{ir}^l\, n_i \leq U_V y_r^l \qquad \forall\, r \in R,\ l \in L \tag{4}$$

$$\sum_{r,\ l} x_{ir}^l \leq 1 \qquad \forall i \in N \tag{5}$$

$$\sum_l y_r^l \leq 1 \qquad \forall\, r \in R \tag{6}$$

$$\sum_l y_r^l \leq \sum_l y_{r-1}^l \qquad \forall\, r > 1 \tag{7}$$

$$\sum_{s\geq r,\ l} x_{is}^l \leq \sum_{p\leq i,\ l} x_{pr-1}^l \qquad \forall\, i > 1,\ r > 1 \tag{8}$$

$$\sum_{r,\ l} y_r^l \leq K \tag{9}$$

$$\begin{aligned}\sum_i x_{ir}^l\, d_i\, B_{lb} + \sum_i x_{ir}^l\, n_i\, B_n + P_r\, B_m \\ + \sum_i x_{ir}^l\, m_i\, n_i\, B_m \leq U_T^l(1 + z_r^l)\end{aligned} \qquad \forall i \in N \tag{10}$$

$$\begin{aligned}\sum_{j,\ l} x_{jr}^l\, (m_{ij} + \left(\frac{m_i^{DC} + m_j^{DC}}{2}\right) \\ \leq P_r + \left(1 - \sum_l x_{ir}^l\right) M\end{aligned} \qquad \forall\, i \in N,\ r \in R \tag{11}$$

$$x_{ir}^l \geq 0 \text{ (integer)} \qquad \forall\, i \in N,\, r \in R, l \in L \qquad (12)$$

$$y_r^l \geq 0 \text{ (integer)} \qquad \forall\, r \in R,\; l \in L \qquad (13)$$

$$z_r^l \geq 0 \text{ (integer)} \qquad \forall\, r \in R,\; l \in L \qquad (14)$$

$$P_r \geq 0 \qquad \forall\, r \in R \qquad (15)$$

$$K \geq 0 \text{ (integer)} \qquad (16)$$

Equation (1) defines the four main cost components of the objective function of the SRP; that is, fleet cost, mileage cost due to distance traveled, labor cost of delivery based on the pounds of inventory to unload, and the overage cost incurred by routes that exceed the route time limit. Constraints (2) ensure that all the stores are visited, while constraints (3) ensure that the truck capacity is not exceeded. Similarly, constraints (4) ensure that no route assignment exceeds the maximum number of stops allowed in a route. With constraints (5) we ensure that no split deliveries are allowed. Constraints (6) define whether a route is assigned to a single driver or a team. Constraints (7) and (8) are used as symmetry-breaking constraints to simplify the assignment process and ensure that the model can find a solution in reasonable time. The fleet size is defined via constraint (9). Constraint (10) is used to calculate the route time incorporating the labor time associated with unloading inventory, a fixed stopping time for each stop, and the estimated travel time. With constraints (11) the model captures the largest distance between a pair of clusters assigned to the same route to incorporate the distance to traveled between clusters in the same route. Lastly, constraints (12) – (16) define the variables and establish non-negativity.

The solution obtained from the bin packing formulation represents an assignment of clusters to routes. However, this assignment is not structured as an actual route, and only gives us an estimate of what the associated cost would be. We use this solution as input for the last step in our solution procedure where the actual routes are built and evaluated before a final solution is obtained. The heuristic used to build the routes is discussed in the following section.

3.3 The Route Building Heuristic

The final step is focused on building the delivery routes based on the assignments obtained from the bin packing formulation. We start by solving a traveling salesman problem (TSP) for each route assignment in the bin packing solution, using the Google's OR-Tools Solver (Google, 2019). These initial routes are then used as input for the "adjusted team savings" algorithm presented in Vigo Camargo & Bozer (2022b). This second step ensures that none of the routes built by the TSP violate the route time limits and decides if any of the routes should be changed between single-driver and team routes. To finalize the routes, we perform a 2-opt exchange over all the constructed routes to evaluate alternative sequencing of the routes and remove any significant crossings within routes that were impacting the total distance traveled. Figure 4-1 summarizes the three steps used to build and finalize the routes for the proposed solution procedure.

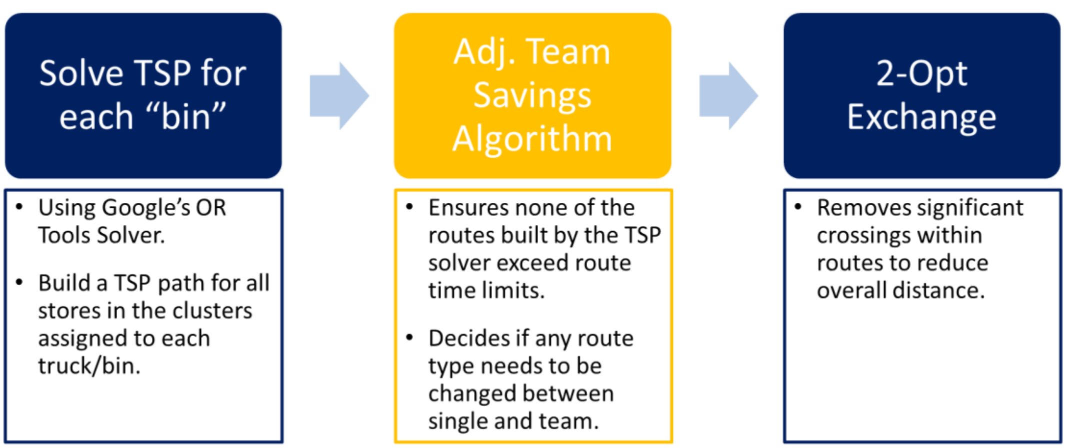


Figure 4-1. The Route Building Heuristic

4. Computational Results

The proposed clustering-based heuristic was coded in Python 3.7.4, and all the problem instances were solved on a Windows 10 computer with a 64-bit 2.50 GHz Intel Core i7 processor and 8GB of RAM. Data from our industry collaborator were used to build two industry-data instances and a baseline to evaluate the performance of the clustering-based heuristic. Additional instances with randomly generated demand data and store locations were also used to study the performance of the heuristic under different network conditions, which we discuss in section 5.2.

4.1 Industry Data

The data set from our industry collaborator is based on a network of 181 stores served by the same DC. It contains the store orders that must be delivered through the planning period, which represents a week. To build the distance matrix, we used the Google Maps API (Google, 2020) and determined the actual road distances between each store and the DC. Truck capacity was set to reflect the actual operations of our industry collaborator, with team routes allowed to carry 35% more inventory than single-driver routes on the same type of trucks. Similarly, route-time limits were set, beyond which an overage cost is incurred, with team routes having a limit that is 30% higher. The parameters used for the cost components in the objective function were also provided by our industry collaborator. (Actual values of the truck capacity, the route-time limits, and the cost parameters are not reported to protect data confidentiality.) While we cannot disclose actual cost values, we can report their relative values. The cost incurred per mile for team routes is 20% higher than single-driver routes, and the cost per pound delivered (i.e., the labor cost) is 35% less for single-driver routes than team routes. If a route exceeds the specified route-time limit, it incurs an overage cost set equal to the daily cost of having to provide an additional truck.

Figure 4-2 shows the locations of the 181 stores and the DC. Given the nature of the fast-food industry, the map clearly shows how some stores form clusters around high-population centers, while other stores are located at distant locations and/or generally isolated. As we discussed earlier, this type of clustering among real-world store locations is the primary motivation behind the development of the clustering-based heuristic we present in this paper. By exploiting the clusters in the network, the clustering-based heuristic we propose for the SRP simplifies the problem and it allows for a solution to be developed using a simple bin-packing approach.

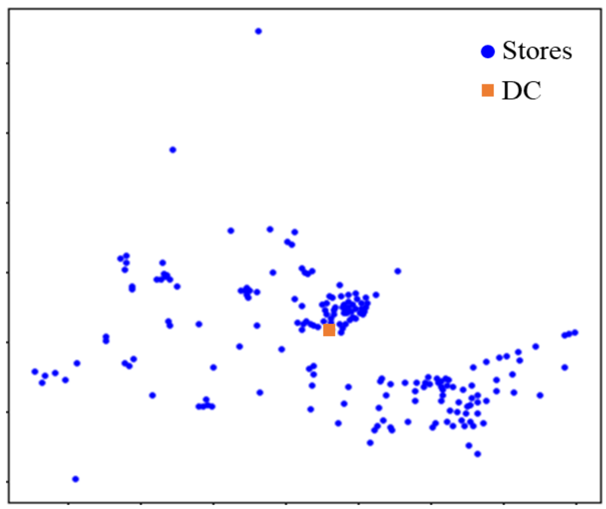


Figure 4-2. Map of Store Locations for our Industry Collaborator

The first step of the heuristic identifies the clusters of stores that have to be visited in each time period. (Not each store is visited in each time period.) Figure 4-3 shows the resulting clusters that were identified for each time period in problem instance 1. Each cluster is colored, and the centroid of each cluster is identified by the store number. For each time period, the routes are built by assigning clusters to trucks using the bin packing model (described in Section 3). We then employ a route improvement procedure (to ensure that no constraints are violated) and a 2-opt exchange procedure (to reduce the distance on each route). Figure 4-4 shows the resulting routes built for each time period in problem instance 1.

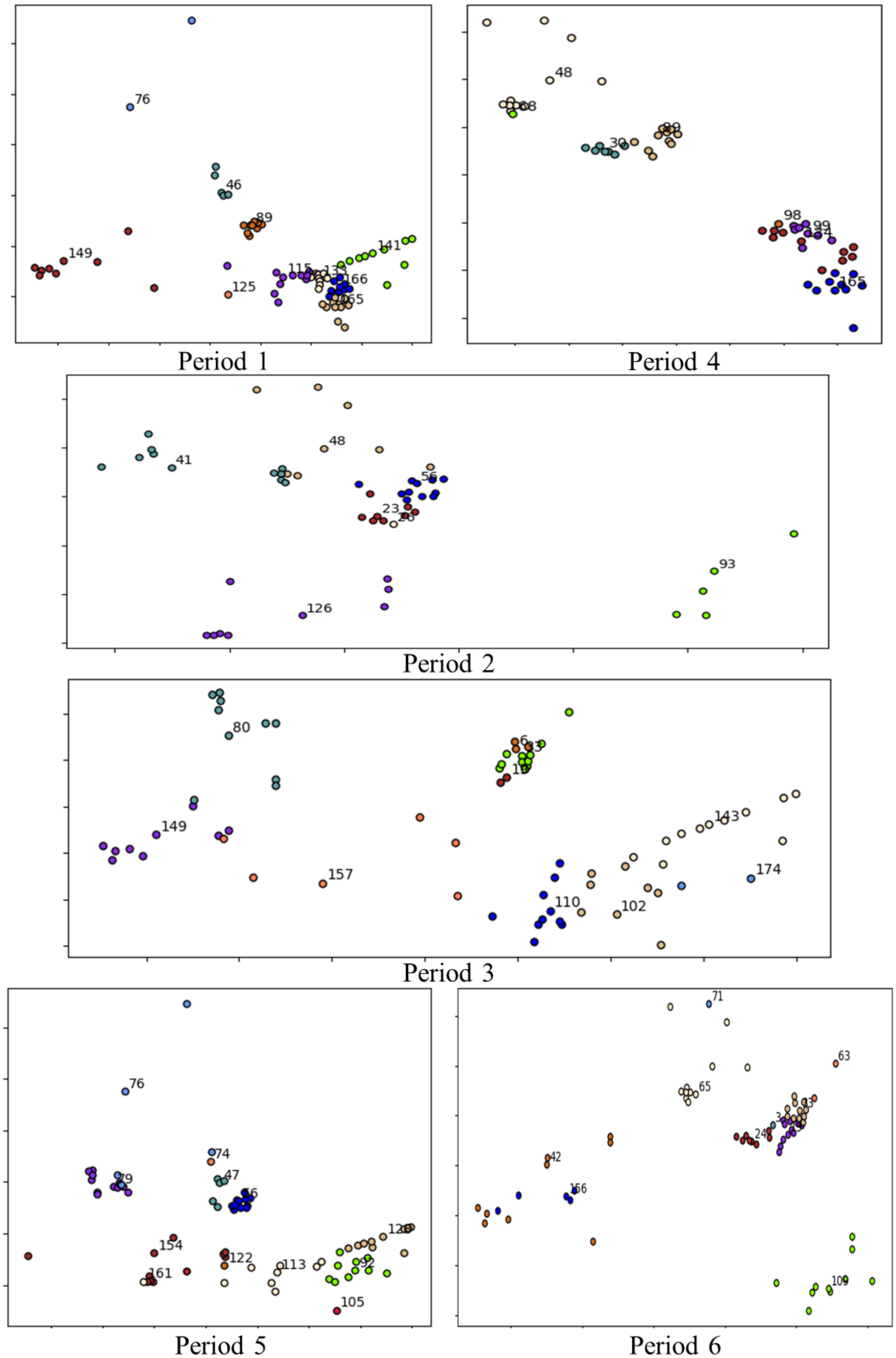


Figure 4-3. Clusters Obtained with Industry Data Instance 1

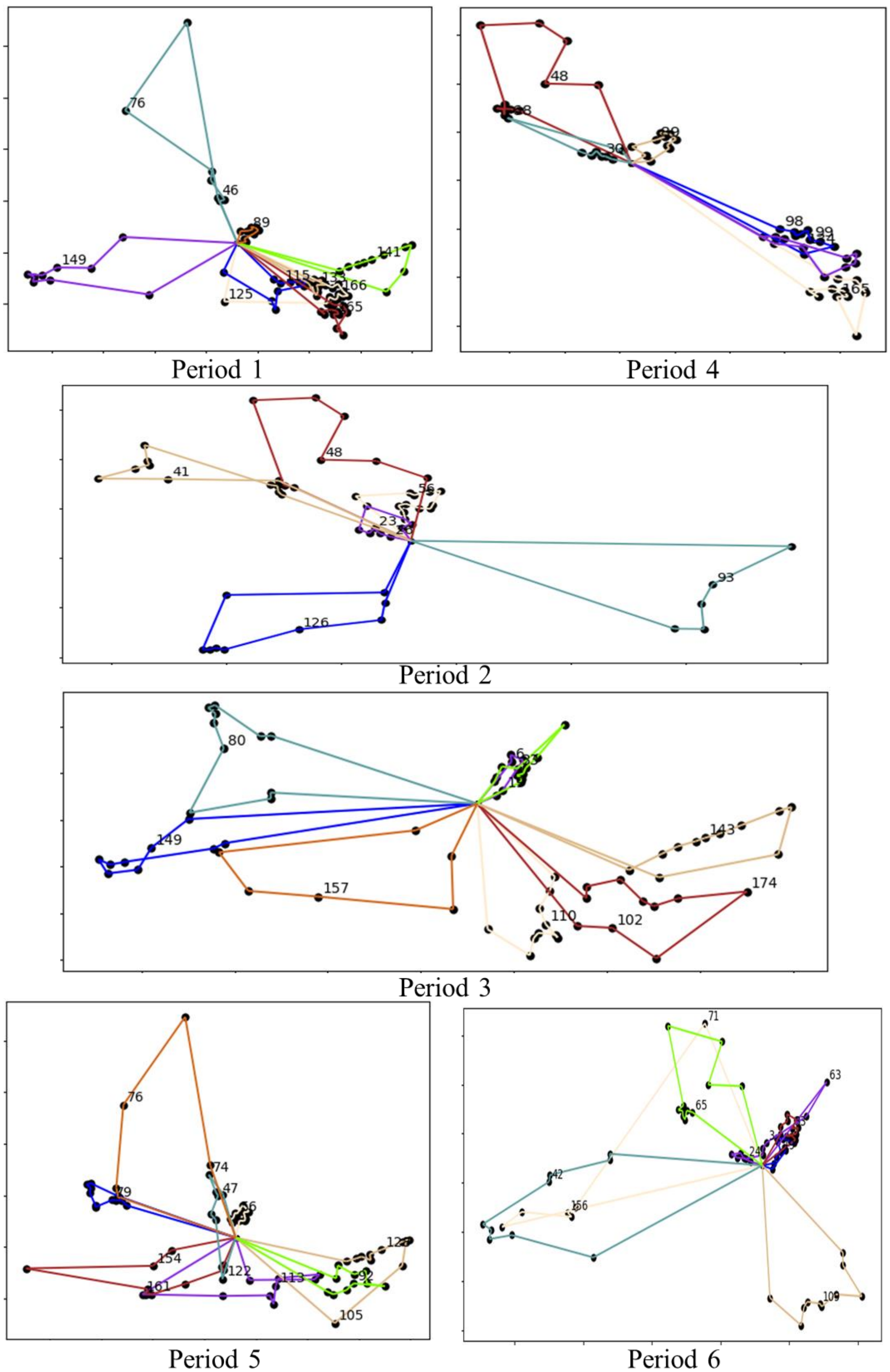


Figure 4-4. Routes Resulting from Industry Data Instance 1

To assess the results obtained with the clustering-based heuristic, we used the routes and delivery plans from our industry collaborator to build an industry baseline. As an additional comparison, we also used the SRP results from Vigo Camargo & Bozer (2022), who employed a set partitioning (SP) model to solve the SRP with the same industry data used in this paper and a computer with the same specifications given above. Table 4-11 shows a summary of the results obtained from the proposed clustering-based heuristic, the industry baseline, and the SP-SRP results for both problem instances.

Table 4-1. Clustering Heuristic vs the Industry Baseline vs SP-SRP

	Instance 1			Instance 2		
	Industry	SP-SRP	Clustering	Industry	SP-SRP	Clustering
Total Cost	$54,283.48	$51,946.13	$46,787.48	$56,284.60	$49,399.04	$48,413.55
Miles Cost	$25,661.31	$23,936.69	$20,689.06	$27,191.51	$22,129.20	$21,449.64
Del Cost	$15,585.85	$13,936.15	$14,099.07	$16,353.05	$13,789.10	$14,964.57
Overage Cost	$4,147.92	$4,296.06	$4,888.62	$2,962.80	$3,703.50	$4,888.62
Truck Cost	$8,888.40	$9,777.24	$7,110.72	$9,777.24	$9,777.24	$7,110.72
Runtime (sec)	-	991.80	115.15	-	393	60.35

For both problem instances, the proposed clustering-based heuristic yields a solution that is lower in total cost than both the industry baseline and the SP-SRP cost, and it does so with significantly shorter runtimes compared to the original SP-SRP solution method. For problem instance 1 (problem instance 2), the clustering-based heuristic obtains a 14% (14%) reduction in total cost compared to the industry baseline, which is also 10% (2%) less than the SP-SRP cost. In both instances, the clustering-based heuristic obtains a solution over 80% faster than the original SP-SRP method used by Vigo Camargo & Bozer (2022). The above results show numerically that the proposed clustering-based heuristic, and the bin packing approach it utilizes, simplifies the problem, which allows it to obtain lower-cost solutions in significantly shorter runtimes.

The solution found by the clustering-based heuristic exhibits an increase in the number of routes that exceed the route-time limit and thus incur an overage cost, with instance 1 increasing by 14%, and instance 2 by 32%. However, the increase in overage cost is balanced out by a 10% reduction in total miles traveled, and a more than 25% reduction in total number of trucks used in both instances. Also, the number of team routes used decreases significantly in the clustering-based solution compared to the industry baseline; the decrease is 86% and 68%, respectively, for problem instances 1 and 2. This is consistent with previous results presented by Vigo Camargo & Bozer (2022), who discussed the potential over-reliance on team routes in the fast-food industry.

4.2 Randomly Generated Problem Instances

To further study the performance of the clustering-based heuristic, we built a series of problem instances based on randomly generated data, using the industry data as a template. These instances are based on 100 stores with orders placed over a planning horizon of a week. We wanted to ensure that clusters, similar to those that appear in real-world scenarios, were included in the network of randomly generated instances. To achieve this, we varied the number of clusters, the size of the clusters, and the maximum distance allowed from the centroid of a cluster, for a total of 27 instances. We generated two different demand data instances and solved all 27 instances using both demand data, for a total of 54 instances.

For the number of clusters, we used 3, 5 and 10. Then we allowed the size of each cluster to be randomly generated with a minimum of three stores and a maximum that varied between 5, 11 and 22. Lastly, for the maximum distance allowed on each cluster, we used the average distance between all the stores in the industry data as our baseline average, and then also included half and double that value as additional parameters. The centroid of each cluster was randomly placed in a 600×600 grid, and the stores within each cluster were placed randomly around the centroid.

Additional stores were randomly placed in the grid, with the DC placed at the center. The distance matrix was generated assuming Euclidean distances between the stores and the DC. Store demand was randomly generated using a normal distribution with an average of 2,030 lbs and a standard deviation of 760.15, based on the demand values from the industry data. Order frequency was also generated using a normal distribution with an average of 2 and a standard deviation of 0.60.

Table 4-2. Clustering Heuristic Performance Comparison

No. of Clust.	Max. Size of Clust.	Max. Dist.	% Diff Soln	% Diff Runtime
3	5	0.5x	-15.5%	-76.7%
		1x	-15.0%	-80.9%
		2x	-13.0%	-74.4%
	11	0.5x	-21.7%	-76.8%
		1x	-16.6%	-70.2%
		2x	-12.0%	-75.0%
	22	0.5x	-6.7%	-85.4%
		1x	-7.9%	-84.8%
		2x	-7.4%	-77.7%
5	5	0.5x	2.3%	-76.3%
		1x	-1.1%	-82.2%
		2x	1.0%	-81.9%
	11	0.5x	-12.7%	-79.5%
		1x	-11.6%	-83.5%
		2x	-10.1%	-73.0%
	22	0.5x	-36.4%	-88.4%
		1x	-23.2%	-88.8%
		2x	-13.8%	-88.5%
10	5	0.5x	-10.0%	-71.1%
		1x	-12.2%	-76.7%
		2x	-11.8%	-76.1%
	11	0.5x	-12.9%	-82.3%
		1x	-15.1%	-79.0%
		2x	-8.8%	-82.6%
	22	0.5x	-8.6%	-88.2%
		1x	-4.6%	-91.1%
		2x	-2.4%	-91.3%

We solved each instance three times using the clustering-based heuristic to obtain an average solution. As a comparison, we also solved each instance using the original SP-SRP method presented by Vigo Camargo & Bozer (2022) and report the average percent difference in total cost and runtime for each instance in Table 2. The actual cost solutions of both methods for all 54 instances are shown in the Appendix.

The results show that the clustering-based heuristic obtains a lower-cost solution for almost every instance, with an average cost reduction of 11%. (Only in the set of instances with 5 clusters of at most 5 stores, the clustering-based heuristic was unable to improve on the baseline with an average solution 2% over the SP-SRP solution.) Furthermore, the clustering-based heuristic obtains solutions significantly faster, with an average runtime 80% faster than that of the original SP-SRP method. Closer examination of the results indicates that, as expected, the clustering-based heuristic performs better with tighter clusters (0.5x), with an average reduction of 13% in total cost for tight clusters in contrast to an 8% reduction for clusters with a larger distance allowed (2x).

Similarly, when larger clusters are allowed (maximum of 22 stores per cluster), the clustering-based heuristic obtains solutions with a 12% cost reduction on average, and for problem instances with a maximum of 5 stores per cluster, it obtains solutions with an 8% cost reduction relative to the original SP-SRP method. Combined with the results obtained from the industry data, our computational results show that the clustering-based heuristic is most efficient when large, dense clusters of stores are present in the network. Such a result is particularly relevant for the proposed heuristic since in many fast-food networks, there is a tendency for clusters to appear in and around populous areas such as city centers due to large customer demand in such areas.

To further explore the performance of the proposed heuristic under real-world scenarios, we built four additional test instances using the actual store locations of our industry collaborator in four different service areas. These four instances are similar in size to the baseline industry instances we used earlier, with the number of stores ranging from 130 to 350. Figure 5 shows the store locations for each test instance. We generated random demand data for each of these instances using the same method discussed above. As a comparison, we once again solved all four instances with the original SP-SRP method as well as with the proposed clustering-based heuristic. Table 4-3 summarizes the results.

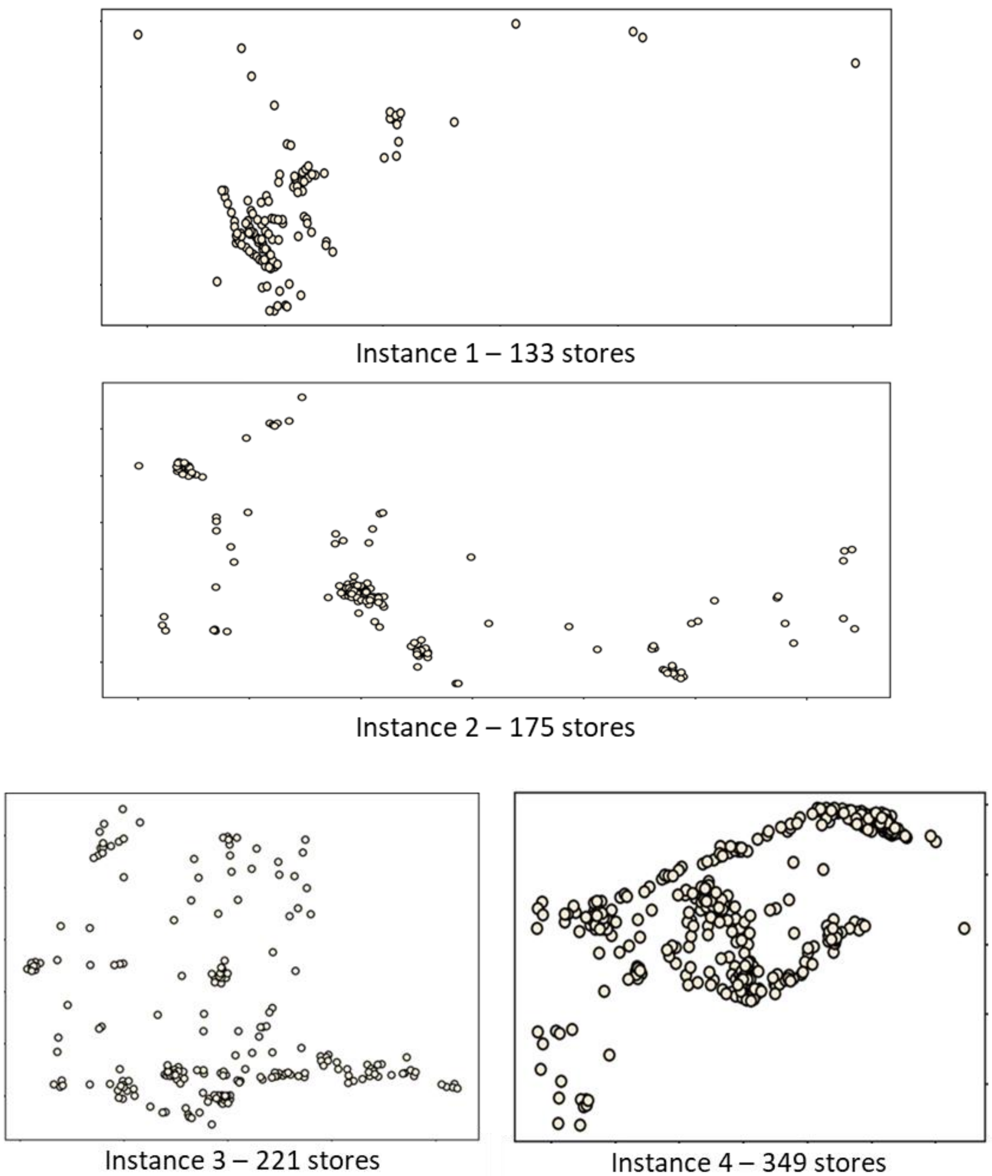


Figure 4-5. Store Locations for Industry Test Instances

Table 4-3. Results with Test Instances Based on Real-World Locations

Inst.	No. of Stores	Avg Clustering Solution	Runtime (Secs)	SP-SRP Solution	Runtime (Secs)	% Diff Soln	% Diff Runtime
1	133	$45,444.63	148.64	$51,849.75	825.31	-12.4%	-82.0%
2	221	$70,829.67	65.95	$120,922.80	1,008.18	-41.4%	-93.5%
3	175	$79,691.41	74.98	$80,112.50	1,725.72	-0.5%	-95.7%
4	349	$101,512.94	149.53	$103,279.76	5,208.17	-1.7%	-97.1%

The additional test results further support the fact that the clustering-based heuristic outperforms the original SP-SRP method, and it does so for all four test instances, based on real-world store locations. The range observed for the percentage differences in the solution values can be attributed to the differences in the composition of the networks. Instance 2, which includes the densest clusters across all four instances (shown in Figure 4-5), yields the highest percentage difference. This is aligned with the results obtained with the randomly generated instances, where tighter clusters yield better results. Furthermore, with an average runtime of 1.6 minutes across all four instances, and an objective function that captures the relevant cost elements, the clustering-based heuristic is a practical tool for decision-makers and analysts who can make a large number of what-if runs to address logistics-related and/or store location-related planning problems.

5. Conclusions

We present a new solution method for the SRP that exploits the clusters found in the store locations of fast-food networks. The proposed heuristic uses a bin-packing formulation to simplify the decision space and find good solutions in short times. Using data obtained from our industry collaborator in the fast-food sector and additional randomly-generated test instances, we evaluated the performance of the proposed heuristic against the original solution method introduced by Vigo

Camargo & Bozer (2022).

The computational results show that the clustering-based heuristic outperforms the original SP-SRP method on the industry data instances with an average cost reduction of 6%. Additional results, using randomly-generated instances, also show that the proposed heuristic can find better solutions than the original SP-SRP with an average cost reduction of 12.5% across all 58 test instances. These results include four test instances with real-world store locations from our industry collaborator. Overall, the proposed heuristic is able to find solutions with an average runtime of 1.60 minutes, which represents a significant reduction of 86%, on average, from the previous method.

The solution method introduced in this study can serve as a planning tool for decision-makers in the logistics field. Considering its short runtime even with a fairly large number of stores, it can be a practical tool to perform what-if analyses and provide critical insights into future store locations and the impact of network composition on logistic challenges. This study focused on the original SRP, which assumes a fixed delivery schedule is given. Future research can incorporate the clustering approach to the flexible SRP, where the delivery schedule is part of the decision space. The short runtimes can be especially valuable in such cases since the decision space is considerably larger for the flexible SRP.

References

Bodin, L. D. (1975) ‘A taxonomic structure for vehicle routing and scheduling problems’, *Computers and Urban Society*, 1(1), pp. 11–29. doi: 10.1016/0305-7097(75)90003-4.

Bramel, J. and Simchi-Levi, D. (1995) 'A Location based Heuristic for General Routing Problems', *Operations Research*, 43(4), pp. 649-660. doi: 10.1287/opre.43.4.649

Cömert, S. E. *et al.* (2017) ‘A new approach for solution of vehicle routing problem with hard time window: an application in a supermarket chain’, *Sadhana - Academy Proceedings in Engineering Sciences*, 42(12), pp. 2067–2080. doi: 10.1007/s12046-017-0754-1.

Dondo, R. and Cerdá, J. (2007) ‘A cluster-based optimization approach for the multi-depot heterogeneous fleet vehicle routing problem with time windows’, *European Journal of Operational Research*, 176(3), pp. 1478–1507. doi: 10.1016/j.ejor.2004.07.077.

Fachini, R. F. and Armentano, V. A. (2020) ‘Logic-based Benders decomposition for the heterogeneous fixed fleet vehicle routing problem with time windows’, *Computers and Industrial Engineering*, 148(October 2019). doi: 10.1016/j.cie.2020.106641.

Ghoseiri, K. and Ghannadpour, S. F. (2010) ‘A hybrid genetic algorithm for multi-depot homogenous locomotive assignment with time windows’, *Applied Soft Computing Journal*, 10(1), pp. 53–65. doi: 10.1016/j.asoc.2009.06.004.

Google (2019) *Google OR-Tools*, *Google*. Available at: https://developers.google.com/optimization.

Google (2020) *Google Maps Platform*, *Google*. Available at: https://cloud.google.com/maps-platform/.

Hamdi-Dhaoui, K., Labadie, N. and Yalaoui, A. (2014) ‘The bi-objective two-dimensional loading vehicle routing problem with partial conflicts’, *International Journal of Production Research*. Taylor & Francis, 52(19), pp. 5565–5582. doi: 10.1080/00207543.2014.885181.

Heßler, K. *et al.* (2022) ‘Bin packing with lexicographic objectives for loading weight- and volume-constrained trucks in a direct-shipping system’, *OR Spectrum*. Springer Berlin Heidelberg, 44(2), pp. 1–43. doi: 10.1007/s00291-021-00628-x.

Hiquebran, D. T. *et al.* (1993) *A revised simulated annealing and cluster-first route-second algorithm applied to the vehicle routing problem, Engineering Optimization*. doi: 10.1080/03052159308941327.

Kao, Y. and Chen, M. (2013) ‘Solving the CVRP problem using a hybrid PSO approach’, in *Computational Intelligence*. Springer, pp. 59–67.

Khanafer, A. *et al.* (2012) ‘The min-conflict packing problem’, *Computers and Operations Research*, 39(9), pp. 2122–2132. doi: 10.1016/j.cor.2011.10.021.

Kim, B. I., Kim, S. and Sahoo, S. (2006) ‘Waste collection vehicle routing problem with time windows’, *Computers and Operations Research*, 33(12), pp. 3624–3642. doi: 10.1016/j.cor.2005.02.045.

Li, K. *et al.* (2014) ‘A two-dimensional bin-packing problem with conflict penalties’, *International Journal of Production Research*. Taylor & Francis, 52(24), pp. 7223–7238. doi: 10.1080/00207543.2014.916826.

MacQueen, J. (1967) ‘Some methods for classification and analysis of multivariate observations’, *Proceedings of the fifth Berkeley symposium on mathematical statistics and probability*, 1(14), pp. 281–297. Available at: http://books.google.de/books?hl=de&lr=&id=IC4Ku_7dBFUC&oi=fnd&pg=PA281&dq=MacQ

ueen+some+methods+for+classification&ots=nNTcK1IdoQ&sig=fHzdVcbvmYJ-lTNHu1HncmOFOkM#v=onepage&q=MacQueen some methods for classification&f=false.

Mor, A. and Speranza, M. G. (2022) ‘Vehicle routing problems over time : a survey’, *Annals of Operations Research*, pp. 1–21. doi: https://doi.org/10.1007/s10479-021-04488-0.

Munien, C. and Ezugwu, A. E. (2021) ‘Metaheuristic algorithms for one-dimensional bin-packing problems: A survey of recent advances and applications’, *Journal of Intelligent Systems*, 30(1), pp. 636–663. doi: 10.1515/jisys-2020-0117.

Qi, M. *et al.* (2012) ‘A spatiotemporal partitioning approach for large-scale vehicle routing problems with time windows’, *Transportation Research Part E: Logistics and Transportation Review*. Elsevier Ltd, 48(1), pp. 248–257. doi: 10.1016/j.tre.2011.07.001.

Sharda, R. *et al.* (2008) ‘The Vehicle Routing Problem: Latest Advances and New Challenges’. Boston, MA: Springer Science+Business Media, LLC (Operations Research/Computer Science Interfaces,1387-666X ;43). Available at: http://dx.doi.org/10.1007/978-0-387-77778-8.

Shin, K. and Han, S. (2011) ‘A Centroid-Based Heuristic Algorithm for the Capacitated Vehicle Routing Problem’, *Computing and Informatics2*, 30, pp. 721–732. doi: 10.1118/1.2241552.

Vigo Camargo, A. and Bozer, Y. A. (2022) *The Store Replenishment Problem with Flexible Delivery Schedule and Limited Store Capacities*.

Vigo Camargo, A. and Bozer, Y. A. (no date) ‘Development and Application of a Cost-Driven Decision Model for Store Replenishment Logistics in the Fast-Food Sector’, *International Journal of Logistics Systems and Management*.

Chapter 5
Conclusions

In this dissertation we introduced the Store Replenishment Problem (SRP) motivated by the logistical challenges encountered in the fast-food supply chain where multiple stores in a known region are replenished out of a distribution center (DC). In each chapter a different aspect or extension of the SRP was studied and a quantitative model was developed to find practical solutions and support the decision-making process. A summary of each chapter, including our conclusions and findings, is presented below.

In Chapter 2, we formally defined the SRP and developed a simple and effective heuristic to minimize costs associated with replenishment logistics. The model uses a cost-driven multi-component objective function that is capable of capturing the trade-offs between the cost incurred due to distance traveled, labor, fleet size, and route time overages. Using data from our industry collaborator, we were able to show that the solutions obtained with the proposed multi-component objective function outperformed one-dimensional objectives (distance-based, or fleet size only) that are commonly used for similar routing problems in the literature. The computational results also provided insights that can significantly impact managerial decisions. Our results showed that longer single-driver routes can be more cost effective than employing a larger number of team routes. Through sensitivity analyses performed on the various cost components, we were able to establish that the distance cost and the route time overage cost have the highest impact on the total cost and solution structure, respectively. This further motivates taking a closer look at how overage costs are determined and the impact that longer single-driver routes can have in the general logistics operations.

In Chapter 3, we present an extension to the SRP (that is, SRP-FS), where the delivery schedule and the amount of inventory to be delivered on each visit are included as decision variables. We formulated the SRP-FS as a MIP with subtour elimination constraints, but due to the complexity of the problem, an alternative solution method was required for large real-world instances. Using a simulated annealing structure, we developed a two-step metaheuristic (SA metaheuristic) that first finds a feasible delivery schedule and then incorporates an adjusted savings algorithm to build corresponding routes and classify them as single-driver or team routes. As part of the analysis carried out in this chapter, we also studied a simpler version of the SRP-FS where only direct deliveries (out-and-back trips) are allowed. Under the direct delivery approach, we were able to develop a heuristic policy that considers the problem as a lot sizing problem with inventory bounds and obtains near-optimal solutions. This method allowed us to easily build initial solutions for the SRP-FS that were then used by the SA metaheuristic as a starting point.

To evaluate the performance of the heuristic, we used a set of randomly generated test instances of various sizes (10, 100, and 350 stores). We solved the test instances assuming infinite truck capacity (ITC) to serve as a baseline and observed an average percentage difference of 7% between the baseline and the SRP-FS solutions. This gap is comparable to the 5% difference we observed between the solutions obtained from directly solving the MIP using CPLEX for the smaller instances and the SA metaheuristic solutions. While these results are limited by the fact that CPLEX was not able to find the optimal solutions for the test instances, the above comparison shows that the ITC baseline can serve as a lower bound to the problem.

The performance of the SA metaheuristic is further supported by the computational results obtained from solving the industry data instances. Comparing the SRP-FS solutions to both the current industry baseline and the SRP solutions obtained in Chapter 2, we were able to show how introducing a flexible delivery schedule significantly outperforms the fixed delivery alternative in all instances. While a flexible delivery schedule has the potential of reducing most of the cost elements associated with replenishment logistics, it is important to note that the implementation would require significant operational changes within the industry that would benefit from input obtained from various stakeholders across the supply chain. The cost reductions attained by using a flexible schedule are reflected in a reduction of the number of trucks required, store visits, distance traveled, and team-routes used. This further motivates the study of how team routes are currently used in industry and the impact a reduction of their use can have in the overall operations.

In Chapter 4, we introduced a new solution method for the SRP that takes advantage of the clusters found in the store networks of fast-food supply chains. By formulating the SRP as a bin-packing problem, we simplify the decision space and are able to find good solutions in short times. The bin-packing formulation assigns clusters of stores to delivery routes in order to minimize costs. These assignments are then used to build actual delivery routes using a routing heuristic. Computational results using industry data from our collaborator showed that the new method outperformed the original set-partitioning approach introduced on Chapter 2, with an average cost reduction of 6%. Across all the test instances we used to evaluate the performance of the new solution method, we found that the bin-packing approach obtains cost reductions of 12.5% on average when compared with the original approach. Furthermore, the clustering-based heuristic had an average runtime of 1.60 minutes, which represents an 86% reduction in runtime from the original method. Considering its short processing time, the method in Chapter 4 can serve as a practical planning tool for decision-makers to consider new store locations and the impact of the network composition on the replenishment logistics. While this chapter was focused on the SRP, future research can extend the proposed approach for the SRP with a flexible delivery schedule, where the simplification of the decision space can be particularly beneficial due to the more complex scenario a flexible schedule introduces to the problem.

Appendix A

Supplemental Information for Chapter 2

Truck contents and route lengths (in time) for instance 2 are shown in Figure A-1 and Figure A-2 for the SRP heuristic and the baseline, respectively.

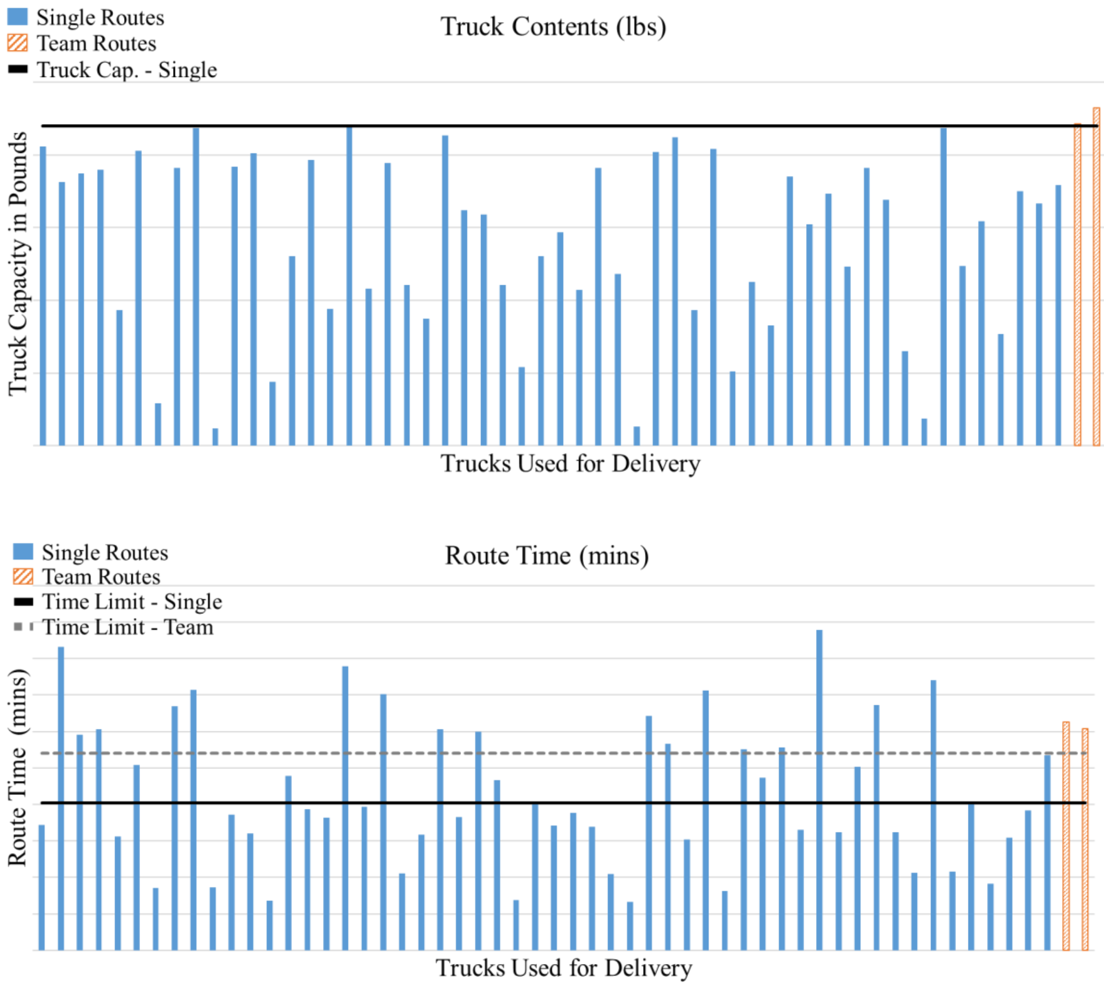
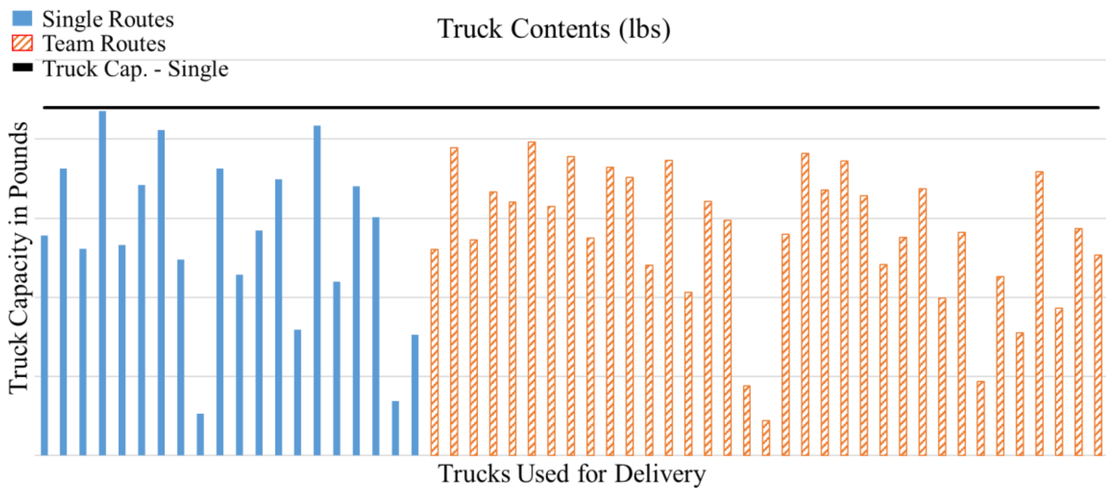


Figure A-1. Truck contents & route time, SRP heuristic-Instance 2.



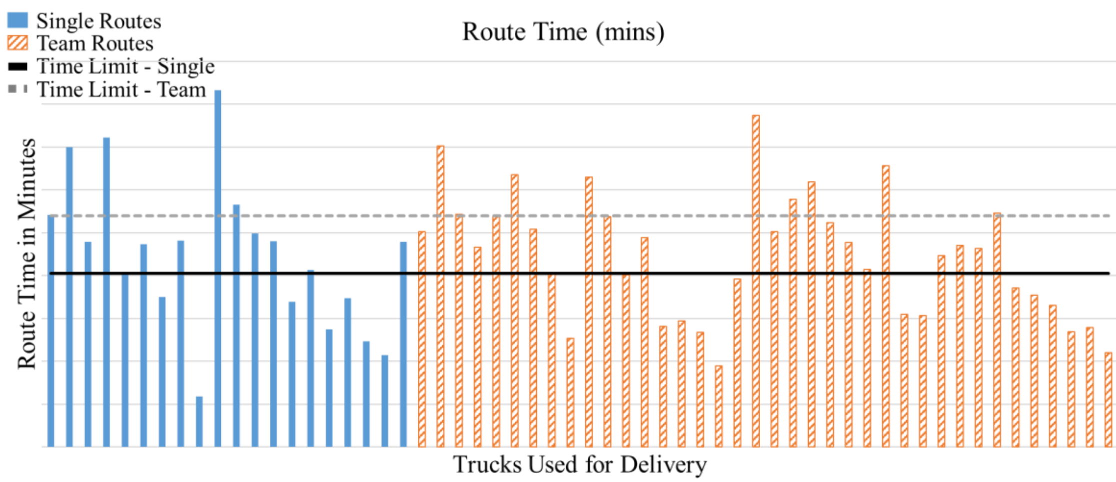


Figure A-2. Truck contents & route time, Baseline-Instance 2.

Figures A-3 – A-5 show plots for the location of stores and DC used for the N10, N100, and N350 random instances respectively.

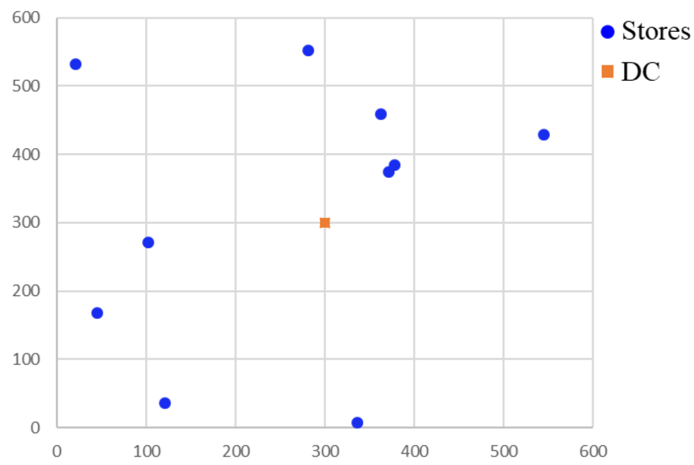


Figure A-3. Plot of Store and DC Location for N10 Instances

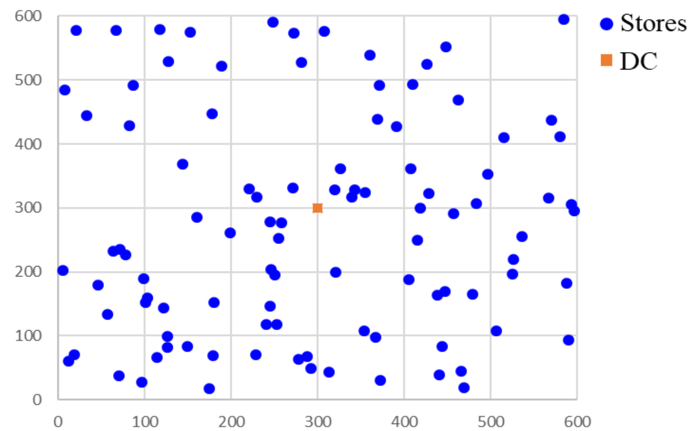


Figure A-4. Plot of Store and DC Location for N100 Instances

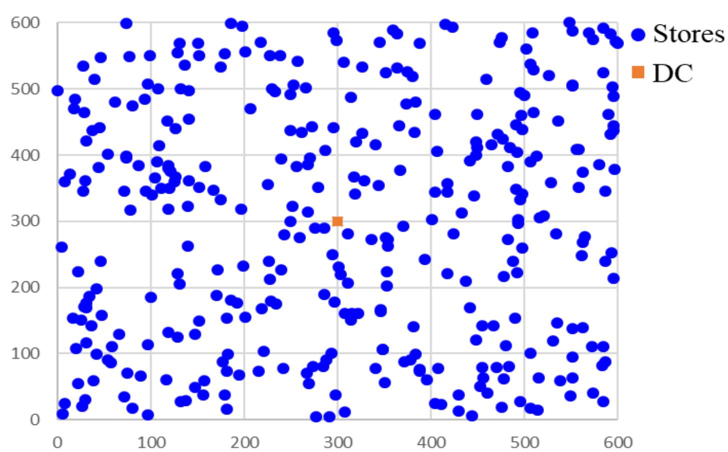


Figure A-5. Plot of Store and DC Location for N350 Instances

Appendix B

Proof of Direct Delivery Optimal Policy from Chapter 3

By contradiction, assume X^* is not optimal. Then there must exist an alternate set of positive integers, $\hat{x}_t \in A^*$, that represents an alternative delivery schedule that is feasible and has fewer total deliveries than X^*. This implies that the set $\hat{x}_t$ meets the following constraints

$$\hat{x}_i^t * U_V \geq \hat{q}_i^t \quad \forall\, i, t \tag{B1}$$

$$U_S \geq \hat{q}_i^t + \hat{I}_i^t \geq d_i^t \quad \forall\, i, t \tag{B2}$$

$$\hat{x}_i^t \geq 0 \;\forall\, i, t \tag{B3}$$

$$\hat{q}_i^t \geq 0, \hat{I}_i^t \geq 0 \tag{B4}$$

and the inequality below is true,

$$\sum \hat{x}_i^t < \sum x_i^t \,. \tag{B5}$$

Inequality B5 implies that there must be at least one period n for which the quantity of inventory delivered to the store in the alternative solution is zero, while in X^* the maximum allowed quantity was delivered. That is to say that for period *n*, we have

$$\hat{q}_i^n = 0 < q_i^n = U_S - I_i^n \,. \tag{B6}$$

WLOG we assume that there is no period where $\hat{q}_i = 0 = q_i$. That is, for no period in the planning horizon, will both solutions make the decision to not make a delivery. If such a period existed, the problem can be modified by removing the period in question without affecting the solution.

Let p be the first time period for which

$$\sum^{p} \hat{x}_i^t < \sum^{p} x_i^t \,. \qquad \textbf{\textit{(A7)}}$$

That is, let period p be the first time period in which the number of deliveries made up to that period in schedule X^* is strictly larger than the deliveries made in the alternative solution. This implies that $\hat{q}_i^p = 0$ and $q_i^p = U_S - I_i^p$. In other words, for period p no inventory was sent to the store in the alternative solution, while a delivery of $U_S - I_i^p$ was made in X^*.

With no deliveries made in period p in the alternative solution, the inventory available for that period must be sufficient to meet the demand, i.e., $\hat{I}_i^t \geq d_i^t$ must be true. Since there was a delivery in period p in solution X^*, then the amount of inventory delivered in the alternative solution must be larger than the amount delivered in solution X^* by period p; that is,

$$\sum^{p-1} \hat{q}^t > \sum^{p-1} q^t \,. \qquad \textbf{\textit{(B8)}}$$

If we examine period $p-1$, we see that there are only two feasible options:

$$\hat{q}^{p-1} \geq q^{p-1} > 0 \quad \text{or} \quad q^{p-1} > \hat{q}^{p-1} = 0 \qquad \textbf{\textit{(B9)}}$$

The above two cases imply that either in the alternative solution the amount delivered was greater than the one made in solution X^*, or that for this period the alternative solution did not make a delivery, while a delivery was made in solution X^*. These are the only two possibilities because we know that if $q^{p-1} = \hat{q}^{p-1} = 0$, we can remove the period without affecting the solution, and that if $q^{p-1} = 0$ when $\hat{q}^{p-1} > 0$, then p could not be the first period where the number of deliveries in solution X^* is larger than in the alternative solution.

With $q^{p-1} > 0$, from the proposed policy it follows that the quantity delivered must be $q^{p-1} = U_S - I^{p-1}$, and by definition of the solution, the inventory available for the selected period p is given by:

$$
\begin{aligned}
I^p &= q^{p-1} + I^{p-1} - r^{p-1} \\
&= U_S - I^{p-1} + I^{p-1} - r^{p-1} \qquad \textbf{\textit{(B10)}} \\
I_p &= U_S - r^{p-1}
\end{aligned}
$$

From the proposed policy, since $q^{p-1} > 0$, then it must mean that either the inventory available in period p is not enough to meet the period's consumption rate and thus

$$r^p > I^p = U_S - r^{p-1}, \qquad \textbf{\textit{(B11)}}$$

or the amount already delivered does not cover the expected store demand by this period:

$$\sum^{p-1} q^t < \sum^{p} d^t \qquad \textbf{\textit{(B12)}}$$

For the first scenario, we can calculate the inventory for period p in the alternative solution A^* as,

$$\hat{I}^p = \hat{q}^{p-1} + \hat{I}^{p-1} - r^{p-1} \qquad \textbf{\textit{(B13)}}$$

Since $\hat{q}^p = 0$, then we must have $\hat{I}^p \geq r^p$, and the following inequality holds:

$$\hat{I}^p = \hat{q}^{p-1} + \hat{I}^{p-1} - r^{p-1} \geq r^p \qquad \textbf{\textit{(B14)}}$$

Combining the above inequalities, we conclude that the inventory for period p in solution A^* must be greater than or equal to the consumption rate for the period, while the inventory for that same period in solution X^* is strictly less than this rate; that is:

$$
\begin{aligned}
&\hat{I}^p \geq r^p > I^p \\
&\hat{I}^p = \hat{q}^{p-1} + \hat{I}^{p-1} - r^{p-1} \geq r^p > U_S - r^{p-1} = I^p \\
&\hat{q}^{p-1} + \hat{I}^{p-1} - r^{p-1} > U_S - r^{p-1} \qquad \textbf{\textit{(B15)}} \\
&\hat{q}^{p-1} + \hat{I}^{p-1} > U_S
\end{aligned}
$$

The above inequality shows how the alternative solution A^* violates the inventory capacity constraint for period $p-1$, which means it is not a feasible solution, and thus X^* must be optimal.

Similarly, for the second scenario we have:

$$\sum^{p-1} q^t < \sum^{p} d^t \qquad \textbf{\textit{(B16)}}$$

Since the alternative solution is not making a delivery on period *p*, then:

$$\sum^{p-1} q^t < \sum^{p} d^t \leq \sum^{p-1} \hat{q}^t \qquad \textbf{\textit{(B17)}}$$

If we subtract the inventory consumption over this period, we can calculate the inventory available for the period:

$$I^p = \sum^{p-1} q^t - r^t < \sum^{p-1} \hat{q}^t - r^t = \hat{I}^p \qquad \textbf{\textit{(B18)}}$$

As we showed before, we can rewrite the inventory calculations as:

$$I^p = U_S - r^{p-1} < \hat{q}^{p-1} + \hat{I}^{p-1} - r^{p-1} = \hat{I}^p, \qquad \textbf{\textit{(B19)}}$$

which again, shows that for this to be true, the inventory of the alternative solution must violate the store capacity constraint:

$$U_S < \hat{q}^{p-1} + \hat{I}^{p-1}, \qquad \textbf{\textit{(B20)}}$$

which proves that the proposition leads to an optimal solution.

Appendix C
Supplemental Information for Chapter 4

The following Tables present detailed solutions for the 54 instances with randomly-generated data used in Section 5.2.

Table C-1. Comparison of Random Cluster Instances with Demand Instance 1.

No. of Clusters	Max. Size of Clusters	Max. Dist	Avg Clustering Solution	Runtime (Secs)	SRP Solution	Runtime (Secs)	% Diff Soln	% Diff Runtime
3	5	0.5x	$ 63,245.52	149.48	$ 77,438.20	535.35	-18.3%	-72.1%
		1x	$ 65,107.87	130.13	$ 78,308.34	580.67	-16.9%	-77.6%
		2x	$ 64,289.84	205.15	$ 74,680.07	611.55	-13.9%	-66.5%
	11	0.5x	$ 63,368.82	172.15	$ 76,476.27	608.43	-17.1%	-71.7%
		1x	$ 63,125.53	200.78	$ 76,786.59	581.62	-17.8%	-65.5%
		2x	$ 64,150.82	183.76	$ 76,277.27	630.90	-15.9%	-70.9%
	22	0.5x	$ 59,128.02	96.48	$ 65,302.83	569.40	-9.5%	-83.1%
		1x	$ 59,894.94	108.41	$ 68,010.14	604.61	-11.9%	-82.1%
		2x	$ 59,728.55	114.46	$ 67,650.34	510.07	-11.71%	-77.6%
5	5	0.5x	$ 63,953.37	201.18	$ 61,114.03	620.67	4.6%	-67.6%
		1x	$ 61,503.56	141.61	$ 61,636.76	625.75	-0.2%	-77.4%
		2x	$ 62,190.46	134.81	$ 61,336.91	629.68	1.4%	-78.6%
	11	0.5x	$ 60,109.64	143.02	$ 72,445.09	525.55	-17.0%	-72.8%
		1x	$ 60,215.11	95.49	$ 71,725.37	539.44	-16.0%	-82.3%
		2x	$ 61,306.09	170.94	$ 70,815.65	556.26	-13.4%	-69.3%
	22	0.5x	$ 57,579.63	65.96	$ 90,487.23	499.75	-36.4%	-86.8%
		1x	$ 58,367.22	58.32	$ 74,049.34	548.48	-21.2%	-89.4%
		2x	$ 58,362.48	56.82	$ 71,252.87	520.35	-18.09%	-89.1%
10	5	0.5x	$ 59,602.73	140.31	$ 63,653.50	523.82	-6.4%	-73.2%
		1x	$ 58,963.09	159.33	$ 64,938.74	580.75	-9.2%	-72.6%
		2x	$ 57,889.49	131.92	$ 66,026.54	582.70	-12.3%	-77.4%
	11	0.5x	$ 58,182.78	90.61	$ 67,101.18	514.54	-13.3%	-82.4%
		1x	$ 59,762.82	121.12	$ 72,728.37	529.77	-17.8%	-77.1%
		2x	$ 61,208.06	85.56	$ 68,178.29	572.41	-10.2%	-85.1%
	22	0.5x	$ 46,652.57	53.54	$ 51,060.34	381.94	-8.6%	-86.0%

		1x	$ 47,763.58	38.06	$ 50,629.07	414.68	-5.7%	-90.8%
		2x	$ 50,654.25	57.46	$ 54,369.43	618.30	-6.83%	-90.7%

Table C-2. Comparison of Random Cluster Instances with Demand Instance 2.

No. of Clusters	**Max. Size of Clusters**	**Max. Distance**	**Avg Clustering Solution**	**Runtime (Secs)**	**SRP Solution**	**Runtime (Secs)**	**% Diff Soln**	**% Diff Runtime**
3	5	0.5x	$ 58,732.66	87.40	67242.84	466.7845	-12.7%	-81.3%
		1x	$ 58,640.17	78.53	67463.37	499.2829	-13.1%	-84.3%
		2x	$ 58,847.03	84.82	66997.47	480.298	-12.2%	-82.3%
	11	0.5x	$ 60,484.04	85.35	82005.87	473.4953	-26.2%	-82.0%
		1x	$ 58,265.19	124.24	68832.1	495.0834	-15.4%	-74.9%
		2x	$ 59,736.65	96.80	64986.91	463.8333	-8.1%	-79.1%
	22	0.5x	$ 53,853.45	52.93	56059.05	431.3504	-3.9%	-87.7%
		1x	$ 55,438.28	54.71	57727.16	440.3037	-4.0%	-87.6%
		2x	$ 56,242.77	84.56	58065.37	380.5483	-3.14%	-77.8%
5	5	0.5x	$ 58,279.35	75.92	58291.73	505.6113	0.0%	-85.0%
		1x	$ 56,333.79	68.90	57525.41	530.1622	-2.1%	-87.0%
		2x	$ 57,167.71	74.10	56864.09	504.3538	0.5%	-85.3%
	11	0.5x	$ 56,319.70	62.51	61440.7	455.7078	-8.3%	-86.3%
		1x	$ 56,144.06	68.62	60433.2	451.3109	-7.1%	-84.8%
		2x	$ 56,813.19	110.90	60979.33	475.5009	-6.8%	-76.7%
	22	0.5x	$ 53,343.03	38.58	83798.74	383.9155	-36.3%	-90.0%
		1x	$ 53,156.43	43.68	71177.29	374.1824	-25.3%	-88.3%
		2x	$ 53,923.70	41.48	59547.04	344.1806	-9.44%	-87.9%
10	5	0.5x	$ 53,433.27	125.67	61808.25	404.0106	-13.5%	-68.9%
		1x	$ 54,354.92	86.59	64040.5	449.5515	-15.1%	-80.7%
		2x	$ 53,863.72	117.96	60644.92	468.3748	-11.2%	-74.8%
	11	0.5x	$ 52,459.14	70.01	59945.7	394.8867	-12.5%	-82.3%
		1x	$ 53,110.83	77.94	60632.55	405.5063	-12.4%	-80.8%
		2x	$ 56,882.05	86.83	61387.06	436.6573	-7.3%	-80.1%
	22	0.5x	$ 43,177.56	25.29	47214.18	261.852	-8.5%	-90.3%
		1x	$ 45,415.99	27.69	47081.17	322.6722	-3.5%	-91.4%
		2x	$ 46,788.19	36.21	45878.18	441.5506	1.98%	-91.8%