Hindawi
Complexity
Volume 2020, Article ID 2189275, 11 pages
https://doi.org/10.1155/2020/2189275

Research Article

Freight Time and Cost Optimization in Complex Logistics Networks

Egemen Sert,[1,2] Leila Hedayatifar,[1] Rachel A. Rigg,[1] Amir Akhavan,[1] Olha Buchel,[1] Dominic Elias Saadi,[1] Aabir Abubaker Kar,[1] Alfredo J. Morales,[1] and Yaneer Bar-Yam[1]

[1]*New England Complex Systems Institute, 277 Broadway, Cambridge, MA 02139, USA*
[2]*Department of Electrical and Electronics Engineering, Middle East Technical University, Ankara, Turkey*

Correspondence should be addressed to Leila Hedayatifar; leila@necsi.edu

Received 12 November 2019; Accepted 3 February 2020; Published 29 February 2020

Guest Editor: Tomas Veloz

Copyright © 2020 Egemen Sert et al. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

The complexity of providing timely and cost-effective distribution of finished goods from industrial facilities to customers makes effective operational coordination difficult, yet effectiveness is crucial for maintaining customer service levels and sustaining a business. Logistics planning becomes increasingly complex with growing numbers of customers, varied geographical locations, the uncertainty of future orders, and sometimes extreme competitive pressure to reduce inventory costs. Linear optimization methods become cumbersome or intractable due to the large number of variables and nonlinear dependencies involved. Here, we develop a complex systems approach to optimizing logistics networks based upon dimensional reduction methods and apply our approach to a case study of a manufacturing company. In order to characterize the complexity in customer behavior, we define a "customer space" in which individual customer behavior is described by only the two most relevant dimensions: the distance to production facilities over current transportation routes and the customer's demand frequency. These dimensions provide essential insight into the domain of effective strategies for customers. We then identify the optimal delivery strategy for each customer by constructing a detailed model of costs of transportation and temporary storage in a set of specified external warehouses. In addition, using customer logistics and the k-means algorithm, we propose additional warehouse locations. For the case study, our method forecasts 10.5% savings on yearly transportation costs and an additional 4.6% savings with three new warehouses.

1. Introduction

Logistics is widely recognized as the most complex among business processes. The challenge of coordinating with multiple suppliers for raw materials and partially finished goods, and the challenge of delivering next-stage finished goods to customers, all in the correct amounts in a timely fashion and in coordination with production processes, despite uncertainty due to independent decision-making of customers, is daunting. These coordination processes are particularly challenging because of the need to optimize costs and maximize customer satisfaction. It is particularly difficult to keep transportation networks optimized when operations span thousands of miles and serve thousands of customers. Logistics is known to be a highly complex challenge that is not amenable to traditional linear optimization strategies due to its high dimensionality and rigidity in the face of limited accuracy and variation of conditions [1–3]. Optimizing a nonlinear system is quite challenging as the output solution is not unique and simply a linear combination of the independent parts [4]. So while mathematical models should be helpful to explore the space of possible strategies and propose optimal solutions when operations become complex [5], solving such models becomes more difficult as the number of strategies and considered variables increases [6].

The issue of complexity in supply chains has been explored from a number of perspectives [7–12]. The literature on these problems, generally divided into the location-routing problem (LRP) [13–15] and the warehouse location problem (WLP) [16–18], provides a range of proposed solutions. For both the LRP and the WLP, there is an objective function to be minimized. The function may consider the

freight cost from production facilities to customers, the storage cost for storing goods in warehouses, and the cost for opening new warehouses. Each approach involves imposing constraints on the objective function to make the optimization more reliable, which makes the solution harder and more time-consuming for large complex systems. Therefore, the LRP and WLP together remain open problems in the field of supply chain management, requiring further improvements in analytical methods.

Solutions to the LRP typically are addressed with capacity constraints on warehouses and/or vehicles, called the capacitated LRP (CLRP) [13, 14, 19, 20]. Capacity can refer to the capacity of warehouses to store goods, the number of vehicles for transport, or the carrying capacity of vehicles. The objective is to find an optimum set of routes that minimizes the total transport distance so that each customer is served with a compatible vehicle and the total demand by customers per route is compatible with the capacity of vehicles on that route [21]. The time window of deliveries is another constraint that can be considered, which suggests a hybrid multiobjective algorithm [22]. The multiechelon LRP (LRP-2E) is another set of solutions to optimize freight costs and delivery time by adding a new layer to the logistic network [23–30], resulting in three layers: production facilities, external warehouses, and customers. Solutions to the WLP typically recommend optimal new warehouse locations to more efficiently serve customers [17, 31]. In most of the problems, a set of potential warehouses with known opening or storage costs is considered. Decisions are made about which warehouses (distribution centers or depots) to keep open to minimize route costs. These methods are categorized into two classes, uncapacitated and capacitated facility location problems [32–34]. For capacitated facility location problems, one more constraint is added to the objective function [35–37]. The uncapacitated facility location problem simplifies to a k-means or k-medians clustering problem [38–40] when the facility opening cost or storage cost is considered to be zero [33].

Here, we show that a simplified parameterized space can provide insight into the optimization challenge and a more detailed quantitative modeling approach that focuses on the relevant details can be successfully applied to real-world optimization with substantial financial benefits for an industrial company. We propose two models to optimize companies' logistics networks, including the route from production facilities to the customers, by using existing warehouses and also recommending additional warehouse locations. To address the LRP, we define a customer space to better clarify the complexity in the logistics of customer-warehouse routes. The space is classified with two strategies: direct and indirect shipment strategies. In the direct strategy, goods are sent to the customer directly from a production facility using box or bulk trucks. In the indirect strategy, in advance of an order by the customer, goods are shipped to an external warehouse near the customer using trains and then "last-mile" shipped by trucks when orders are placed. Our methods identify the strategy for each customer that is most cost-effective and enables delivery to the customer within a predefined time interval. The choice of strategies and vehicles depends on the frequency of orders and amount of demand from customers. To address the WLP, in addition to optimization over existing facilities, we identified potential additional warehouse locations using the k-means algorithm weighted by the customer demand quantity. With these new warehouses, we estimate that savings can be further increased. We apply these methods to a medium-sized American manufacturing company with a particular logistics network, consisting of multiple production facilities, external warehouses, and customers along with three types of shipment methods (box truck, bulk truck, and train).

The rest of this paper is organized as follows: In Section 2, we describe our methodology, including the design of a customer space, a mathematical model to characterize customers and determine favorable strategies for each customer type, and a method to optimize warehouse locations. In Section 3, we describe our results that demonstrate effective optimization of shipment and storage costs. In Section 4, we summarize our conclusions.

2. Methodology and Framework

2.1. Customer Space. In order to develop a general understanding of the assignment of strategies to customers and the effectiveness of each strategy, we first created a descriptive model of customer characteristics named the "customer space" (see Figure 1). Each customer is characterized by two variables: the distance of the most used shipment route from the customer to the production facility and the customer demand frequency. The demand frequency is the ratio of the total quantity ordered by the customer to the customer life span using historical corporate data. The expected relationship between these two variables and the choice of strategies is as follows:

(i) The direct strategy is most effective for (1) customers close to production facilities, regardless of demand frequency, or (2) customers who order rarely, regardless of distance, as illustrated by the blue region in Figure 1(b). For close customers, maintaining an external warehouse is unnecessary given that the proximity of customers ensures rapid delivery. For low-demand customers, the uncertainty of order arrivals makes it inefficient to plan ahead, and shipping directly is a practical solution.

(ii) The indirect strategy becomes optimal when the customer's distance to production facilities is long and orders are frequent above a certain level, as illustrated by the green region in Figure 1(b). When both demand and distance are large enough, the certainty of ordering behavior supports the replenishment of inventory in external facilities before the customer even places the next order. Cheaper, slower transportation alternatives are possible between production facilities and external warehouses. When the customer places the next order, the goods will already be at the external warehouse and can be rapidly delivered to the customer. This indirect strategy may reduce transportation-

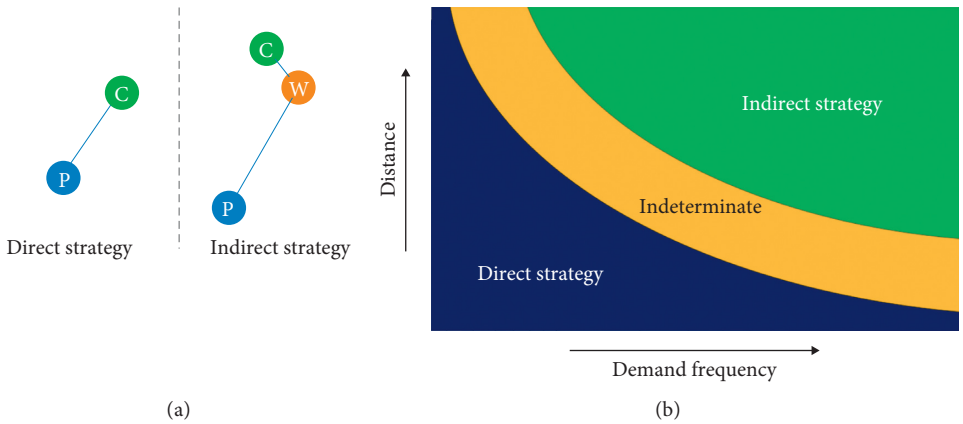

Figure 1: Shipping strategies. (a) The direct strategy (left) consists of shipping goods from the production facility (blue circle) to customers (green circle). The indirect strategy (right) delivers goods first to an external warehouse (orange circle), where goods are stored and subsequently shipped to customers (green circle). (b) Overview of the customer space and the associated optimal transportation strategies.

associated costs while preserving or even improving customer satisfaction.

(iii) The best strategy for customers with intermediate distance and intermediate demand will depend upon details of the freight and storage cost information, as illustrated by the yellow region in Figure 1(b).

2.2. Optimization via Route Strategies. To solve the problem of choosing the best storage and transportation strategy for each customer, we first constructed a model of the costs of shipment and storage to decide between direct and indirect strategies. The better strategy depends on the direct delivery time and on analysis of cost of shipment and storage. We defined the direct delivery time as the time between the shipment of a good and its delivery to the customer. According to corporate policy, the maximum delivery time for finished goods is set to two days for customer satisfaction. Delivery time is calculated using truck speeds of 70 miles per hour and 8 hours of driving per day and railcar speeds of 49 miles per hour and 24 hours of travel per day. If the time of direct delivery is more than two days, adequate customer satisfaction requires using the indirect strategy as an imposed constraint.

The mathematical model evaluates the costs of the direct and indirect strategies and includes a production facility (P), external warehouse (W), and customer (C), as illustrated in Figure 2. The potential costs include c_d, the cost of shipment from P to C; c_w, the cost of shipment from P to W; c_s, the cost of storage at W; and c_o, the cost of shipment from W to C. The freight costs c_d, c_w, or c_o must also be multiplied by the number of shipments n_d, n_w, or n_o, respectively. The number of shipments depends on the demand from the customer. The customer's expected demand over a year is estimated to be the demand frequency multiplied by the days in a year. We considered the number of shipments in a year to be the ratio of total demand to the shipment carrying capacity of trucks and railcars. The cost J for a given strategy π is then

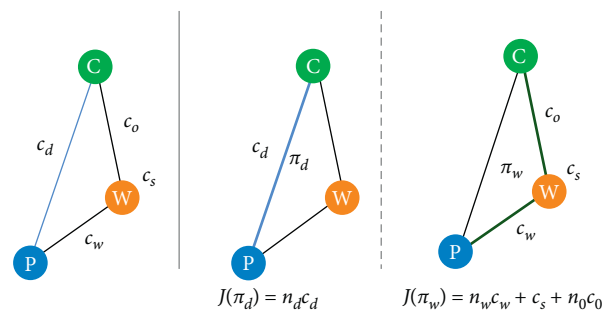

Figure 2: Shipping strategies of a product from a production facility (P) to a customer (C), with an option to use an external warehouse (W). Strategies are associated with freight and storage costs. Costs c_d, c_w, and c_o represent the freight costs associated with the respective labelled transportation links. The cost c_s denotes the storage cost of the material in the external warehouse (W). The number n denotes the number of times the shipment is made. Every strategy π is associated with a cost $J(\pi)$.

determined for the direct strategy as $J(\pi_d) = n_d c_d$ and the indirect strategy as $J(\pi_w) = n_w c_w + c_s + n_o c_o$.

Storage and freight costs depend on various parameters in the model. We calculated these costs directly based upon detailed descriptions of those costs that vary between shippers and warehouses. The storage cost c_s depends on (1) the storage facility type s, (2) the quantity that is stored q (inventory cost), (3) the time the quantity is stored t, and (4) loading u and unloading w events, giving $c_s = S(s, q, t, u, w)$. The freight cost $c_f \in \{c_d, c_w, c_o\}$ depends on (1) the carrier type s', (2) the distance the goods are sent d, and (3) the quantity of the goods q', giving the relationship $c_f = F(s', d, q')$. In order to calculate the actual cost based upon the company data, we extracted existing routes along with their associated distances from historical data and incorporated specific storage costs.

Finally, we defined savings for strategies as follows: Each customer i should have an optimal shipping cost, designated C_i, which also includes storage costs if present. Each customer has a current shipment route (designated route 0),

which has a known cost $C0_i$. We then independently calculated the lowest cost route (designated route 1), which has a cost $C1_i$. We calculated $C1_i$ by examining nearest warehouses and incorporating storage costs and transportation costs. Finally, we compared the current cost to our calculated costs, and if $C1_i < C0_i$, then the preferred cost, C_i, equals $C1_i$ or otherwise $C_i = C0_i$. From this, we calculated total percent savings (S) for all customers as a percentage: $S = 100 * (1 - (\sum_{i=1}^{N} C_i)/(\sum_{i=1}^{N} C0_i))$. Here, N is the total number of customers.

2.3. Optimization via Additional Warehouses. Aside from the existing external warehouses, we identified prospective locations for new warehouses for additional savings. In order to determine potential locations, we used the k-means algorithm [38–40] to find the optimum locations for the warehouses that best match the locations of customers to minimize the freight cost across all customers, C_f. Freight cost of transporting demands from the j-th warehouse to the i-th customer, A_{ij}, is the price of consumed fuel by the vehicles shipping the goods, but here it was defined as direct incurred shipping cost for the shipments in the database. It is weighted based on the overall amount of demands by customers shipped from a warehouse, D_{ij}, and is a function of the Euclidean distance, d_{ij}, between the i-th customer and the j-th warehouse:

$$\begin{aligned} & C_f = \sum_{i=1}^{N} \sum_{j=1}^{M} A_{ij} x_{ij} \\ \text{minimize} \quad & A_{ij} = F^p \times R \times D_{ij} \times d_{ij} \\ & d_{ij} = \left\| w_j - c_i \right\| \end{aligned} \tag{1}$$

$$\begin{aligned} \text{subject to} \quad & x_{ij} \in \{0, 1\}, \quad \forall i \in C, j \in W, \\ & \sum_{i=1}^{N} x_{ij} = 1, \quad \forall j \in W, \end{aligned} \tag{2}$$

where the variable x_{ij} equals 1 if the customer i is served by the warehouse j and equals 0 if it is not and N and M are the number of customers and warehouses. We assigned customer demand weights according to $W_i = \lceil \sum_{k=0}^{n_i} Q_k / Q_0 \rceil$, where n_i is the number of orders by the customer i, Q_k is the quantity of the order k by the customer i, and Q_0 is an industry standard measure for a significant customer volume. The brackets $\lceil x \rceil = \text{ceil}(x)$ indicate the smallest integer greater than x. In fact, Q_0 corresponds to the average shipment size by standard vehicles. So, $D_{ij} = W_i$ if $x_{ij} = 1$; otherwise, it is 0. In the calculation of d_{ij}, c_i and w_j refer to the geographical location of customers and warehouses, respectively. Equation (2) indicates that each customer is only connected to one warehouse. Here, F^p refers to the fuel price and R refers to the average fuel consumption rate by vehicles. For simplicity, we considered one type of vehicle with a fixed shipment size.

We use the k-means algorithm to aggregate the customer locations into k disjoint groups or clusters and find a centroid C_k for each group to minimize the average squared distance between the centroid and customer locations within each group. To consider the weight of customer demands, we assigned W_i points to the location of each customer i. The number of groups to be found is a parameter of the analysis. The algorithm is an iterative refinement technique that starts from random locations for centroids and updates the location of centroids in each iteration until reaching an optimum location for all the centroids. We considered the centroid to be an approximate optimum location for a warehouse assigned to the customers of a group. The freight cost from warehouses to customers inside the groups decreases as the number of centroids increases and slowly converges to zero. We determined the optimum number of centroids from the deceleration in the freight cost. We compared the location of currently active warehouses with the location of centroids, identifying the best locations for the additional warehouses to decrease the transportation costs. The k-means analysis dramatically reduces the number of candidate locations to be considered for cost optimization.

We added the new warehouse locations proposed by our analysis to the system. Since we cannot know the storage cost of a theoretical warehouse, we used three representative storage costs (high, medium, and low cost) based on existing warehouses to model storage costs for the proposed warehouses. We calculated potential savings for each proposed warehouse using the three cost levels.

3. Results

We tested our model on a dataset from a medium-sized manufacturing company with more than fifteen years of customer orders. The company and its customers are located primarily in the US. The logistics network has about 15 production facilities and more than 30 external warehouses and serves more than 2000 customers. The majority of the customers have not ordered more than 10 times, and due to the lack of data on these customers' ordering patterns, the direct shipment method is always chosen by the company (as discussed in Section 2.1). Therefore, we excluded customers with 10 or fewer orders from our analysis. Customers who have ordered more than 10 times may benefit from either the direct or the indirect strategy, so we chose these customers (≈300 customers) for analysis. Our goal was to find the strategy that minimizes the total cost for each customer. First, we estimated the total cost of each strategy from freight costs of each shipment and associated storage costs, where present. In addition to estimating the total cost of each strategy for existing warehouses, we proposed additional warehouse locations by performing clustering on the geographical location data of customers weighted by the total quantity ordered.

In order to estimate freight costs, we observed the freight costs of direct and indirect strategies used for each customer including the freight options and location data. The indirect strategy is associated with multiple shipments: one shipment from the production facility to the external warehouse and one or multiple subsequent shipments from the warehouse to the customer. Each shipment is made with one of the

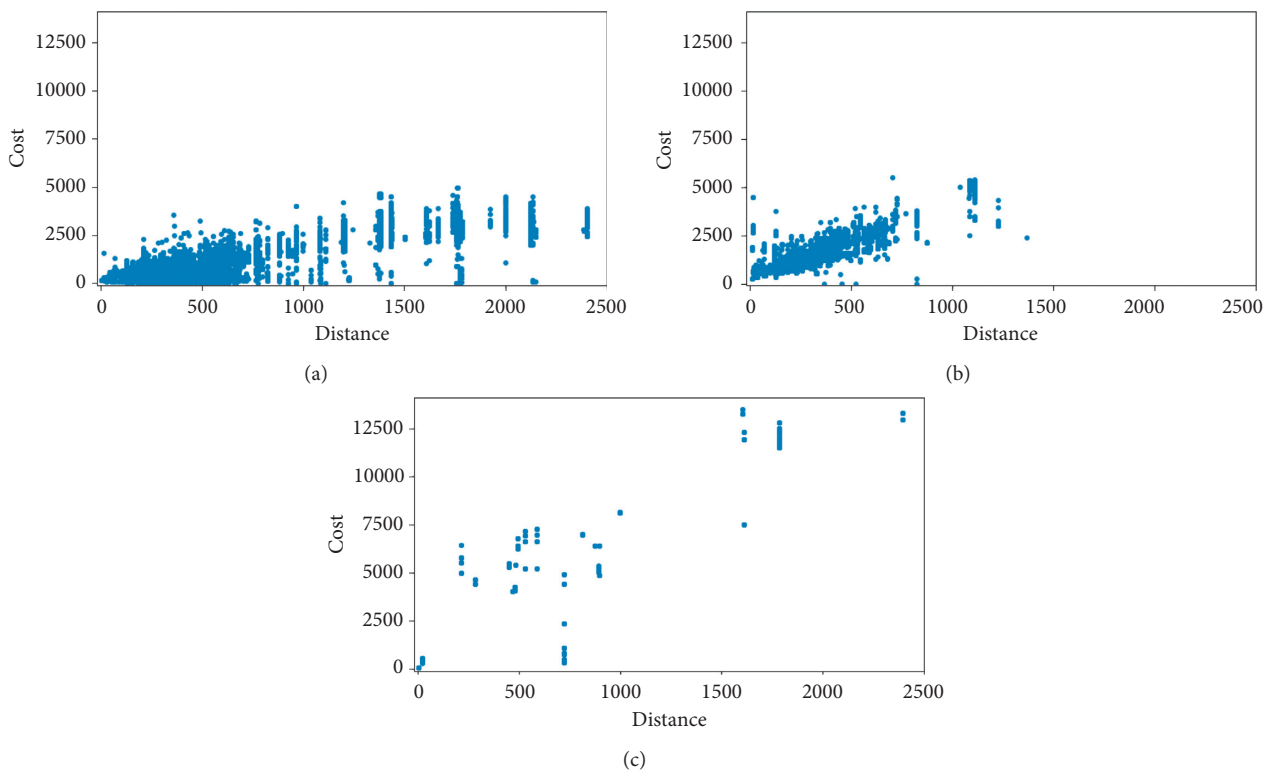

Figure 3: Cost versus shipping distance using actual data for three transportation options: (a) box truck; (b) bulk truck; (c) train.

following three types of vehicles: box truck, bulk truck, and train. The costs of each type are given in Figures 3(a)–3(c) for box truck, bulk truck, and train shipments, respectively. High variance in freight costs is due to different carriers having different pricing structures. Carriers may charge by distance, quantity, or both. For better classification, we obtained the freight costs of carriers by performing multiple linear regression analysis on each carrier's historical data to extrapolate the freight costs for possible routes. We constrained the parameters of the regression model to be positive. The decomposition of the freight costs by carriers is shown in Appendix in Figures 7–9 for box truck, bulk truck, and train shipments, respectively. The plots indicate that different carriers may be specialized for different distances, different quantities, and particular customers.

In the indirect strategy, the storage cost of external warehouses is a key factor in addition to the freight cost. The benefit of using an external warehouse depends on (1) distances from the warehouse to the production facility and the customer and (2) storage costs. The shipping distance affects associated freight costs. The storage pricing is unique to each individual warehouse and is determined based upon the quantity stored, the duration of the storage, and the number of loading and unloading events that occur. Detailed cost specifications were provided by the company for each external warehouse in use. These were used for calculating the cost of the indirect strategy.

We have identified a change in strategy by the company over time in the customer space, as shown in Figure 4. Figure 4(a) shows strategy decisions for all years of the dataset, and Figure 4(b) shows decisions for only the last two months. Green x's denote customers for which the indirect strategy has been used at least once, while blue dots denote customers for which only the direct strategy has been used. As shown in Figure 4(a), for many customers, the indirect strategy has been used at least once regardless of demand frequency, except for demand frequency below 600 lbs/day, in which case only the direct strategy is used. For customers with large distances from production facilities, the company has used the indirect strategy for customers with high demand frequency but not for customers with low demand frequency. Meanwhile, the data for the last two months in Figure 4(b) show a change in corporate strategy, with a significant drop in the number of customers serviced by the indirect strategy. We can infer from the graph that the key variable used for strategy selection is still demand frequency, with a higher demand frequency increasing the chance of using the indirect strategy.

After fitting our model to the historical company data, we extracted optimal strategies for servicing customers. Then, we identified which strategy should optimally be used across the customer space (Figure 5). Blue dots denote customers correctly serviced with the direct strategy. Yellow triangles denote customers serviced with the direct strategy that would benefit from the indirect strategy. Green x's denote customers correctly serviced with the indirect strategy. Magenta squares denote customers serviced with the indirect strategy that would benefit from changing external warehouses. Finally, red stars denote customers serviced with the indirect strategy that would benefit from using

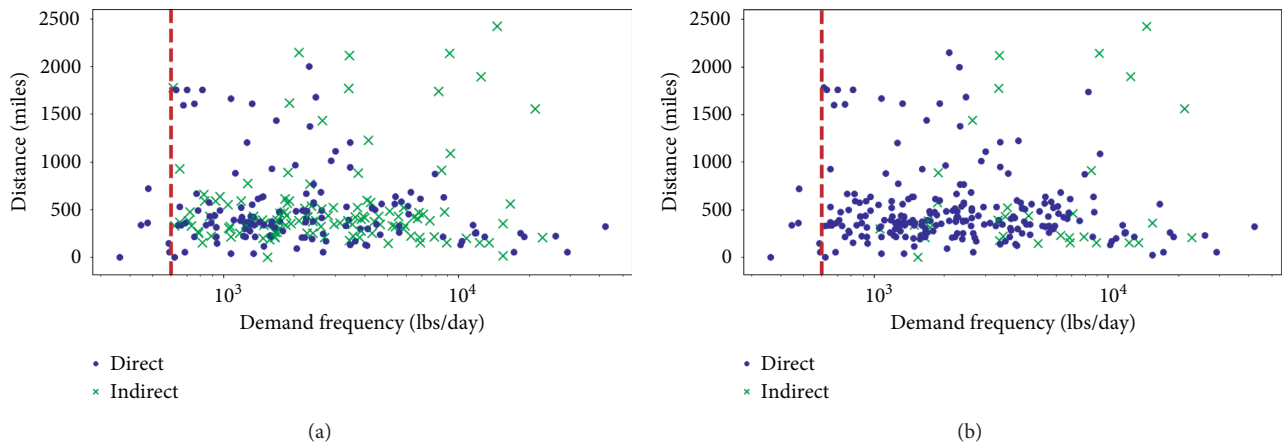

Figure 4: Distance versus demand frequency. Green x's denote customers that were serviced by the indirect strategy at least once, and blue dots denote customers that were only serviced by the direct strategy. Red dotted lines represent a demand frequency of 600 lbs/day; customers with a lower demand frequency only received goods by the direct strategy, per company policy. (a) All historical company data. (b) The last two months of data.

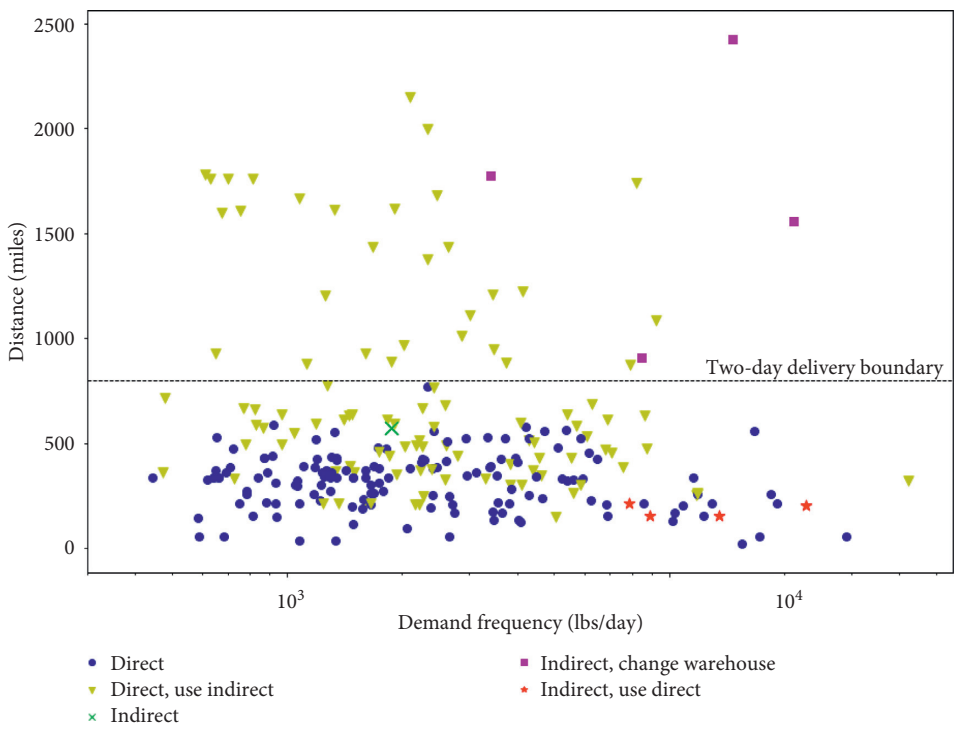

Figure 5: Optimal strategies for the customers shown on the customer space (distance versus demand frequency). Blue circles and green x's show customers correctly serviced by direct and indirect strategies. The rest shows the serviced strategies that must be changed to optimize the freight cost.

the direct strategy. We calculated the potential savings by comparison of the historical strategies with the proposed ones. In total, the model predicts 10.5% savings if shipments and warehouses are optimized over the current options.

In addition to the analysis of the existing warehouses, we incorporated optimum locations for extra warehouses to increase the savings. Figure 6 shows the results of warehouse optimization using the k-means algorithm. The algorithm identifies warehouse locations that minimize the freight costs, C_f, from customers to their nearest warehouse (Figure 6(a)). The algorithm takes as the input the number of warehouses to be determined. When the number of warehouses is below 10, adding any new warehouses leads to a sharp decrease in the value of C_f, but the effect slows down for larger numbers of warehouses. The orange line shows the actual freight cost for the demands by customers which is comparable with the freight cost from a single warehouse to the customers. Note that the total number of the company's production facilities and warehouses is more than 45. In many of the shipments, the company served a customer

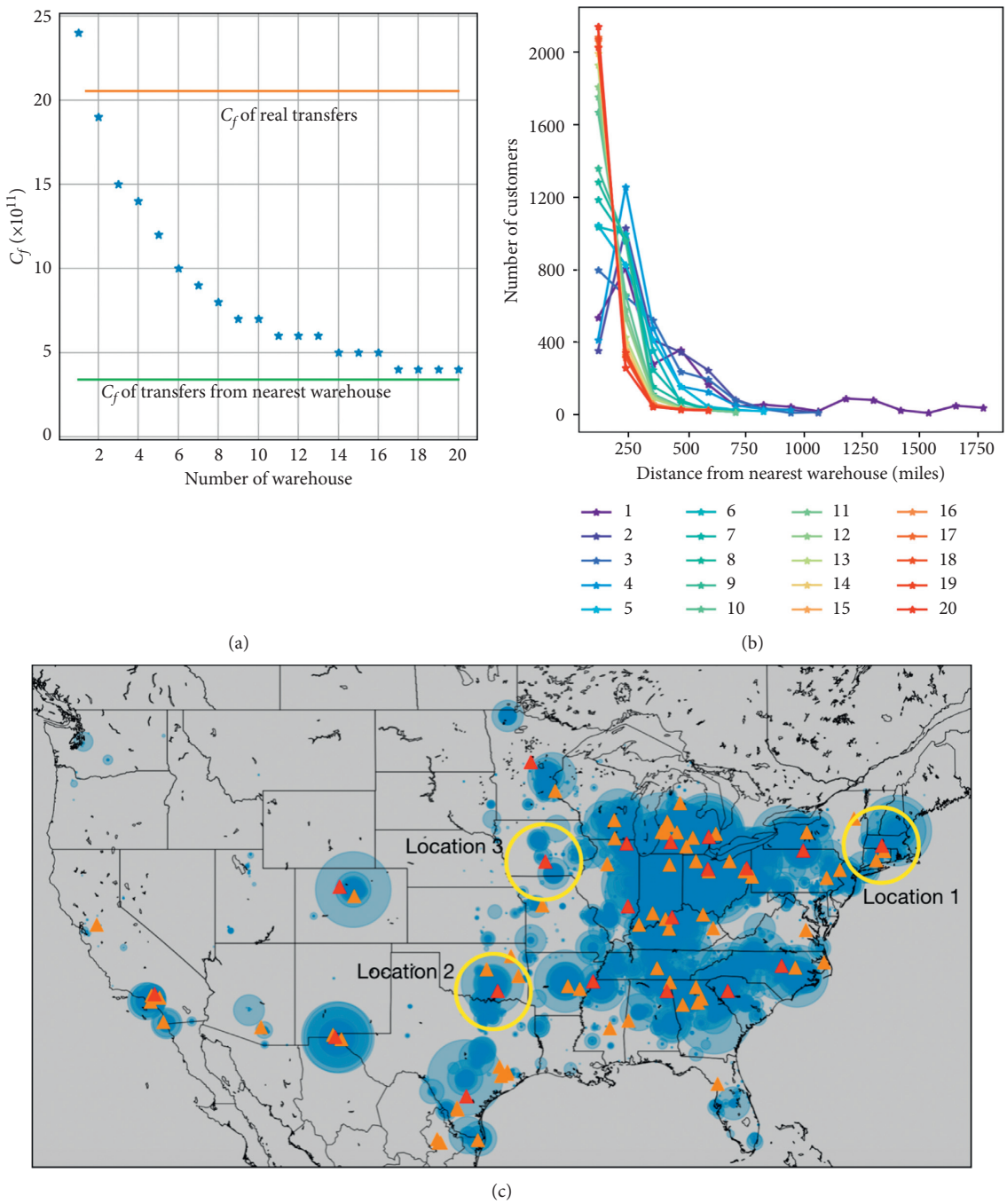

Figure 6: Optimal warehouse locations proposed by the k-means algorithm. (a) Freight cost decreases by increasing the number of optimized warehouses. Orange and green lines show the actual freight cost and the freight cost of transfers from the current nearest company warehouses. (b) Distribution of customers based on their distance from the nearest warehouse. (c) A network of customers (blue circles) and current (orange triangles) and optimized (red triangles) warehouse locations. Yellow circles denote geographical areas where candidate warehouses should be placed but that currently have inactive warehouses (Locations 1 and 2) or no warehouse (Location 3).

Table 1: Potential cost savings if additional warehouses are utilized.

Location	Storage cost (%)		
	High	Medium	Low
Location 1	10.5	10.5	11.0
Location 2	10.5	10.5	10.7
Location 3	10.7	11.5	14.4
All	10.7	11.5	15.1

The first three rows represent the total corporate savings of optimized delivery with the addition of one warehouse. The final line is the total savings with all three of the additional warehouses. The labels high, medium, and low denote the reference values used to calculate storage costs for the additional locations.

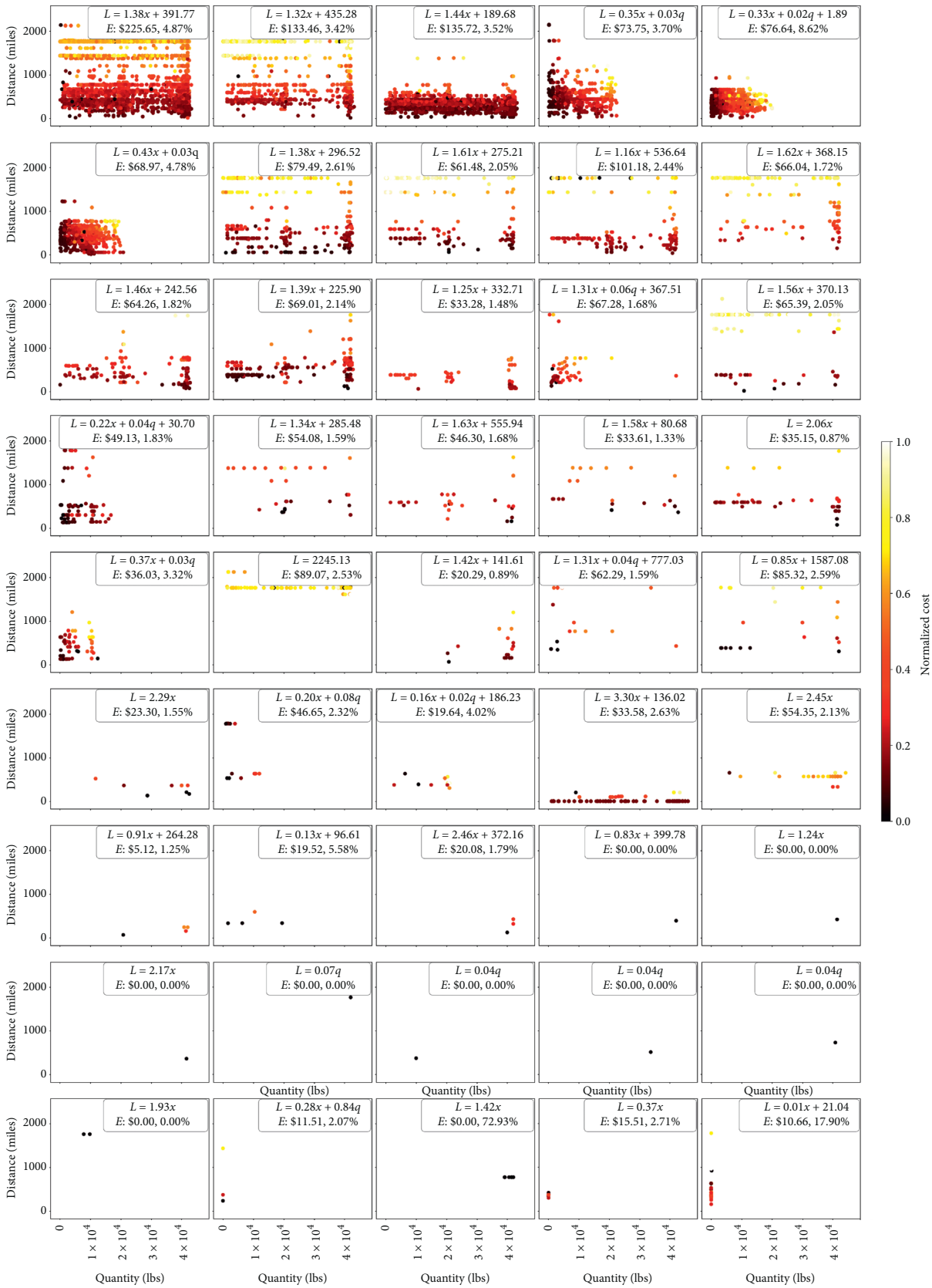

Figure 7: Distance versus quantity scatter plots for box truck carriers. Color indicates freight cost (key on right), normalized to [0, 1] for each carrier to have a minimum cost of 0 and a maximum cost of 1. Cost equations are obtained by performing regression analysis on each carrier. Weights are constrained to have positive values. In each cost equation (inset), L indicates the cost equation of the carrier, x denotes the distance in miles, q denotes the quantity in pounds, and E denotes the absolute error of the cost regression in dollars and as a percentage. Carrier names are omitted to preserve anonymity.

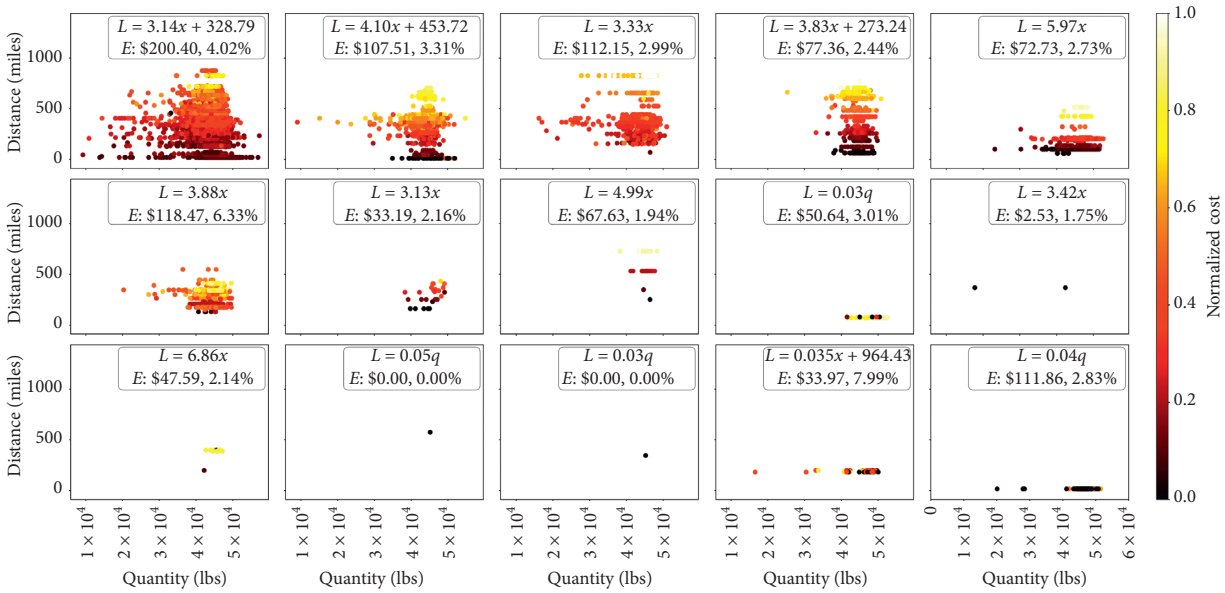

Figure 8: Distance versus quantity scatter plots for bulk truck carriers. Color indicates freight cost (key on right), normalized to [0, 1] for each carrier to have a minimum cost of 0 and a maximum cost of 1. Cost equations are obtained by performing regression analysis on each carrier. Weights are constrained to have positive values. In each cost equation (inset), L indicates the cost equation of the carrier, x denotes the distance in miles, q denotes the quantity in pounds, and E denotes the absolute error of the cost regression in dollars and as a percentage. Carrier names are omitted to preserve anonymity.

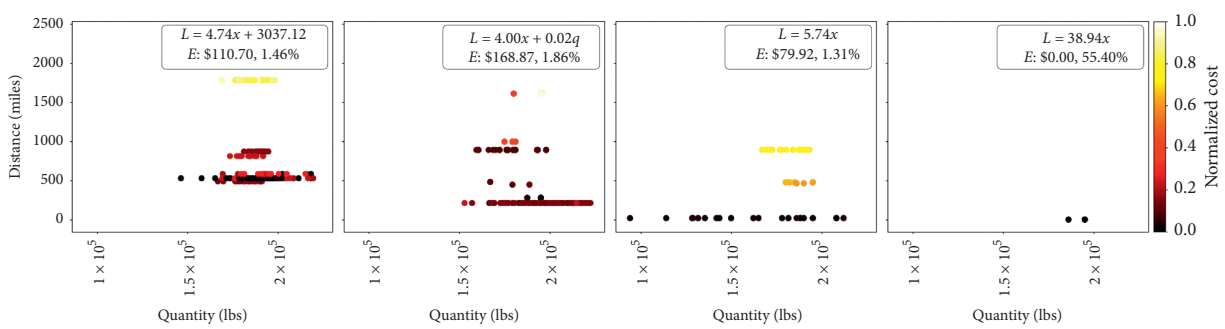

Figure 9: Distance versus quantity scatter plots for train carriers. Color indicates freight cost (key on right), normalized to [0, 1] for each carrier to have a minimum cost of 0 and a maximum cost of 1. Cost equations are obtained by performing regression analysis on each carrier. Weights are constrained to have positive values. In each cost equation (inset), L indicates the cost equation of the carrier, x denotes the distance in miles, q denotes the quantity in pounds, and E denotes the absolute error of the cost regression in dollars and as a percentage. Carrier names are omitted to preserve anonymity.

from a very large distance. The green line shows the freight cost if the customers have been served from their nearest warehouse which is comparable with the freight cost of shipping from 20 optimally located warehouses. In Figure 6(b), we indicate the distribution of customers based on their distance from the nearest optimized warehouse. In the presence of one optimized warehouse, most of the customers have a distance larger than 250 miles. However, adding the second and third optimized warehouses drastically reduces the distance between customers and nearest warehouses. The distances change gradually for higher numbers of warehouses.

Figure 6(c) shows the locations of all company facilities and external warehouses (orange triangles) and customers (blue circles) around the US. We randomized the location of actual warehouses and customers for confidentiality. The size of the circles is proportional to the total order quantity by each customer, such that the more the orders a customer has placed, the larger the size of the circles. The figure shows the optimum warehouse locations (red triangles) recommended by the k-means algorithm for 20 warehouses. Some of the k-means recommended locations are not located near active warehouses, revealing significant potential cost savings.

After examining recommended warehouse locations, we identified three as particularly relevant for cost savings. The areas for the optimal warehouses are shown as yellow circles in Figure 6(c). Two of these areas (Locations 1 and 2) include previously active but currently inactive warehouses, and a third one (Locations 3) does not have either a current or a

previously active warehouse. Since we do not know the storage costs associated with the new warehouses, we used cost information of three actively used warehouses that are known to have high, medium, and low storage pricing rates for the same amount of goods being stored for the same amount of time. As a general example, storing 1,000 lbs of goods for a month would cost at a high-cost warehouse, at a medium-cost warehouse, and at a low-cost warehouse. The potential total savings including the new warehouse locations range from 10 to 15%, as shown in Table 1.

4. Conclusions

In summary, we have developed a method of characterizing the customer space and a mathematical model that provides recommendations for optimizing shipment routes of a logistics network. This is a multiscale approach to the logistics high-dimensional optimization problem. Firstly, we begin by projecting onto a low-dimensional space. We then identify a first-order boundary between strategies. Secondly, we incorporate details due to other dimensions to refine the solutions. Customer spaces also help give an aggregate view of customer behaviors and characteristics. They allow policymakers to compare customers and develop strategies based on the aggregate behavior of the system as a whole.

In particular, based on the customer space of demand frequency versus distance from the production facility, we analyzed two strategies: direct and indirect shipments. Each strategy applies to an area of the customer space with an indeterminate boundary between them. Specific company policies determine the location of the boundary generally. Moreover, detailed properties of each customer can affect the specific strategy used for that customer.

We also used the k-means algorithm to find the optimized location of warehouses based on the location of customers and their demands. The accuracy of the optimization can be improved by updating the conventional k-means algorithm to consider the capacity of warehouses and further details about customers. Still, using this optimization method, companies are able to define the locations of next potential warehouses even without details that can be determined only once they are in operation.

We have applied this analysis to a case study of a manufacturing company with particular constraints. We showed that these optimizations can provide considerable cost savings and improved service quality and customer satisfaction for the company.

Many papers have been published in location-routing problem (LRP) and warehouse location problem (WLP) fields; a few are mentioned in Introduction, but they are still open problems. It has been challenging to find solutions that are applicable to large companies with thousands of customers. While considering more constraints in the calculation of freight cost can improve the accuracy of the outputs, it would increase the complexity and make the solutions much more difficult if not impossible for large systems. Our approach has been shown to work for a company with more than 2000 customers. The future work may further improve the optimization by adding additional constraints such as a limitation on the number of customers assigned to each facility in addition to a limitation on distance. Overall, we showed that, through a targeted approach to data analysis, we can build a heuristic understanding of the customer space and develop specific descriptive and prescriptive models to yield significant savings.

Appendix

In this section, we show company data on freight costs for a year of box truck, bulk truck, and train shipments. Figures 7–9 show freight costs of the individual shipment types decomposed by the carrier. In total, there are 45 box truck carriers, 15 bulk truck carriers, and 4 train carriers. Figure 7 shows scatter plots of individual freight events by distance and quantity. The color represents cost normalized by the maximum freight cost for that carrier (yellow denotes the maximum freight cost and black denotes the minimum freight cost). Figures 8 and 9 show similar plots for bulk truck and train carriers. The plots indicate that carriers are selected based on the quantity shipped and the freight distance. Regression results for cost functions are shown in the legend of each subplot. The symbol x denotes the distance, and q denotes the quantity.

Data Availability

Data are available at http://www.necsi.edu/customer/data.

Conflicts of Interest

The authors declare that they have no conflicts of interest.

Acknowledgments

We thank Irving Epstein and William Glenney for feedback and Matthew Hardcastle for proofreading the manuscript. The writing of this paper was supported by NECSI funds.

References

[1] V. L. M. Spiegler, M. M. Naim, D. R. Towill, and J. Wikner, "A technique to develop simplified and linearised models of complex dynamic supply chain systems," *European Journal of Operational Research*, vol. 251, no. 3, pp. 888–903, 2016.

[2] M. T. Melo, S. Nickel, and F. Saldanha-Da-Gama, "Facility location and supply chain management—a review," *European Journal of Operational Research*, vol. 196, no. 2, pp. 401–412, 2009.

[3] Z. Zhang, Xu Wang, Q. Guo, Z. Li, and Y. Wu, "Simulation and analysis of the complex behavior of supply chain inventory system based on third-party logistics management inventory model with no accumulating of unsatisfied demand," *Complexity*, vol. 2019, Article ID 3194093, 18 pages, 2019.

[4] D. P. Atherton and G. M. Siouris, "Nonlinear control engineering," *IEEE Transactions on Systems, Man, and Cybernetics*, vol. 7, no. 7, pp. 567-568, 1977.

[5] Y. Bar-Yam, "Complexity rising: from human beings to human civilization, a complexity profile," in *Encyclopedia of Life Support Systems*, United Nations, Oxford, UK, 2002.

[6] S. Zokaee, A. Jabbarzadeh, B. Fahimnia, and S. J. Sadjadi, "Robust supply chain network design: an optimization model with real world application," *Annals of Operations Research*, vol. 257, no. 1-2, pp. 15–44, 2017.

[7] F. Isik, "Complexity in supply chains: a new approach to quantitative measurement of the supply chain complexity," in *Supply Chain Management*, P. Li, Ed., pp. 417–432, InTech, Rijeka, Croatia, 2011.

[8] J. Aelker, T. Bauernhansl, and H. Ehm, "Managing complexity in supply chains: a discussion of current approaches on the example of the semiconductor industry," *Procedia CIRP*, vol. 7, pp. 79–84, 2013.

[9] C.-Y. Cheng, T.-L. Chen, and Y.-Y. Chen, "An analysis of the structural complexity of supply chain networks," *Applied Mathematical Modelling*, vol. 38, no. 9-10, pp. 2328–2344, 2014.

[10] T. Kučera and J. Chocholáč, "Outsourcing as an innovative approach to logistics services of warehousing: a case study," in *The 10th International Days of Statistics and Economics: Conference Proceedings*, Prague, Czech Republic, 2016.

[11] T. Kucera, "Application of the activity-based costing to the logistics cost calculation for warehousing in the automotive industry," *Communications-Scientific Letters of the University of Zilina*, vol. 21, no. 4, pp. 35–42, 2019.

[12] T. Kučera, "Logistics cost calculation of implementation warehouse management system: a case study," *MATEC Web of Conferences*, vol. 134, Article ID 00028, 2017.

[13] G. Nagy and S. Salhi, "Location-routing: issues, models and methods," *European Journal of Operational Research*, vol. 177, no. 2, pp. 649–672, 2007.

[14] C. Prodhon and C. Prins, "A survey of recent research on location-routing problems," *European Journal of Operational Research*, vol. 238, no. 1, pp. 1–17, 2014.

[15] M. Schneider and M. Drexl, "A survey of the standard location-routing problem," *Annals of Operations Research*, vol. 259, no. 1-2, pp. 389–414, 2017.

[16] Y. Abo-Elnaga, B. El-Sobky, and L. Al-Naser, "An active-set trust-region algorithm for solving warehouse location problem," *Journal of Taibah University for Science*, vol. 11, no. 2, pp. 353–358, 2017.

[17] S. Coniglio, J. Fliege, and R. Walton, "Facility location with item storage and delivery," in *Optimization and Decision Science: Methodologies and Applications*, A. Sforza and C. Sterle, Eds., pp. 287–294, Springer International Publishing, Cham, Switzerland, 2017.

[18] O. Vera, D. Arturo, E. Olivares-Benitez et al., "Combined use of mathematical optimization and design of experiments for the maximization of profit in a four-echelon supply chain," *Complexity*, vol. 2018, Article ID 8731027, 25 pages, 2018.

[19] J. G. Villegas, C. Prins, C. Prodhon, A. L. Medaglia, and N. Velasco, "A grasp with evolutionary path relinking for the truck and trailer routing problem," *Computers & Operations Research*, vol. 38, no. 9, pp. 1319–1334, 2011.

[20] S. H. H. Doulabi and S. Abbas, "Lower and upper bounds for location-arc routing problems with vehicle capacity constraints," *European Journal of Operational Research*, vol. 224, no. 1, pp. 189–208, 2013.

[21] J. G. Villegas, C. Prins, C. Prodhon, A. L. Medaglia, and N. Velasco, "A matheuristic for the truck and trailer routing problem," *European Journal of Operational Research*, vol. 230, no. 2, pp. 231–244, 2013.

[22] K. C. Tan, Y. H. Chew, and L. H. Lee, "A hybrid multi-objective evolutionary algorithm for solving truck and trailer vehicle routing problems," *European Journal of Operational Research*, vol. 172, no. 3, pp. 855–885, 2006.

[23] V. R. Ghezavati, M. S. Jabal-Ameli, and A. Makui, "A new heuristic method for distribution networks considering service level constraint and coverage radius," *Expert Systems with Applications*, vol. 36, no. 3, pp. 5620–5629, 2009.

[24] M. Eskandarpour, S. H. Zegordi, and E. Nikbakhsh, "A parallel variable neighborhood search for the multi-objective sustainable post-sales network design problem," *International Journal of Production Economics*, vol. 145, no. 1, pp. 117–131, 2013.

[25] K. Subulan, A. S. Taşan, and A. Baykasoğlu, "Designing an environmentally conscious tire closed-loop supply chain network with multiple recovery options using interactive fuzzy goal programming," *Applied Mathematical Modelling*, vol. 39, no. 9, pp. 2661–2702, 2015.

[26] N. Labadi, C. Prins, and M. Reghioui, "A memetic algorithm for the vehicle routing problem with time windows," *RAIRO—Operations Research*, vol. 42, no. 3, pp. 415–431, 2008.

[27] B. N. L. Kechmane and A. Baalal, "Optimization of a two-echelon location lot-sizing routing problem with deterministic demand," *Mathematical Problems in Engineering*, vol. 2018, Article ID 2745437, 12 pages, 2018.

[28] O. Belgin, I. Karaoglan, and F. Altiparmak, "Two-echelon vehicle routing problem with simultaneous pickup and delivery: mathematical model and heuristic approach," *Computers & Industrial Engineering*, vol. 115, pp. 1–16, 2018.

[29] S. R. Kancharla and G. Ramadurai, "Multi-depot two-echelon fuel minimizing routing problem with heterogeneous fleets: model and heuristic," *Networks and Spatial Economics*, vol. 19, no. 3, pp. 969–1005, 2019.

[30] R. Cuda, G. Guastaroba, and M. G. Speranza, "A survey on two-echelon routing problems," *Computers & Operations Research*, vol. 55, pp. 185–199, 2015.

[31] C. Monthatipkul, "A non-linear program to find an approximate location of a second warehouse: a case study," *Kasetsart Journal of Social Sciences*, vol. 37, no. 3, pp. 190–201, 2016.

[32] D. B. Shmoys, É. Tardos, and K. Aardal, "Approximation algorithms for facility location problems," in *Proceedings 29th ACM Symposium on Theory of Computing*, pp. 265–274, El Paso TX, USA, May 1997.

[33] J. Vygen, "Approximation Algorithms for Facility Location Problems (Lecture Notes)," 2005.

[34] Z. Drezner and H. W. Hamacher, *Facility Location: Applications and Theory*, Springer, New York, NY, USA, 2014.

[35] E. H. L. Aarts and J. Karel Lenstra, *Local Search in Combinatorial Optimization*, Princeton University Press, Princeton, NJ, USA, 2003.

[36] B. Manthey and M. B. Tijink, "Perturbation resilience for the facility location problem," *Operations Research Letters*, vol. 46, no. 2, pp. 215–218, 2018.

[37] B. Korte and J. Vygen, "Facility location," in *Combinatorial Optimization*, pp. 629–665, Springer, New York, NY, USA, 2018.

[38] S. Lloyd, "Least squares quantization in pcm," *IEEE Transactions on Information Theory*, vol. 28, no. 2, pp. 129–137, 2006.

[39] D. Arthur and S. Vassilvitskii, "*K*-means++: the advantages of careful seeding," in *Proceedings of the 18th Annual ACM-SIAM Symposium on Discrete Algorithms*, New Orleans, LA, USA, 2007.

[40] J. B. MacQueen, "Some methods for classification and analysis of multivariate observations," in *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability*, vol. 1, pp. 281–297, University of California Press, Berkeley, Calif, 1967.