

## ROBUST OPTIMIZATION FOR SUPPLY CHAIN WITH ROUTING PROBLEM: A LEARNING-BASED APPROACH

HAMI TALEBI, MEHRAN KHALAJ\*, DAVOOD JAFARI,  
PARISA MOUSAVI AHRANJANI AND AMIR HOSSEIN KAMALI DOLATABADI

**Abstract.** Effective supply chain management is crucial for businesses to remain competitive in today's dynamic market. Despite extensive research, there is a lack of integrated approaches that simultaneously address resource allocation, routing, and delivery scheduling under uncertain conditions. This study develops a hybrid framework that combines robust optimization, simulated annealing, and reinforcement learning to enhance supply chain operations in complex networks involving fixed suppliers, distribution centers, and customers. Empirical results from rigorous testing demonstrate significant efficiency improvements and adaptability across diverse scenarios. A real-world case study from the logistics sector highlights the practical benefits, achieving notable cost savings and operational robustness. Sensitivity analysis further underscores the model's capability to adapt to parameter variations. These findings emphasize the value of incorporating learning-based strategies into supply chain optimization, offering a powerful tool for organizations to address uncertainty and enhance decision-making efficiency.

**Mathematics Subject Classification.** 90B06.

Received August 13, 2024. Accepted December 23, 2024.

### 1. INTRODUCTION

Efficient supply chain management is a critical aspect of maintaining competitiveness in today's rapidly evolving business environment. At its essence, supply chain optimization focuses on the strategic allocation of resources across interconnected networks comprising suppliers, distribution centers (DCs), and customers. This problem involves assigning DCs to suppliers, allocating customers to DCs, and designing optimal vehicle routes to meet demand while adhering to various constraints. These challenges are influenced by several factors, including fluctuating demand patterns, the limited capacities of vehicles and DCs, and stringent delivery timeframes imposed by customers.

The inclusion of soft time windows for customer deliveries adds a temporal dimension to the problem, introducing penalties for deviations from preferred delivery times. Balancing customer satisfaction, operational efficiency, and resource utilization within these constraints is a complex task requiring robust optimization methodologies. Moreover, the dynamic nature of supply chains, influenced by disruptions and uncertainties in demand, further complicates decision-making and underscores the need for adaptive and scalable approaches.

---

*Keywords.* Supply chain, robust optimization, simulated annealing, reinforcement learning, vehicle routing problem.

Department of Industrial Engineering, Parand and Robat Karim Branch Islamic Azad University, Parand, Iran.

\*Corresponding author: [gh.khalaj@iaau.ac.ir](mailto:gh.khalaj@iaau.ac.ir)

To address these challenges, this paper introduces a comprehensive framework that integrates simulated annealing with reinforcement learning to solve complex, large-scale supply chain optimization problems. The proposed methodology efficiently explores the solution space, generating optimal or near-optimal solutions that account for temporal constraints and uncertainties in demand. This framework aims to enhance decision-making capabilities for real-world supply chain operations by providing a robust and practical tool for optimization.

To demonstrate the applicability of the proposed framework, we conduct rigorous testing on a variety of test problems, utilizing custom-generated datasets inspired by real-world complexities and prior research. The results are benchmarked against an optimal solver, showcasing the effectiveness of the learning-based strategies in achieving superior performance across various scenarios. Furthermore, sensitivity analyses are performed to assess the framework's adaptability and robustness under changing conditions, offering actionable insights into its practical utility.

A case study drawn from the logistics sector further illustrates the practical implications of our approach. The study focuses on a major e-commerce platform aiming to optimize its supply chain operations, which include a network of fixed suppliers, multiple distribution centers, and a diverse customer base. Key challenges addressed in the study include fluctuating demand patterns, capacity limitations, and the need to adhere to delivery timeframes. The insights derived from this real-world scenario highlight the scalability and efficacy of the proposed framework in tackling the multifaceted challenges of modern supply chain systems.

The remainder of this paper is structured as follows: Section 2 reviews relevant literature. Section 3 defines the problem and presents the mathematical formulations. Robust optimization formulations are detailed in Section 4. Section 5 describes the proposed solution methodology. Numerical results and their analyses are presented in Section 6, while a real-world case study is detailed in Section 7. Section 8 provides managerial insights. Finally, Section 9 concludes the paper, summarizing the key contributions and outlining directions for future research.

## 2. LITERATURE REVIEW

In this section we review the literature of supply chain, logistics and routing problem areas, robust optimization area and researches of learning-based solution methods.

### 2.1. Supply chain, logistics and routing problem areas

Incorporation of routing into supply chain problems is one of the important subfields in the realm of supply chain optimization problems. Javid and Azad [1] presented a simultaneous optimization of location, allocation, capacity, inventory, and routing decisions in a stochastic supply chain system. They hybridized Tabu Search and Simulated Annealing method for the solution method of their problem. Lee *et al.* [2] studied a supply chain network design problem, with decisions of location of facilities, allocation of facilities, and routing. They showed practical applications of their problem and provided a heuristic for the solution method. Schmid *et al.* [3] addressed extensions of classical vehicle routing problem in the context of supply chain management. They focused on extensions of vehicle routing problems with respect to lot-sizing, scheduling, packing, batching, inventory and inter-modality.

Vehicle routing in supply chains are studied for different product types. Awad *et al.* [4] reviewed recent studies in distribution of temperature sensitive products in cold supply chains. Musavi and Bozorgi-Amiri [5] presented a hub location-vehicle scheduling model for perishable products in a food supply chain. They adopted a non-dominated sorting genetic algorithm-II (NSGA-II) to solve their multi-objective model. Also, Vehicle routing for biomass supply chain is addressed by Cao *et al.* [6]. Tavana *et al.* [7] provided a location-inventory-routing model for green supply chains with low-carbon emissions. They also considered uncertainty in their model.

Addressing time window in routing decisions is also considered by researchers. Govindan *et al.* [8] introduced a two-echelon location-routing problem with time-windows in a perishable food supply chain network design problem. Iassinovskaia *et al.* [9] considered time windows and simultaneous pickup and delivery in a supply chain with returnable transport items.

Also, some researchers addressed resiliency in their vehicle routing and supply chain design problem [10–12].

## 2.2. Robust optimization

The field of robust optimization has flourished over the years, yielding a multitude of models that have significantly contributed to its advancement. Among the prominent methodologies, the works pioneered by Mulvey *et al.* [13], Ben-Tal *et al.* [14] and Bertsimas and Sim [15]. These models have not only gained popularity but have also demonstrated their efficacy in addressing complex optimization challenges, solidifying their position as seminal contributions to the field.

Robust optimization has found wide-ranging applications across various domains, evident in numerous papers and studies. One of the examples is the work by Pishvaei *et al.* [16]. They proposed a robust optimization model for handling the uncertain data in a closed-loop supply chain network design problem. Also, Rahbari *et al.* [17] presented two robust models for uncertainty of travel times and freshness of products in a vehicle routing and scheduling problem with cross-docking. Ala *et al.* [18] applied a robust possibilistic mixed-integer linear programming approach to design a multi-objective blood supply chain network, optimizing facility location and distribution decisions to reduce costs and delivery times while addressing supply-demand uncertainties. Goli *et al.* [19] developed a possibilistic programming model and simulation-based solution method for optimizing organ transplant supply chains, focusing on center location, allocation, and distribution under shipment time uncertainty to improve cost efficiency and patient satisfaction.

Habibzadeh Boukani *et al.* [20] considered fixed setup cost parameter and capacity of each hub, as uncertain parameters, in capacitated single and multiple allocation hub location problems. They tackled the uncertainties with robust optimization. Varas *et al.* [21] considered the problem of scheduling production under uncertainty in a scheduling production problem for a sawmill. They analyzed the behavior of the robust solutions with respect to uncertainty using Monte Carlo simulation. Lotfi *et al.* [22] developed a robust stochastic multi-objective programming model for closed-loop supply chain network design, incorporating conditional value at risk to address demand fluctuations and enhance resilience. By the same author, Lotfi *et al.* [23] introduced a robust bi-level stochastic optimization framework for supply chain network design, incorporating viability and sustainability policies to enhance resiliency and manage emission constraints.

In recent years, robust optimization has emerged into data-driven robust optimization [24]. Many researchers have contributed in this field. One example is the paper by Shang *et al.* [25]. They provided a novel uncertainty set using kernel learning for data-driven robust optimization. This method is used by many researchers. Musavi and Bozorgi-Amiri [26] used kernel-based data-driven robust optimization in a hub location-routing problem with travel time uncertainty. Mohseni *et al.* [27] proposed a data-driven two-level transactive energy management framework. They used data-driven robust optimization approach with an uncertainty set constructed using the robust kernel density estimation. Li *et al.* [28] proposed a two-stage data-driven set based robust optimization for a near-zero carbon emission power plant production problem. Also, Lotfi *et al.* [29] proposed a viable supply chain network design model integrating open innovation, blockchain technology, and a data-driven robust optimization approach to enhance antifragility, sustainability, and agility in managing disruptions.

## 2.3. Learning-based solution methods

In an effort to bolster the efficacy of solution methods, particularly those based on meta-heuristics, researchers have recently integrated learning-based techniques into classical methodologies. These advancements predominantly draw inspiration from the realm of machine learning, notably the principles rooted in reinforcement learning. Karimi-Mamaghan *et al.* [30] presented a learning-based metaheuristic based on NSGA-II,  $k$ -Means clustering method, and an Iterated Local Search algorithm for a hub location problem under congestion. Cheng *et al.* [31] addressed a Random-Forest-based metaheuristic in a parallel scheduling problem.

Seyyedabbasi *et al.* [32] used reinforcement learning to hybridize three version of Grey Wolf Optimizer meta-heuristics for solving global optimization problems. Qin *et al.* [33] introduced a reinforcement learning-based

hyper-heuristic for a vehicle routing problem. Also, de Santiago Junior *et al.* [34] proposed a reinforcement learning-based hyper-heuristic to solve multi-objective problems.

In the realm of integrating reinforcement learning techniques, q-learning has emerged as a pivotal approach adopted by numerous researchers. Gölcük and Ozsoydan [35] proposed a q-learning and hyper-heuristic based meta-heuristic. Xi and Lei [36] introduced a q-learning based solution method for a distributed two-stage hybrid flow shop scheduling problem. Ni *et al.* [37] addressed a q-learning based artificial bee colony algorithm for unmanned vehicle path planning problem.

Recently, a progressive trend among researchers involves delving deeper into the realm of deep reinforcement learning to elevate the efficacy of solution methodologies. Zhang *et al.* [38] introduced a meta-learning-based deep reinforcement learning approach to handle complex multiobjective optimization problems by training a meta-model to derive submodels for various subproblems, resulting in a more rapid convergence to the Pareto front. Kallestad *et al.* [39] proposed a deep reinforcement learning hyper-heuristic framework, enhancing heuristic selection compared to Adaptive Large Neighborhood Search and Uniform Random Selection in solving combinatorial optimization problems.

## 2.4. Gap analysis

This paper stands at the intersection of several dynamic domains within supply chain optimization, robust optimization, and learning-based methodologies. However, despite the comprehensive exploration of these areas, certain gaps persist in the current landscape:

While robust optimization techniques are effectively applied to address uncertainties within supply chains, there remains a notable gap concerning the nuanced handling of uncertain demand patterns.

The modification of simulated annealing through reinforcement learning techniques marks an innovative approach in this study. However, the depth of exploration into the adaptability and scalability of this modified simulated annealing method remains somewhat limited. While acknowledging the incorporation of reinforcement learning, a more comprehensive analysis of its efficacy in diverse and complex supply chain scenarios would provide a deeper understanding of its potential and limitations. Investigating the performance of this learning-based simulated annealing across a wider spectrum of supply chain complexities could bridge this existing gap.

The paper presents a compelling case study involving a prominent online store in Iran. While this offers valuable insights into the practical application of the proposed framework within a specific context, the scope's geographical and sectoral limitations present a gap concerning the framework's generalizability.

In essence, this paper makes significant contributions by utilizing robust optimization with uncertain demand considerations, adapting simulated annealing through learning-based approaches, and presenting a practical case study. To address the gaps identified, Table 1 summarizes the key aspects of existing research.

## 3. PROBLEM DEFINITION

### 3.1. Problem statement

The core challenge addressed in this study revolves around the optimization of supply chain operations involving fixed suppliers, distribution centers (DCs), and customers. This multifaceted problem encompasses the allocation of DCs to suppliers, the assignment of customers to DCs, and the establishment of efficient routes for vehicles traversing between DCs and customers, as well as among customers themselves. While the routes from suppliers to DCs remain predetermined, the critical decision lies in determining the optimal allocation of resources to meet demand at each juncture of the supply chain.

Moreover, the problem necessitates a careful consideration of vehicle allocation to each DC, taking into account the limited capacities of both vehicles and DCs. This entails the delicate balance of ensuring that the resources are optimally utilized without overburdening any specific node. The complexities arise from the

TABLE 1. Comparison of related works based on key aspects.

Paper	Year	Routing	Uncertainty	Robust optimization	Time window constraints	Learning-based methods	Reinforcement learning
[1]	2010	✓	✓				
[2]	2010	✓					
[16]	2011		✓	✓			
[3]	2013	✓					
[8]	2014	✓	✓		✓		
[20]	2016		✓	✓			
[9]	2017	✓	✓		✓		
[17]	2019	✓	✓	✓			
[4]	2021	✓	✓				
[6]	2021	✓					
[7]	2021	✓	✓				
[36]	2022					✓	✓
[39]	2023					✓	✓
[40]	2024					✓	✓
This study		✓	✓	✓	✓	✓	✓

interplay of various factors, such as the varying demand patterns from customers, differing capacities of the DCs, and the limitations of the vehicles in terms of volume.

Adding to the intricacy of the problem, customers present soft time windows, which further complicate the scheduling of deliveries. The determination of optimal delivery times must account for considerations of earliness and lateness, with penalties associated with deviations from specified time frames. This temporal dimension introduces an additional layer of complexity to the optimization process, as it necessitates a fine-tuned balancing act between meeting customer expectations and efficiently utilizing the available resources.

In Figure 1, a graphical representation of the problem is provided, illustrating the network of fixed suppliers, distribution centers, and customers, as well as the various routes connecting them. This visual aid offers a concise overview of the intricate logistical relationships that need to be optimized. The primary objective of this study is to develop a comprehensive framework that addresses these complexities. Through this endeavor, we aim to achieve a streamlined, cost-effective, and customer-centric supply chain operation that maximizes resource utilization while meeting delivery expectations.

### 3.2. Problem formulation

#### Indices

- $i, j \in \{0, 1, 2, \dots, N + 1\}$  Index for customers
- $k \in \{1, 2, \dots, C\}$  Index for distribution centers
- $u \in \{1, 2, \dots, S\}$  Index for suppliers

#### Parameters

- VE Vehicle usage cost
- VC Vehicle capacity
- $d_i$  Demand of customer  $i$
- $A_i$  Start of time window at customer  $i$
- $B_i$  End of time window at customer  $i$
- $CC_k$  Capacity of distribution center  $k$
- $M$  Big number

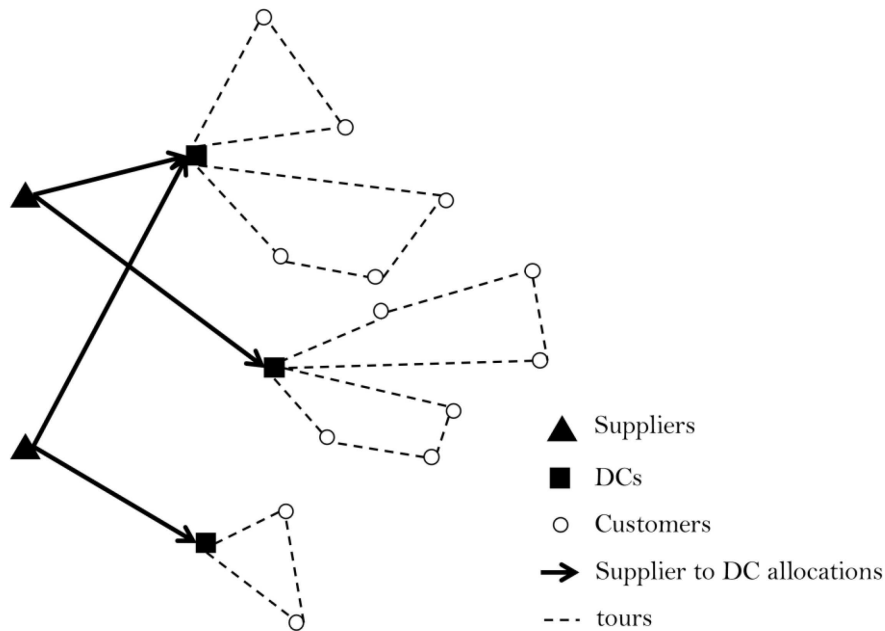


FIGURE 1. Graphical representation of the problem.

- $TC_{ij}$  Transportation cost between customer  $i$  and  $j$   
 $TC_{ki}$  Transportation cost between distribution center  $k$  and customer  $i$   
 $TC_{uk}$  Transportation cost between supplier  $u$  and distribution center  $k$

*Positive variables*

- $at_i$  Arrival time of delivery at customer  $i$   
 $load_i$  Load of vehicle in delivery of customer  $i$   
 $l_i$  Lateness time of delivery at customer  $i$   
 $e_i$  Earliness time of delivery at customer  $i$   
 $\beta_k$  Sum of products consolidated in distribution center  $k$   
 $\theta_{uk}$  Flow of products from supplier  $u$  to distribution center  $k$

*Integer variable*

- $v_k$  Number of vehicles allocated to distribution center  $k$

*Binary variables*

- $x_{ijk}$  Equal to 1 if customer  $i$  is served before customer  $j$ , which both nodes allocated to distribution center  $k$ , otherwise 0  
 $y_{ik}$  Equal to 1 if customer  $i$  is allocated to distribution center  $k$ , otherwise 0

*Objective function*

$$\text{Min} \sum_{u=1}^S \sum_{k=1}^C TC_{uk} \theta_{uk} + VE \sum_{k=1}^C v_k + \sum_{i=1}^N at_i + \sum_{i=1}^N e_i + \sum_{i=1}^N l_i. \quad (1)$$

Constraints

$$\sum_{k=1}^C y_{ik} = 1 \quad \forall i = 1, 2, \dots, N \tag{2}$$

$$2x_{ijk} \leq y_{jk} + y_{ik} \quad \forall i, j = 1, 2, \dots, N, \quad k = 1, 2, \dots, C \tag{3}$$

$$\sum_{\substack{i=0 \\ i \neq j}}^N x_{ijk} = y_{jk} \quad \forall j = 1, 2, \dots, N, \quad k = 1, 2, \dots, C \tag{4}$$

$$\sum_{\substack{j=1 \\ i \neq j}}^{N+1} x_{ijk} = y_{ik} \quad \forall i = 1, 2, \dots, N, \quad k = 1, 2, \dots, C \tag{5}$$

$$\sum_{j=1}^N x_{0,jk} = v_k \quad \forall k = 1, 2, \dots, C \tag{6}$$

$$\sum_{i=1}^N x_{i,N+1,k} = v_k \quad \forall k = 1, 2, \dots, C \tag{7}$$

$$at_i \geq TC_{ki}x_{0,ik} \quad \forall i = 1, 2, \dots, N, \quad k = 1, 2, \dots, C \tag{8}$$

$$at_j + M(1 - x_{ijk}) \geq TC_{ij} + at_i \quad \forall i, j = 1, 2, \dots, N, \quad k = 1, 2, \dots, C, \quad i \neq j \tag{9}$$

$$\text{load}_i \geq \tilde{d}_i x_{0,ik} \quad \forall i = 1, 2, \dots, N, \quad k = 1, 2, \dots, C \tag{10}$$

$$\text{load}_j + M(1 - x_{ijk}) \geq \tilde{d}_i + \text{load}_i \quad \forall i, j = 1, 2, \dots, N, \quad k = 1, 2, \dots, C, \quad i \neq j \tag{11}$$

$$\text{load}_i \leq VC \quad \forall i = 1, 2, \dots, N \tag{12}$$

$$at_i \leq B_i + l_i \quad \forall i = 1, 2, \dots, N \tag{13}$$

$$at_i + e_i \geq A_i \quad \forall i = 1, 2, \dots, N \tag{14}$$

$$\sum_{i=1}^N y_{ik} \tilde{d}_i \leq \beta_k \quad \forall k = 1, 2, \dots, C \tag{15}$$

$$\sum_{u=1}^S \theta_{uk} = \beta_k \quad \forall k = 1, 2, \dots, C \tag{16}$$

$$\beta_k \leq CC_k \quad \forall k = 1, 2, \dots, C \tag{17}$$

$$at_i, \text{load}_i, e_i, l_i, \beta_k, \theta_{uk} \geq 0, v_k \in \mathbb{Z}, x_{ijk}, y_{ik} \in \{0, 1\} \quad \forall i, j = 0, 1, 2, \dots, N + 1, \quad k = 1, 2, \dots, C. \tag{18}$$

Objective function (1) consists of five parts. Part one calculates total costs of transportations between suppliers and distribution centers. Part two calculates cost of vehicle usages in distribution centers. Part three calculates cost of transportations between customers. Part four calculates cost of earliness in deliveries and part five calculates cost of lateness in deliveries.

Equation (2) ensures that each customer should be allocated to only one distribution center. Equation (3) specifies that each vehicle inside distribution center  $k$  will server customers  $i$  and  $j$  only if these customers are assigned to distribution center  $k$ . Equations (4) and (5) are classic constraints of a vehicle routing problem. These equations examine whether each vehicle enters and leaves the assigned customer only once. Equations (6) and (7) verify that the start and end of vehicle routes within distribution centers depend on the number of vehicles in those distribution centers. Equations (8) and (9) calculate the arrival time of vehicles at customers. Equations (10) and (11) calculate the load of vehicles when serving customers. Equation (12) ensures that load of vehicles do not exceed capacity of them. Equations (13) and (14) calculate the earliness and lateness of orders.

Equation (15) calculates the total amount of products consolidated at each distribution center. Equation (16) calculates the number of products from each supplier to distribution centers. Equation (17) ensures that number of products at each distribution center do not exceed capacity of them. Finally, equation (18) shows the type of variables in the proposed model.

#### 4. ROBUST FORMULATION

We suggest employing a robust optimization approach to address uncertainties in demand. The objective of this robust optimization is to design routes that factor in the fluctuations in demand. The robust optimization formulation is inspired by the work of Bertsimas and Sim [15].

Assume the optimization problem is as follows:

$$\begin{aligned} \min \quad & cx \\ \text{s.t.} \quad & \sum_j \widetilde{a}_{ij} x_j \leq b_i, \quad \forall i, \quad \forall \widetilde{a}_{ij} \in J_i \\ & x \in X, \quad x_j \geq 0. \end{aligned} \quad (19)$$

Or:

$$\begin{aligned} \min \quad & cx \\ \text{s.t.} \quad & \max_{\widetilde{a}_{ij} \in J_i} \left( \sum_j \widetilde{a}_{ij} x_j \right) \leq b_i, \quad \forall i \\ & x \in X, \quad x_j \geq 0. \end{aligned} \quad (20)$$

The parameter  $\widetilde{a}_{ij}$  is subject to uncertainty and falls within the range  $[\overline{a}_{ij} - \widehat{a}_{ij}, \overline{a}_{ij} + \widehat{a}_{ij}]$  where  $\overline{a}_{ij}$  is the nominal value and  $\widehat{a}_{ij}$  is the maximum deviation from the nominal value.  $J_i$  represents the set of uncertain parameters for constraint  $i$ . Thus, we can state:

$$\eta_{ij} = \frac{\widetilde{a}_{ij} - \overline{a}_{ij}}{\widehat{a}_{ij}} \quad (21)$$

where the variable  $\eta_{ij}$  ranges from  $-1$  to  $1$ . Problem (20) can be expressed in the following manner:

$$\begin{aligned} \min \quad & cx \\ \text{s.t.} \quad & \sum_j \overline{a}_{ij} x_j + \max_{\eta_{ij}} (\widehat{a}_{ij} \eta_{ij} x_j) \leq b_i, \quad \forall i \\ & \sum_j \eta_{ij} \leq \Gamma_i, \quad \forall i \\ & 0 \leq \eta_{ij} \leq 1, \quad \forall j \in J_i \\ & x_j \geq 0, \quad \forall j \end{aligned} \quad (22)$$

where we establish a budget uncertainty  $\Gamma_i$  to restrict the accumulation of deviation. This implies that we assume the uncertain parameter variation cannot surpass a defined threshold  $\Gamma_i$ . For the protection function in (22) we have:

$$\begin{aligned} \max_{\eta_{ij}} \quad & (\widehat{a}_{ij} \eta_{ij} x_j) \\ \text{s.t.} \quad & \sum_j \eta_{ij} \leq \Gamma_i, \quad \forall i \end{aligned}$$

$$\begin{aligned} 0 \leq \eta_{ij} \leq 1, & \quad \forall j \in J_i \\ x_j \geq 0, & \quad \forall j. \end{aligned} \tag{23}$$

The dual of (23) is:

$$\begin{aligned} \min \Gamma_i z_i + \sum_{j \in J_i} p_{ij} \\ \text{s.t. } \Gamma_i + p_{ij} \geq \widehat{a}_{ij} x_j, & \quad \forall i, j \in J_i \\ p_{ij} \geq 0, & \quad \forall i, j \in J_i \\ z_i \geq 0, & \quad \forall i, \\ x_j \geq 0, & \quad \forall j \in J_i. \end{aligned} \tag{24}$$

Then the robust counterpart of (22) will be:

$$\begin{aligned} \min cx \\ \text{s.t. } \sum_j \overline{a}_{ij} x_j + \Gamma_i z_i + \sum_{j \in J_i} p_{ij} \leq b_i, & \quad \forall i \\ \Gamma_i + p_{ij} \geq \widehat{a}_{ij} x_j, & \quad \forall i, j \in J_i \\ p_{ij} \geq 0, & \quad \forall i, j \in J_i \\ z_i \geq 0, & \quad \forall i, \\ x_j \geq 0, & \quad \forall j \in J_i. \end{aligned} \tag{25}$$

In our problem, the fluctuation in demand is constrained within a range of  $[\overline{d}_i - \widehat{d}_i, \overline{d}_i + \widehat{d}_i]$ , where  $\overline{d}_i$  represents the base demand of customer  $i$ , which serves as a reference point for evaluating fluctuations in demand and  $\widehat{d}_i$  signifies the highest allowable deviation in demand at customer  $i$ . So, we can state  $\phi_i = \frac{\widehat{d}_i - \overline{d}_i}{\overline{d}_i}$ .

Using the formulations mentioned above, we obtain a robust formulation of constraints (10), (11) and (15) by:

$$\overline{d}_i x_{0,ik} + \Gamma_i z'_i + p'_i \leq \text{load}_i \quad \forall i = 1, 2, \dots, N, \quad k = 1, 2, \dots, C \tag{26}$$

$$\Gamma_i + p'_i \geq \widehat{d}_i x_{0,ik} \quad \forall i = 1, 2, \dots, N, \quad k = 1, 2, \dots, C \tag{27}$$

$$\overline{d}_i + \Gamma_i z''_i + p''_i + \text{load}_i \leq \text{load}_j + M(1 - x_{ijk}) \quad \forall i, j = 1, 2, \dots, N, \quad k = 1, 2, \dots, C, \quad i \neq j \tag{28}$$

$$\Gamma_i + p''_i \geq \widehat{d}_i \quad \forall i = 1, 2, \dots, N \tag{29}$$

$$\sum_{i=1}^N y_{ik} \overline{d}_i + \sum_{i=1}^N (\Gamma_i z'''_i + p'''_i) \leq \beta_k \quad \forall k = 1, 2, \dots, C \tag{30}$$

$$\sum_{i=1}^N (\Gamma_i + p'''_i) \geq \sum_{i=1}^N y_{ik} \widehat{d}_i \quad \forall k = 1, 2, \dots, C \tag{31}$$

$$z'_i, z''_i, z'''_i, p'_i, p''_i, p'''_i \geq 0 \quad \forall i = 1, 2, \dots, N. \tag{32}$$

Based on the above formulations, the robust mixed integer linear programming model is as follows:

$$\text{Min } \sum_{u=1}^S \sum_{k=1}^C \text{TC}_{uk} \theta_{uk} + \text{VE} \sum_{k=1}^S v_k + \sum_{i=1}^N at_i + \sum_{i=1}^N e_i + \sum_{i=1}^N l_i$$

s.t.

$$(2)-(9), (12)-(14), (16)-(18), (26)-(32).$$

DC 1 routes	2	3	0	5	7
DC 2 routes	6	0	1	4	

FIGURE 2. Example of solution representation.

## 5. SOLUTION METHODOLOGY

The proposed model is constructed around a Vehicle Routing Problem (VRP) framework, which provides a formal mathematical representation for the optimization of distribution routes. It is important to note that VRP is a well-known NP-hard problem, implying that finding an exact optimal solution is computationally infeasible for larger instances, limiting its applicability in real-world scenarios. This restriction necessitates the exploration of alternative methodologies capable of handling large-scale instances within reasonable time frames. To overcome this challenge, we employ a Learning-based Simulated Annealing (LSA) algorithm, a powerful metaheuristic approach that leverages both the principles of simulated annealing and the adaptability of learning. By integrating these techniques, our approach aims to efficiently navigate the solution space, yielding optimal or near-optimal solutions for complex, large-scale VRP instances, thus providing a practical and effective tool for real-world supply chain optimization.

### 5.1. Simulated annealing

This section presents the simulated annealing algorithm developed to solve the proposed model. Simulated Annealing (SA) is a powerful metaheuristic optimization technique inspired by the process of annealing in metallurgy. It is particularly well-suited for tackling combinatorial optimization problems, including the VRP, due to its ability to efficiently explore solution spaces and escape local optima [41].

In order to consciously select a neighborhood structure to generate a new solution, three reinforcement learning approaches are employed. These approaches are designed to learn and assess the value and efficiency of each available neighborhood structure. By leveraging the principles of reinforcement learning, the algorithm dynamically adapts its exploration strategy, focusing on those moves that are most likely to lead to improved solutions. This adaptive nature enhances the algorithm's ability to effectively navigate the complex solution space of large-scale VRP instances, ultimately contributing to the attainment of high-quality solutions. Through the integration of simulated annealing with reinforcement learning, our approach synergistically combines the strengths of both techniques, providing a robust and versatile tool for addressing real-world supply chain optimization challenges.

*Solution representation.* The representation of the solution in our proposed model employs  $C$  number of vectors, which collectively encapsulate the allocations and routing decisions. Each vector corresponds to the routes of a specific DC, effectively representing both the allocation of customers to each DC and the subsequent routes taken to service these allocated customers. Within each vector, the presence of zero values serves as delimiters, demarcating the boundaries between different vehicles within the DC's fleet. The allocation problem from suppliers to DCs is treated as a secondary optimization stage, easily solved as an LP problem. Furthermore, additional constraints, such as capacity limitations, are incorporated into the objective function as violation variables. Figure 2 shows a solution representation. In this example with 2 DCs, nodes 2, 3, 5 and 7 are allocated to DC 1. First vehicle in this DC serves nodes 2 and 3 and second vehicle serves node 5 then node 7.

In this paper, we have custom-designed seven distinct neighborhood search structures to systematically navigate and exploit solution spaces. These tailored search structures play a pivotal role in enhancing the efficiency and effectiveness of our optimization approach:

- **Insertion Intra:** this neighborhood search structure focuses on intra-route operations. It involves the insertion of a customer within an existing route of a single vehicle.

- **Insertion Inter:** the Insertion Inter structure extends the scope to inter-route operations. It entails the transfer of a customer from one vehicle’s route to another.
- **Swap Intra:** the Swap Intra structure focuses on intra-route operations, similar to the Insertion Intra structure. However, in this case, the operation involves swapping the positions of two customers within the same route of a single vehicle.
- **Swap Inter:** Swap Inter broadens the scope to inter-route operations. It involves the exchange of customers between the routes of two different vehicles.
- **Reversion Intra:** the Reversion Intra structure addresses intra-route operations. It focuses on reversing segments of a route within a single vehicle’s itinerary.
- **Vehicle Addition:** Vehicle Addition introduces a different dimension to the optimization process. It involves the addition of an entirely new vehicle to the fleet.
- **Vehicle Removal:** conversely, Vehicle Removal considers the removal of a vehicle from the fleet. This operation evaluates the impact of downsizing the fleet on the overall solution.

## 5.2. Learning-based simulated annealing

In meta-heuristic algorithms, a critical consideration lies in the strategic selection of neighborhood search structures (SS) for generating new solutions. Frequently, SSs are chosen at random, lacking a clear indication of their suitability. However, a more informed approach involves assessing the performance of each SS and making selections based on this evaluation. For instance, when a SS is employed without resulting in any improvement, it suggests that the chosen neighborhood search may not be well-suited for the current state and thus warrants some form of penalty. Conversely, if a SS leads to a substantial enhancement, it indicates its efficacy and merits a corresponding reward. This principle is framed within a learning context, where the model adapts and refines its understanding of the value of each action by tracking the associated penalties and rewards.

Consider the value function for each action, denoted by  $Q(a)$  for all  $a$  in the set  $A$ , is pivotal. Additionally, let  $SS(A)$  represent the frequency of selecting action  $A$  (which corresponds to a specific neighborhood search structure). Further, let’s envision a circumstance where upon taking an action, the system receives a reward  $R$ , and  $\alpha$  denotes the learning rate. Within this context, the formulations for each of the three proposed learning variants can be succinctly presented as follows:

- **Adaptive Approach:** this method is a dynamic update rule fundamental that estimates the value function of each action by averaging the rewards actually received, providing an efficient use of memory resources. This approach updates value functions iteratively, allowing the algorithm to adapt its estimations over time and effectively capture changes in the environment:

$$Q(a) = Q(a) + \frac{R - Q(a)}{SS(a)}.$$

- **Dynamic Environment Adaptation:** this approach is designed to address dynamic environments, where the rewards associated with actions may vary over time. While the adaptive approach method assumes a stationary environment with a consistent probability distribution of rewards, this assumption can be easily violated:

$$Q(a) = Q(a) + \alpha[R - Q(a)].$$

- **Q-Learning:** this approach is an off-policy reinforcement learning algorithm designed to ascertain optimal or near-optimal solutions within Markov Decision Process problems. For effective application, it necessitates the delineation of states, actions, and rewards.

However, the definition of states takes on a distinct character in Q-Learning. It hinges on quantifying the number of instances in which the current solution fails to exhibit improvement. Specifically, if  $UI_t$  denotes the count of times the current solution proves unimproved, the state variable is discretized into distinct intervals,

each corresponding to a different state. These states serve as a reflection of the optimization process and guide the algorithm's decision-making.

$$s_t = \begin{cases} 1, & 0 \leq \text{UI}_t < \zeta_1 \\ 2, & \zeta_1 \leq \text{UI}_t < \zeta_2 \\ 3, & \zeta_2 \leq \text{UI}_t < \zeta_3 \\ 4, & \zeta_3 \leq \text{UI}_t < \zeta_4 \\ 5, & \zeta_4 \leq \text{UI}_t. \end{cases}$$

In this dynamic environment, when the system state is  $s_t$  and action  $a_t$  is chosen, and a corresponding reward  $r_t$  is received, the action-value function is updated using a temporal-difference learning rule. This update considers both the immediate reward ( $r_t$ ) and the estimated value of the subsequent state ( $s_{t+1}$ ), weighed by the discount factor  $\gamma \in (0, 1)$ . The update rule ensures that the algorithm learns to anticipate and evaluate the long-term consequences of its actions, contributing to the overall refinement of the action-value estimates.

$$Q(s_t, a_t) = Q(s_t, a_t) + \alpha \left[ r_t + \gamma \max_{b \in A} Q(s_{t+1}, b) - Q(s_t, a_t) \right].$$

The selection of neighborhood search structures is governed by the epsilon-greedy strategy. This approach strikes a balance between exploration and exploitation, allowing the algorithm to occasionally choose a random structure (exploration) while predominantly favoring the selection of structures that have shown higher promise in previous iterations (exploitation).

$$\text{SS}(a_t) = \begin{cases} \text{Random Selection}, & \text{rand}(0, 1) < \epsilon \\ \underset{s_t \in S}{\text{argmax}}(Q(s_t, a_t)), & \text{Otherwise.} \end{cases}$$

Figure 3 illustrates the flowchart of our proposed learning-based SA algorithm, which integrates reinforcement learning principles with traditional SA optimization.

## 6. NUMERICAL RESULTS

### 6.1. Performance evaluation

In order to comprehensively evaluate the performance of the proposed algorithms, it is imperative to subject these methods to rigorous testing across a variety of test problems. Given that there exists no established standard test problem specifically tailored for our proposed model, the datasets are meticulously generated at random, drawing inspiration from data employed in prior research conducted by Cordeau *et al.* [42]. The dataset used from [42] encompasses location of customers and DCs, demands and time windows. These custom-generated datasets serve as the foundation for our evaluation, reflecting real-world complexities and nuances. To ensure the integrity of the experimental setup, other essential parameters are systematically generated. These parameters are generated as follows:  $\text{VE} = 2000$ ,  $\text{VC} = 70$  and  $\text{CC}_k \sim U\left(\frac{\sum_i d_i}{N}, \frac{1.2 \times \sum_i d_i}{N}\right)$ .

It's important to note that all proposed algorithms are meticulously coded and executed in the Julia programming language. The computations are performed on a computing system powered by an Intel Core i5 processor, running at 2 GHz, and equipped with 16 GB of memory. To validate the quality of the obtained results, the Gurobi solver is employed to ascertain optimal solutions.

The size of the test problems is determined by the specific counts of suppliers, DCs, and customers involved in each scenario. In our computational results, the size of each test problem is succinctly represented as a tuple (S, C, N), where S, C, and N denote the respective quantities of suppliers, DCs, and customers in the given instance.

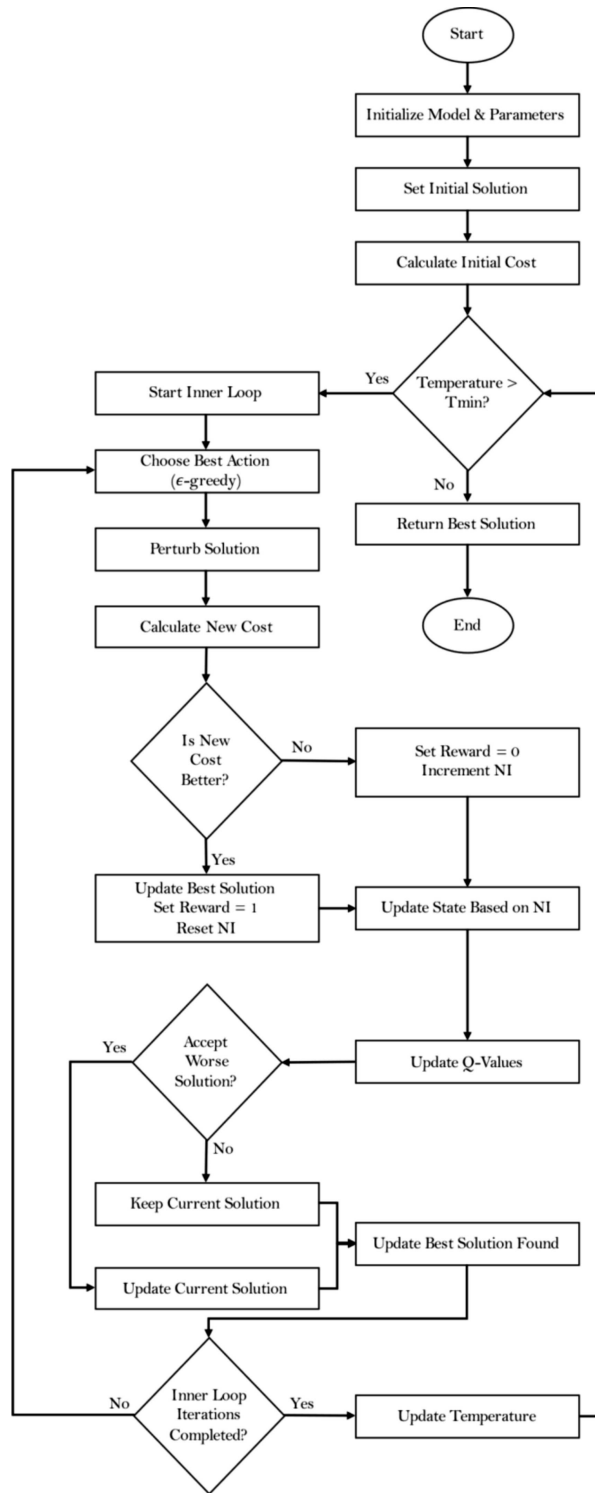


FIGURE 3. Flowchart of the learning-based SA algorithm.

The test problems serve as the foundation for comparing various solution methodologies. Due to the stochastic nature of metaheuristic algorithms in solution selection and moves, each problem instance is subjected to 20 replications for robustness in the evaluation process. The methods under evaluation include:

- Optimal solver.
- Simulated annealing without learning approach (SA).
- Simulated annealing with “adaptive approach” learning method (SA-AA).
- Simulated annealing with “Dynamic Environment Adaptation” learning method (SA-DEA).
- Simulated annealing with Q-Learning method (SA-QL).

Table 2 offers a comprehensive assessment of algorithm performance, delving into their effectiveness based on respective objective values. Notably, the integration of learning approaches yields a discernible enhancement in the overall efficiency of the SA algorithm, underscoring the significance of incorporating these adaptive techniques.

Upon closer examination of the numerical results, it becomes apparent that, for problem instances of smaller and moderate scales, the various learning-based SA algorithms produce comparable outcomes. This suggests that, in scenarios characterized by relatively constrained complexities, the choice of learning strategy may have a more subtle impact on the final solution. However, as we transition into larger-scale problem instances, the distinctive advantage of Q-Learning becomes apparent, exhibiting a superior performance that positions it as the preferred choice for such formidable challenges.

Turning our attention to Table 3, we delve deeper into the algorithms’ performance, this time gauging their computational efficiency. This crucial metric allows us to assess not only the quality of solutions but also the speed at which they are derived. It’s worth noting that while the incorporation of learning-based strategies may introduce a marginal increase in computational time, this investment is often outweighed by the considerable improvements in solution quality.

These findings collectively underscore the versatility and robustness of the proposed learning-based SA algorithms across a diverse range of problem sizes. Their stability across multiple replications, an aspect exemplified by the 20 replications per problem instance, fortifies their credibility and efficacy in tackling the intricate challenges of supply chain optimization.

## 6.2. Sensitivity analysis on robustness

In this section, we undertake a sensitivity analysis to evaluate how variations in specific parameters impact the robustness of our proposed optimization framework. This analysis provides crucial insights into the system’s adaptability and performance under changing conditions.

Figure 4 showcases the impact of alterations in total costs on the robustness of our optimization approach. Notably, there is a clear increasing trend observed in robustness levels as total costs rise. This suggests that, within the context of this analysis, higher costs correspond to increased levels of robustness. This could indicate that, in this scenario, investing in higher-cost resources may lead to more resilient and adaptable solutions.

Figure 5 illustrates the sensitivity of the optimization framework to changes in the number of vehicles in the network. We observe an increasing trend in robustness levels as the number of vehicles is augmented. However, it’s important to note that while there is an upward trajectory, the slope of the trend is relatively moderate. This implies that, within the examined range, the impact of vehicle count on robustness is notable but not overwhelmingly influential. This may suggest that other factors also play significant roles in determining the optimal fleet size.

Figure 6 delves into the sensitivity of the optimization framework with regard to the transportation costs between suppliers and distribution centers. In the lower levels of robustness (0.1 to 0.4), there isn’t a substantial deviation observed. However, as we move beyond this range, there is a noticeable and consistent increasing trend. This suggests that, at lower levels of robustness, the system may be less sensitive to variations in these transportation costs. As robustness increases, however, the impact of these costs becomes more pronounced.

TABLE 2. Objective function comparison across all algorithms.

#	Size	Optimal	SA	SA-AA	SA-DEA	SA-QL
1	(2, 2, 10)	14 194.16	<b>14 194.16</b>	<b>14 194.16</b>	<b>14 194.16</b>	<b>14 194.16</b>
2	(2, 3, 10)	15 311.26	15 337.83	<b>15 311.26</b>	15 337.83	<b>15 311.26</b>
3	(2, 2, 15)	19 646.81	19 761.54	19 679.82	<b>19 646.81</b>	19 679.82
4	(2, 3, 15)	24 656.54	<b>24 656.54</b>	24 709.69	24 717.02	<b>24 656.54</b>
5	(2, 4, 15)	24 533.81	<b>24 533.81</b>	24 618.69	<b>24 533.81</b>	24 696.49
6	(3, 4, 20)	30 869.42	<b>30 869.42</b>	31 058.97	<b>30 989.09</b>	31 413.60
7	(3, 4, 25)	33 733.78	<b>33 733.78</b>	35 746.38	35 916.44	35 753.81
8	(3, 4, 30)	40 444.32	40 748.29	<b>40 555.50</b>	40 708.44	40 752.94
9	(3, 5, 30)	40 134.33	41 249.13	40 777.00	<b>40 522.75</b>	40 871.95
10	(3, 5, 35)	46 956.22	<b>47 166.14</b>	47 878.05	47 813.64	47 784.27
11	(4, 5, 35)	42 661.51	<b>43 650.75</b>	44 268.55	44 038.84	44 026.69
12	(3, 5, 40)	51 013.40	51 633.60	<b>51 159.94</b>	51 272.83	51 565.69
13	(3, 5, 45)	55 212.11	55 851.47	56 013.09	56 235.08	<b>55 790.47</b>
14	(4, 5, 45)	53 011.56	55 686.67	55 990.27	55 928.23	<b>55 550.69</b>
15	(3, 5, 50)	61 835.63	<b>61 835.63</b>	62 000.93	63 020.92	61 877.97
16	(3, 5, 55)	64 158.91	67 578.70	67 863.68	68 236.92	<b>67 505.03</b>
17	(4, 5, 55)	68 119.43	70 190.46	70 523.58	70 289.65	<b>70 130.58</b>
18	(4, 5, 60)	70 415.14	72 450.16	72 442.59	72 533.00	<b>72 404.20</b>
19	(4, 6, 65)	74 142.22	75 645.87	76 089.52	76 145.30	<b>75 477.01</b>
20	(4, 6, 70)	69 431.45	80 276.88	80 933.69	81 500.40	<b>80 124.97</b>
21	(5, 6, 75)	82 243.31	85 211.33	86 495.53	87 496.04	<b>84 942.89</b>
22	(5, 5, 80)	–	95 863.90	96 541.05	96 646.18	<b>94 925.55</b>
23	(5, 5, 85)	–	98 804.82	98 706.36	100 492.69	<b>98 367.77</b>
24	(5, 6, 90)	–	105 367.21	106 198.51	106 882.37	<b>104 668.88</b>
25	(5, 7, 100)	–	131 268.97	131 398.47	131 919.72	<b>130 913.92</b>

**Notes.** Bold values indicate the best performance among the compared methods (SA, SA-AA, SA-DEA, SA-QL)

This insight could guide decision-making in terms of supplier selection and transportation cost management strategies.

### 7. CASE STUDY

In this section, a comprehensive case study addressing the proposed problem will be presented, offering a real-world application that illustrates the nuances and potential solutions.

The case study was conducted utilizing real-world data sourced from one of the largest online commercial stores in Iran, renowned for its extensive market reach and consumer base. The datasets employed in this study were specific to the Tehran province, providing a representative sample reflective of the region’s diverse market dynamics and consumer behaviors. Included within this dataset were the geographic coordinates of suppliers, DCs, and customers, alongside essential demand data encompassing customer-specific time windows and their corresponding demands. Furthermore, the dataset encompassed crucial information regarding the capacities of vehicles and DCs, crucial for analyzing logistics and supply chain dynamics within the Tehran province.

The dataset comprises a total of 1887 customer nodes, 13 DC nodes, and 2 supplier nodes, showcasing a comprehensive network of key stakeholders within the logistics framework. The spatial arrangement and distribution of these nodes are visually depicted in Figure 7, providing an illustrative representation of their strategic placements within the Tehran province. The demands of customers are segmented across distinct time windows of 9–12 AM, 12–3 PM, 3–6 PM, and 6–9 PM, reflecting the diversified temporal patterns within which customer demands are prevalent.

TABLE 3. Computational time (in seconds) comparison across all algorithms.

#	Size	Optimal	SA	SA-AA	SA-DEA	SA-QL
1	(2, 2, 10)	0.05	0.31	0.38	0.41	0.37
2	(2, 3, 10)	0.05	0.41	0.54	0.39	0.46
3	(2, 2, 15)	0.15	0.42	0.43	0.32	0.45
4	(2, 3, 15)	0.15	0.43	0.44	0.41	0.47
5	(2, 4, 15)	0.17	0.44	0.49	0.46	0.67
6	(3, 4, 20)	4.33	0.54	0.53	0.43	0.56
7	(3, 4, 25)	0.46	0.61	0.56	0.40	0.50
8	(3, 4, 30)	8.21	0.54	0.68	0.42	0.63
9	(3, 5, 30)	9.59	0.52	0.55	0.41	0.53
10	(3, 5, 35)	10.16	0.51	0.56	0.42	0.59
11	(4, 5, 35)	8.44	0.50	0.56	0.86	0.56
12	(3, 5, 40)	161.62	0.51	0.67	0.54	0.60
13	(3, 5, 45)	194.81	0.57	0.64	0.59	0.67
14	(4, 5, 45)	209.60	0.59	0.65	0.50	0.59
15	(3, 5, 50)	558.84	0.53	0.61	0.50	0.61
16	(3, 5, 55)	801.91	0.59	0.65	0.59	0.67
17	(4, 5, 55)	850.51	0.61	0.57	0.49	0.54
18	(4, 5, 60)	1614.77	0.55	0.58	0.47	0.55
19	(4, 6, 65)	1992.21	0.59	0.77	0.60	0.68
20	(4, 6, 70)	2801.59	0.65	0.70	0.66	0.72
21	(5, 6, 75)	4980.10	0.72	0.72	0.62	0.75
22	(5, 5, 80)	–	0.66	0.73	0.59	0.70
23	(5, 5, 85)	–	0.66	0.72	0.61	0.72
24	(5, 6, 90)	–	0.73	0.81	0.66	0.81
25	(5, 7, 100)	–	0.76	0.84	0.70	0.85

The case study is resolved utilizing the proposed solution method expounded upon in Section 5, demonstrating its applicability and efficacy in addressing the complex logistical challenges inherent in Tehran’s diverse market landscape.

For this case study, the dataset is approached with different levels of robustness, strategically tailored to navigate the inherent uncertainties embedded within the demands data. Figure 8 visually represents the outcomes of the case study, delineating the solutions for the vehicle routes under two distinct scenarios: one without zero level of robustness and the other employing a full level of robustness. These illustrations vividly depict the stark differences in route optimization strategies when confronted with varying degrees of uncertainty within the dataset.

In the case study’s robustness analysis, the implementation of a full level of robustness reveals a striking shift in the network’s configuration. The optimization strategy, designed to accommodate uncertainties in demand data, prompts a notable consolidation of routes, directing them towards a more centralized hub of distribution. This centralization is pivotal in leveraging the efficiency of the logistics network. By funneling deliveries through a concentrated set of DCs, the system gains resilience against fluctuations in demand. Consequently, this approach not only streamlines logistical operations but also fortifies the network against potential disruptions or fluctuations in customer demands.

Moreover, the application of a full level of robustness facilitates a broader distribution of demand among multiple DCs. This dispersion plays a pivotal role in enhancing the network’s adaptability and resilience. By decentralizing the delivery points across an expanded array of DCs, the system ensures a more balanced and responsive approach to fulfilling customer demands. This distribution strategy minimizes the risk associated with overburdening specific DCs while ensuring a more equitable and efficient allocation of resources. As a

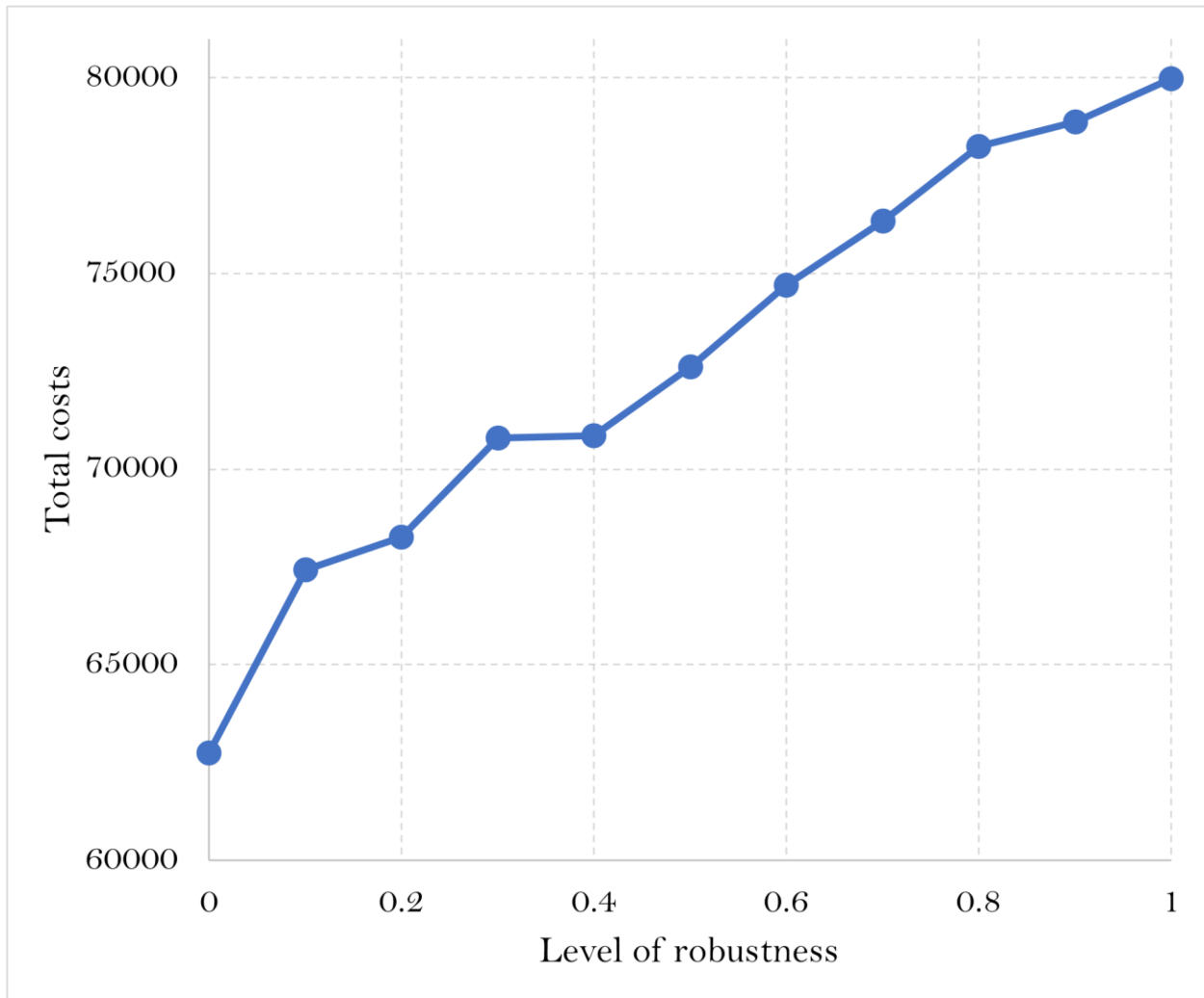


FIGURE 4. Changes in level of robustness to total costs.

result, the network becomes more adept at managing varying demand patterns, thereby elevating its overall performance and responsiveness to the dynamic market landscape.

An additional crucial facet analyzed within this case study is the impact of vehicle cost (VE) on the logistics framework. This parameter significantly influences both the number of vehicles employed within the network and the cumulative transportation durations. Figure 9 effectively illustrates the fluctuations in total transportation costs and total vehicle costs concerning variations in the vehicle cost parameter.

The depiction in Figure 8 illuminates the direct correlation between vehicle cost and the incurred expenses in transportation. As the vehicle cost parameter undergoes alterations, the total transportation costs and overall vehicle expenses experience corresponding shifts. This visual representation underscores the sensitivity of the logistics system to changes in vehicle costs, elucidating how adjustments in this parameter can directly influence the economic aspects of the network, thereby impacting the number of vehicles utilized and the resultant total transportation expenses.

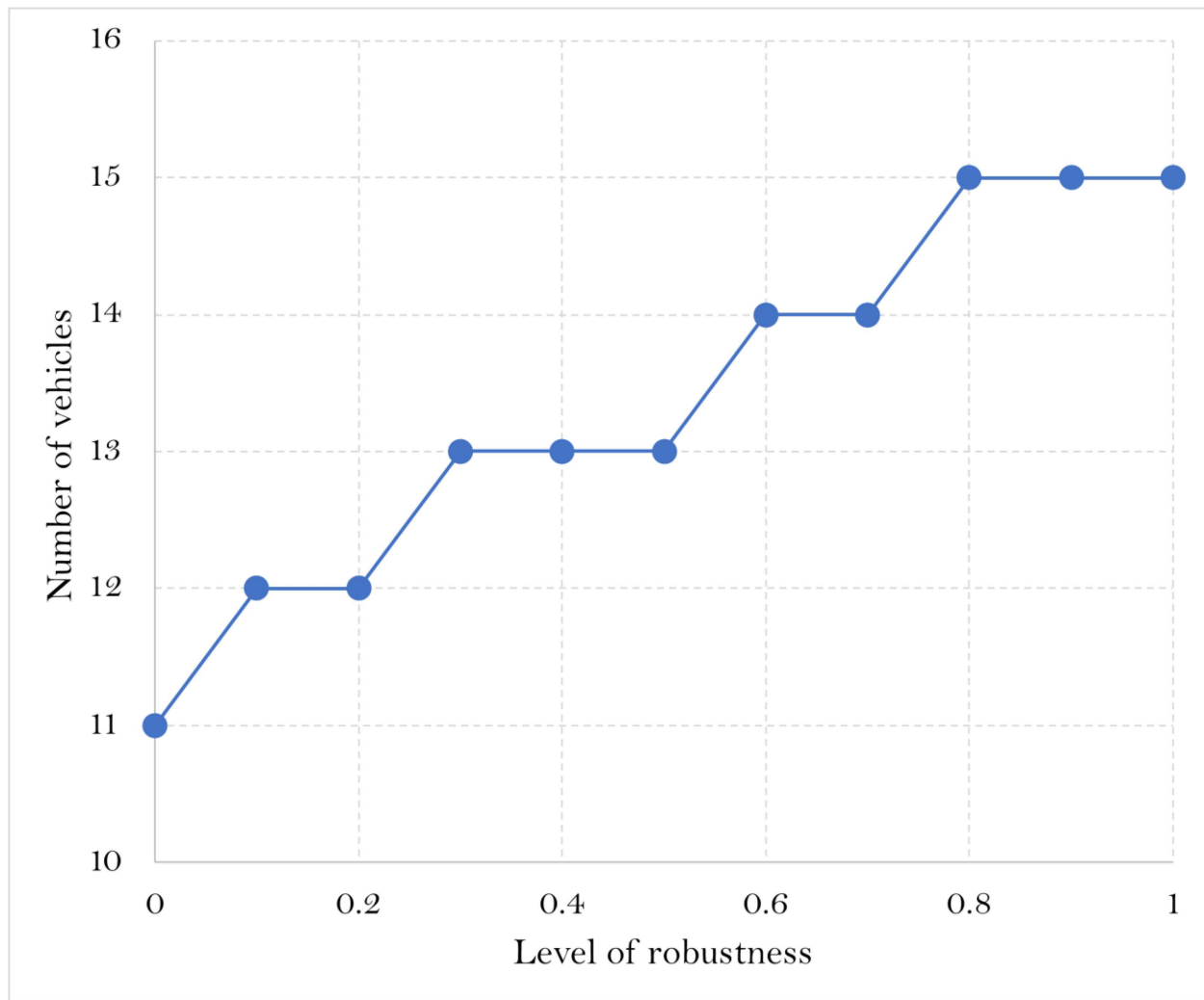


FIGURE 5. Changes in level of robustness to number of vehicles in the network.

The observable trend reveals a compelling relationship between the increase in VE and its subsequent effects on the logistics network. With a rise in VE, signifying the utilization of more expensive vehicles, the total vehicle costs understandably surge. Simultaneously, an intriguing counterbalancing effect emerges as the total transportation costs exhibit a decrease. This decrease in transportation expenses is attributed to the availability of higher-cost vehicles, which, despite inflating the overall vehicle costs, manage to curtail travel times due to their enhanced capabilities or efficiencies.

However, amidst this direct correlation, a pivotal observation surfaces: there exists an optimal value for the VE that significantly benefits the network. Beyond a certain threshold, the incremental increase in vehicle cost no longer proportionally reduces transportation expenses, leading to diminished returns. This inflection point signifies the optimal value for VE within the logistics framework, delineating a balance where the trade-off between higher vehicle costs and reduced transportation expenses reaches an equilibrium, optimizing the network's cost-effectiveness and operational efficiency.

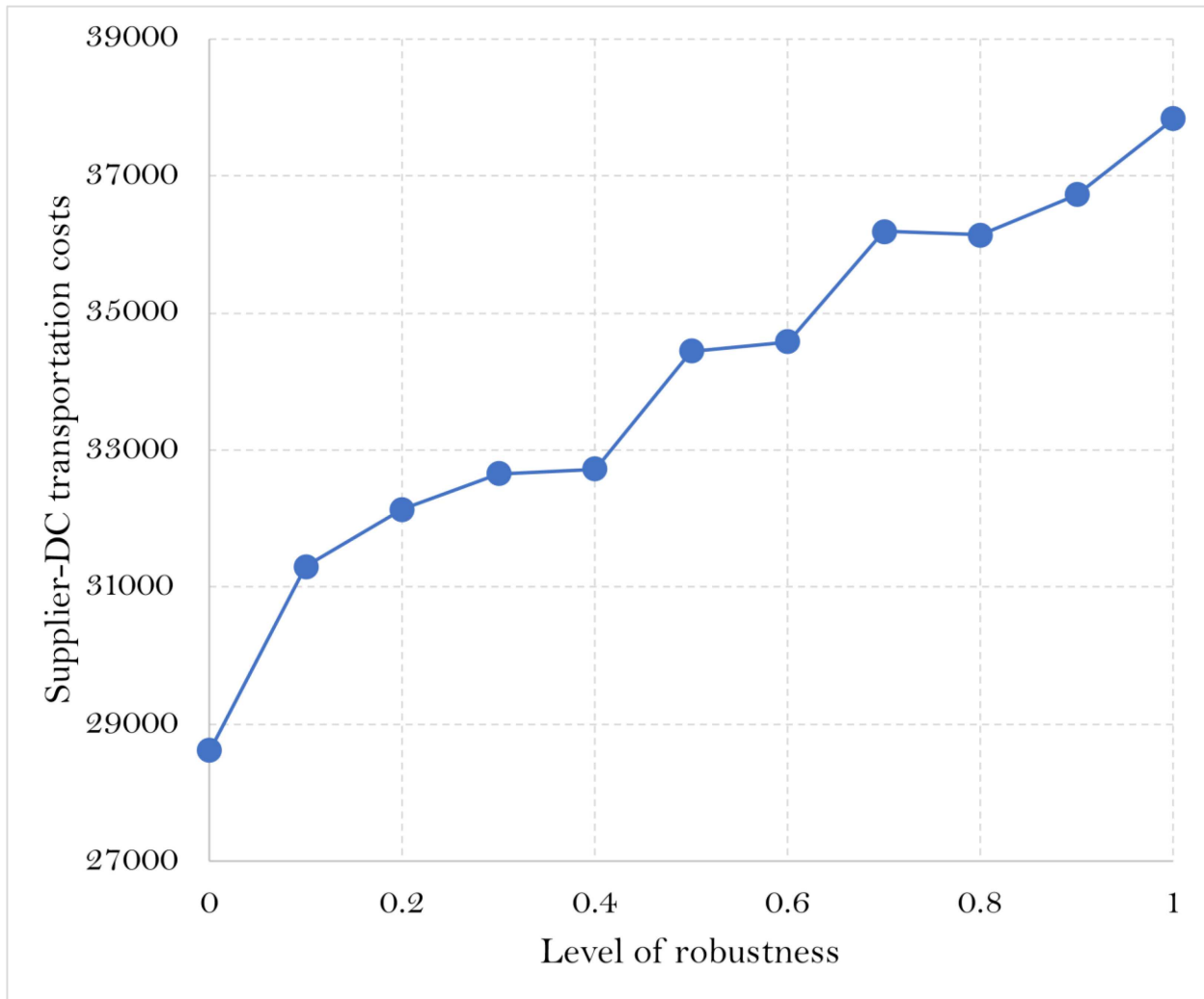


FIGURE 6. Changes in level of robustness to supplier-to-DC transportation costs.

## 8. MANAGERIAL INSIGHTS AND PRACTICAL IMPLICATIONS

The findings from this study offer several crucial insights for supply chain managers and practitioners. First, our hybrid framework demonstrates that the integration of learning-based strategies with traditional optimization methods yields superior results, particularly in large-scale operations. This is evidenced by the Q-Learning enhanced simulated annealing algorithm's superior performance in complex scenarios, suggesting that organizations should consider investing in advanced analytics capabilities rather than relying solely on conventional optimization approaches.

The case study results from Tehran's online commercial sector reveal a critical trade-off between robustness and operational costs that managers must carefully consider. The analysis shows that while higher robustness levels require increased investment in resources (as seen in the total cost sensitivity analysis), they provide better protection against demand uncertainties through strategic distribution center utilization. This finding challenges

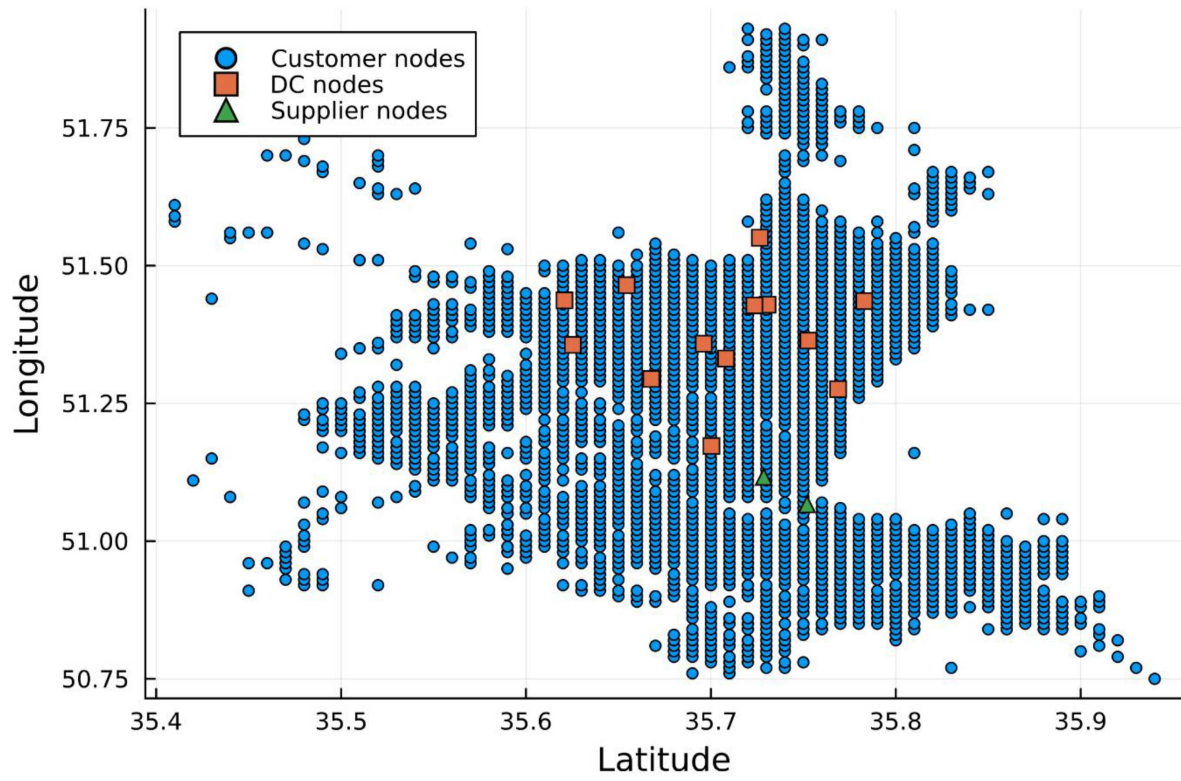


FIGURE 7. Spatial distribution of nodes in Tehran province's logistics network, featuring customer, DC, and supplier nodes for the case study.

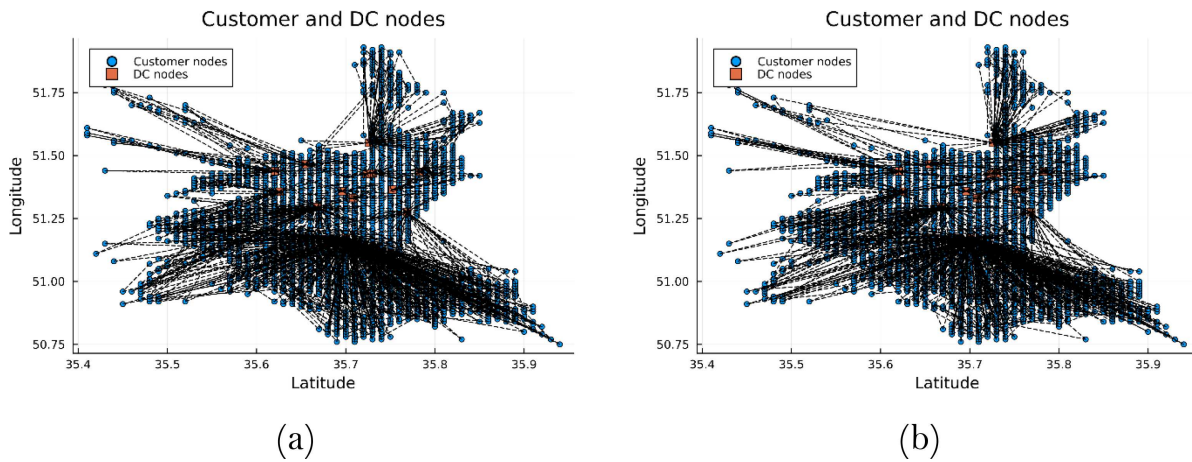


FIGURE 8. Route of vehicle in case study for: (a) zero robustness level and (b) full robustness level.

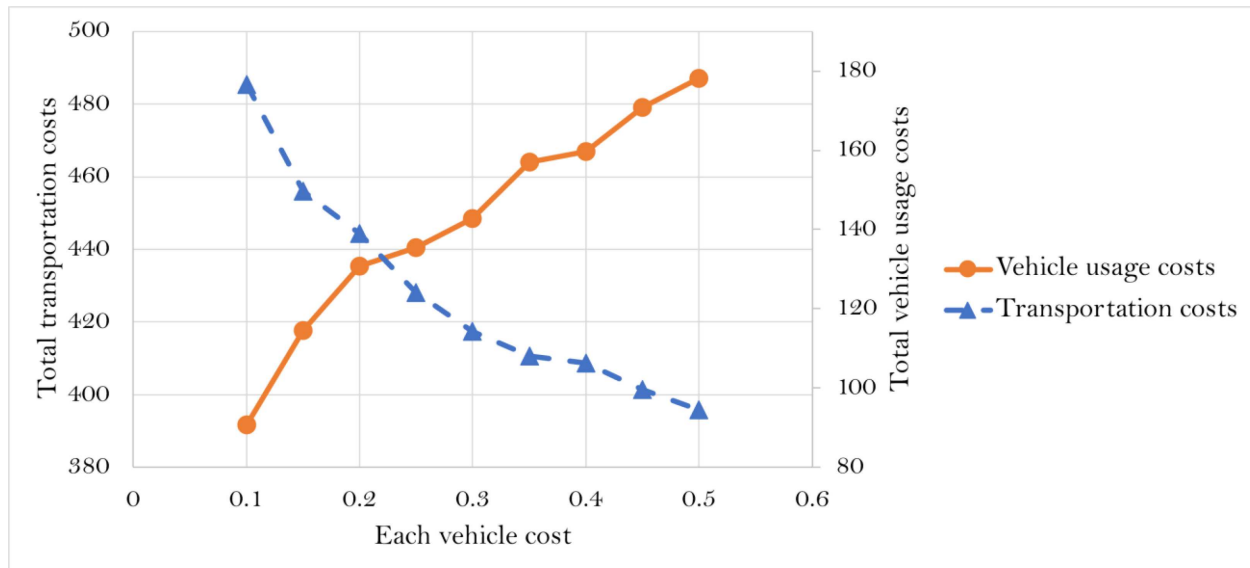


FIGURE 9. Changes in total transportation costs and total vehicle costs according to each vehicle cost.

the traditional cost-minimization approach, suggesting that managers should evaluate their risk tolerance and market volatility when determining optimal resource allocation strategies.

Perhaps most significantly, our vehicle cost analysis reveals an optimal investment point beyond which additional spending yields diminishing returns. This insight provides managers with a practical decision-making framework for fleet investment strategies, demonstrating that the most expensive solutions are not necessarily the most efficient. Organizations can use this finding to optimize their vehicle fleet composition by identifying the sweet spot between vehicle capabilities and operational costs, potentially leading to significant cost savings while maintaining service quality.

## 9. CONCLUSIONS

In this study, we have developed a comprehensive and innovative framework for optimizing supply chain operations involving fixed suppliers, distribution centers, and customers. This research addresses the multifaceted challenges of resource allocation, vehicle routing, and customer satisfaction within specified time windows, proposing a robust optimization approach integrated with SA and reinforced by learning-based strategies such as Q-Learning. This hybrid methodology enhances both solution quality and computational efficiency, enabling decision-makers to address the complexities of modern logistics networks effectively.

- **Scientific Value:** the integration of learning-based techniques with robust optimization. Our findings reveal that adaptive methods, particularly Q-Learning, significantly improve the performance of the SA algorithm in larger-scale instances, demonstrating their scalability and efficacy in solving high-dimensional supply chain problems.
- **Advancement:** by addressing the stochastic nature of disruptions, our framework advances the state of the art in supply chain optimization, providing a reliable and adaptable tool for real-world applications.
- **Computational Experiments:** through extensive performance evaluations and sensitivity analyses, we demonstrated the robustness and versatility of the proposed framework. These results offer critical managerial insights, such as the trade-offs between cost parameters and robustness levels, as well as the interplay between vehicle costs, transportation expenses, and network efficiency.

- **Case Study:** the applicability of our framework was validated through a real-world case study involving a prominent e-commerce network in Tehran, Iran. This case study highlighted the framework's ability to handle large-scale, realistic logistics challenges, including dynamic customer demands and spatially distributed nodes.
- **Managerial Insights:** by integrating robust optimization strategies, the proposed method not only ensures operational efficiency but also enhances resilience against uncertainties, making it a valuable tool for supply chain management.
- **Limitations and Future Work:** while the research contributes significantly to the field, we acknowledge limitations, such as the focus on single-objective optimization. Future work could extend this to multi-objective optimization, incorporating environmental and social sustainability, as well as real-time dynamic demand patterns and emerging technologies like blockchain for traceability and artificial intelligence for predictive analytics.

#### DATA AVAILABILITY STATEMENT

No new data/codes were created or analyzed in this study.

#### REFERENCES

- [1] A.A. Javid and N. Azad, Incorporating location, routing and inventory decisions in supply chain network design. *Transp. Res. Part E Logistics Transp. Rev.* **46** (2010) 582–597.
- [2] J.-H. Lee, I.-K. Moon and J.-H. Park, Multi-level supply chain network design with routing. *Int. J. Prod. Res.* **48** (2010) 3957–3976.
- [3] V. Schmid, K.F. Doerner and G. Laporte, Rich routing problems arising in supply chain management. *Eur. J. Oper. Res.* **224** (2013) 435–448.
- [4] M. Awad, M. Ndiaye and A. Osman, Vehicle routing in cold food supply chain logistics: a literature review. *Int. J. Logistics Manage.* **32** (2021) 592–617.
- [5] M. Musavi and A. Bozorgi-Amiri, A multi-objective sustainable hub location-scheduling problem for perishable food supply chain. *Comput. Ind. Eng.* **113** (2017) 766–778.
- [6] J.X. Cao, Z. Zhang and Y. Zhou, A location-routing problem for biomass supply chains. *Comput. Ind. Eng.* **152** (2021) 107017.
- [7] M. Tavana, H. Tohidi, M. Alimohammadi and R. Lesansalmasi, A location-inventory-routing model for green supply chains with low-carbon emissions under uncertainty. *Environ. Sci. Pollut. Res.* **28** (2021) 50636–50648.
- [8] K. Govindan, A. Jafarian, R. Khodaverdi and K. Devika, Two-echelon multiple-vehicle location-routing problem with time windows for optimization of sustainable supply chain network of perishable food. *Int. J. Prod. Econ.* **152** (2014) 9–28.
- [9] G. Iassinovskaia, S. Limbourg and F. Riane, The inventory-routing problem of returnable transport items with time windows and simultaneous pickup and delivery in closed-loop supply chains. *Int. J. Prod. Econ.* **183** (2017) 570–582.
- [10] M. Yavari, H. Enjavi and M. Geraeli, Demand management to cope with routes disruptions in location-inventory-routing problem for perishable products. *Res. Transp. Bus. Manage.* **37** (2020) 100552.
- [11] M.M. Nasiri, H. Mousavi and S. Nosrati-Abarghooee, A green location-inventory-routing optimization model with simultaneous pickup and delivery under disruption risks. *Decis. Anal. J.* **6** (2023) 100161.
- [12] F. Rayat, M. Musavi and A. Bozorgi-Amiri, Bi-objective reliable location-inventory-routing problem with partial backordering under disruption risks: a modified AMOSA approach. *Appl. Soft Comput.* **59** (2017) 622–643.
- [13] J.M. Mulvey, R.J. Vanderbei and S.A. Zenios, Robust optimization of large-scale systems. *Oper. Res.* **43** (1995) 264–281.
- [14] A. Ben-Tal, L. El Ghaoui and A. Nemirovski, Robust Optimization. Vol. 28. Princeton University Press (2009).
- [15] D. Bertsimas and M. Sim, The price of robustness. *Oper. Res.* **52** (2004) 35–53.
- [16] M.S. Pishvaei, M. Rabbani and S.A. Torabi, A robust optimization approach to closed-loop supply chain network design under uncertainty. *Appl. Math. Modell.* **35** (2011) 637–649.
- [17] A. Rahbari, M.M. Nasiri, F. Werner, M. Musavi and F. Jolai, The vehicle routing and scheduling problem with cross-docking for perishable products under uncertainty: two robust bi-objective models. *Appl. Math. Modell.* **70** (2019) 605–625.

- [18] A. Ala, V. Simic, N. Bacanin and E.B. Tirkolaei, Blood supply chain network design with lateral freight: a robust possibilistic optimization model. *Eng. App. Artif. Intell.* **133** (2024) 108053.
- [19] A. Goli, A. Ala and S. Mirjalili, A robust possibilistic programming framework for designing an organ transplant supply chain under uncertainty. *Ann. Oper. Res.* **328** (2023) 493–530.
- [20] F. Habibzadeh Boukani, B. Farhang Moghaddam and M.S. Pishvaei, Robust optimization approach to capacitated single and multiple allocation hub location problems. *Comput. Appl. Math.* **35** (2016) 45–60.
- [21] M. Varas, S. Maturana, R. Pascual, I. Vargas and J. Vera, Scheduling production for a sawmill: a robust optimization approach. *Int. J. Prod. Econ.* **150** (2014) 37–51.
- [22] R. Lotfi, Z. Sheikhi, M. Amra, M. AliBakhshi and G.-W. Weber, Robust optimization of risk-aware, resilient and sustainable closed-loop supply chain network design with Lagrange relaxation and fix-and-optimize. *Int. J. Logistics Res. App.* **27** (2024) 705–745.
- [23] R. Lotfi, F. Shoushtari, S.S. Ali, S.M.R. Davoodi, M. Afshar and M.M. Sharifi Nevisi, A viable and bi-level supply chain network design by applying risk, robustness and considering environmental requirements. *Cent. Eur. J. Oper. Res.* (2024) 1–29.
- [24] D. Bertsimas, V. Gupta and N. Kallus, Data-driven robust optimization. *Math. Program.* **167** (2018) 235–292.
- [25] C. Shang, X. Huang and F. You, Data-driven robust optimization based on kernel learning. *Comput. Chem. Eng.* **106** (2017) 464–479.
- [26] M. Musavi and A. Bozorgi-Amiri, Data-driven robust optimization for hub location-routing problem under uncertain environment. *J. Ind. Syst. Eng.* **15** (2024) 109–129.
- [27] S. Mohseni, M.S. Pishvaei and R. Dashti, Privacy-preserving energy trading management in networked microgrids via data-driven robust optimization assisted by machine learning. *Sustain. Energy Grids Networks* **34** (2023) 101011.
- [28] Y. Li, Y. Sun, J. Liu, C. Liu and F. Zhang, A data driven robust optimization model for scheduling near-zero carbon emission power plant considering the wind power output uncertainties and electricity-carbon market. *Energy* **279** (2023) 128053.
- [29] R. Lotfi, R. Hazrati, S. Aghakhani, M. Afshar, M. Amra and S.S. Ali, A data-driven robust optimization in viable supply chain network design by considering Open Innovation and Blockchain Technology. *J. Clean. Prod.* **436** (2024) 140369.
- [30] M. Karimi-Mamaghan, M. Mohammadi, A. Pirayesh, A.M. Karimi-Mamaghan and H. Irani, Hub-and-spoke network design under congestion: a learning based metaheuristic. *Transp. Res. Part E: Logistics Transp. Rev.* **142** (2020) 102069.
- [31] C.-Y. Cheng, P. Pourhejazy, K.-C. Ying, S.-F. Li and C.-W. Chang, Learning-based metaheuristic for scheduling unrelated parallel machines with uncertain setup times. *IEEE Access* **8** (2020) 74065–74082.
- [32] A. Seyyedabbasi, R. Aliyev, F. Kiani, M.U. Gulle, H. Basyildiz and M.A. Shah, Hybrid algorithms based on combining reinforcement learning and metaheuristic methods to solve global optimization problems. *Knowl.-Based Syst.* **223** (2021) 107044.
- [33] W. Qin, Z. Zhuang, Z. Huang and H. Huang, A novel reinforcement learning-based hyper-heuristic for heterogeneous vehicle routing problem. *Comput. Ind. Eng.* **156** (2021) 107252.
- [34] V.A. de Santiago Jr., E. Özcan and V.R. de Carvalho, Hyper-heuristics based on reinforcement learning, balanced heuristic selection and group decision acceptance. *Appl. Soft Comput.* **97** (2020) 106760.
- [35] İ. Gölcük and F.B. Ozsoydan, Q-learning and hyper-heuristic based algorithm recommendation for changing environments. *Eng. App. Artif. Intell.* **102** (2021) 104284.
- [36] B. Xi and D. Lei, Q-learning-based teaching-learning optimization for distributed two-stage hybrid flow shop scheduling with fuzzy processing time. *Complex Syst. Model. Simul.* **2** (2022) 113–129.
- [37] X. Ni, W. Hu, Q. Fan, Y. Cui and C. Qi, A Q-learning based multi-strategy integrated artificial bee colony algorithm with application in unmanned vehicle path planning. *Expert Syst. App.* **236** (2024) 121303.
- [38] Z. Zhang, Z. Wu, H. Zhang and J. Wang, Meta-learning-based deep reinforcement learning for multiobjective optimization problems. *IEEE Trans. Neural Netw. Learn. Syst.* **34** (2022) 7978–7991.
- [39] J. Kallestad, R. Hasibi, A. Hemmati and K. Sörensen, A general deep reinforcement learning hyperheuristic framework for solving combinatorial optimization problems. *Eur. J. Oper. Res.* **309** (2023) 446–468.
- [40] Z. Zhang, Z. Shao, W. Shao, J. Chen and D. Pi, MRLM: a meta-reinforcement learning-based metaheuristic for hybrid flow-shop scheduling problem with learning and forgetting effects. *Swarm Evol. Comput.* **85** (2024) 101479.
- [41] S. Kirkpatrick, C.D. Gelatt Jr. and M.P. Vecchi, Optimization by simulated annealing. *Science* **220** (1983) 671–680.
- [42] J.F. Cordeau, M. Gendreau and G. Laporte, A tabu search heuristic for periodic and multi-depot vehicle routing problems. *Networks: An Int. J.* **30** (1997) 105–119.