A Distribution Network Design Model Using Data Classification and Fleet Optimization

Mohammad Amirahmadi [a], Hamid Esmaeeli [a*], Kia Parsa [a], Amin mostafaee [b]

m.amirahmadi@iau.ac.ir, hamidesmaeeli@iau.ac.ir, kiaparsa@iau.ac.ir, a_mostafaee@iau.ac.ir

[a] Department of Industrial Engineering, NT.C., Islamic Azad University, Tehran, Iran

[b] Department of Mathematics, NT.C., Islamic Azad University, Tehran, Iran

* Corresponding author:

Email address: hamidesmaeeli@iau.ac.ir (Hamid Esmaeeli)

Abstract

This study seeks to bridge the existing gaps in previous researches by introducing a comprehensive data-driven network design model. The process begins with an in-depth analysis of customer demand, utilizing unsupervised learning algorithms to gain valuable insights into consumer behavior. This analysis will help identify demand levels across various geographical regions while uncovering patterns that fluctuate over time. These insights will serve as essential inputs for the network design model. To facilitate effective data classification and analysis, the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm will be employed, enabling accurate estimation of customer demand based on innovative parameters. Building upon these findings, a new mathematical model will be created that incorporates fleet optimization constraints. Importantly, during this modeling process, emphasis will be placed not only on optimizing the number, location, and capacity of facilities but also on refining fleet types and their compositions to enhance overall efficiency. Due to the complexity of the model, it will be solved using various numerical case problems. Due to the complexity of the model, it will be solved using various numerical case problems. The results demonstrate that the proposed data-driven model achieves an average profit improvement of 10-15% compared to traditional non-clustered approaches. Furthermore, the model yields noticeable cost savings of approximately 8-12% in transportation and fleet-related expenses. Furthermore, the integrated nature of the model allows for an examination of key parameters to extract valuable managerial insights, demonstrating the synergy between data-driven clustering and mathematical optimization for distribution network design.

Keywords: Machine-learning, Distribution systems, Fleet optimization, DBSCAN algorithm, Revenue management, Demand pattern recognition.

1-Introduction

Supply chains are becoming increasingly complex, and companies must adopt innovative methods to optimize their operations. These challenges are further intensified by growing uncertainties in demand and logistics Juris (2020). It can be said that in any supply chain, efficient management at each level of the chain, as well as comprehensive management across all levels, leads to efficiency and effectiveness in the flow of materials, products, services, and information among various levels Perl and Sirisoponsilp (1988). The design of a supply chain network determines the structure and configuration of a network, as well as its costs and service quality Ziari and Taleizadeh (2025). Supply chain network design involves a wide range of decisions, including long-term considerations such as evaluating the number, capacity, and location of supply chain facilities, medium-term planning such as distribution, transportation, and inventory management policies, and finally, short-term planning to meet customer demands Janjevic, Merchán et al. (2021).

The network design problem is one of the most important strategic decisions, which involves determining the number and locations of raw material suppliers, manufacturing plants, intermediate warehouses, and distribution centers, along with selecting distribution channels from suppliers to customers and determining the transportation size between facilities over a broad time horizon Li, Han et al. (2023). To design a supply chain network, the entire network has been divided into three subnetworks: the internal network, the distribution network, and the external network. A common goal in designing a distribution network is to create a system with minimal costs while meeting the demands of retailers without exceeding the capacities of warehouses and factories. This issue always involves creating balances between the system cost components, such as setup and operational costs of factories and internal and external transportation costs.

Creating a distribution system that integrates data mining tools is crucial in today's data-driven environment. As the volume and variety of data continue to grow, employing these tools to analyze data and identify valuable insights has become indispensable Hung Lau (2012). Implementing a distribution system enriched with data mining capabilities can lead to substantial enhancements in performance and optimization of distribution activities. Through the examination of demand trends, purchasing patterns, customer behavior, and other relevant factors, organizations can achieve more precise planning for the distribution of goods and services, which in turn improves strategic decision-making in logistics Shi, Qiu et al. (2020). Utilizing data mining tools in distribution networks can facilitate the optimization of warehouses, scheduling of routes, management of inventory, forecasting of demand, and overall resource efficiency. These tools are adept at processing detailed data and applying various algorithms aimed at improving the operational efficiency and effectiveness of distribution systems, ultimately boosting customer satisfaction while minimizing costs throughout the distribution chain Chen, Wu et al. (2016).

The need for a cohesive and agile distribution system that responds to demand data is increasingly recognized, particularly given the disruptions caused by the pandemic in supply and distribution networks. Additionally, as supply chains globalize and vertical integration declines, unforeseen events in any segment of the chain can trigger far-reaching consequences. Enhancing the efficiency of the transportation fleet can lead to significant cost savings related to fuel, labor, maintenance, and repairs. By optimizing transport routes and schedules, the delivery times can be reduced, thereby elevating service quality. Moreover, incorporating sustainable transportation methods can mitigate environmental pollution, contributing to ecological preservation Wang, Gunasekaran et al. (2018).

These considerations are vital in designing distribution systems that can effectively respond to customer needs. Therefore, this paper seeks for simultaneously analyzing demand trends and refining fleet composition to fill existing gaps in the literature, facilitating the design of a responsive system capable of catering to a larger customer base. This initiative aims to enhance performance by introducing a novel design model. Utilizing concepts of location, optimizing communication processes, and improving transportation fleet management within the proposed model will be essential for the development of an effective final framework considering uncertainty.

The article is organized as follows: Section 2 reviews the current literature on network design problems. In the Section 3, the proposed model is presented, followed by the Section 4, where the mathematical model is solved and determined various options. To illustrate the model application, a numerical example is implemented in Section 5. Finally, the results and outline directions for future research is discussed in Section 6.

2-Literature review

To gain insight into the existing literature and perform a comprehensive review of prior research, it is important to provide a concise introduction to the most significant and pertinent studies in this area. This section will focus on those key studies. From the earliest researches on this issue, Xin-hua and Jin (2007) explored a demand fulfillment issue characterized by uncertain demand and unlimited capacity at distribution centers. Their paper examined various costs including facility location, transportation, inventory holding, and safety stock. Miranda and Garrido (2008) assessed a three-tier system that involved the location of distribution centers (wholesalers), the order quantities from wholesalers based on the desired service level, and the anticipated normally distributed demand. This study accounted for maximum order capacity and inventory capacity for each wholesaler, utilizing a modified Lagrangian relaxation method to solve the problem. Van Wijk, Adan et al. (2012) and Shavandi and Bozorgi (2012) developed inventory location models considering Poisson and fuzzy demand scenarios, respectively. Also Jha, Somani et al. (2012) addressed location problems without incorporating inventory policies, focusing on facilities with limited capacity. The study assumed deterministic demand and considered a multi-product flow, aiming to minimize location and transportation costs.

Shahabi, Unnikrishnan et al. (2014) present a three-level inventory location problem that considers demand interactions. Their analysis covers the location of factories and wholesalers, the allocation of wholesalers to factories, and the assignment of retailers to wholesalers. Additionally, the model assesses safety stock levels and the amount of inventory held at each wholesaler, aiming to reduce inventory, transportation, and location costs. They calculate transportation costs between factories and wholesalers, as well as between wholesalers and retailers, based on Euclidean distances and assess order times based on these calculated distances. Gzara, Nematollahi et al. (2014) investigate a comprehensive decision-making model that consists of two sets of decisions. The first set pertains to strategic choices about locating distribution centers and assigning them to customers, while the second set focuses on tactical and operational decisions regarding inventory levels at each distribution center. Traditionally, these decisions were considered separately, but the authors propose viewing them simultaneously. The primary objective of this model is to minimize overall costs, including expenses associated with opening distribution centers, transportation, and holding inventory.

Ahmadi-Javid and Hoseinpour (2015) explore the location problem to maximize the revenue of a distribution network within a supply chain, dealing with price-sensitive demand and a multi-product flow. This issue is approached using mixed-integer nonlinear programming and solved with the Lagrangian relaxation method for both limited and unlimited capacity scenarios. Behnamian, Fatemi Ghomi et al. (2018) studied location issues related to truck allocation and scheduling in multiple terminals, although their two-stage model did not account for delivery time windows for customers. Simultaneously, Theophilus, Dulebenets et al. (2021) focused on truck scheduling in a cold chain-related cross-docking terminal to reduce costs, yet they did not incorporate storage capacity constraints for temporary storage in their proposed model.

In the realm of urban rapid service systems, there has been an examination of vehicle and hub capacities Wu, Qureshi et al. (2022). To lower costs, routing and location-inventory problems for perishable products in a three-level supply chain in China were analyzed using a mixed-integer linear programming (MILP) model, although service levels were not a focal point in this analysis Song and Wu (2023). Moreover, for waste management within a two-level supply chain, Caramia and Pizzari (2022) addressed location-allocation issues. Following this, Fahmy, Zaki et al. (2023) utilized a mixed-integer nonlinear programming (MINLP) model to examine facility location-allocation within a supply chain, specifically to reduce costs associated with perishable products. Other research has focused on location-allocation problems in various fields, including disaster response, healthcare, food supply chains, and facility planning. Additionally, Hasani Goodarzi, Zegordi et al. (2021) investigated the location of cross-docking warehouses through a MINLP model at an automotive company in the Middle East.

Simultaneous collection and delivery with network design models are considered by Vincent, Aloina et al. (2023), Alavijeh, Steen et al. (2024) and Kidd, Darvish et al. (2024). Additionally, Shi, Lin et al. (2022) researched the simultaneous collection and delivery of medical supplies via drones during the COVID-19 pandemic to shorten delivery times and lower the risk of patient contact. Alongside this, Hosseini-Motlagh, Farahmand et al. (2022) studied vehicle routing to optimize speed and cover traffic while considering simultaneous product collection and delivery. Finally, a precise algorithm for vehicle optimization with simultaneous collection and delivery was introduced by Che and Zhang (2023), which assumes that customer collection demands are random. Aldossary (2024) presented a deep learning and hybrid optimization framework for managing electric vehicle (EV) fleets in smart cities. It focused on predicting charging demand using IoT data and optimizing routes in real-time, significantly reducing travel distance compared to other algorithms. Kweon, Kim et al. (2024) addressed a real-world parcel delivery network problem involving multiple hubs and the consolidation of small-sized parcels. They developed solution methods, including a mixed-integer programming model and heuristic algorithms, to optimize the flow of parcels from origin to destination. Farahani, Zegordi et al.

(2025) proposed a stochastic programming model to optimize fleet size and routing with a mix of autonomous and conventional vehicles for a drug distribution system under human resource disruptions (e.g., a pandemic). The model demonstrated significant cost savings compared to deterministic approaches. Liu and Shi (2025) designed a cold chain distribution network that incorporates a heterogeneous fleet of refrigerated trucks under carbon reduction policies. They developed a multi-objective optimization model to simultaneously decide on routes, fleet composition, facility location, and charging station placement.

As summarized in **Table 1**, the existing literature on distribution network design is rich with models addressing facility location, inventory management, and routing, often under uncertainty. However, a synthesis of recent works reveals persistent gaps. First, while some studies incorporate demand uncertainty, few implement data mining techniques for granular, spatial-temporal demand pattern recognition and clustering as a direct input for network design. Secondly, optimization models frequently treat fleet composition as a given or secondary factor, rather than a primary decision variable integrated with facility location and capacity planning. Finally, there is a scarcity of models that simultaneously co-optimize the data-driven demand segmentation, multi-echelon facility infrastructure, and the heterogeneous fleet required for fulfillment.

This research seeks to bridge these gaps by introducing a comprehensive model that integrates these often-decoupled decisions. Our contributions are threefold:

1- We employ the DBSCAN clustering algorithm to move beyond traditional deterministic/stochastic demand assumptions, extracting actionable insights from customer data to identify geographical demand clusters and patterns.
2- We develop a novel MILP model that explicitly determines the optimal number and type of vehicles required for both mid-mile and last-mile transportation, alongside deciding facility locations and capacities. (In particular, last-mile costs refer to the expenses associated with delivering products from the final distribution center to end customers. These costs typically include vehicle operating expenses, driver labor, and fuel consumption for the final stage of delivery. For instance, in parcel logistics, the last-mile cost often represents the most expensive segment of the distribution process due to fragmented demand and frequent stops).
3- We propose a unified framework where the outputs of the data analysis (demand clusters) directly inform the inputs of the optimization model, creating a closed-loop, responsive design process for multi-period, multi-product distribution networks.

By addressing these elements, our approach facilitates the design of a more agile, efficient, and cost-effective distribution network capable of responding to real-world demand complexities.

Table 1. Related literature on this subject

Study	Focus Area	Demand Treatment	Fleet Consideration	Key Limitations Addressed in This Study
Miranda and Garrido (2008)	Location-Inventory	Stochastic (Normal)	Not Considered	No data-driven pattern analysis, no fleet optimization
Shahabi, Unnikrishnan et al. (2014)	Location-Inventory	Stochastic (Correlated)	Not Considered	No integration of clustering for demand segmentation
Gzara, Nematollahi et al. (2014)	Location-Inventory	Deterministic	Not Considered	Static demand view, no fleet decisions
Behnamian, Fatemi Ghomi et al. (2018)	Truck Scheduling	Deterministic	Scheduling only	No demand clustering, no holistic fleet *composition*
Wang, Gunasekaran et al. (2018)	Network Design	Big Data (Conceptual)	Not Considered	Conceptual, no specific algorithm or integrated model
Fahmy, Zaki et al. (2023)	Location-Allocation	Deterministic	Not Considered	No data mining, fleet not optimized
Song and Wu (2023)	Location-Routing	Deterministic	Homogeneous Fleet	No demand clustering, heterogeneous fleet not considered
Vincent, Aloina et al. (2023)	Routing	Deterministic	Given Fleet	Focus on routing only, not integrated network design
Aldossary (2024)	Routing & Scheduling	Real-time (IoT)	Optimized (EV-specific)	Focus is on real-time routing of a homogeneous EV fleet, not on integrated facility location and strategic network design
Kweon, Kim et al. (2024)	Hub Network Design	Deterministic	Not Considered	Focuses on parcel consolidation and hub flows; fleet composition and vehicle types are not decision variables
Farahani, Zegordi et al. (2025)	Fleet Size & Mix	Stochastic	Optimized (AV/CV Mix)	Focus is on tactical fleet composition under disruption for a given network; does not integrate strategic facility

Table 1. Related literature on this subject

Study	Focus Area	Demand Treatment	Fleet Consideration	Key Limitations Addressed in This Study
				location or data-driven demand clustering
Liu and Shi (2025)	Network Design	Deterministic	Optimized (Heterogeneous)	Focuses on a specific cold chain application with carbon policies; demand is not analyzed through data mining/clustering techniques
This Study	**Integrated Network Design**	**Data-Driven (DBSCAN Clustering)**	**Optimized Composition & Size**	**Integrates clustering, facility location, and heterogeneous fleet optimization.**

3-Problem statement

This paper focuses on the design of a supply and distribution network for a company that manages the receipt of various products from suppliers and their delivery to customers. The network is structured to address different requirements for storage and transportation, leading to distinct operational sections: processing centers operate independently for different product categories, while receiving and distribution activities are managed collectively. As depicted in **Fig 1**, the proposed distribution network encompasses suppliers, processing centers, sorting facilities, distribution channels, and end customers. Products are transported to distribution centers through various means, including the company's own transportation fleets and those of suppliers, before being directed to processing centers. Once the items are processed into parcels, they move to sorting centers and are eventually sent through distribution channels for delivery to customers.

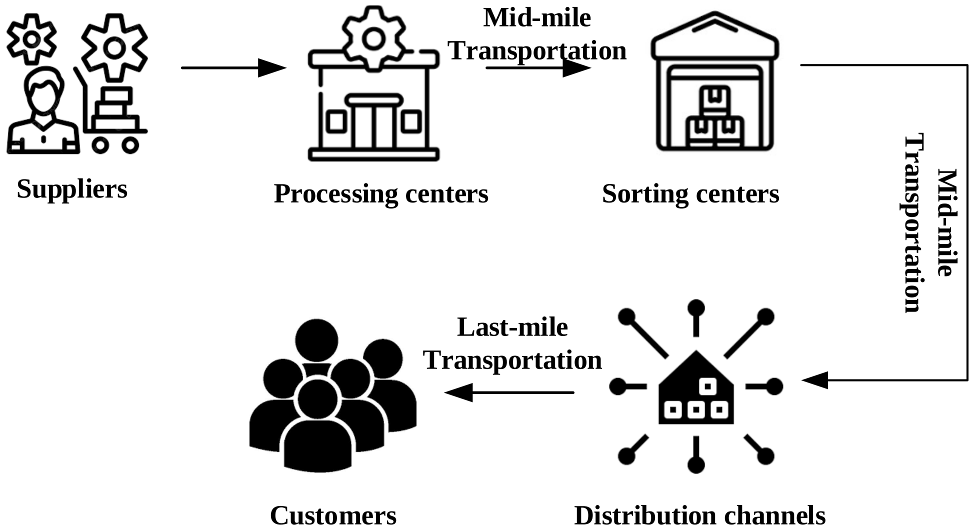

Fig 1. The proposed multi-echelon distribution network structure. (This structure illustrates the physical network that the mathematical model in Section 3.1 is designed to optimize)

The primary aim of this design challenge is to establish an efficient network that addresses facility location and capacity planning, defines connections between facilities, and specifies the product flow within the network, all with the objective of minimizing costs. The model developed in this project represents a multi-product, multi-period supply chain network, integrating distribution centers, processing centers, and sorting facilities. The flow of goods is initiated by customer orders, which are aggregated by their geographical city. The final destination for delivery is the customer located within a city. To effectively respond to customer demands, it is crucial to analyze demand patterns, identify trends, and cluster customers for efficient distribution network planning. Data mining techniques will be employed to explore these demand patterns and understand customer behavior. Additionally, optimizing the fleet composition and determining the appropriate number of each vehicle type for last-mile connections is essential. By analyzing customer behavior and discovering demand clusters, these segments will serve as input for refining the distribution network design. The design of the proposed network is governed by a set of specific assumptions that define the problem's policies, limitations, and operational rules. These assumptions are critical for structuring the mathematical model and are outlined as follows:

- **Facility Capacity and Establishment:** The capacity for each facility type i (e.g., processing center, distribution center) is quantifiable and predetermined at different levels j (e.g., small, medium, large). A facility can only be established in a city a if it is practically feasible, which is indicated by the parameter I_{ia}. Establishing a facility incurs a fixed construction or mortgage cost (F_{ijat}) if built, or a rental cost (G_{ijat}) if operated by a contractor, depending on the contract type.
- **Transportation and Fleet Considerations:** The transportation fleet is heterogeneous, consisting of multiple vehicle types v with varying capacities (${ca_l}^p$), costs, and environmental impacts. The selection of the appropriate vehicle type and the optimization of routes are crucial goals, aiming to minimize costs while maintaining sufficient responsiveness. A key environmental and service limitation is the maximum permissible

transfer time ($Time_{ii'}$) between two facilities i and i'. This constraint ensures that routes are not only cost-effective but also adhere to service time promises and indirectly mitigate environmental impact by discouraging excessively long routes.

- **Flow and Allocation Constraints:** The flow of products within the network is contingent on prior allocation decisions. A facility must be officially allocated to another facility for any product flow to occur between them. Furthermore, each customer's demand in a city must be fully satisfied by the facilities allocated to that demand point, ensuring complete coverage.
- **Demand and Data-Driven Input:** Customer demand ($De^{p}_{aa't}$) is a known input parameter, derived from the data-driven analysis described in Section 4. The patterns and clusters identified through the DBSCAN algorithm inform the geographical and temporal distribution of this demand data used in the model.

By formalizing these assumptions, the model captures the core trade-offs between infrastructure investment (facility location and capacity), operational efficiency (transportation and fleet costs), and service/environmental constraints (transfer time limits)

3.1. Mathematical modeling

The model comprises the following sets, parameters, and decision variables. The proposed multi-echelon network, as depicted in Fig. 1, consists of several specific facility types (suppliers, processing centers, sorting facilities, distribution centers). However, for the purpose of generalized modeling and formulation, these specific types are abstracted into a unified set of facilities i ∈ I. The connections and flows between these different facility types are governed by the model's constraints and cost parameters. This abstraction allows the model to be applied to various network configurations while maintaining the core optimization of location, allocation, and fleet composition across all levels.

Table 2. Sets, parameters, and decision variables for the mathematical model

Sets	
i	Set of facilities, $i = \{1, \dots I\}$
j	Set of capacity levels, $j = \{1, \dots, J\}$
a	Set of geographical city zones (demand points), $a = \{1, \dots, A\}$
p	Set of products, $p = \{1, \dots, P\}$
v	Set of vehicles, $v = \{1, \dots, V\}$
t	Set of time periods, $t = \{1, \dots, T\}$
Parameters	
$Tr^{ii'}_{aa't}$	Transportation costs per trailer from facility *i* in city *a* to facility *i'* in city *a'* in the time period *t*
Tc^{p}_{it}	Transportation costs per parcel of product type *p* from facility *i* to a customer in time period *t*
$To^{aa'}_{vt}$	Transporting cost from origin city *a* to destination city *a'* when transporting with vehicle type *v* in the time period *t*

$De_{aa't}^{p}$	Demand of product p in from city a to city a' in the time period t (In this study, the parameter $De_{aa't}^{p}$ is obtained by aggregating customer-level demand after clustering customers by location and demand behaviour using the DBSCAN algorithm. Specifically, customers are clustered using geographic coordinates (latitude and longitude) and demand features. For each cluster, we compute the centroid and aggregate customer demand for each product *p* and time period *t*. The aggregated cluster demand is then used to populate $De_{aa't}^{p}$ for the network design model).
F_{ijat}	Mortgage cost of the facility *i* with capacity level *j* in city *a* in the time period *t*
G_{ijat}	Rent cost of the of the facility *i* with capacity level *j* in city *a* in the time period *t*
E_{ijt}	Cost of equipping capacity level *j* of the processing facilitation type *i* in time period *t*
C_{ij}	Capacity level *j* of facility *i*
${ca_l}^p$	Capacity of vehicle *l* based on product *p*
$Time_{ii'}$	Maximum transfer time between facility *i* and *i'*
$dis_{vaa'}$	Time distance from origin city *a* to destination city *a'* when transporting with vehicle type *v*
I_{ia}	1, If it is possible to establish facility *i* in the city *a;* 0 otherwise.
M	Big M
Decision Variable	
Y_{ijat}	1, If facility *i* with capacity level *j* in the city *a* is opened in time period *t*; 0 otherwise.
$X_{aa't}^{pii'}$	1, If facility *i* in the city *a* is assigned to facility *i'* in the city *a'* in time period *t*; 0 otherwise.
$Q_{aa't}^{pii'}$	Quantity of product type *p* shipped from facility *i* in the city *a* to facility *i'* in the city *a'* at time period t.
$M_{aa't}^{pi}$	Quantity of product type *p* shipped from facility *i* in the city *a* to customers in the city *a'* at time period t.
$N_{vaa't}$	Required number of vehicle type *v* for transporting from the city *a* to the city *a'* at time period t.

In this paper, capacity levels refer to the maximum processing or handling ability of facilities (e.g., processing centers, sorting facilities, or distribution centers) within a given time period. For example, the capacity of a sorting facility may be expressed in terms of the maximum number of parcels it can process per day. This definition ensures that capacity constraints in the model reflect practical operational limits. In addition, the model tracks the flow of products $Q_{aa't}^{pii'}$ between a facility i in city a and a facility i' in city a'. This structure is essential for accurately calculating transportation costs and times, which are inherently dependent on the geographical distance between city pairs (a, a'), not just the facility types (i, i'). The variable

$N_{vaa\prime t}$ for the number of vehicles required is also defined over city pairs for the same reason. Now, the mathematical model can be developed as follows.

$$\mathrm{Z} = \min[\sum_{t}\sum_{i}\sum_{j}\sum_{a}(F_{ijat} + G_{ijat} + E_{ijt}).Y_{ijat}$$

$$+\sum_{p}\sum_{i}\sum_{a}\sum_{a\prime}\sum_{t} Tc_{it}^{p}M_{aa\prime t}^{oi} + \sum_{p}\sum_{i}\sum_{i\prime}\sum_{a}\sum_{a\prime}\sum_{v}\sum_{t} Tr_{aa\prime t} \times \frac{Q_{aa\prime t}^{pii\prime}}{ca_{v}{}^{p}} \tag{1}$$

$$+\sum_{v}\sum_{a}\sum_{a\prime}\sum_{t} To_{vt}^{aa\prime}N_{vaa\prime t}\ ]$$

$s.t.$

$$\sum_{j} X_{ijat} \leq I_{ia} \qquad (\forall i, a, t) \tag{2}$$

$$dis_{vaa\prime}.X_{aa\prime t}^{pii\prime} \leq Time_{ii\prime}.\sum_{j} Y_{ijat} \qquad (\forall p, v, i, i', a, a', t) \tag{3}$$

$$\sum_{a} X_{aa''t}^{pii'} \leq 1 \qquad (\forall p, i, i', a', t) \tag{4}$$

$$Q_{aa\prime t}^{pii\prime} \leq M.X_{aa\prime t}^{oii\prime} \qquad (\forall p, i, i', a, a', t) \tag{5}$$

$$\sum_{i,a} M_{aa't}^{pi} = \sum_{ka} De_{aa't}^{p} \qquad (\forall a', p, t) \tag{6}$$

$$\sum_{i,a} Q_{aa't}^{pii'} = \sum_{a''} M_{a'a''t}^{pi'} \qquad (\forall i', a', p, t) \tag{7}$$

$$\sum_{i',a'} Q_{aa't}^{pii'} \leq \sum_{j} C_{ij}Y_{ijat} \qquad (\forall p, i, a, t) \tag{8}$$

$$\sum_{v} ca_{v}{}^{p}N_{laa't} > Q_{aa't}^{pii'} \qquad \forall (i, i', a, a', p, t) \tag{9}$$

$$N_{vaa\prime t} \in Z \qquad \forall (v, a, a', t) \tag{10}$$

$$Y_{ijat}, X_{aa\prime t}^{pii\prime} \in \{0,1\} \qquad (\forall p, i, i', j, a, a', t) \tag{11}$$

$$Q_{aa\prime t}^{pii\prime}, M_{aa\prime t}^{pi} \in R^{+} \qquad (\forall p, i, i', a, a', t) \tag{12}$$

The objective function (1) aims to minimize costs associated with establishing facilities, equipping them, transportation, distribution, and offsetting carbon emissions through various strategies, while also accounting for risk costs. In the objective function, the carbon emission term is derived by multiplying activity levels (e.g., transported quantities and processing volumes) with standardized emission factors. For transportation, emission factors are applied per ton-kilometer, based on widely used logistics sustainability studies and published databases. The model then assigns a monetary cost to these emissions, which can be offset through investment in carbon credits or equivalent mitigation measures. By internalizing this cost, the model allows the decision-maker to evaluate trade-offs between profitability and sustainability. Thus, the offset term represents the economic effort required to neutralize emissions generated by the supply chain operations considered in this study. Constraints (2) ensure that facilities are located in cities where construction is practical and feasible. Constraints (3) maintain the necessary coverage radius for effective facility allocation. Constraints (4) detail the allocation of facilities to ensure they meet specific allocation criteria. Constraints (5) specify that flow within the network is contingent upon prior allocations being made. Meanwhile, constraints (6) establish coverage requirements for urban demand, ensuring that all areas are adequately serviced. Constraints (7) reflect the flow balance in the distribution network, ensuring that incoming and outgoing flows at each facility are appropriately aligned. Constraints (8) outline the capacity limits of the facilities that need to be established, ensuring they are not overwhelmed. Additionally, constraints (9) determine the number of vehicles required based on the number of parcels transported and the capacities of these vehicles. Finally, constraints (10)-(12) define the allowable ranges for the decision variables, ensuring that all solutions fall within practical and operational limits. Collectively, these components work together to generate an efficient and sustainable distribution network.

4. Solution method

In this section, to effectively tackle the proposed issue, it is important to first acknowledge the specific challenges involved. Prior to developing any solutions, a deep dive into customer demand trends and their geographical spread is necessary. This analysis serves as a crucial foundation for sound decision-making. Adopting a data-driven methodology allows for the leveraging of both historical and current data, revealing patterns, shifts, and relationships in customer demand across various locations. Utilizing advanced techniques such as data mining enables the extraction of insights into the spatial and temporal variations of demand. With a solid grasp of these demand dynamics and geographical distributions, the network design model can then be developed and implemented.

The mathematical model formulated in Section 3.1 is a multi-period, multi-product Mixed-Integer Linear Programming (MILP) model. This class of models is known to be NP-hard, making them computationally challenging to solve for real-world large-scale instances due to the combination of integer decision variables (e.g., facility opening, vehicle number) and continuous variables (e.g., product flow). To solve this model efficiently, the General Algebraic Modeling System (GAMS) was employed as a high-level platform for formulating and solving optimization problems. Within GAMS, the CPLEX solver was selected. CPLEX is a powerful, state-of-the-art solver specifically designed for linear, mixed-integer, and quadratic programming problems. Its significance in this study stems from several key capabilities:

- *Advanced MILP Techniques*: CPLEX utilizes a suite of sophisticated algorithms, including branch-and-cut, cutting planes, and heuristics, to effectively navigate the solution space of complex MILP models and find optimal or near-optimal solutions.

- *Handling Large-Scale Problems*: It is highly efficient at managing problems with a large number of constraints and variables, which is characteristic of our network design model encompassing multiple facilities, products, time periods, and city pairs.

- *Robustness and Reliability*: As an industry-standard solver, CPLEX provides proven robustness and reliability for strategic optimization problems, ensuring the results obtained are credible and suitable for deriving managerial insights.

The integration of this computational framework was essential for the feasible and timely resolution of the model, enabling the extensive numerical analysis and sensitivity tests presented in Section 5.

4.1. Clustering with DBSCAN algorithm

This section presents a transformative approach to demand analysis, shifting away from established structural changes and counting methods Ziari (2024). Traditional strategies often depend on static optimization and set limits that do not accurately capture consumer behavior or fully integrate pricing analysis features Cao, Hu et al. (2023). A key advantage of this new model is its flexibility, enabling it to adapt smoothly to slight changes in demand as well as unexpected events, like global crises or economic downturns. Such adaptability is essential in today's rapidly evolving marketplace, where consumer behavior can change dramatically due to external influences. The capacity to swiftly modify pricing strategies allows retailers to preserve operational efficiency, even in unpredictable situations Paramita and Hariguna (2024). This ability is particularly significant as it allows retailers to continuously adjust their approaches in response to immediate demand fluctuations. To implement the Density-Based Spatial Clustering of Applications with Noise (DBSCAN) algorithm, the following steps are outlined, as illustrated in **Fig 2**. As it can be seen, this algorithm tries to identify Clusters by grouping together points that are closely packed in dense regions while separating points in sparse areas. Then, it handles noise and effectively classifies outliers or noise points that do not belong to any cluster. This algorithm is also capable of detecting arbitrarily shaped clusters unlike some other clustering methods. It can properly find clusters of varying shapes and sizes by just two parameters, epsilon (the neighborhood radius) and minimum points (the minimum number of points needed to form a dense region) Wei, Gao et al. (2024).

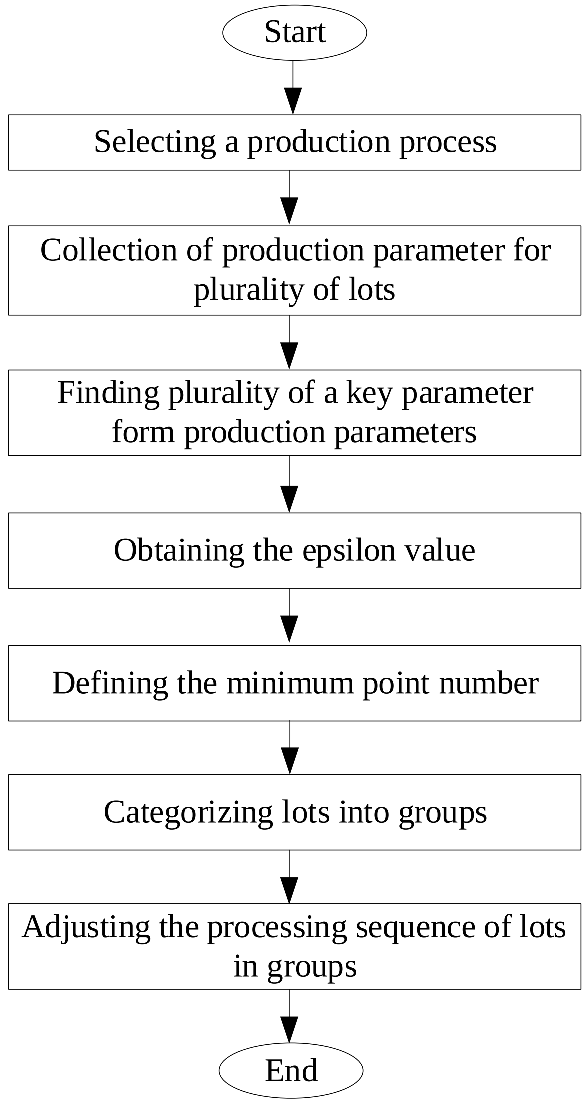

Fig 2. Flowchart of the DBSCAN algorithm Miao, Yuan et al. (2022). (The process starts with the input of customer location data, identifies core points within a radius ε that have a minimum number of neighbors, and expands clusters from these points. The key outputs are dense clusters of customer demand and noise points (outliers). These resulting clusters define the geographical demand zones a and their parameters, which serve as the critical data-driven input for the distribution network optimization model presented in Section 3.1).

In conclusion, this groundbreaking approach transforms the landscape of demand analysis and forecasting. By implementing a data-driven, autonomous model that accurately mirrors consumer behavior and adapts to changes in the market, retailers can create customized network structures that lead to more precise demand segmentation and improved profitability. As the retail environment continues to develop, adopting such advanced techniques will be essential for maintaining a competitive advantage.

Retailers who utilize these sophisticated analytical methods can not only meet current market demands but also foresee future trends, positioning themselves for enduring success in an unpredictable economic climate. This adaptability fosters a more responsive network structure capable of adjusting to fluctuations in consumer behavior. By continuously applying the algorithm across different time periods and making iterative updates to reflect evolving demand patterns, this model effectively clusters geographic demand into useful inputs for optimizing distribution network design. Developed using Python, the outcomes of this approach will be showcased in the following section, where they will be compared with traditional methodologies. This comparative study seeks to deliver valuable insights into the effectiveness and benefits of the new approach.

4.2 Demand data, preprocessing, DBSCAN implementation and clustering outputs

The computational experiments use the dataset referenced in Jabbarzadeh, Haughton et al. (2018) and the setup described in Section 5: 200 customers, 5 suppliers, 2 processing centers, 2 sorting facilities, 15 candidate distribution centers, 20 product types and 5 time periods. The raw customer-level dataset contains for each customer: city identifier, latitude, longitude, and demand by product and time period. When necessary, missing values in demand were imputed using median imputation. For clustering we construct per-customer features intended to capture both spatial proximity and demand intensity. The features used are: (1) geographic coordinates (latitude, longitude); and (2) aggregated demand intensity per customer (total demand across products and across the considered time horizon, or product-specific demand when product clusters are required). Before clustering, features are standardized (zero mean, unit variance) using standard scaling to ensure that spatial coordinates and demand have comparable influence.

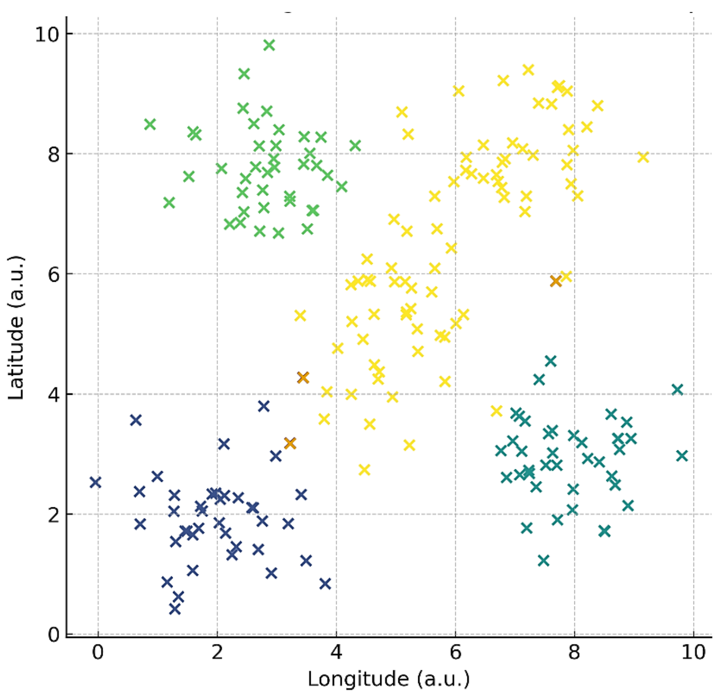

Fig 3. DBSCAN clustering of customers' demand

We implement DBSCAN using the scikit-learn library (Python). DBSCAN requires two parameters: epsilon (ε), the neighborhood radius, and min_samples, the minimum number of points to form a dense region. We select ε by inspecting the k-distance graph (k = min_samples) and choose min_samples = 5 as a default for this size of dataset. In the numerical example reported in Section 5, the DBSCAN call used the values ε = 0.6 and min_samples = 5 after standardization. DBSCAN assigns each customer either to a cluster id (0,1, 2, ...) or to noise (label = −1). For each non-noise cluster c we compute: (i) cluster centroid location (mean latitude and longitude of cluster members), and (ii) aggregated demand by product and time (sum of customer demands within the cluster for each product p and each time period t). The aggregated cluster demand becomes $De_{aa\prime t}^{p}$ the input used in the network model. Noise points (if any) are treated as singleton clusters (each noise point is kept as its own cluster) so that all demand is represented in the model. Using cluster aggregation reduces problem dimensionality and yields stabilized demand inputs which the model then uses to define flows and routing. **Fig 3** shows the DBSCAN clustering result for the customer set (clusters colored, noise marked).

5. Computational results

To assess the practicality and effectiveness of the proposed model, data from Jabbarzadeh, Haughton et al. (2018) was employed, supplemented by the generation of several random parameters. The model was analyzed using the CPLEX solver, which is part of the GAMS optimization suite. The framework developed consists of a network involving five suppliers, 200 customers, two processing centers, two sorting facilities, and 15 distribution centers. Additionally, the model incorporates 20 different product types and spans across 5 distinct time periods. This comprehensive structure allows for a thorough evaluation of the model's capabilities in real-world applications. This section focuses on analyzing the influence of critical parameters within the model and how they affect its performance and outcomes. The goal is to provide valuable insights for managers, decision-makers, and retailers in forecasting inputs and anticipating potential changes in scenarios. Important aspects such as data-driven optimization, the capacity of processing centers, and annual demand will be explored in the context of the proposed model. By evaluating how variations in these parameters impact the model's behavior, useful insights can be generated to illustrate anticipated outcomes and support informed decision-making in practice.

It should be noted that the numerical experiments in this study are based on generated/historical demand data and not on a live, industrial-scale dataset. While the adopted data is sufficient to demonstrate the applicability and advantages of the proposed data-driven model, it does not fully capture all sources of uncertainty, variability, and behavioral patterns present in real-world supply chains. Consequently, the generalizability of the numerical results is subject to these limitations. Future work will focus on applying the methodology to larger and more diverse real-world datasets to further validate and extend the insights obtained here.

5.1.1. Analyses on annual demand

One crucial analysis that plays a significant role in logistics is the assessment of annual demand. This analysis is vital for understanding how fluctuations in demand can influence both operational and strategic costs. Within this framework, a key constraint is the requirement for the system to effectively meet demand while minimizing logistics expenses across a well-organized network. This highlights that any variations in demand can have substantial effects on customer expectations and the overall reliability of the network. Therefore, examining the sensitivity of demand becomes critical for making informed management choices that align with the goals of the organization. To provide a thorough perspective, the annual demand analysis has been conducted at three specific levels: 5%, 10%, and 15%. Each of these levels reflects different strategies that the organization may consider to achieve a balance between customer satisfaction and the associated strategic and operational costs. The results from these analyses not only indicate potential performance under different scenarios but also shed light on the trade-offs between the design of the network and transportation expenses. The findings of this extensive analysis are presented in **Tables 3** and **4**, illustrating how an increase in annual demand impacts strategic and operational costs. This analysis serves not only as a tool for strategic planning but also provides decision-makers with critical insights that can enhance delivery capabilities while optimizing logistical and distribution networks.

Table 3. Analysis on the annual demand rate

Increase in annual demand=10%		**Time period**				
		1	**2**	**3**	**4**	**5**
Construction	Set-up costs	164.7	314.2	557.9	856.3	1081.7
	Facilitation costs	2.7	3.8	5.2	7.5	9.3
Transportation	Mid-mile costs	318.8	468.6	753.1	1016.9	1202.7
	Last-mile costs	429.2	618.6	990.8	1576.0	1950.1

Table 4. Analysis on the annual demand rate

Increase in annual demand =15%		**Time period**				
		1	**2**	**3**	**4**	**5**
Construction	Set-up costs	206.5	391.2	687.4	1016.7	1302.9
	Facilitation costs	3.3	4.6	6.3	9.1	11.5
Transportation	Mid-mile costs	396.1	573.3	932.5	1260.3	1471.3
	Last-mile costs	521.8	744.2	1195.3	1896.6	2346.7

5.1.2. Analyses on data-driven approach

To effectively assess the influence of data-driven optimization on distribution network design, we implemented two distinct models across a variety of samples. The first was a conventional model that did not consider clustering or trend analysis within the data. In contrast, the second model employed a data-driven approach. The findings from these analyses are illustrated in **Fig 4**. As shown in **Fig 4**, the data-driven optimization method, coupled with customer demand clustering through the DBSCAN algorithm prior to the network design phase, consistently yielded higher profits across all evaluated samples compared to the traditional model. This advantage can be attributed to the data-driven methodology's capability to process and analyze extensive datasets, allowing it to uncover significant patterns and trends that traditional approaches often miss. By incorporating clustering techniques, the data-driven model enables a more precise segmentation of customer demand, facilitating a far more effective distribution strategy. In summary, when comparing the two models across various samples, the data-driven approach resulted in an average profit increase of approximately 10% in the distribution networks designed. Furthermore, the ability of this model to adapt to real-time data and shifting customer behavior not only boosts profitability but also enhances operational efficiency and improves overall customer satisfaction. Consequently, the integration of data-driven optimization is a vital advancement towards developing a more agile and effective distribution network, positioning organizations to respond better to market fluctuations and customer needs. This ongoing commitment to integrating data insights will ultimately drive long-term success in the competitive landscape of distribution logistics.

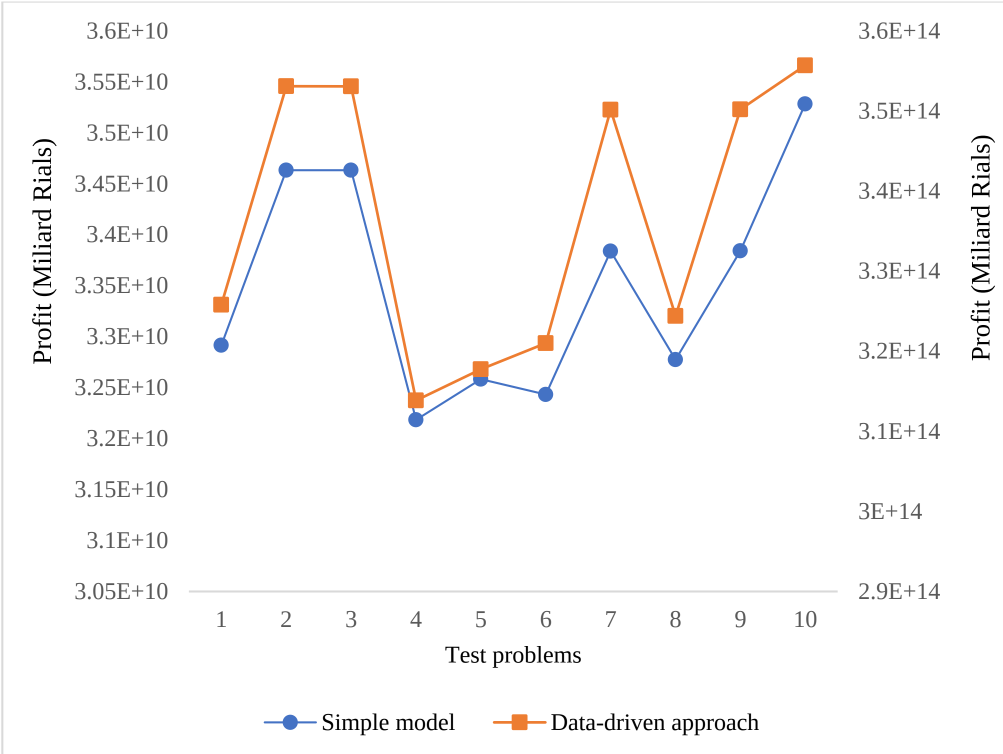

Fig 4. Profit of the basic model versus the data-driven model

5.1.3. Analyses on distribution capacity

In this section, a sensitivity analysis has been conducted on the capacity of distribution centers, increasing the capacity by intervals of 10 percent up to a total increase of 50 percent. As shown in **Fig 5**, the overall profit of the distribution system improves as distribution capacity increases. However, it is important to note that initially, up to a 30 percent increase in capacity significantly enhances profit levels. Beyond this point, as capacity continues to rise, the rate of profit improvement begins to decrease, and the percentage changes notably diminish. This indicates that increasing capacity beyond a certain threshold becomes ineffective.

Improving profit in the proposed distribution network is significantly attributed to increased distribution capacity. Hence, when the network enhances its capacity, it can efficiently handle larger volumes of goods, leading to faster order fulfillment and reduced lead times. This responsiveness enhances customer satisfaction, fostering loyalty and potentially attracting new clients, which ultimately boosts sales revenue. By optimizing transportation resources, companies can reduce per-unit costs, leading to higher profit margins. Overall, expanding distribution capacity not only streamlines operations but also presents opportunities for growth and improved profitability in a competitive market.

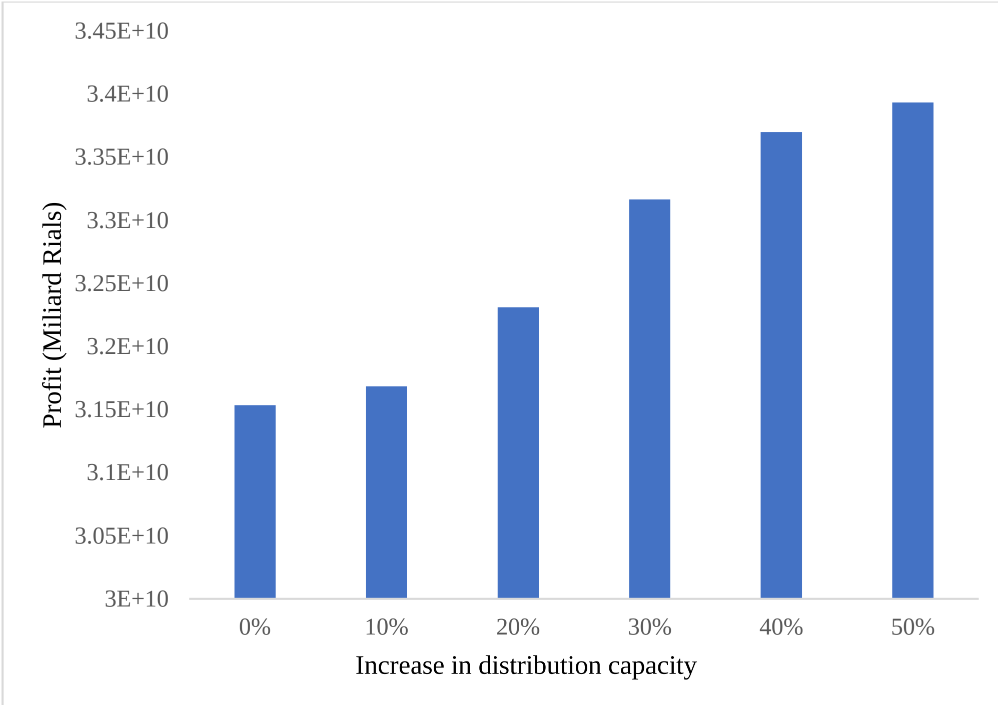

Fig 5. Profit changes based on increasing distribution capacity

These findings have significant implications for real-world managerial decision-making and strategic planning. The analysis, based on real-world data patterns Jabbarzadeh, Haughton et al. (2018), provides a data-driven framework for investment decisions. The initial sharp rise in profit up to a 30% capacity increase strongly justifies strategic investments in expansion for networks operating near or at capacity. Managers should interpret this as a clear signal to allocate capital for scaling operations, as it leads to substantial returns through improved efficiency and market responsiveness. However, the subsequent onset of diminishing returns beyond the 30% threshold is a critical strategic insight. It indicates a saturation point where the cost of additional capital investment outweighs the incremental revenue benefits. For managers, this finding is crucial for avoiding over-investment and inefficient capital allocation. This could involve:

1- Dynamic Resource Allocation: Reallocating existing fleet and inventory between distribution centers based on real-time demand patterns from the clustering analysis, rather than simply building more space.
2- Process Optimization: Investing in technologies or process re-engineering within existing facilities to increase throughput (e.g., better warehouse management systems, automation) before committing to brick-and-mortar expansion.
3- Strategic Flexibility: Considering the use of temporary or contract-based storage solutions to handle demand peaks, thus maintaining agility without incurring the fixed costs of permanent capacity expansion.

Therefore, the model acts as a powerful decision-support tool. It helps managers identify the 'sweet spot' for capacity investment and provides a quantitative basis for shifting strategy from capital-intensive expansion to operational excellence and optimization once that point is reached.

6. Conclusions

This study proposed a novel data-driven model designed to remain responsive under uncertainty, and compared it against a classical approach. The results demonstrate clear improvements in both performance and robustness. The design of distribution networks plays a vital role in the global landscape, as these systems are essential for delivering services efficiently and maintaining high-quality standards for end consumers. To excel in this area, it is important to integrate various strategic decisions, such as optimizing fleet types, automating transportation processes, and employing customer behavior. These elements can significantly improve both fleet efficiency and overall distribution system performance. Traditional approaches to network design often concentrate primarily on the strategic placement of facilities aimed at minimizing supply chain costs. However, this focus frequently neglects crucial factors, such as emerging demand trends, customer density, and evolving consumer behaviors. This study contributes to this field by demonstrating the significant benefits of integrating data-driven demand clustering with a holistic optimization model that co-optimizes facilities and fleet. To build a more versatile distribution system that better addresses customer needs, it is imperative to integrate these often-overlooked elements into the design framework. By analyzing demand patterns alongside optimizing fleet composition, existing gaps in the literature can be filled, resulting in a system that is much more responsive and capable of reaching a broader customer base effectively. In addition, the incorporation of carbon emission offsets into the objective function highlights the environmental dimension of the model. By explicitly assigning a cost to emissions and requiring offsets, the framework encourages the design of logistics networks that are not only cost-efficient but also environmentally responsible. This contributes to the broader agenda of sustainable supply chain management by quantifying the trade-off between economic performance and carbon neutrality.

This exploration introduces an innovative design model aimed at enhancing performance through the application of strategic location planning, flow optimization, and transport fleet management. The proposed model is data-driven and emphasizes fleet optimization, formulated as a Mixed-Integer Linear Programming (MILP) problem. To effectively categorize customers with diverse characteristics, the DBSCAN clustering technique is employed within this framework. A numerical case study, previously utilized by Jabbarzadeh, Haughton et al. (2018), serves as a testing ground, allowing for the extraction of meaningful managerial insights. It is proved that the model is capable of delivering on average 15% greater profit using data-driven approaches and it as assured that the model has noticeable cost savings on transportation and fleet optimization. Moreover, it is showed that crucial parameters including demand annual increase rate or distribution capacities can be implemented in managerial decision-making processes for the better management of distribution networks. Finally, while the numerical analysis highlights the advantages of the data-driven approach, it is important to recognize that the dataset used is generated/historical and does not fully represent the complexity of real supply chain environments. This limitation may affect the extent to which the results can be generalized. We consider the application of the proposed methodology to real-world, large-scale datasets as a promising direction for future research.

While this study provides a framework for integrating data-driven demand analysis with network and fleet optimization, it opens several avenues for future research. Our model uses historical data for clustering and planning. A significant advancement would be the development of a dynamic, closed-loop system that integrates real-time data streams (e.g., IoT sensors, live traffic updates, fluctuating fuel prices) to enable adaptive re-optimization of routing and inventory allocation during the operation period, moving from a static plan to a responsive, living network. The current model optimizes a network in isolation. Incorporating game-theoretic elements or agent-based modeling to simulate the actions and reactions of competing distribution networks would provide a more realistic landscape. This would allow for optimizing strategies not just for operational efficiency, but for market share capture and competitive advantage in a contested environment. In addition, the model's sensitivity to parameters like demand and capacity lays the groundwork for researching robustness. Future models could explicitly incorporate stochastic scenarios for large-scale disruptions and optimize network design for resilience, perhaps by maximizing service level under disruption or minimizing worst-case scenario costs. And finally, for truly large-scale national or global network implementations, the development of efficient tailored heuristic or meta-heuristic algorithms would be valuable to solve the resulting complex problems within reasonable computation time, facilitating even more granular analysis.

References

Ahmadi-Javid, A. and P. Hoseinpour (2015). "Incorporating location, inventory and price decisions into a supply chain distribution network design problem." Computers & Operations Research 56: 110-119.

Alavijeh, N. M., et al. (2024). "Capacity limitation based local flexibility market for congestion management in distribution networks: Design and challenges." International Journal of Electrical Power & Energy Systems 156: 109742.

Aldossary, M. (2024). "Enhancing Urban Electric Vehicle (EV) Fleet Management Efficiency in Smart Cities: A Predictive Hybrid Deep Learning Framework." Smart Cities 7(6): 3678-3704.

Behnamian, J., et al. (2018). "Location-allocation and scheduling of inbound and outbound trucks in multiple cross-dockings considering breakdown trucks." Journal of Optimization in Industrial Engineering 11(1): 51-65.

Cao, M., et al. (2023). "Research on variable weight CLIQUE clustering algorithm based on partial order set 1." Journal of Intelligent & Fuzzy Systems(Preprint): 1-13.

Caramia, M. and E. Pizzari (2022). "Clustering, location, and allocation in two stage supply chain for waste management: A fractional programming approach." Computers & Industrial Engineering 169: 108297.

Che, Y. and Z. Zhang (2023). "An Integer L-shaped algorithm for vehicle routing problem with simultaneous delivery and stochastic pickup." Computers & Operations Research 154: 106201.

Chen, X., et al. (2016). "Data-driven DG capacity assessment method for active distribution networks." IEEE Transactions on Power Systems 32(5): 3946-3957.

Fahmy, S. A., et al. (2023). "Optimal locations and flow allocations for aggregation hubs in supply chain networks of perishable products." Socio-Economic Planning Sciences 86: 101500.

Farahani, M., et al. "A Stochastic Programming Model to Mitigate Disruption Effects on the Drug Distribution System Under an Autonomous Vehicle Fleet."

Gzara, F., et al. (2014). "Linear location-inventory models for service parts logistics network design." Computers & Industrial Engineering 69: 53-63.

Hasani Goodarzi, A., et al. (2021). "Reliable cross-docking location problem under the risk of disruptions." Operational Research 21: 1569-1612.

Hosseini-Motlagh, S.-M., et al. (2022). "A multi-path traffic-covering pollution routing model with simultaneous pickup and delivery." Computers & Industrial Engineering 173: 108644.

Hung Lau, K. (2012). "Demand management in downstream wholesale and retail distribution: a case study." Supply Chain Management: An International Journal 17(6): 638-654.

Jabbarzadeh, A., et al. (2018). "Closed-loop supply chain network design under disruption risks: A robust approach with real world application." Computers & Industrial Engineering 116: 178-191.

Janjevic, M., et al. (2021). "Designing multi-tier, multi-service-level, and multi-modal last-mile distribution networks for omni-channel operations." European Journal of Operational Research 294(3): 1059-1077.

Jha, A., et al. (2012). "Minimizing transportation cost of a joint inventory location model using modified adaptive differential evolution algorithm." The International Journal of Advanced Manufacturing Technology 60: 329-341.

Juris, J. S. (2020). Networking futures: The movements against corporate globalization, Duke University Press.

Kidd, M. P., et al. (2024). "A relax-and-restrict matheuristic for supply chain network design with facility location and customer due date flexibility." Transportation Research Part E: Logistics and Transportation Review 182: 103370.

Kweon, O., et al. (2024). "Parcel delivery network optimization problem considering multiple hubs and consolidation of small-sized parcels." Computers & Industrial Engineering 191: 110113.

Li, Y., et al. (2023). "Data-driven distributionally robust scheduling of community integrated energy systems with uncertain renewable generations considering integrated demand response." Applied Energy 335: 120749.

Liu, Y. and X. Shi (2025). "Cold chain distribution network design with heterogeneous fleet under carbon reduction policy application effect." Transportation Planning and Technology 48(5): 979-1002.

Miao, X., et al. (2022). Topology Identification in LV Distribution Network Based on SAE and Modified DBSCAN. 2022 IEEE/IAS Industrial and Commercial Power System Asia (I&CPS Asia), IEEE.

Miranda, P. A. and R. A. Garrido (2008). "Valid inequalities for Lagrangian relaxation in an inventory location problem with stochastic capacity." Transportation Research Part E: Logistics and Transportation Review 44(1): 47-65.

Paramita, A. S. and T. Hariguna (2024). "Comparison of K-Means and DBSCAN Algorithms for Customer Segmentation in E-commerce." Journal of Digital Market and Digital Currency 1(1): 43-62.

Perl, J. and S. Sirisoponsilp (1988). "Distribution networks: facility location, transportation and inventory." International Journal of Physical Distribution & Materials Management 18(6): 18-26.

Shahabi, M., et al. (2014). "A three level location-inventory problem with correlated demand." Transportation Research Part B: Methodological 69: 1-18.

Shavandi, H. and B. Bozorgi (2012). "Developing a location–inventory model under fuzzy environment." The International Journal of Advanced Manufacturing Technology 63: 191-200.

Shi, X., et al. (2020). "Early anomaly detection and localisation in distribution network: A data-driven approach." IET Generation, Transmission & Distribution 14(18): 3814-3825.

Shi, Y., et al. (2022). "A bi-objective optimization model for the medical supplies' simultaneous pickup and delivery with drones." Computers & Industrial Engineering 171: 108389.

Song, L. and Z. Wu (2023). "An integrated approach for optimizing location-inventory and location-inventory-routing problem for perishable products." International Journal of Transportation Science and Technology 12(1): 148-172.

Theophilus, O., et al. (2021). "Truck scheduling optimization at a cold-chain cross-docking terminal with product perishability considerations." Computers & Industrial Engineering 156: 107240.

Van Wijk, A., et al. (2012). "Approximate evaluation of multi-location inventory models with lateral transshipments and hold back levels." European Journal of Operational Research 218(3): 624-635.

Vincent, F. Y., et al. (2023). "The vehicle routing problem with simultaneous pickup and delivery and occasional drivers." Expert Systems with Applications 214: 119118.

Wang, G., et al. (2018). "Distribution network design with big data: Model and analysis." Annals of Operations Research 270: 539-551.

Wei, Z., et al. (2024). "Adaptive marine traffic behaviour pattern recognition based on multidimensional dynamic time warping and DBSCAN algorithm." Expert Systems with Applications 238: 122229.

Wu, Y., et al. (2022). "Adaptive large neighborhood decomposition search algorithm for multi-allocation hub location routing problem." European Journal of Operational Research 302(3): 1113-1127.

Xin-hua, L. and Q. Jin (2007). Joint Decision Problem for Inventory Control and Facility Location. 2007 IEEE International Conference on Automation and Logistics, IEEE.

Ziari, M. (2024). "A Data-Driven pricing Model for Distribution Systems Considering Competition." Advances in Industrial Engineering.

Ziari, M. and A. A. Taleizadeh (2025). "Integrated Data-driven and Artificial Intelligence Framework to develop Digital Twins in Distribution System of Supply Chains: A Real Industrial Case." International Journal of Production Economics: 109743.