Learning from Data to Simplify Supply Chain Network Design

by

Hanbo Wang

Submitted on May 8th, 2026 in Partial Fulfillment of the

Requirements for the Degree of Master of Applied Science in Supply Chain Management

at the Massachusetts Institute of Technology

ABSTRACT

In today's dynamic market environment, solving large-scale supply chain network design problems within a practical timeframe poses a significant computational challenge. Given the massive scale of real-world data, Mixed-Integer Linear Programming (MILP) models often fail to provide solutions quickly and may even encounter out-of-memory errors. To address this challenge, this study develops a systematic simplification framework that renders large-scale supply chain network design problems computationally tractable while largely preserving solution quality. The framework simplifies the input data through product aggregation and customer clustering. To validate that this simplification does not compromise solution quality, we apply the same MILP model to both the original and simplified versions of small-scale datasets and compare the outputs. The results confirm that our simplification framework works effectively, thereby allowing us to confidently apply it to solve large-scale, real-world supply chain network design problems. This framework offers practitioners a computationally efficient tool for rapid, data-driven network planning decisions.

Capstone Advisor: Dr. Selene Silvestri

Title: Research Scientist

Capstone Co-Advisor: Dr. Sarah Schaumann

Title: Postdoctoral Associate

TABLE OF CONTENTS

1. INTRODUCTION 3
   1.1 Motivation 3
   1.2 Problem Statement and Key Questions 3
   1.3 Project Goals and Expected Outcomes 3
2. STATE OF THE PRACTICE 4
   2.1 The Model for Network Planning 4
   2.2 Product Aggregation for Dimensionality Reduction 5
   2.3 Spatial Clustering and Error Correction 5
   2.4 Data Scaling in Distance-Based Algorithms 7
   2.5 Summary 7
3. METHODOLOGY 7
   3.1 Framework Development 8
   3.2 Validation 8
   3.3 Full-Scale Application 11
4. RESULTS AND DISCUSSION 11
   4.1 Implementation of the Simplification Framework 12
   4.2 Validation Results on the Small-Scale Dataset 12
   4.3 Application to the Full-Scale Network 15
   4.4 Discussion and Limitations 17
5. RECOMMENDATIONS 17
   5.1 Management Recommendations 17
   5.2 Future Work 18
REFERENCES 19

1. INTRODUCTION

This research is sponsored and supported by the MIT Center for Transportation & Logistics (MIT CTL).

In today's highly competitive and dynamic market environment, ensuring a timely supply chain response is a major challenge for companies. This study aims to provide companies with a simplification framework that can be used to solve supply chain network design (SCND) problems in a rapid and accurate manner.

1.1 Motivation

Currently, supply chain planners employ state-of-the-art optimization solvers to solve network planning problems. However, solvers face a critical computational bottleneck because real-world supply chain networks comprise numerous distribution centers, thousands of customers, multiple product types, and multiple transportation modes. Feeding fully detailed information into SCND models either results in solver failures or requires long computation times. This computational limitation significantly impedes practical decision-making, particularly when companies need to timely re-evaluate their network configuration in response to changing market conditions, such as shifts in customer demand or disruptions to distribution center operations.

Current approaches to large-scale supply chain optimization often address computational demands by developing specialized algorithms tailored to specific network configurations (Garcia & You, 2015). However, such case-specific solutions lack generalizability. There is therefore a need for a more universal simplification framework that reduces data complexity before solving. This framework should be sufficiently detailed to enable the model results to reflect the real network structure with high financial fidelity, while remaining computationally tractable within a practical time limit.

This project is motivated by the practical business need to bridge this gap. It provides a systematic, data-driven simplification framework that enables companies to solve full-scale network models within a limited time. This research focuses on the supply chain segment from Distribution Centers (DCs) to customers, aiming to resolve the trade-off between solution time and planning accuracy in supply chain network design.

1.2 Problem Statement and Key Questions

To address computational bottlenecks, this project develops a structured, data-driven dimensionality reduction framework. The research is guided by two core questions and is executed through a systematic methodological approach.

Key Research Questions

- Simplification framework: What systematic data aggregation techniques can be applied to effectively reduce the number of decision variables for SCND models?
- Feasibility and Scalability: How can the feasibility of this simplification framework be rigorously tested on a smaller-scale dataset to ensure it remains robust, accurate, and fully applicable when deployed on larger-scale data?

1.3 Project Goals and Expected Outcomes

The primary objective of this project is twofold. First, it aims to develop and validate a data-driven dimensionality reduction framework designed to resolve computational bottlenecks in large-scale supply chain network models. Second, it seeks to apply this proposed framework

to solve a full-scale, real-world optimization problem, delivering an optimal network configuration that satisfies all customer demands within a time limit.

The expected outcome of this research is a scalable and validated simplification framework for solving SCND problems. By bridging the gap between mathematical precision and computational efficiency, this study enables practitioners to simplify massive supply chain design data before feeding it into models. Ultimately, by accelerating the planning process, this framework supports more informed decision-making, thereby reducing operational risk and improving the reliability of customer fulfillment under dynamic market conditions.

2. STATE OF THE PRACTICE

The primary research problem addressed in this study is how to effectively reduce the computational complexity of large-scale SCND while maintaining high solution accuracy. To achieve the objectives of this research, this chapter synthesizes prior research and industry practices on data dimensionality reduction approaches designed to overcome computational bottlenecks in SCND.

2.1 The Model for Network Planning

Mixed-Integer Linear Programming (MILP) is a widely used approach to solving SCND problems (Arampantzi et al., 2018). This approach mathematically formulates customer demands and operational constraints in the supply chain to determine optimal solutions. A classic example of such formulations is the transportation problem, which seeks the minimum-cost allocation of commodity flows from a set of suppliers to a set of customers, subject to supply capacity and demand satisfaction constraints (Williams, 2013).

The formulation of this problem involves defining decision variables, an objective function, and constraints.

Tables 1 and 2 present the sets and parameters used in the transportation problem model.

Symbol	Definition
C	The set of customers
D	The set of distribution centers

Table 1. Sets of the Transportation Problem Model

Symbol	Domain	Definition
Cost_{dc}	$\text{Cost}_{dc} \in \mathbb{R}^+$	The transportation cost from each distribution center $d \in D$ to each customer $c \in C$
Cap_{d}	$\text{Cap}_{d} \in \mathbb{R}^+$	The operational capacity of each distribution center $d \in D$

Table 2. Parameters of the Transportation Problem Model

The decision variable $x_{dc} \geq 0$ denotes by the continuous volume transported from the distribution center d to the customer c.

The objective function (1) minimizes total transportation cost:

$$\text{Minimize } Z = \sum_{c \in C, d \in D} \text{Cost}_{dc} \cdot x_{dc} \quad (1)$$

This objective is subject to two primary constraints:

Capacity Constraints:

$$\sum_{c \in C} x_{dc} \leq \text{Cap}_d \qquad \forall d \in D \quad (2)$$

Demand Constraints:

$$\sum_{d \in D} x_{dc} = \text{Demand}_c \qquad \forall c \in C \quad (3)$$

Capacity Constraints (2) ensure the total outbound flow does not exceed the operational capacity of each distribution center.

Demand Constraints (3) guarantee that every customer's exact demand is fulfilled.

In this model, all decision variables are continuous, rendering this a standard Linear Programming (LP) formulation. When the model additionally incorporates integer decision variables, such as integer shipment quantities, the problem extends to a Mixed-Integer Linear Programming (MILP) formulation.

However, the existing literature demonstrates that as the network scale expands, MILP solvers encounter computational bottlenecks (Tang et al., 2025). When a network simultaneously incorporates large numbers of distribution centers, thousands of customer nodes, various SKUs, multiple time periods, and multiple transportation modes, the combinatorial scale of decision variables grows exponentially. Models may then fail to obtain optimal solutions within acceptable time frames or encounter out-of-memory errors. Consequently, how to effectively reduce the model scale while largely preserving the solution quality is one of the challenges in SCND research.

2.2 Product Aggregation for Dimensionality Reduction

The number of products constitutes a significant source of model complexity in network planning. In practical business contexts, enterprises often maintain thousands of individual Stock Keeping Units (SKUs). Treating each SKU as an independent decision variable significantly increases the total number of variables, often exceeding the solver capacity (van Elzakker et al., 2018).

The network planning practice for McDonald's China serves as a representative case study (Tang et al., 2025). Facing the challenge of modelling more than 2,000 individual SKUs, the research team consolidates functionally similar or interchangeable SKUs into product category groups, compressing the SKU count to approximately 150 groups. This practice demonstrates that product aggregation constitutes an effective dimensionality reduction approach that balances solution quality and computational feasibility.

2.3 Spatial Clustering and Error Correction

The number of customers also constitutes a source of model complexity. Spatial clustering techniques are widely applied to reduce the complexity arising from the number of customers (Imai et al., 2023).

However, conventional spatial clustering algorithms rely on geographic distance as the similarity metric, disregarding the significant variation in freight volumes across customer nodes (Wen et al., 2016). This causes cluster centroids to fail to accurately reflect the transportation cost structure of the actual logistics network. To resolve this issue, Cheng & Zheng (2024) propose a demand-weighted distance method, incorporating the freight volume of each customer node as a weight in the calculation of the distance by K-means. This draws cluster centroids towards high-volume nodes, thereby aligning the spatial distribution of simplified demand nodes more closely with the true transportation cost structure (Figure 1).

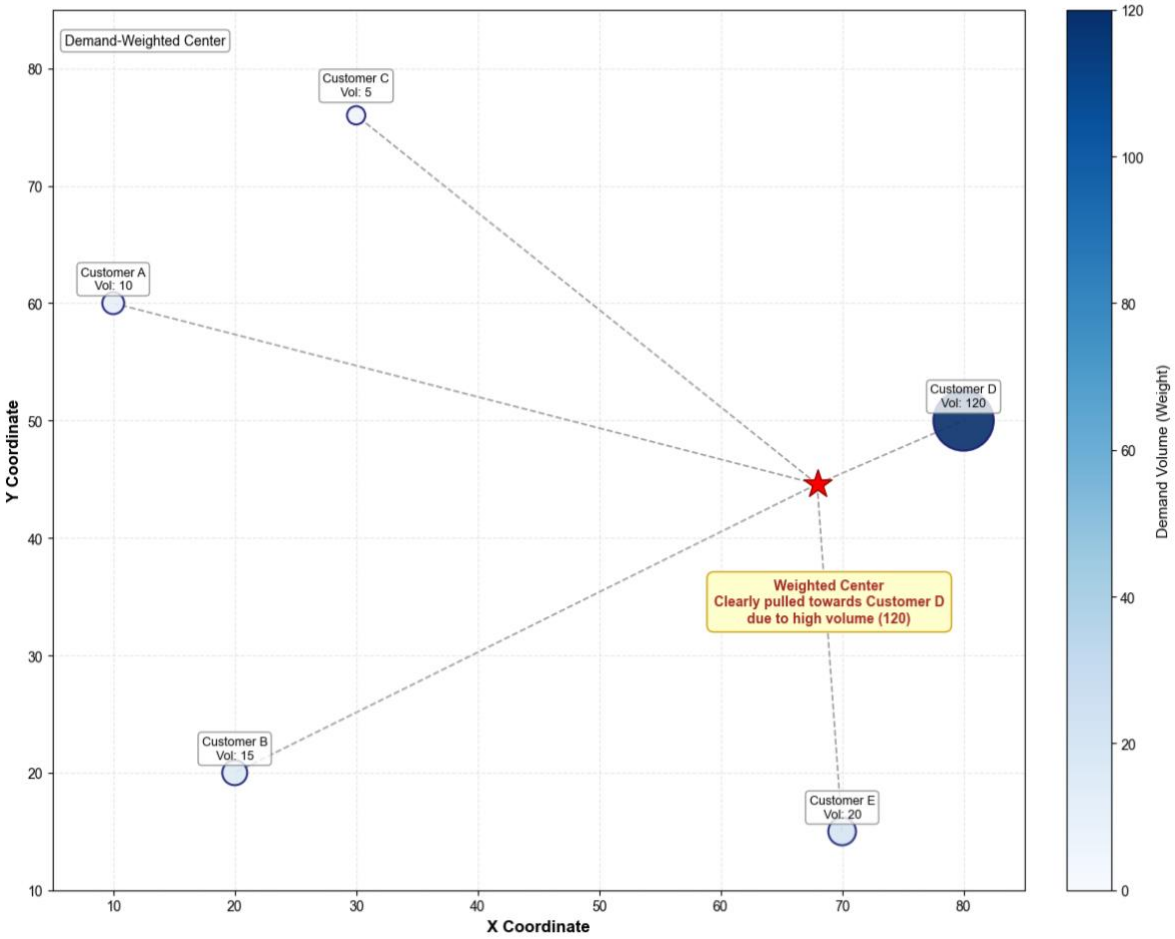


Figure 1. Demand-weighted Clustering Method

After this clustering method addresses the complexity arising from the number of customers, it still introduces cost discrepancies. Mapping multiple dispersed customers to a single centroid consolidates the transportation costs of multiple delivery routes into the cost of a single virtual direct route, causing the model's aggregate network transportation costs to be underestimated. This phenomenon is defined as "spatial aggregation error" (Hillsman & Rhoda, 1978).

Imai et al. (2023) propose a solution to this problem. After the MILP model yields the optimal solution, the virtual centroids are disaggregated back to individual customer nodes, and transportation costs are recalculated for each customer based on their assigned DC. This largely eliminates the cost underestimation introduced during the clustering phase.

2.4 Data Scaling in Distance-Based Algorithms

As established in the data mining literature, distance-based algorithms such as K-means are sensitive to extreme numerical disparities (Han et al., 2011). In the context of demand-weighted clustering for SCND, this issue is particularly pronounced: freight volumes and geographic coordinates can differ by several orders of magnitude, causing the clustering outcome to be dominated by volume rather than spatial distribution. As a result, cluster centroids may fail to accurately represent the true geographic structure of the customer base, undermining the spatial fidelity of the simplified network. It is therefore essential to ensure that the numerical scales of input variables are appropriately balanced before applying K-means clustering.

2.5 Summary

In summary, the current literature and industry practices establish MILP as one of the primary approaches to solving supply chain network planning problems. However, the massive volume of data in large-scale networks exceeds the processing capacity of solvers, creating a severe computational bottleneck. Prior research has introduced various dimensionality reduction techniques, such as demand-weighted spatial clustering and product aggregation. However, these methods are often implemented in isolation, and no existing work offers a systematic approach that integrates these techniques to achieve rapid solution times without significantly compromising solution quality.

3. METHODOLOGY

The methodology developed in this study is structured into three steps, designed to fulfill two primary objectives: to develop and validate a data-driven dimensionality reduction framework that resolves the computational bottlenecks of large-scale models, and to apply this framework to solve a full-scale, real-world optimization problem.

The systematic approach follows these three core steps (Figure 2):

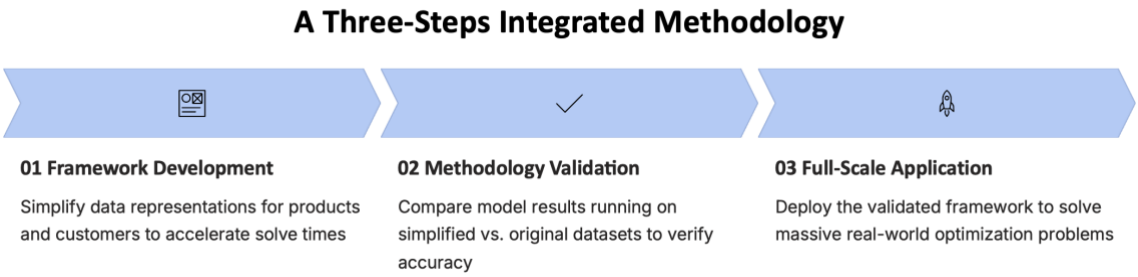


Figure 2. The Structure of Methodology

Step 1: Framework Development: By integrating the dimensionality reduction methods identified in the State of the Practice, this study constructs a comprehensive and systematic simplification framework.

Step 2: Methodology Validation: This step conducts a series of rigorous numerical analyses to demonstrate that the model using processed data preserves solution quality within a limited solving time.

Step 3: Full-Scale Application: This step applies the framework to the complete real-world supply chain network to deliver an optimal solution within practical computational time limits.

3.1 Framework Development

This section details the first step of the framework development methodology introduced in Figure 2. As illustrated in Figure 3, the simplification framework is implemented in three sequential steps, comprising product aggregation, customer clustering, and cost correction.

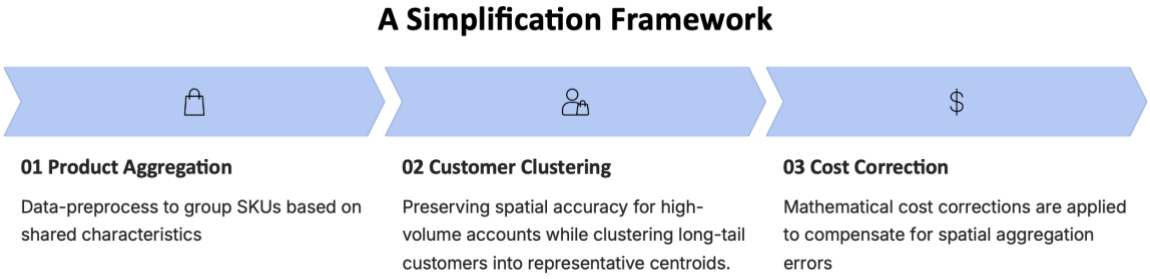


Figure 3. The Steps of the Simplification Framework

3.1.1 Product Aggregation

The first dimension of complexity in SCND is the variety of products. As outlined in Step 1 of Figure 3, and based on the product aggregation practice established in Section 2.2, this study first determines whether full detail at the SKU level is necessary and, where appropriate, aggregates individual SKUs into representative product families.

3.1.2 Customer Clustering

The second dimension of complexity comes from the number of customers. As outlined in Step 2 of Figure 3, and drawing upon the spatial clustering techniques and data scaling considerations reviewed in Sections 2.3 and 2.4, clustering is applied to reduce the number of customer nodes to a computationally manageable scale. To address the scaling effect in demand-weighted K-means, critical customers are intentionally isolated to preserve their full granularity and ensure relatively precise spatial mapping.

For the remaining long-tail customers, a demand-weighted k-means algorithm is then applied to group their locations, mathematically drawing cluster centroids toward relatively high-volume nodes. In this way, thousands of minor delivery points are replaced by a limited number of representative cluster centroids, significantly compressing the number of customer nodes and enabling the solver to find the optimal solution in a fraction of the time.

3.1.3 Cost Correction

Although clustering substantially reduces computational time, it inevitably introduces spatial aggregation errors. As illustrated in Step 3 of Figure 3, and based on the cost correction practices reviewed in Section 2.3, a post-optimization cost correction is performed to compensate for the underestimated costs.

3.2 Validation

As outlined in Figure 2, validating the framework is the second step. To this end, a network planning model is first established to obtain the results required for the validation analysis. As reviewed in Section 2.1, the MILP model satisfies this requirement. After the model is established, a comparative evaluation approach is employed to assess the effectiveness of the simplification framework. By feeding both the simplified data and the unprocessed data into the same MILP model separately and comparing the model outputs, the effectiveness and accuracy of the simplification framework can be rigorously verified.

3.2.1 MILP Model Formulation

Building upon the transportation problem model reviewed in Section 2.1, this study formulates the following MILP model.

Tables 3 and 4 present the sets and parameters used in the MILP model.

Notation	Description
W	The set of Distribution Centers (DCs)
C	The set of customers
A	The set of arcs from W to C
P	The set of products
T	The set of Times
Z	The set of storage conditions
M	The set of transportation modes
P_z	The set of products that require storage condition Z

Table 3. Sets of the Model

Notation	Domain	Description
s_w^z	$s_w^z \in \mathbb{R}^+$	The capacity of each DC $w \in W$ under a specific storage condition $z \in Z$
$\gamma_c^{t,p}$	$\gamma_c^{t,p} \in \mathbb{R}^+$	The demand of a specific customer $c \in C$ for a specific product $p \in P$ at a specific time $t \in T$.
f_{wc}^m	$f_{wc}^m \in \mathbb{R}^+$	The fixed cost of each transportation mode $m \in M$ on different arcs $(w, c) \in A$
v_{wc}^m	$v_{wc}^m \in \mathbb{R}^+$	The variable cost of each transportation mode $m \in M$ on different arcs$(w, c) \in A$
k_m	$k_m \in \mathbb{Z}^+$	The capacity of each transportation mode

Table 4. Parameters of the Model

Table 5 presents the variables used in the MILP model.

Notation	Domain	Description
$x_{wc}^{t,p,m}$	$x_{wc}^{t,p,m} \in \mathbb{R}_{\geq 0}$	The flow volume from each DC $w \in W$ to each customer $c \in C$ at a specific time $t \in T$ for a specific product $p \in P$ via a transportation mode $m \in M$
$h_{wc}^{t,m}$	$h_{wc}^{t,m} \in \mathbb{Z}_{\geq 0}$	The number of trucks from each DC $w \in W$ to each customer $c \in C$ at a specific time $t \in T$ via a transportation mode $m \in M$

Table 5. Variables of the Model

The objective function (4) minimizes total transportation cost:

$$\text{Minimize} \sum_{w \in W} \sum_{c \in C} \sum_{t \in T} \sum_{m \in M} \left[h_{wc}^{t,m} \times f_{wc}^{m} + \sum_{p \in P} \left(x_{wc}^{t,p,m} \times v_{wc}^{m} \right) \right] \tag{4}$$

This objective is subject to the following three constraints:

$$\sum_{c \in C} \sum_{p \in P_z} \sum_{m \in M} x_{wc}^{t,p,m} \leq s_w^z, \qquad \forall w \in W, z \in Z, t \in T \tag{5}$$

$$\sum_{w \in W} \sum_{m \in M} x_{wc}^{t,p,m} = \gamma_c^{t,p}, \qquad \forall c \in C, t \in T, p \in P \tag{6}$$

$$\sum_{p \in P} x_{wc}^{t,p,m} \leq h_{wc}^{t,m} \times k_m, \qquad \forall (w,c) \in A, t \in T, m \in M \tag{7}$$

Constraint (5) ensures that the total flow out of each DC does not exceed its storage capacity. Constraint (6) ensures that the total flow received by each customer equals their demand. Constraint (7) ensures that the total flow assigned to each transportation mode does not exceed its capacity.

3.2.2 Comparative Analysis

As the full-scale dataset is computationally intractable for solvers, the validation is executed on a solvable subset of the original data. This simplification framework is considered valid only if the MILP results from aggregated or clustered data match those obtained from the unprocessed data, or if the deviation remains within an acceptable range.

To ensure the effectiveness of each step in the simplification framework, validation proceeds sequentially. First, the accuracy of product aggregation is verified by comparing the MILP results for aggregated SKUs with those for the original, disaggregated SKUs. Only if the results of this step are confirmed to be consistent within the acceptable threshold can the aggregated SKUs serve as the product data input for the subsequent validation.

Second, the customer clustering method is validated. The initial MILP results for clustered customers are subjected to a post-optimization cost correction to accurately reflect the true costs. These corrected costs are then evaluated against the results obtained from solving the model with the original customers. This sequential structure prevents error compounding and ensures that the accuracy of each distinct method is assessed independently.

3.3 Full-Scale Application

Once validated on a small-scale dataset, the framework is applied to the full-scale dataset, as illustrated in Step 3 of Figure 2. When the model is run on the full-scale data, the solution time and optimality gap of the model using the data processed by the simplification framework, as well as those of the model using the unprocessed data, are recorded separately. Those records are used to demonstrate the effectiveness of the simplification framework.

Sections 3.1 through 3.3 collectively complete the development of the simplification framework, fulfilling the first objective of this project. The numerical validation and full-scale application results are presented in Section 4.

4. RESULTS AND DISCUSSION

Based on the simplification framework introduced in the previous section, this chapter details its application to a real-world dataset provided by the project sponsor. Analysis of the provided data reveals that the target network serves the healthcare industry. As outlined in Table 6, the customer nodes represent individual hospitals, the DCs serve as warehouses that fulfil these healthcare demands, and the SKUs denote the specific pharmaceutical products required by different customers.

Entity	Count	Key Attributes
Customer	5,127	Geographical coordinates (latitude/longitude), city, state, zip code; facility type (hospital ownership, emergency services)
DCs	40	Location (city, state, latitude/longitude), operational status, scale category; storage capacity by type (rack, flammable, temperature-controlled, secured, total) in pallets
Products (SKUs)	2,410	Unit cost (per pallet), storage area requirement (e.g., rack, flammable)
Transportation lanes	DC × Customer	Distance (miles), travel time (hours); FTL: capacity, fixed cost; LTL: capacity, fixed cost, variable cost (per kg)
Time periods	12	Monthly planning horizons
Demand	Customer × SKU × 12	Monthly demand volume (pallets) per customer-product pair across 12 time periods

Table 6. Detailed Dataset Information

As illustrated in Figure 4, the network comprises a vast number of granular customer nodes (blue points) and distribution centers (red pins). This mapping highlights the significant computational burden of the full-scale problem, underscoring the necessity of the dimensionality-reduction framework developed in this study.

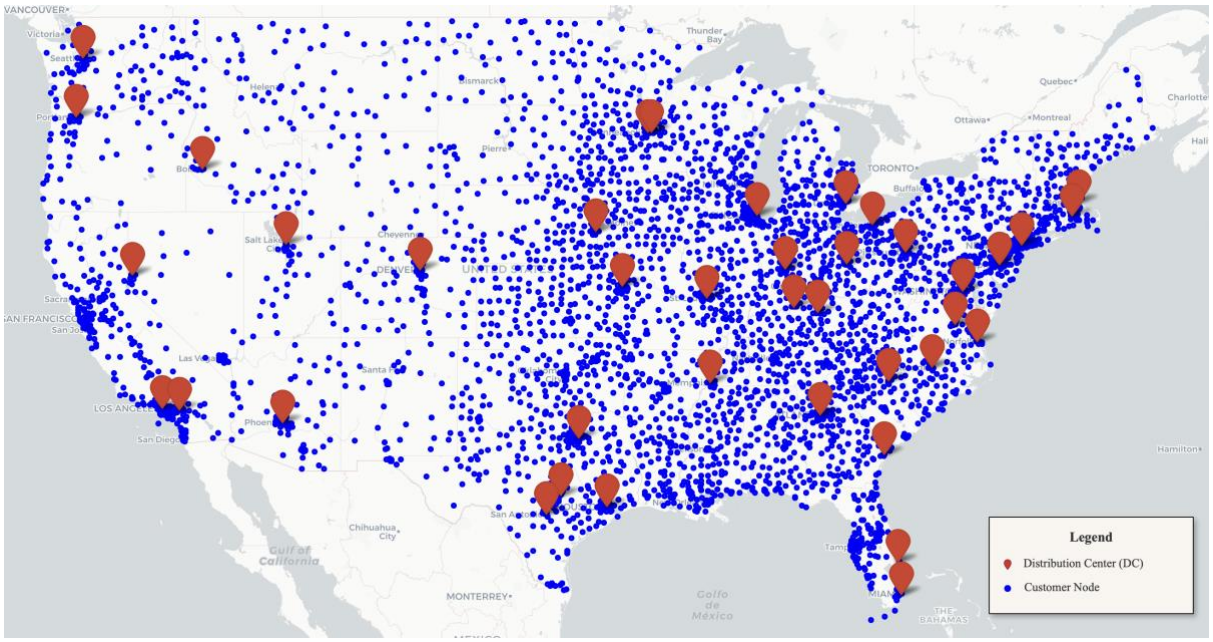


Figure 4. Spatial Distribution of Original Customer Nodes and DCs

4.1 Implementation of the Simplification Framework

Following the framework established in Section 3.1, three simplification steps are applied sequentially.

First, products are aggregated by storage condition, as DC capacity constraints in this network are governed by storage type rather than individual SKU attributes. This aggregation is shown to preserve solution quality on the small-scale dataset (see Section 4.2.1).

Second, customers are segmented by demand concentration: the top 22% of customers, who account for 80% of total demand, are retained as independent nodes, while the remaining long-tail customers are clustered.

Third, a post-optimization cost correction is applied to compensate for the cost underestimation introduced by clustering. The combined effect of customer clustering and the subsequent cost correction is examined in Section 4.2.2.

Once each simplification step has been verified to preserve solution quality on the small-scale dataset, the integrated framework is then applied to solve the full-scale problem. The corresponding results are presented in Section 4.3.

4.2 Validation Results on the Small-Scale Dataset

The validation results are organized into two parts, corresponding to the two simplification steps: product aggregation (Section 4.2.1), and customer clustering with the subsequent cost correction (Section 4.2.2).

4.2.1 Validation Results of Product Aggregation

After aggregating the 2,410 SKUs into four representative categories by storage conditions, the MILP model yields solutions identical to those obtained from the original disaggregated data. Table 7 compares the outputs before and after product aggregation across four test scenarios. In the most complex scenario, the number of decision variables is reduced by over 99%, dropping from approximately 48 million to 120,000. Critically, this reduction does not compromise solution accuracy. Across all scenarios, not only does the Truck Gap remain zero, but total cost is also identical. These results validate the first simplification step. The

aggregated SKUs are therefore used as the product data input for the subsequent validation of customer clustering.

Scenario	Customer	DC	Total Demand	Total Cost	Trucks	Before Aggregation Decision Variables	After Aggregation Decision Variables	Truck Gap
1	20	10	19,627	901,305	865	5,980,800	24,000	0
2	50	5	29,705	1,537,652	1,471	9,144,000	30,000	0
3	100	8	69,154	3,211,139	3,798	38,476,800	96,180	0
4	100	10	69,154	3,211,139	3,798	48,096,000	120,000	0

Table 7. Results of the Product Aggregation Analysis

4.2.2 Validation Results of Customer Clustering

To ensure that the framework is applicable to the full dataset, the demand-weighted k-means clustering algorithm requires the specification of the number of clusters, k. Rather than selecting a case-specific value, this study defines the compression ratio as the number of clusters divided by the total number of long-tail customers, and systematically evaluates its effect on solution accuracy and computational time across networks of varying scales.

4.2.2.1 Determining the Compression Ratio

Prior research has established an optimality gap of 0.5% to 2.0% as an acceptable stopping criterion (Tang, 2018). To ensure solution accuracy, this study adopts a relatively strict threshold of 0.5%. The MILP model is solved across varying numbers of customers (ranging from 100 to 2,000) and different cluster sizes (ranging from 10 to 200). The solutions are then benchmarked against the results obtained using the unprocessed data in the same MILP model, yielding the following pattern. As illustrated in Figure 5, when the compression ratio is excessively low (< 4%), the simplified data fundamentally under-represents the geographic distribution of customers, leading to a highly volatile error distribution. However, a substantial reduction in variance is observed in the 5%–6% range. In this zone, the median error falls below the 0.5% accuracy threshold. Compression ratios beyond 5% continue to reduce the error, but with diminishing marginal returns. This establishes a baseline: a compression ratio of approximately 5.0% is sufficient to achieve robust accuracy regardless of the underlying customer base size.

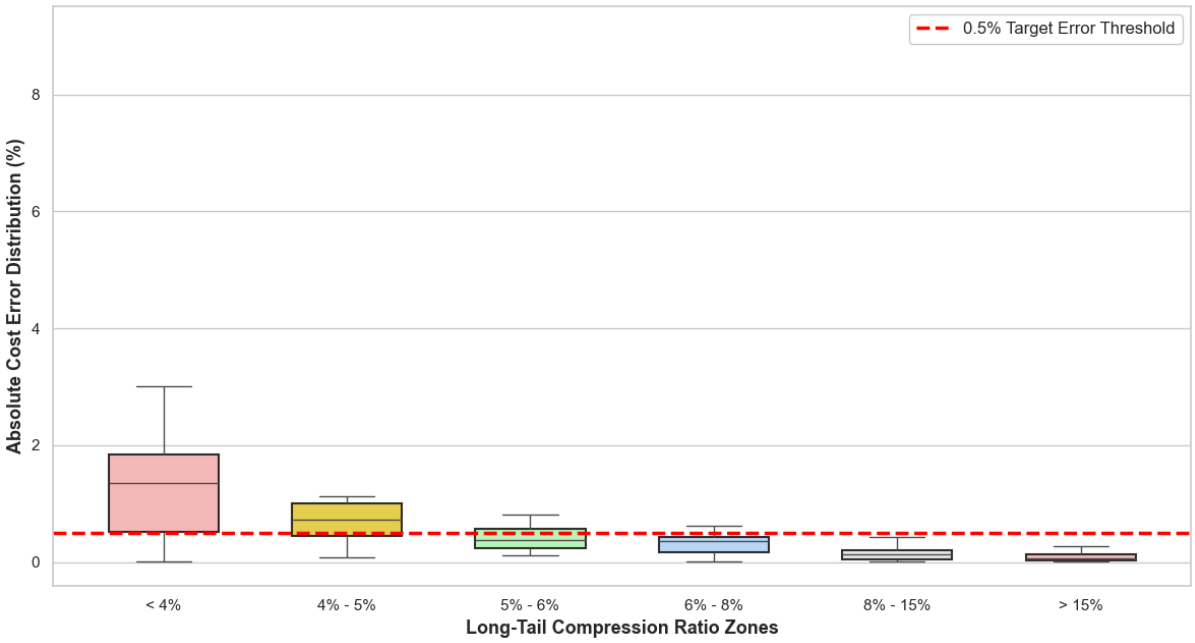


Figure 5. Accuracy Distribution by Compression Ratio

4.2.2.2 Solution Time Vs. Accuracy Trade-Off

Based on the compression ratio analysis, it is necessary to examine the trade-off between solution time and model accuracy (Figure 6). Within the 4%–6% compression ratio interval, the solution time remains relatively stable, while the error drops significantly. Beyond this range, the gains in accuracy exhibit diminishing marginal returns, whereas the solution time increases rapidly. Consequently, a compression ratio of approximately 5% represents a favorable operating point.

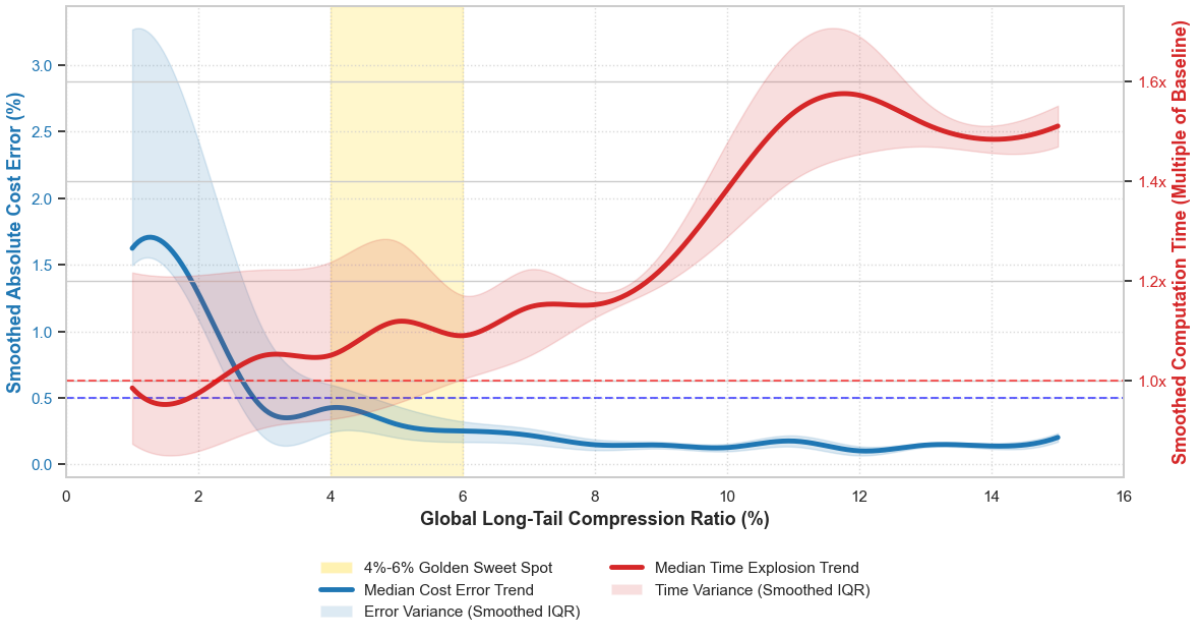


Figure 6. Accuracy Vs. Computation Tradeoff

4.2.2.3 Out-of-Sample Validation

To further validate the robustness of the 5% compression ratio, out-of-sample tests are conducted on large-scale networks with 2,000 to 3,500 customers (Figure 7). By applying a 5%

compression ratio and limiting the solution time to one hour, the relative error for all samples remains within 0.5%.

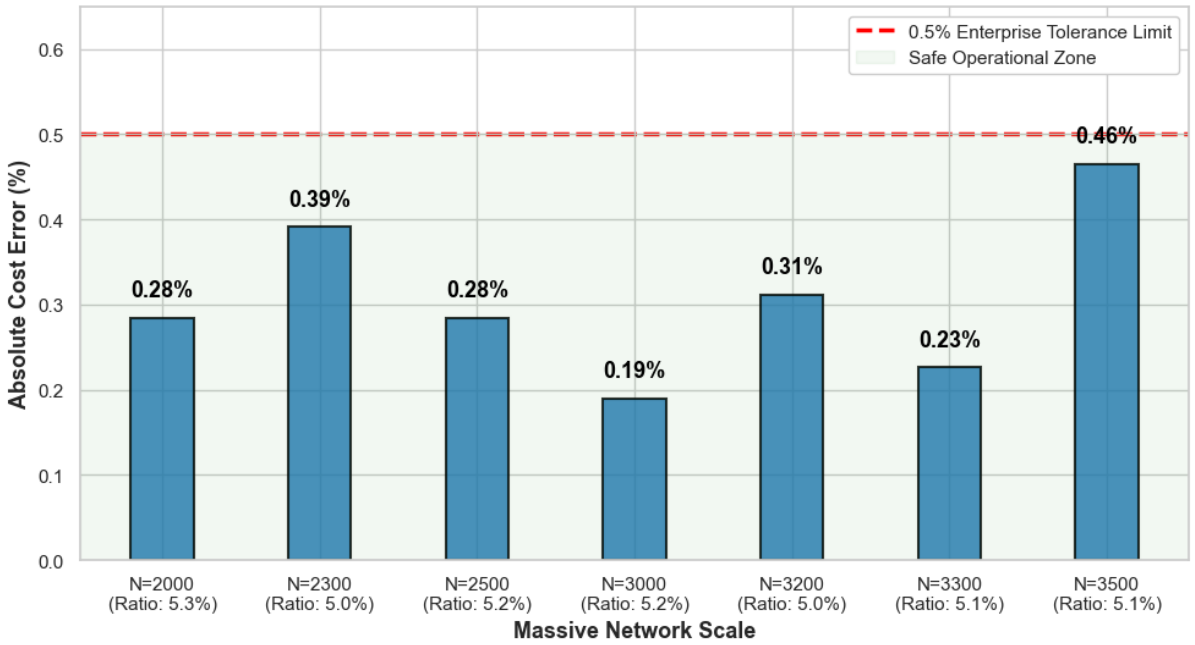


Figure 7. Out of Sample Validation: Accuracy at ~5% Compression Ratio

Overall, incorporating product aggregation, isolating the core customers who represent 80% of demand, and clustering the remaining long-tail customers at a 5% compression ratio have been validated as feasible in terms of both solution time and quality. We are therefore confident in applying this simplification framework to the full-scale dataset.

4.3 Application to the Full-Scale Network

This section reports the results of applying the simplification framework to the full-scale network. The MILP model is solved using the Gurobi Optimizer, comparing two configurations: the fully simplified dataset (with both product aggregation and customer clustering applied) against a baseline with only product aggregation applied.

Compared with the dense spatial distribution shown in Figure 4, the post-clustering customer distribution in Figure 8 exhibits a substantial reduction in network complexity.

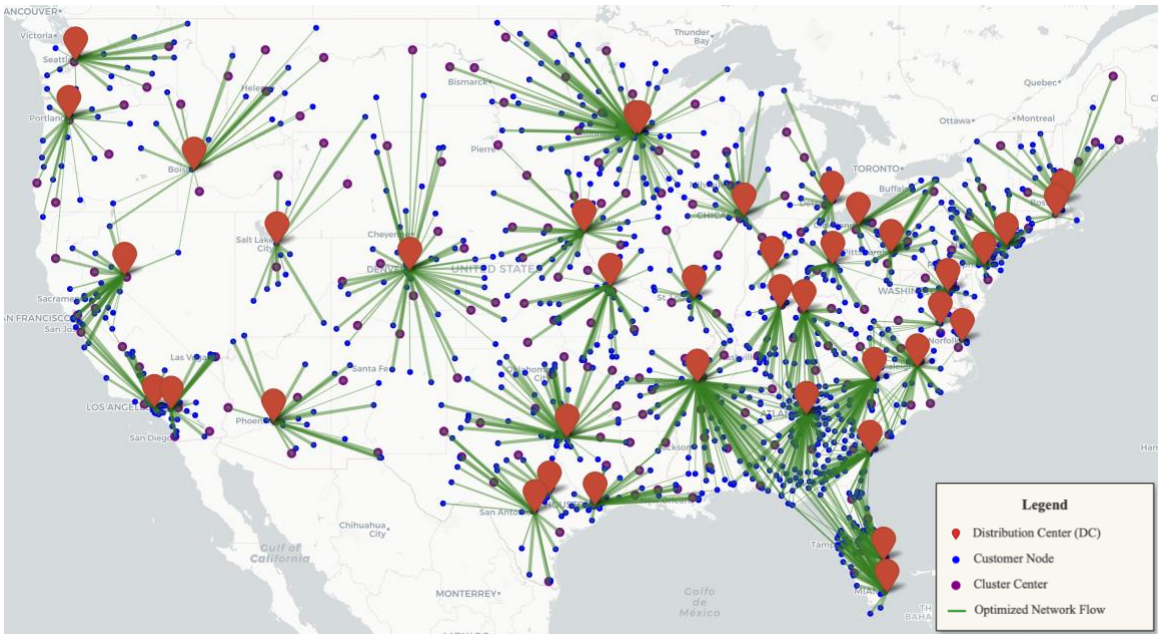


Figure 8. Optimized Supply Chain Network Configuration

Based on the results in Figure 9, the optimization process for the baseline configuration, with only product aggregation applied, encounters difficulties. It progresses slowly and ultimately fails due to an out-of-memory (OOM) error at 4,342 seconds (approximately 1.2 hours). At the point of failure, an acceptable optimality gap has not been reached.

In contrast, the model utilizing the fully simplified dataset, with both product aggregation and customer clustering applied, avoids this computational bottleneck. It terminates successfully without memory failures and, within the 2-hour time limit, converges to an optimality gap of 0.199%.

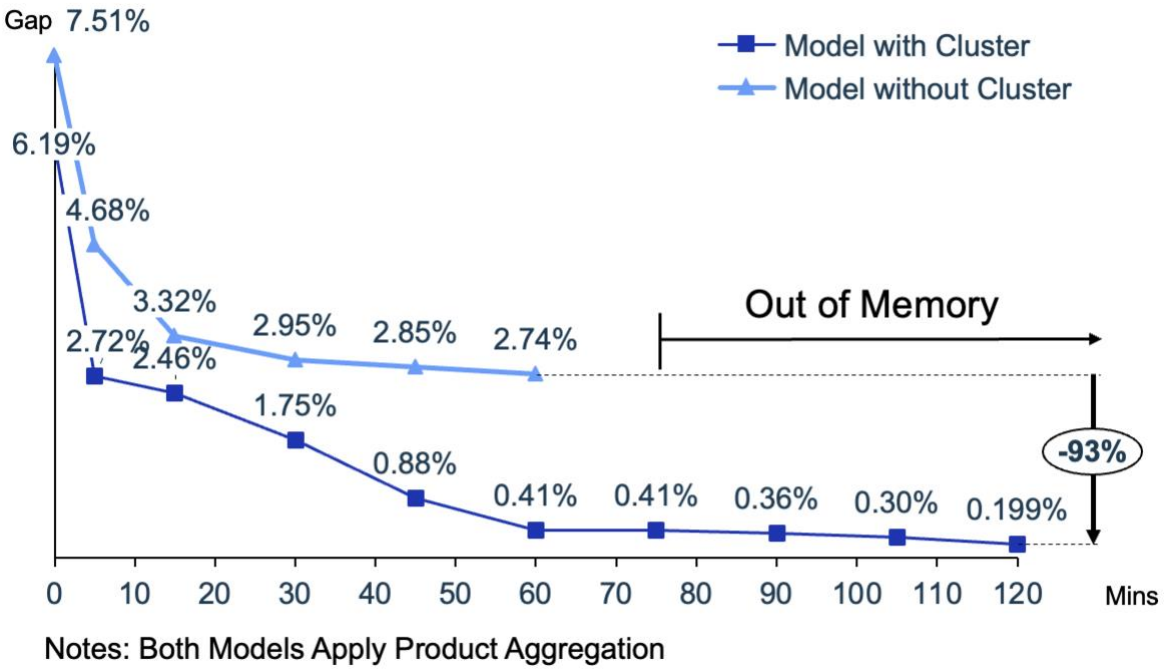


Figure 9. Convergence of the Optimality Gap Over Time

Through the validation and comparative analysis of the full dataset, the simplification framework proposed in this study is demonstrated to be feasible, enabling the model to handle the full-scale problem and generate results within a reasonable timeframe.

4.4 Discussion and Limitations

The results of this study carry significant implications for companies seeking to apply this simplification framework in practice. The validated simplification framework resolves a bottleneck that previously rendered full-scale network design computationally intractable, enabling the solver to converge to an optimality gap of 0.199% within the 2-hour time limit.

Beyond resolving the immediate computational barrier, the framework provides direct strategic value. By reducing the number of decision variables while maintaining solution quality within an accepted error margin, the framework enables companies to conduct rapid scenario analysis and network replanning. In a volatile market environment where fuel prices, customer demand patterns, and logistics costs fluctuate significantly, the ability to re-evaluate and adjust the distribution network configuration within hours directly reduces the risk of financially suboptimal network decisions at scale.

Despite these contributions, two limitations of the current framework should be acknowledged. The product aggregation based on storage conditions is determined by existing network constraints and confirmed through empirical validation. Since the network planning problem in this project only covers the segment from the DC to the customer and does not involve upstream stages such as manufacturing, product aggregation for an end-to-end network planning scenario should also be evaluated on a case-by-case basis.

In addition, this study identifies 5% as the optimal compression ratio for long-tail customer clustering. However, this parameter is derived from a single company's dataset. Networks with different industry characteristics, geographic distributions, or varying degrees of customer demand concentration may yield a different optimal compression ratio. Therefore, when transferring this framework to other contexts, it is recommended to conduct a detailed analysis of the compression ratio before adopting it as a universal parameter.

5. RECOMMENDATIONS

This capstone project addresses a critical computational challenge in large-scale supply chain network design: when planning networks that encompass thousands of customer nodes, dozens of distribution centers, and thousands of product variants, state-of-the-art optimization solvers frequently fail due to excessive computation time or out-of-memory errors. To resolve this, the project develops and validates a systematic, data-driven simplification framework that reduces the dimensionality of the dataset while preserving solution quality. Validation on a smaller-scale dataset confirms that the framework maintains solution accuracy within 0.5%, and a full-scale application demonstrates that the solver, previously unable to produce a feasible result, now converges to an optimality gap of 0.199% within 2 hours. Based on these findings, this chapter provides management recommendations for implementing the framework and outlines directions for future research.

5.1 Management Recommendations

The simplification framework proposed in this project has been validated. In practice, management can apply this simplification framework during the network planning process to reduce solution time. By facilitating rapid calculations, this framework enables management

to make timely decisions regarding the supply chain network, even in the face of volatile market conditions.

From a strategic perspective, the ability to solve large-scale network design problems in a timely manner supports several key supply chain objectives. It enables the company to evaluate alternative network configurations in a timely manner, thereby reducing the risk of financially suboptimal distribution decisions. More broadly, faster and more reliable network planning strengthens the company's ability to maintain service levels and customer fulfillment rates, which are foundational to customer retention and long-term competitiveness. By empowering planners to iterate on network designs within a single working session rather than waiting hours for an inconclusive result, this framework supports more agile and data-driven supply chain management at the strategic level.

5.2 Future Work

The simplification framework established in this project has resolved the computational bottleneck for the segment from DCs to customers. Moving forward, future research can draw upon this framework to address end-to-end network planning that encompasses upstream stages such as procurement and manufacturing, with appropriate adjustments for the added complexity of those stages. Furthermore, current planning models have not yet integrated the complex scenarios encountered in cross-border supply chains, such as customs clearance and health and safety regulations. Therefore, future research on the simplification framework could further explore more complex scenarios within cross-border supply chains. Such advancements can further empower multinational corporations to respond swiftly to global macroeconomic changes.

REFERENCES

Arampantzi, C., Minis, I., & Dikas, G. (2018). A strategic model for exact supply chain network design and its application to a global manufacturer. *International Journal of Production Research*, *57*(5), 1371–1397.

Cheng, X., & Zheng, F. (2024). Research on the application of weighted distance K-means clustering algorithm with capacity constraint in express service location. *2024 12th International Conference on Information Systems and Computing Technology (ISCTech)*, 1–5.

Garcia, D.J. and You, F. (2015). Supply Chain Design and Optimization: Challenges and Opportunities. Computers & Chemical Engineering, 81(81), pp.153–170.

Han, J., Kamber, M., & Pei, J. (2011). *Data mining: Concepts and techniques* (3rd ed.). Morgan Kaufmann.

Hillsman, E. L., & Rhoda, R. (1978). Errors in measuring distances from populations to service centers. *The Annals of Regional Science*, *12*(3), 74–88.

Imai, R. A. M., Barbieri da Cunha, C., & Sauter Guazzelli, C. (2023). An analysis of the impact of demand aggregation on the solution quality for facility location problems. *Transportes*, *31*(3), Article e2801.

Tang, C.-H. (2018). Two-stage stochastic modeling of transportation outsourcing plans for transshipment centers. *A Quarterly Journal of Operations Research*, 16, 67–94.

Tang, S., Wang, L., Shi, Y., Li, A., Lin, K., Xiang, C., Niu, S., Zhou, S., Liu, M., & Tang, H. (2025). McDonald's China adopts operations research for network design. *INFORMS Journal on Applied Analytics*, *55*(1), 36–47.

van Elzakker, M. A. H., Zondervan, E., Raikar, N. B., Hoogland, H., & Grossmann, I. E. (2018). *An SKU decomposition algorithm for the tactical planning in the FMCG industry (Version 1)*. Carnegie Mellon University.

Wen, R., Yan, W., & Zhang, A. N. (2016). Weighted clustering of spatial pattern for optimal logistics hub deployment. In *2016 IEEE International Conference on Big Data (Big Data)* (pp. 3186–3187).

Williams, H. P. (2013). Model building in mathematical programming (5th ed.). Wiley.