

Data-driven robust service network design problem: Balancing robustness and conservatism

Lingpu Zhang ^a, Min Huang ^{a,*}, Yuxin Zhang ^a, Ziteng Wang ^b, Miao Li ^a,
Xingwei Wang ^c

^a College of Information Science and Engineering, Northeastern University Shenyang, Liaoning, 110819, China

^b Department of Industrial and Systems Engineering, Northern Illinois University, DeKalb, IL 60115, USA

^c College of Computer Science and Engineering, Northeastern University Shenyang, Liaoning, 110169, China

ARTICLE INFO

Keywords:

Service network design
Uncertainty
Data-driven robust optimization
Column-and-constraint generation
Support vector clustering

ABSTRACT

The rapid growth of e-commerce platforms has given rise to convenient online shopping methods, but platform operators usually don't have sufficient capacity to deliver these online orders. Hence, based on the estimated demand for each type of commodity, logistics service providers generally require providing freight services to fulfill customer demands while maintaining their own profitability. However, each origin-destination demand is impacted by various factors and is inherently uncertain. This poses significant challenges for logistics service providers seeking to develop an economically viable and efficient freight service network. Given that most research on service network design problems rarely utilizes historical data, in this paper, we propose a data-driven robust service network design problem, aiming to answer: *is it possible to obtain more economically efficient decisions when we design a service network based on historical data?* To achieve this, a support vector clustering method is employed to adaptively capture a compact convex uncertainty set from the historical demand data, allowing explicit control over the conservatism of decisions. Based on the obtained uncertainty set, a novel data-driven robust optimization model is developed. Considering the NP-hard nature of this problem, we propose a customized column-and-constraint generation algorithm that effectively solves the model. Compared to the Benders decomposition algorithm, it significantly improves computational efficiency. Numerical experiments demonstrate that the proposed method outperforms traditional robust optimization approaches in mitigating demand uncertainty, exhibiting robustness while being less conservative. Meanwhile, we provide valuable managerial insights to decision-makers through parameter sensitivity analysis. Results reveal that neither overly aggressive nor excessively conservative strategies are optimal, and adjusting the clustering parameter allows decision-makers to tailor the service network to their risk tolerance for more cost-effective operations under demand uncertainty.

1. Introduction

In the past decade, the rapid advancement of e-commerce platforms has provided consumers with a highly convenient method for purchasing goods. On the one hand, this has significantly increased the freight transportation volume for logistics service providers, and this growth is beneficial for enhancing the company's overall revenue. For example, in 2024 alone, China's express delivery volume exceeded 174.5 billion parcels, with revenue surpassing 1.4 trillion CNY¹. However, on the other hand, this also allows customers to have the flexibility to modify their order quantities at any time or even return items, which complicates companies' ability to accurately predict customer demand due to

the inherent volatility influenced by various factors. In this case, handling such demand volatility has become one of the major challenges facing leading logistics service providers (Bai et al., 2014; Mirzaei et al., 2025). The size of the utilizing fleet that is too large relative to demand will result in wasted assets; conversely, the size of the utilizing fleet that is too small will fail to meet demand, leading to reduced customer satisfaction or reliance on expense-intensive temporary scheduling. Therefore, it is crucial to account for this uncertainty during the planning phase rather than attempting to remediate it during operations.

Addressing this challenge is particularly important in the context of medium/long-term planning in e-commerce logistics, which is inherently a less-than-truckload (LTL) operation or small parcel delivery

* Corresponding author.

E-mail addresses: 18840799141@163.com (L. Zhang), mhuang@mail.neu.edu.cn (M. Huang), neuzyx@163.com (Y. Zhang), zwang3@niu.edu (Z. Wang), 3308074761@qq.com (M. Li), wangxw@mail.neu.edu.cn (X. Wang).

¹ https://www.gov.cn/lianbo/bumen/202504/content_7018267.htm

<https://doi.org/10.1016/j.eswa.2026.131970>

Received 19 September 2025; Received in revised form 4 February 2026; Accepted 3 March 2026

Available online 11 March 2026

0957-4174/© 2026 Published by Elsevier Ltd.

due to the small size of individual order parcels (Boland et al., 2017). Under this transportation model, it is virtually impossible for logistics service providers to develop individual, direct transportation solutions for each customer. Considering profitability and asset utilization, it is necessary to coordinate routes in both time and space to consolidate commodities that could potentially be transported by the same truck. This issue has long been addressed by operations research optimization methods: *service network design* (SND) (see Andersen et al., 2011; Hewitt, 2022; Pedersen et al., 2009). As a tactical decision, the logistics service providers aim to develop a scheduling table that specifies the routes for services (e.g., trucks and trains) to be repeatedly executed during each *scheduling length* (a week) within the *planning cycle* (half a year or a quarter). Decision-makers hope to deliver the goods at the lowest possible cost while maintaining a high service level for customers (see Liu et al., 2023; Teypaz et al., 2010). The medium/long-term and repetitive nature of these schedules makes advance consideration of demand uncertainty essential for two key reasons: First, the services provided involve numerous stakeholders, such as platform operators, shippers, and customers, thereby making modifying existing scheduling plans often difficult; second, if the estimation of uncertainty is inaccurate, cumulative losses continuously may grow as operating times extend. According to recent reports, LTL freight constituted over 30% of China's road transportation freight², underscoring the critical importance for companies to effectively address these challenges when designing service networks.

In fact, research has been done on SND under demand uncertainty. Specifically, the SND problem has been examined through the lenses of stochastic optimization and robust optimization (RO) by Lium et al. (2009) and Wang and Qi (2020), respectively. They use scenario sets representing demand or the demand mean and the maximum deviation to characterize uncertainty, respectively. Their findings indicate that the stochastic solutions enhance transportation flexibility via path sharing, thereby reducing the company's operational costs associated with handling uncertain demand. However, for the former, stochastic optimization methods require obtaining the true probability distribution of customer demands and can't avoid the negative impact of worst-case scenarios; for the latter, robust optimization models that rely solely on the mean demand and maximum deviation tend to yield overly conservative decisions. Although recent studies (Scherr et al., 2022) and (Mirzaei et al., 2025) further emphasize the importance of considering demand uncertainty, there are still two critical questions that remain without answers: (i) **To what extent can decisions that are almost not based on historical data be effectively implemented in real-world contexts, and how effective can such decisions be deemed?** (ii) **Is it feasible to derive robust yet less conservative decisions by leveraging historical demand data?** To address these questions, this paper investigates the SND problem through the framework of data-driven robust optimization (DDRO).

In the context of traditional RO studies, it is assumed that the true data distribution is not accessible. Hence, a number of researchers have devised well-natured uncertainty sets for characterizing parameter uncertainty (Bertsimas & Sim, 2004). Then, RO makes decisions under worst-case scenarios within the predefined uncertainty set, which directly influence decision performance. Utilizing an uncertainty set that is excessively large in relation to the true data can result in considerable resource wastage, whereas employing a set that is overly restrictive may undermine the robustness of the decisions made. Therefore, it is crucial to explore methodologies for adaptively constructing uncertainty sets based on observable historical data, thereby ensuring that decisions are both robust and cost-effective. The approach known as data-driven robust optimization (DDRO) offers a viable solution to this challenge (Bertsimas et al., 2018). Alexandre et al. (2020) carves out the uncertainty set by forming a convex hull of the worst-case scenario based on historical data, dynamically adjusting the scope of the set. However,

when e-commerce platforms conduct large-scale promotions, such as the Tmall 618 event, demand can experience irregular fluctuations. If such demand fluctuations are all taken into account, the decisions tend to become overly conservative. To avoid this solution, it is necessary to allow for certain outliers, and support vector clustering (SVC) provides a feasible approach for this adjustment. As an extended support vector machine (Cortes & Vapnik, 1995), SVC is an unsupervised machine learning method aimed at obtaining the smallest sphere that encloses the data and controlling the proportion of outliers, which has been applied in numerous studies (such as Ruchika & Pattem, 2023 and Sun et al., 2025). According to the research by Shang et al. (2017), SVC can effectively capture the uncertainty set by solving a quadratic programming problem, significantly reducing the conservatism of RO. The resulting uncertainty set is a compact convex hull, which is tractable and can be directly incorporated into the optimization model.

Hence, this study introduces the DDRO SND problem utilizing the SVC method. Initially, an uncertainty set is developed by analyzing the historical data through the SVC method. Following this, a two-stage robust optimization model is formulated. Given the NP-hard characteristics of the problem and the efficiency of the column-and-constraint generation (CCG) algorithm (Zeng & Zhao, 2013) in solving such problems, a customized CCG algorithm is designed to address the proposed two-stage RO model.

In summary, this paper contributes in threefold: 1. Develops a novel two-stage robust optimization model for the service network design problem by capturing demand uncertainty through the support vector clustering method, thereby providing decision support for logistics service providers. To address the challenges in solving the model, a customized column-and-constraint generation algorithm is designed. Computational results compared to the classic Benders decomposition algorithm demonstrate that the CCG method can rapidly generate high-quality solutions. 2. Compared with decision-making based on budget uncertainty sets, the proposed method shows robust and cost-effective decision outcomes, highlighting its ability to maintain cost-effectiveness and decision robustness. 3. Through extensive numerical experiments, valuable managerial insights are provided for logistics service providers, mainly including: both overly conservative and completely non-conservative risk-taking attitudes lead to profit declines. When accurate demand distribution is unavailable, a compromise risk tolerance attitude should be maintained. Additionally, the methodology proposed in this paper allows for parameter adjustments, effectively assisting enterprises in reducing costs while simultaneously providing probabilistic assurance that expected demand will be met.

The remainder of the paper is structured as follows. Section 2 reviews the relevant literature. In Section 3, we detail the service network design problem and introduce the process of constructing uncertainty sets through support vector clustering; based on this, a two-stage robust optimization model is developed. Section 4 designs a CCG algorithm to solve the proposed model. In Section 5, numerical experiments are conducted to demonstrate the effectiveness of the developed model and algorithms. In Section 6, we provide managerial insights for logistics service providers. Section 7 concludes the paper and discusses several future research directions.

2. Literature review

The service network design problem has been explored in a large body of literature (see Andersen et al., 2009a; Ghamlouche et al., 2003; Hewitt, 2022; Kris et al., 2013) as a tactical planning problem, typically formulated as a fixed-cost multi-commodity network design problem with limited capacities (Ghamlouche et al., 2003, 2004), primarily applied to address the challenge of long-distance transportation for various raw materials or finished products. Existing research has explored this classic combinatorial optimization problem from multiple perspectives. Recently, Crainic and Rei (2025) provided a comprehensive review of related studies. This section briefly summarizes the existing literature

² <https://www.chinabgao.com/info/1254225.html>

in two aspects: First, the SND in the deterministic environment; second, the SND considering uncertainty.

2.1. SND in deterministic environment

Initially, the main challenge faced by freight service providers has been fulfilling demand that fluctuates cyclically. Freight demand has typically been higher on weekdays and lower on weekends. This is because transportation demand is tied to production supply chains during this time, and workday product activities are more frequent than on weekends. Logistics service providers have been able to manage this regular demand with relative ease. Consequently, most research has been conducted under deterministic settings. Crainic and Rousseau (1986) concentrated on the transportation of commodities between origin-destination pairs and proposed the first formula and solution method for the service network design problem. Given the complexity of the problem and computational constraints, researchers have developed various metaheuristic algorithms. For example, to effectively address large-scale SND instances considering fixed costs and capacity constraints, Ghamlouche et al. (2003) and Ghamlouche et al. (2004) proposed a heuristic solution method based on cycle-based neighborhood structures and a heuristic solution method based on path relinking and cycle-based neighborhood structures, respectively. Later, Yaghini et al. (2015) developed a neighborhood search algorithm for cutting-plane, further improving the efficiency of solving the SND problem. These researchers have explored variants of the SND problem from different perspectives and proposed multiple solution algorithms based on problem characteristics, yet they have neglected asset management and efficient utilization.

As logistics companies have become gradually concerned about optimizing the efficiency of asset utilization, the integration of asset management within the service network design problem has garnered widespread attention. These assets, such as human resources, trucks, terminals/containers, etc., determine various dimensions of service quality, operational efficiency, and cost-effectiveness associated with the carrier's offerings. Managing assets is essential to the overall performance of the service network (see Crainic & Rei, 2025). Pedersen et al. (2009) were the first to incorporate design-balanced (asset-balanced) constraints into service network design problems. These constraints require that the quantity of assets entering and exiting each node must remain equal, resulting in a circular structure for assets throughout the entire planning cycle. By introducing such constraints, logistics service providers can comprehensively evaluate asset utilization across multiple cycles. Based on this, Andersen et al. (2009b) further explored multi-fleet collaboration in SND and discussed the potential of cooperative services in enhancing resource utilization. Andersen et al. (2009a) also compared the computational efficiency of arc-based, path-based, and cycle-based approaches for modeling service network design problems with design-balanced constraints. Experimental results on a set of randomly generated small-scale instances indicate that the cycle-based solution method provides significantly stronger bounds than the other two approaches, potentially leading to substantial reductions in solution time. Meanwhile, Andersen et al. (2011) proposed a branch-and-bound algorithm capable of efficiently handling SND instances with design-balanced constraints. Recent studies, Liu et al. (2023) and Scherr et al. (2018), have addressed service network design problems under design-balanced constraints: the former focuses on a service network design problem with consistent multiple trips, while the latter addresses a service network design problem for a mixed autonomous fleet.

In summary, existing studies under deterministic settings have explored service network design problems from various perspectives, providing highly valuable formulas, solution methods, and managerial insights. However, modeling using the mean or estimated values of parameters fails to reflect the nature of real-world problems (Bai et al., 2014). In practical settings, implementing the outputs of these models may lead to severe negative consequences.

2.2. SND considering uncertainty

The rapid growth of e-commerce has led to dramatic fluctuations in demand, which makes freight companies and researchers more concerned with the SND problem under demand uncertainty. Lium et al. (2009) and Bai et al. (2014) employed scenario sets to characterize uncertainty and developed a two-stage stochastic optimization (SO) model. The objective is to minimize the average construction cost over all scenarios. They found that compared to deterministic solutions, the stochastic solution enhances scheduling flexibility and robustness by leveraging consolidation and multi-commodity path sharing. Recently, Mirzaei et al. (2025) introduced a hybrid two-stage stochastic opportunity-constrained programming model that tackled the dual uncertainties of demand fluctuations and service capacity disruptions. An accelerated Benders decomposition algorithm has been developed to solve the proposed model. Considering that stochastic optimization methods require knowledge of the true probability distribution of demand and cannot avoid the significant negative impacts caused by worst-case scenarios, Wang and Qi (2020) developed the robust SND formulation. They developed a two-stage RO model and designed a CCG to solve the proposed model. The results emphasize the advantage of robust optimization methods in hedging against uncertain risks. Based on Wang and Qi (2020), to overcome the conservatism of the robust model, Xiang et al. (2022) proposed an almost robust optimization (see Baron et al., 2019) for addressing the SND problem. By setting a risk threshold, it allows some demands to not be fully met.

However, the methods proposed by the above research struggle to adaptively generate solutions based on historical data to hedge against corresponding risks. Failure to respond promptly to changes in demand and adjust decisions accordingly can lead to various negative impacts, such as excessive construction costs or expensive outsourcing costs. This, combined with logistics service providers' desire to minimize the negative impact of demand fluctuations in the highly competitive freight market, promotes us to use DDRO's approach (see Bertsimas et al., 2018) to address the SND problem. We aim to adaptively adjust the uncertainty set based on historical data to make more cost-effective decisions. As demonstrated by the studies conducted by Wu and Hifi (2021), Mohsen et al. (2024) and Long et al. (2025), data-driven robust optimization techniques are capable of integrating real-time data with area knowledge to dynamically respond to environmental variations, a functionality that traditional robust optimization approaches lack. Furthermore, when e-commerce platforms hold promotional events, demand may surge dramatically. To eliminate these outliers that could lead to overly conservative decision-making, an SVC approach is adopted. This method, on the one hand, controls the proportion of outliers by adjusting parameters, thereby influencing the robustness of the solution. On the other hand, the captured uncertainty set is in the form of a compact convex hull, making it easy to integrate into optimization models for solving (Shang et al., 2017).

A large number of scholars have developed effective algorithms for solving the SND problem (refer to Andersen et al., 2011; Boland et al., 2017; Marshall et al., 2020; Pedersen et al., 2009). In this paper, we focus on the algorithm for solving the two-stage RO model. Employing a two-stage robust method introduces more computational challenges, as Ben-Tal and Goryashko (2004) demonstrated that solving a two-stage robust model is typically NP-hard. Hence, some studies tend to use heuristic algorithms to solve the SND. However, heuristic algorithms often struggle to reach the optimal solution. To efficiently obtain the optimal solution for the two-stage robust model, the CCG algorithm (An & Zeng, 2015; Zeng & Zhao, 2013) has been proposed and proven more effective than the Benders decomposition algorithm in solving the robust SND problem (see Wang & Qi, 2019; Zhang et al., 2024). This motivates us to develop a customized CCG algorithm to solve the proposed model.

To highlight the contributions of this paper, we summarize the most relevant works in Table 1. Research related to robust service network design, which relies only on the average demand and maximum

Table 1
Most relevant works in SND.

Paper	Model		Uncertainty				Algorithm		Asset Management	
	arc-based	path-based	Determ	SO	RO	DDRO	exact	heuristic	Yes	No
Ghamlouche et al. (2003)	✓		✓					✓		✓
Ghamlouche et al. (2004)	✓		✓					✓		✓
Pedersen et al. (2009)	✓		✓					✓		
Lium et al. (2009)	✓			✓					✓	
Andersen et al. (2011)		✓	✓				✓		✓	
Bai et al. (2014)	✓			✓					✓	
Boland et al. (2017)	✓		✓				✓			✓
Hewitt et al. (2019)		✓		✓				✓	✓	
Wang and Qi (2020)	✓				✓		✓		✓	
Xiang et al. (2022)	✓				✓		✓		✓	
Liu et al. (2023)	✓		✓					✓	✓	
Truden and Hewitt (2024)	✓		✓						✓	
Zhang et al. (2024)	✓				✓		✓		✓	
Mirzaei et al. (2025)	✓			✓			✓		✓	
This paper	✓					✓	✓		✓	

deviation, does not sufficiently leverage the available data. This often results in overly conservative solutions. There has been no research to develop robust service network design solutions using data-driven frameworks. To address this gap and simultaneously explore what data-driven approaches can offer logistics service providers, in this paper, we first study a DDRO SND problem, develop a two-stage RO model using SVC to capture the uncertainty set, and design the customized CCG algorithm to solve the proposed model.

3. Problem description and mathematical formulation

In this section, we first present the problem description and the traditional two-stage RO model based on the budget uncertainty set. To overcome its limitations in conservatism and data adaptability, we then introduce a novel data-driven approach: the SVC method is employed to adaptively construct a compact uncertainty set from historical demand, with its conservatism controllable via parameter. Building on this set, the two-stage DDRO formulation is developed. Fig. 1 provides a comprehensive understanding of the entire workflow, highlighting how the data-driven uncertainty set integrates with the two-stage robust optimization framework. The implementation details of the CCG algorithm are presented in Section 4.

3.1. Problem description

The volume of goods ordered by customers on e-commerce platforms is usually far smaller than the capacity of delivery trucks. It is unprofitable and impractical for logistics service providers to develop a dedicated, individual transportation solution for each customer. Therefore, logistics service providers tend to consolidate commodities with the same origin-destination or even different origin-destinations onto a single truck for delivery (Crainic, 2000). For convenience, it is advisable to treat all goods with the same origin-destination pair as a single commodity k , and let $\mathcal{K}$ denote the set of commodities transported by the logistics service provider. For each $k \in \mathcal{K}$, it possesses the following attributes: o_k, d_k , indicates the origin and destination of the commodity k ; b_k, e_k , defines the earliest departure time and latest arrival time for the commodity k ; and the nominal demand of the commodity k is represented as $\bar{q}_k$. For each commodity k , its transportation demand may persist throughout a quarter or even half a year. Simultaneously, due to various factors, the volume of demand q_k remains uncertain. Hence, logistics service providers must coordinate truck routes across time and space to develop the scheduled table, which is repeatedly executed during each *scheduling length* (a week) within the *planning cycle* (half a year or a quarter), aiming at minimal cost to meet fluctuating commodity demand within specified time constraints. That is, a service network design problem needs to be addressed.

The SND problem is often formulated as a time-expanded network, denoted $\mathcal{G}(\mathcal{N}, \mathcal{E})$, where $\mathcal{N}$ is the set of nodes that can be served and $\mathcal{E}$ is the set of arcs connecting these nodes. Obviously, time information is included in the arc $\mathcal{E}$. Given that e-commerce platforms typically promise same-day or next-day delivery, we adopt the assumption (Bai et al., 2014; Lium et al., 2009) that the movement of the freight between any nodes at adjacent moments takes the same number of periods. We assume that a transport cycle is discretized into T time periods, denoted by $t = 0, 1, \dots, T-1 \in \mathcal{T}$. At each period t , there exists a copy of the nodes $\mathcal{N}$. Based on the above assumptions, the departure period t^- for freight arriving at the next node at period t is given by:

$$t^- = \begin{cases} t-1, & \text{if } t \geq 1 \\ T-1, & \text{if } t = 0 \end{cases}$$

Accordingly, it $\mathcal{E}$ consists of two types of arcs: non-holding arcs $\mathcal{E}^m$ and holding arcs $\mathcal{E}^s$. $\mathcal{E}^m$ are of the form $((i, t^-), (j, t))$, indicating that the movement of freight is between nodes i and j ; $\mathcal{E}^s$ are of the form $((i, t^-), (i, t))$, denoting that the freight is held at node i . The movement of freight between nodes over time will incur costs associated with assets such as drivers. Denote by f_{ij} the fixed cost associated with moving freight from i to j , and the unit cost of freight movement c_{ij} from i to j , $i, j \in \mathcal{N}^t$. These costs are determined in advance by the logistics service providers. The capacity of trucks is represented by u . Given that fulfilling all potential customer demands often requires substantial asset investment, logistics service providers have budgetary constraints on the maximum size of their fleets, and the maximum number of trucks is denoted by V . Additionally, considering the fluctuations in demand, some unserved demands are met through temporary, expense-intensive scheduling, referred to as outsourcing, the unit outsourcing cost by λ .

The definition of decision variables is given below:

- x_{ij}^t : indicates the service frequency of trucks between nodes i and j during period t . $x_{ij}^t \in \{0, 1, 2, \dots\}$.
- y_{ij}^{kt} : indicates the transportation volume of commodity k between nodes i and j during the time period t . $y_{ij}^{kt} \geq 0$.
- $Q(k)$: indicates the amount of outsourcing required for a commodity k in the optimal commodity flow.

Based on the above description, an instance for SND is shown in Fig. 2. Nodes of different colors represent commodities with distinct origin-destination pairs, with the destination marked on Fig. 2. The instance involves $|\mathcal{K}| = 2$, one is delivered from node 2 to node 1 with departure and terminal times of $(0, 3)$, while the other is delivered from node 0 to node 1 with corresponding times of $(2, 4)$. The trucks continuously follow a circular route to complete the transport task.

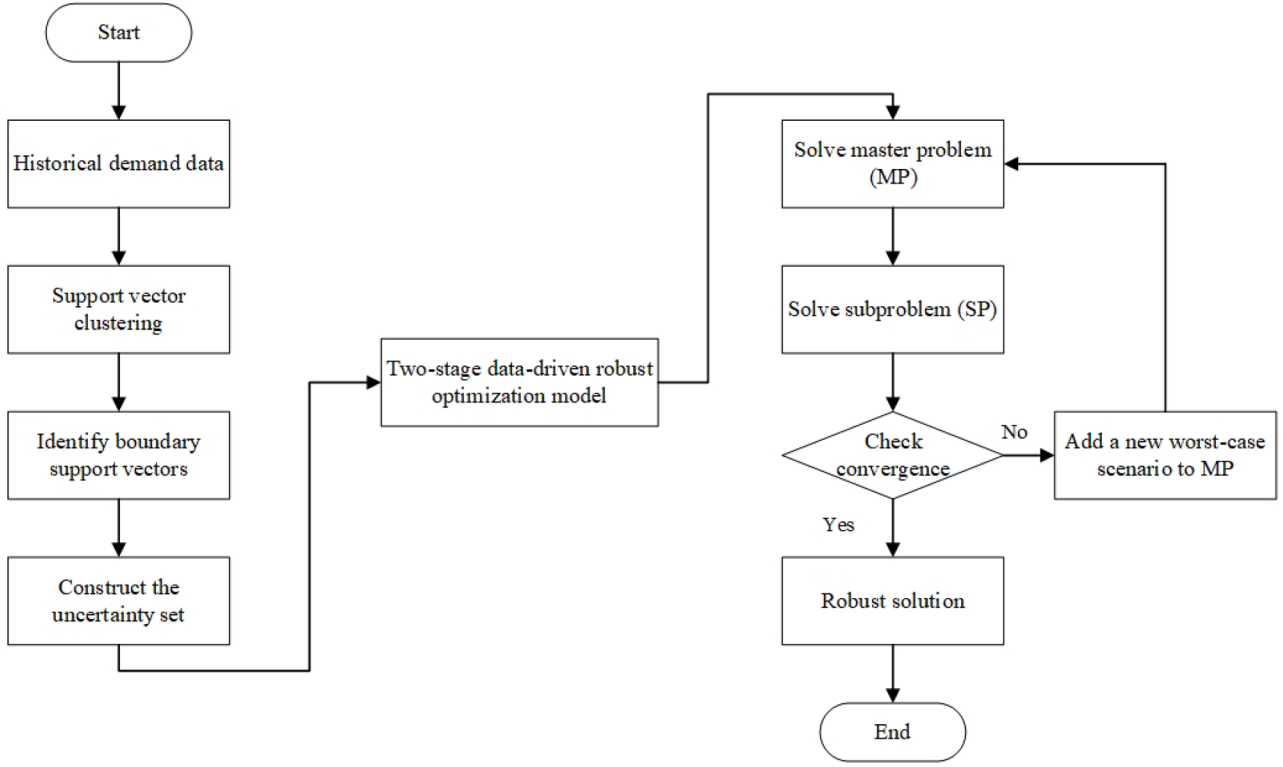


Fig. 1. Flowchart of a data-driven robust optimization method for solving service network design problems.

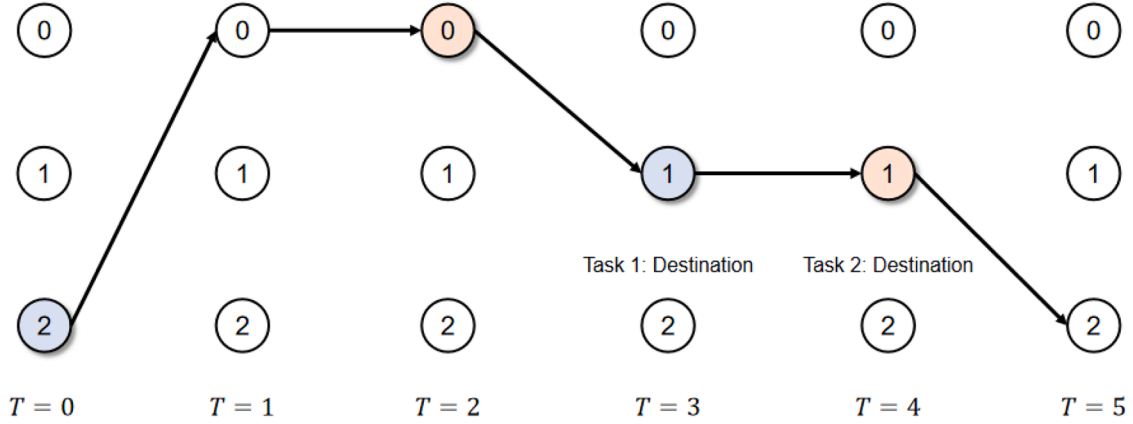


Fig. 2. A SND instance.

3.2. RO formulation

To hedge the risk associated with uncertain demand, [Lium et al. \(2009\)](#) and [Bai et al. \(2014\)](#) describe uncertainty in terms of scenarios and study SND under demand uncertainty. However, the stochastic programming approach could not avoid the negative impact of the worst-case scenario and needed to know prior probabilities of scenarios. To overcome these limitations, considering computational complexity and control over the conservativeness of the solution, [Wang and Qi \(2020\)](#) proposed a two-stage RO model based on a budget uncertainty set to tackle uncertainty-related risks. This is the baseline model studied in this paper as follows:

(RO) Stage 1:

$$\min_x \sum_{i \in \mathcal{T}} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} f_{ij} \cdot x_{ij}^t + \max_p \Omega(x, p) \quad (1a)$$

$$\text{s.t.} \quad \sum_{j \in \mathcal{N}} x_{ji}^t = \sum_{j \in \mathcal{N}} x_{ij}^t, \quad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}, \quad (1b)$$

$$\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} x_{ij}^t \leq V, \quad \forall t \in \mathcal{T}, \quad (1c)$$

$$x_{ij}^t \in \{0, 1, 2, \dots\}, \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \quad (1d)$$

where $\Omega(x, p)$ is the optimal value of the second-stage problem. Given the first-stage decision x , the second-stage problem can be defined as follows:

Stage 2:

$$\Omega(x, p) := \min_{y, Q} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} c_{ij} \cdot y_{ij}^{kt} + \sum_{k \in \mathcal{K}} \lambda \cdot Q(k) \quad (1e)$$

$$\text{s.t.} \quad \sum_{k \in \mathcal{K}} y_{ij}^{kt} \leq u \cdot x_{ij}^t, \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \quad (1f)$$

$$\sum_{j \in \mathcal{N}} y_{ij}^{kt} - \sum_{j \in \mathcal{N}} y_{ji}^{kt}$$

Table 2
List of notations used in the SND model.

Notion	Meaning
Sets	
$\mathcal{N}, \mathcal{A}$	The set of nodes and arcs in the physical network.
$\mathcal{T}$	The total number of periods within a scheduled length and period $t \in \{0, 1, 2, \dots, \mathcal{T} - 1\}$.
$\mathcal{K}$	The set of commodities for carriers.
S	The set of extreme points of the uncertainty set identified by the SVC method.
Parameters	
o_k, d_k	The origin and destination for the commodity k .
b_k, e_k	The departure time and delivery deadline for the commodity k .
q_k	The actual demands for the commodity k .
$\bar{q}_k, \hat{q}_k$	The nominal demand and maximum deviation for the commodity k .
D_k^s	The demand of the commodity k at the s th vertex in the uncertainty set identified by SVC.
t^-	The departing period for a truck arriving at time period t . i.e. $t^- = t - 1, \text{ if } t \geq 1; \mathcal{T} - 1, \text{ otherwise}$.
f_{ij}, c_{ij}	The fixed cost and unit transportation cost from i to j .
u	Truck capacity of the carrier.
V	Maximum number of trucks for the carrier.
λ	The unit commodity outsourcing cost.
Decision variables	
x_{ij}^t	The service frequency of trucks. $x_{ij}^t \in \{0, 1, 2, \dots\}$.
y_{ij}^{kt}	The transportation volume of the commodity k . $y_{ij}^{kt} \geq 0$.
$Q(k)$	The amount of outsourcing required for the commodity k . $Q(k) \geq 0$.

$$= \begin{cases} \bar{q}_k + p_k \cdot \hat{q}_k - Q(k), & \text{if } i = o_k, t = b_k \\ -\bar{q}_k - p_k \cdot \hat{q}_k + Q(k), & \text{if } i = d_k, t = e_k \\ 0, & \text{otherwise} \end{cases}, \forall i \in \mathcal{N}, \forall k \in \mathcal{K}, \forall t \in \mathcal{T}, \quad (1g)$$

$$\sum_{k \in \mathcal{K}} p_k \leq \Gamma, \quad (1h)$$

$$0 \leq p_k \leq 1, \quad \forall k \in \mathcal{K}, \quad (1i)$$

$$y_{ij}^{kek} = 0, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \forall k \in \mathcal{K}, \quad (1j)$$

$$y_{ij}^{kt} \in \mathbb{R}^+, Q(k) \in \mathbb{R}^+, \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \forall k \in \mathcal{K}. \quad (1k)$$

In stage 1, the objective function (1a) is designed to minimize the worst-case cost, which includes the fixed cost, transportation cost, and outsourcing cost; Constraint (1b) is the asset balance equation, while Constraint (1c), as a budget constraint, restricts the maximum number of trucks. In stage 2, the objective function $\Omega(x, p)$ is the worst-case minimum transportation cost and outsourcing cost; Constraint (1f) serves as a capacity constraint, ensuring that the total flow does not exceed the network capacity; Constraint (1g) ensures the flow conservation equations; Constraint (1h) and Constraint (1i) portray uncertainty; p_k indicates the degree of fluctuation of the commodity k , controlled by the conservatism parameter Γ ; Constraint (1j) ensures that each commodity can reach its destination before the deadline, and Constraint (1d) and Constraint (1k) define the domain of the decision variables.

The uncertainty set used for the RO model is as follows:

$$\mathcal{U}_1 = \left\{ q \mid q_k = \bar{q}_k + p_k \cdot \hat{q}_k, \sum_{k \in \mathcal{K}} p_k \leq \Gamma, 0 \leq p_k \leq 1, \forall k \in \mathcal{K} \right\}, \quad (2)$$

where $\bar{q}_k$ is the average demand of the commodity k , and the notion $\hat{q}_k \geq 0$ is the maximum deviation, and the notion $\Gamma \in [0, |\mathcal{K}|]$ is the budget of uncertainty, controlling the robustness level.

Although the previously described model effectively mitigates risks related to demand uncertainty, its methodology is marked by a pronounced degree of conservatism. Furthermore, the model does not provide for the adaptive adjustment of the range of sets according to demand data, which makes it difficult to guarantee the reliability of decisions in the real world.

3.3. DDRO formulation

In the set (2), nominal demand and demand deviations cannot adaptively self-tune based on historical data. This lack of flexibility can make it difficult to hedge effectively against dynamically changing demands. Decisions may be too conservative or even invalid when faced with real data. Hence, we aim to dynamically adjust the uncertainty set based on historical data, allowing for less conservative yet risk-hedging decisions.

Alexandre et al. (2020) defines the uncertainty set as a set of convex hulls representing the historical demands of the scenarios. For convenience, we illustrate this construction method using the historical demand of two commodities, as shown in Fig. 3.

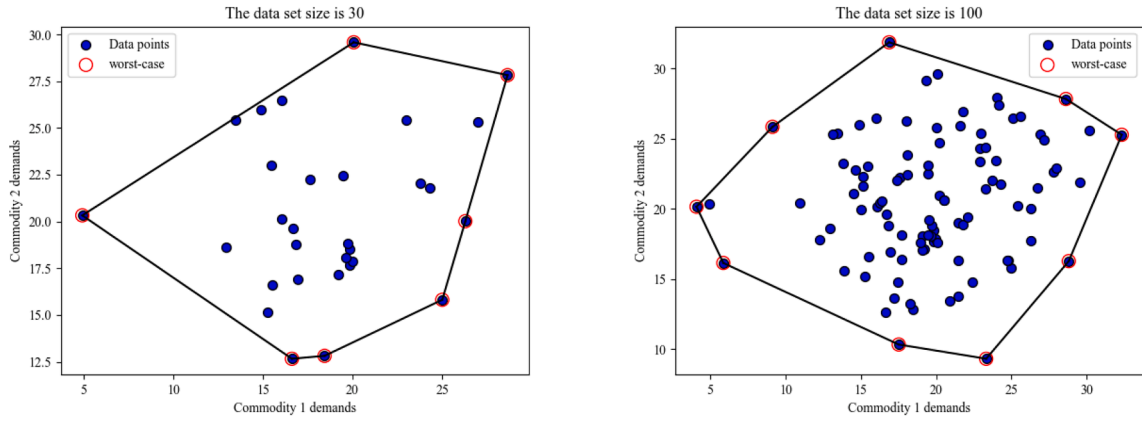
In Fig. 3, we consider that varying amounts of historical data will form different convex hulls, i.e., uncertainty sets that dynamically adapt based on historical data. We set that the historical data of demands is represented as D_k^i , where i denotes the i th data item, and D_k^i denotes the demand quantity for the commodity k in the i th data item. For the sake of convenience in developing the formula, it is referred to as x^i and there are $|\mathcal{D}|$ data in total. Then, we have access to a set of potential worst-case scenarios D_k^s derived from historical data, where $s \in S$ the vertex set S represents the s th vertex and D_k^s represents the demand quantity for the commodity k in the s th vertex. Hence, the uncertainty set $\mathcal{U}_2$ is defined as follows:

$$\mathcal{U}_2 = \left\{ q \mid q_k = p^s \cdot D_k^s, \sum_{s \in S} p^s = 1, 0 \leq p^s \leq 1, \forall k \in \mathcal{K}, \forall s \in S \right\}, \quad (3)$$

However, the presence of outliers in the data is inevitable. For example, there might be a surge in demand during events such as Tmall 618 or other promotional campaigns organized by e-commerce platforms. If such outliers are included in $\mathcal{U}_2$, decisions will tend to be more conservative, resulting in wasted assets. Therefore, it is essential to allow some points to fall outside the set. A natural idea is to construct the uncertainty set based on historical data by some algorithm, and decision-makers can control the number of outliers by adjusting the parameters. The SVC provides the mechanism to achieve this.

The SVC algorithm aims to identify the minimum enclosing sphere that contains the majority of the samples. Specifically, it solves the following optimization problem:

$$\min_{a, R, \xi} R^2 + \frac{1}{N\nu} \sum_i \xi_i \quad (4a)$$



(a) A convex hull formed when the sample point is 30 (b) A convex hull formed when the sample point is 100

Fig. 3. Uncertainty set considering different amounts of historical data.

$$s.t. \quad \|\Phi(x^i) - a\|^2 \leq R^2 + \xi_i, \quad \forall i \in |D|, \quad (4b)$$

$$\xi_i \geq 0, \quad \forall i \in |D|. \quad (4c)$$

The objective function (4a) minimizes the hypersphere radius R and penalizes anomalies by ξ_i . Additionally, N denotes the number of historical demand data points, and the parameter v controls the maximum proportion of outliers. Constraint (4b) indicates that demand data points x^i whose distance from the hypersphere center a exceeds the radius R should be penalized. This constraint facilitates the construction of a minimal-volume hypersphere within the feature space, encompassing the majority of data points while allowing for the presence of some outliers. Constraint (4c) implies that the penalty is non-negative. Subsequently, the Lagrangian function of the problem is given by:

$$L(R, a, \xi, \zeta, \theta) = R^2 + \frac{1}{Nv} \sum_i \xi_i - \sum_i \zeta_i \cdot (R^2 + \xi_i - \|\Phi(x^i) - a\|^2) - \sum_i \theta_i \cdot \xi_i \quad (5)$$

where the variables ζ_i and the variables θ_i are Lagrange multipliers.

We define the kernel matrix $K(x^i, x^j) = \Phi(x^i)\Phi(x^j)$. In the context of e-commerce logistics, the demand distribution for each commodity between origin and destination is relatively continuous, exhibiting a unimodal or near-unimodal fluctuation pattern. According to Shang et al. (2017), the weighted generalized intersection kernel (WGIK) function can effectively handle this situation. Hence, we define the kernel function as $K(x^i, x^j) = \sum_k l_k - \|Q(x^i - x^j)\|_1$, where $x^i, x^j \in D_k^i$. The weighting matrix $Q = \Sigma^{-\frac{1}{2}}$ is formed by utilizing the covariance matrix Σ of the observations in the dataset. Meanwhile, l_k needs to be sufficiently large in order to ensure that the kernel matrix $K(x^i, x^j)$ is positive definite. The parameter l_k satisfies $l_k > \max_i \{q_k^T x_i\} - \min_i \{q_k^T x_i\}$. Then, (5) converts into its dual problem via the Karush-Kuhn-Tucker (KKT) condition.

$$\min_{\zeta} \sum_i \sum_j \zeta_i \zeta_j K(x^i, x^j) - \sum_i \zeta_i K(x^i, x^i) \quad (6a)$$

$$s.t. \quad 0 \leq \zeta_i \leq \frac{1}{Nv}, \quad \forall i \in |D|, \quad (6b)$$

$$\sum_i \zeta_i = 1, \quad \forall i \in |D|. \quad (6c)$$

According to the KKT conditions, $\zeta_i = 0$ and $\theta_i = \frac{1}{Nv}$ hold if x^i is inside the sphere; $0 < \zeta_i < \frac{1}{Nv}$ and $0 < \theta_i < \frac{1}{Nv}$ indicates that x^i is on the boundary, i.e., $\|\Phi(x^i) - a\|^2 = R^2$; $\zeta_i = \frac{1}{Nv}$ and $\theta_i = 0$ is an anomaly, i.e., $\|\Phi(x^i) - a\|^2 > R^2$. Observations that are satisfying $\zeta_i > 0$ are support vectors (SV), while those with $0 < \zeta_i < \frac{1}{Nv}$ are boundary support vectors

(BSV). Hence, we obtain:

$$SV = \{i | \zeta_i > 0, \forall i \in |D|\} \quad (7)$$

and

$$BSV = \{i | 0 < \zeta_i < \frac{1}{Nv}, \forall i \in |D|\} \quad (8)$$

In that case, the radius R can be determined as the distance from the center a to any boundary support vector x^i :

$$R^2 = K(x^i, x^i) - 2 \sum_n \zeta_n K(x^i, x^n) + \sum_m \sum_n \zeta_m \zeta_n K(x^m, x^n), \forall i \in BSV \quad (9)$$

Using the obtained R , the uncertainty set can be defined as:

$$\mathcal{U}(x) = \{x | K(x, x) - 2 \sum_n \zeta_n K(x, x^n) + \sum_m \sum_n \zeta_m \zeta_n K(x^m, x^n) \leq R^2\} \quad (10)$$

It is worth noting that Eq. (10) describes a closure that excludes data from outliers. We still form a convex hull such as Eq. (3) to obtain the set of the worst-case scenario, as shown in Fig. 4. The parameter v controls the percentage of outliers, which is an upper bound of outliers. i.e., at least $(1 - v) \cdot 100\%$ of the data is inside the convex hull. The set defined by Alexandre et al. (2020) is a special case in our study.

For the reader's convenience, we summarize the sets, parameters, and decision variables required for modeling in Table 2. The DDRO model as defined by SVC processing $\mathcal{U}_2$ is as follows:

(DDRO) Stage 1:

$$\min_x \sum_{i \in \mathcal{T}} \sum_{j \in \mathcal{N}} \sum_{k \in \mathcal{K}} f_{ij} \cdot x_{ij}^{kt} + \max_p \Omega(x, p) \quad (11a)$$

$$s.t. \quad \sum_{j \in \mathcal{N}} x_{ji}^{kt} = \sum_{j \in \mathcal{N}} x_{ij}^{kt}, \quad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}, \quad (11b)$$

$$\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} x_{ij}^{kt} \leq V, \quad \forall t \in \mathcal{T}, \quad (11c)$$

$$x_{ij}^{kt} \in \{0, 1, 2, \dots\}, \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \quad (11d)$$

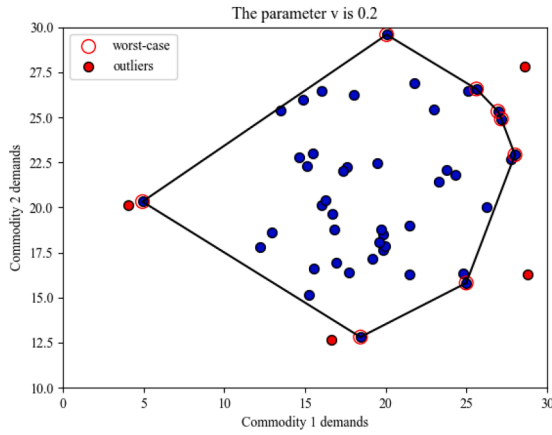
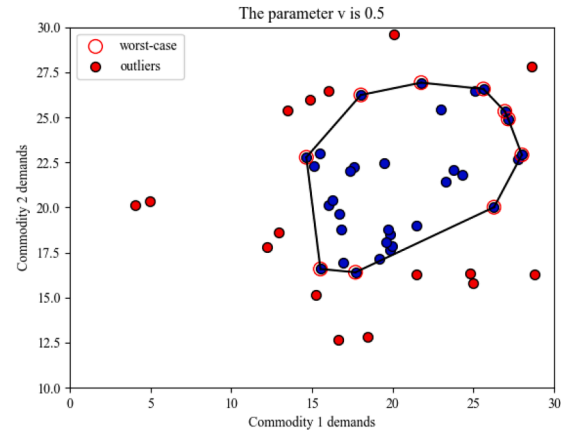
where $\Omega(x, p)$ is the optimal value of the second-stage problem.

Stage 2:

$$\Omega(x, p) := \min_{y, Q} \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} c_{ij} \cdot y_{ij}^{kt} + \sum_{k \in \mathcal{K}} \lambda \cdot Q(k) \quad (11e)$$

$$s.t. \quad \sum_{k \in \mathcal{K}} y_{ij}^{kt} \leq u \cdot x_{ij}^{kt}, \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \quad (11f)$$

$$\begin{aligned} & \sum_{j \in \mathcal{N}} y_{ij}^{kt} - \sum_{j \in \mathcal{N}} y_{ji}^{kt} \\ &= \begin{cases} q_k - Q(k), & \text{if } i = o_k, t = b_k \\ -q_k + Q(k), & \text{if } i = d_k, t = e_k \\ 0, & \text{otherwise} \end{cases}, \forall i \in \mathcal{N}, \forall k \in \mathcal{K}, \forall t \in \mathcal{T}, \end{aligned} \quad (11g)$$

(a) A convex hull formed when parameter v is 0.2(b) A convex hull formed when parameter v is 0.5Fig. 4. Uncertainty set using different parameter v .

$$\sum_{s \in S} p^s = 1 \quad (11h)$$

$$q_k = \sum_{s \in S} p^s \cdot D_k^s, \quad \forall k \in \mathcal{K}, \quad (11i)$$

$$0 \leq p^s \leq 1, \quad \forall s \in S, \quad (11j)$$

$$y_{ij}^{ke} = 0, \quad \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \forall k \in \mathcal{K}, \quad (11k)$$

$$y_{ij}^{kt} \in \mathbb{R}^+, Q(k) \in \mathbb{R}^+, \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \forall k \in \mathcal{K}. \quad (11l)$$

In the DDRO model, the objective function (11a) is to minimize the fixed cost in the first stage, the transportation cost and outsourcing cost in the second stage; Constraint (11b) indicates asset balance equations, which ensure the number of trucks entering and exiting the same node is equal in each adjacent time period; Constraint (11c) limits the number of trucks used. $\Omega(x, p)$ is the worst-case minimum transportation cost and outsourcing cost; Constraint (11f) serves as a capacity constraint, ensuring that the total flow does not exceed the network capacity; Constraint (11g) is the flow conservation equation; Constraint (11h), Constraint (11i) and Constraint (11j) define a scenario-based uncertainty set, which is identified via SVC; Constraint (11k) ensures that each commodity can reach its destination before the deadline, and Constraint (11d) and Constraint (11l) define the domain of the decision variables.

4. A solution method

It can be observed that in Section 3.3, the proposed DDRO service network design model is a two-stage robust optimization model. The two-stage robust optimization model is classified as the NP-hard problem (Ben-Tal & Goryashko, 2004), making it difficult to solve. A large number of scholars have been working on finding an efficient algorithm. For example, the Benders decomposition algorithm. However, the Benders decomposition algorithm faces challenges in computational speed when applied to large-scale instances. Zeng and Zhao (2013) developed a CCG algorithm to overcome the slow convergence of the Benders decomposition, which improves the speed and accuracy of solving the two-stage robust optimization model (see Wang & Qi, 2019 and Zhang et al. (2024)). In this paper, the customized CCG algorithm is designed for the proposed model. The following sections provide the specific reformulation and algorithmic framework.

4.1. Sub-problem and master problem

In Eq. (11a), the decision variables x, p are fixed when making second-stage internal decisions. That is, given any fixed $(\bar{x}, \bar{p})$, the interior minimization problem can be converted into its dual maximization formulation. At this stage, the second-stage problem becomes a

max-max problem, i.e., a single-layer maximization. By introducing dual variables α, β, γ , the constraints (11f), (11g), and (11k) are incorporated into Eq. (11a). The dual problem of the second-stage internal problem is formulated as follows:

$$(\text{DRP}(\bar{x}, \bar{p})) \quad \max_{\alpha, \beta, \gamma} - \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \{i\}} \sum_{t \in \mathcal{T}} u_{ij}^t \alpha_{ij}^t + \sum_k (\beta_{o_k}^{kbk} - \beta_{d_k}^{kek}) \cdot q_k \quad (12a)$$

$$\text{s.t.} \quad \beta_{o_k}^{kbk} - \beta_{d_k}^{kek} \leq \lambda, \quad \forall k \in \mathcal{K}, \quad (12b)$$

$$\beta_i^{kt-} - \beta_j^{kt-} \leq \begin{cases} c_{ij} + \alpha_{ij}^t, & \text{if } i \neq j \\ c_{ij}, & \text{if } i = j \end{cases}, \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{T} \setminus \{e_k\}, \forall i, j \in \mathcal{N}, \forall k \in \mathcal{K}, \quad (12c)$$

$$\beta_i^{kek} - \beta_j^{kt} + \gamma_{ij}^k \leq \begin{cases} c_{ij} + \alpha_{ij}^t, & \text{if } i \neq j \\ c_{ij}, & \text{if } i = j \end{cases}, \quad \forall t \in \mathcal{T}, \forall i, j \in \mathcal{N}, \forall k \in \mathcal{K}, \quad (12d)$$

$$\alpha_{ij}^t \geq 0, \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \quad (12e)$$

$$\sum_{s \in S} \bar{p}^s = 1 \quad (12f)$$

$$q_k = \sum_{s \in S} \bar{p}^s \cdot D_k^s, \quad \forall k \in \mathcal{K}, \quad (12g)$$

$$0 \leq \bar{p}^s \leq 1, \quad \forall s \in S. \quad (12h)$$

The second-stage problem can be reformulated as:

$$\max_{p, \alpha, \beta, \gamma} - \sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N} \setminus \{i\}} \sum_{t \in \mathcal{T}} u_{ij}^t \alpha_{ij}^t + \sum_{k \in \mathcal{K}} (\beta_{o_k}^{kbk} - \beta_{d_k}^{kek}) q_k$$

$$\text{s.t.} \quad \text{Constraints (12b) – (12e)},$$

$$\sum_{s \in S} p^s = 1, \quad (13)$$

$$q_k = \sum_{s \in S} p^s \cdot D_k^s, \quad \forall k \in \mathcal{K},$$

$$0 \leq p^s \leq 1, \quad \forall s \in S.$$

which is presented in a tractable form. Notably, as discussed by Wang and Qi (2019) and Wang and Qi (2020), an optimal solution exists for Model (13). According to Alexandre et al. (2020), $p^s \in \{0, 1\}$, $\forall s \in S$.

In the two-stage RO model, all wait-and-see decisions are included in the second-stage model. According to the CCG framework, we define a subproblem (SP) to identify the worst-case scenario. The SP is expressed as

$$\begin{aligned}
\text{SP} \quad & \max_{p, \alpha, \beta, \gamma} - \sum_{i \in \mathcal{N}'} \sum_{j \in \mathcal{N} \setminus \{i\}} \sum_{t \in \mathcal{T}} u \bar{x}_{ij}^t \alpha_{ij}^t + \sum_{k \in \mathcal{K}} (\beta_{ok}^{kbk} - \beta_{dk}^{kek}) q_k \\
\text{s.t.} \quad & \text{Constraints (12b) - (12e),} \\
& \sum_{s \in S} p^s = 1, \\
& q_k = \sum_{s \in S} p^s \cdot D_k^s, \quad \forall k \in \mathcal{K}, \\
& p^s \in \{0, 1\}, \quad \forall s \in S.
\end{aligned} \tag{14}$$

We solve the SP to identify the worst-case scenario. Simultaneously, it is added to the master problem (MP) to obtain the total cost under the worst-case scenario. Let the auxiliary decision variable η represent an evaluation of the worst-case costs. Then, the MP is given as

$$\text{MP} \quad \min_x \sum_{i \in \mathcal{N}'} \sum_{j \in \mathcal{N}} \sum_{t \in \mathcal{T}} f_{ij} x_{ij}^t + \eta \tag{15a}$$

$$\sum_{j \in \mathcal{N}} x_{ji}^t = \sum_{j \in \mathcal{N}'} x_{ij}^t, \quad \forall i \in \mathcal{N}, \forall t \in \mathcal{T}, \tag{15b}$$

$$\sum_{i \in \mathcal{N}'} \sum_{j \in \mathcal{N}} x_{ji}^t \leq V, \quad \forall t \in \mathcal{T}, \tag{15c}$$

$$x_{ij}^t \in \{0, 1, 2, \dots\}, \quad \forall t \in \mathcal{T}, \forall i \in \mathcal{N}, \forall j \in \mathcal{N}, \tag{15d}$$

$$\eta \geq \sum_{i \in \mathcal{N}'} \sum_{j \in \mathcal{N}} \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} c_{ij} \cdot (y_{ij}^{kt})^h + \lambda \cdot \sum_{k \in \mathcal{K}} (Q(k))^h, \quad h = 1, 2, \dots, H \tag{15e}$$

$$\begin{aligned}
& \sum_{k \in \mathcal{K}} (y_{ij}^{kt})^h \leq u x_{ij}^t, \quad \forall i, j \in \mathcal{N}, \forall t \in \mathcal{T}, \\
& h = 1, 2, \dots, H,
\end{aligned} \tag{15f}$$

$$\begin{aligned}
& - \sum_{j \in \mathcal{N}'} (y_{ok}^{kt})^h + \sum_{j \in \mathcal{N}} (y_{okj}^{kbk})^h = q_k - (Q(k))^h, \quad \forall k \in \mathcal{K}, \forall t \in \mathcal{T}, \\
& h = 1, 2, \dots, H,
\end{aligned} \tag{15g}$$

$$\begin{aligned}
& - \sum_{j \in \mathcal{N}'} (y_{dk}^{kt})^h + \sum_{j \in \mathcal{N}} (y_{dkj}^{kek})^h = -q_k + (Q(k))^h, \quad \forall k \in \mathcal{K}, \forall t \in \mathcal{T}, \\
& h = 1, 2, \dots, H,
\end{aligned} \tag{15h}$$

$$\begin{aligned}
& - \sum_{j \in \mathcal{N}'} (y_{ji}^{kt})^h + \sum_{j \in \mathcal{N}} (y_{ij}^{kt})^h = 0, \quad \forall i \in \mathcal{N}, \forall k \in \mathcal{K}, \forall t \in \mathcal{T}, \\
& h = 1, 2, \dots, H, (i, t) \neq (o_k, b_k), (d_k, e_k),
\end{aligned} \tag{15i}$$

$$q_k = \sum_{s \in S} (p^s)^h \cdot D_k^s, \quad \forall k \in \mathcal{K}, \tag{15j}$$

$$(y_{ij}^{kek})^h = 0, \quad \forall i, j \in \mathcal{N}, \forall k \in \mathcal{K}, h = 1, 2, \dots, H, \tag{15k}$$

$$(y_{ij}^{kt})^h \in \mathbb{R}^+, \quad \forall i, j \in \mathcal{N}, \forall k \in \mathcal{K}, \forall t \in \mathcal{T}, h = 1, 2, \dots, H, \tag{15l}$$

$$(Q(k))^h \in \mathbb{R}^+, \quad \forall k \in \mathcal{K}, h = 1, 2, \dots, H. \tag{15m}$$

4.2. Algorithmic framework

In Section 4.1, we give a reformulation of the two-stage RO model. The model is decomposed into MP and SP. The worst-case scenario is identified by solving the SP and is subsequently added to the MP (i.e., constraint (15d)-(15m)); the decision variable x , obtained from the MP, is then used in the SP until the algorithm converges (see Zeng & Zhao, 2013). The framework of the CCG algorithm is as Algorithm 1. According to Zeng and Zhao (2013), the CCG algorithm will converge in a finite number of iterations when the extreme points are finite. In this paper, the uncertain set is a convex hull identified by SVC from historical data with a finite number of vertices. Obviously, meet above the convergence theorem.

5. Numerical experiment

In this section, numerical experiments were conducted to validate the proposed model and algorithm, primarily for a threefold purpose: 1.

Algorithm 1: The CCG algorithm

Input: All parameters and sets in SND problem, and the uncertainty set identified via SVC.

Output: A SND scheme x

- 1 **Initialize.** Iteration $h \leftarrow 1$, upper bound $UB = \infty$, lower bound $LB = 0$, termination condition $\epsilon = 10^{-5}$, $S' \leftarrow \emptyset$.
- 2 **Solve SP.** Generate $\bar{x}$, then solve SP to obtain the optimal solution $(p^s)^h$ and the optimal objective function value sp^* . Set $UB \leftarrow \min\{UB, f\bar{x} + sp^*\}$. Then, add the SP optimal solution $(p^s)^h$ to the scenario set $S' \leftarrow S' \cup \{(p^s)^h\}$ (i.e., constraints 15e - 15m).
- 3 **while** $(UB - LB)/UB > \epsilon$: **do**
- 4 **Solve MP.** Solve MP to get the optimal objective function value mp^* and the optimal solution x^h based on S' ; then, set $LB \leftarrow \max\{LB, mp^*\}$.
- 5 **Solve SP.** Fix $\bar{x} = x^h$, then solve SP to obtain the optimal solution $(p^s)^h$ and the optimal objective function value sp^* . Set $UB \leftarrow \min\{UB, f x^h + sp^*\}$.
- 6 **Expansion scenario set.** Adding the SP optimal solution $(p^s)^h$ to the scenario set $S' \leftarrow S' \cup \{(p^s)^h\}$ (i.e., constraints (15e) - (15m)). Set $h \leftarrow h + 1$.
- 7 **end**
- 8 **Output** SND scheme x .

By comparing with the Benders decomposition algorithm, demonstrate advantages of the proposed CCG algorithm in both solution time and accuracy. 2. By comparing with robust optimization models based on budget uncertainty sets, show the effectiveness of the DDRO model in cost savings. 3. By conducting sensitivity analysis on model parameters, valuable managerial insights are provided to logistics service providers. All experiments were conducted on an Intel(R) Core(TM) i5-14400F processor, operating at 2.50 GHz and supported by 16 GB of RAM with NVIDIA GeForce RTX 4060 Ti. We use Python invoking Gurobi 11.0.3 as a coding platform.

5.1. Generating instances

According to Wang and Qi (2020) and Xiang et al. (2022), we generate 12 groups of different instances with synthetic data. The scale of these instances is determined by combinations of parameters: $|\mathcal{N}'| \in \{6, 12\}$, $|\mathcal{K}| \in \{8, 16, 24\}$, and $|\mathcal{T}| \in \{5, 10\}$. Each instance is named using its structure parameters. For example, an instance denoted as K8N6T5 indicates there are 8 commodities, 6 transport nodes, and 5 time periods. Other parameters are generated in each simulation trial as follows: $c = 150$, $f_{ij} \sim U(100, 200)$, $u = 20$, $c_{ij} \sim U(10, 30)$ and $V \in \{1, 2, 3, \dots, 10\}$.

5.2. Algorithm performance

In this section, we evaluate the performance of the developing CCG algorithm compared to the Benders decomposition. The performance of both algorithms is assessed using the following metrics: 1) Programming Time (CPU): The average time taken by an algorithm to solve a group of instances of a given scale, the maximum runtime is set to 7200 seconds. It is important to note that if the maximum time limit is reached before an iteration is completed, the algorithm should not terminate immediately. Instead, it should continue until that iteration is finished; 2) Optimality gap when the algorithm is terminated, defined as $GAP = (UB - LB)/(UB) \times 100\%$, where UB and LB are obtained by solving SP and MP, respectively; 3) The number of iterations, denoted as $Iter$; 4) The output objective function value when the algorithm is terminated, denoted as UB ; 5) The objective value gap between the CCG and the Benders decomposition, defined as $diff^* = (UB_{CCG} - UB_{BD})/(UB_{CCG}) \times 100\%$.

Table 3
Performance of the CCG method and the Benders decomposition algorithm.

Instance	CCG				Benders				
	<i>UB</i>	<i>iter</i>	<i>CPU</i>	<i>GAP</i>	<i>UB</i>	<i>iter</i>	<i>CPU</i>	<i>GAP</i>	<i>diff</i> *
K8N6T5	4890	1	1	0	4890	5	4522.3	0	0
K8N6T10	6883	1	1	0	8965	17	7253.4	50.2%	-30.2%
K8N12T5	4810	1	2.2	0	7295	8	7358	61.9%	-51.6%
K8N12T10	6810	1	14.7	0	9814	5	8153.9	47.3%	-44.1%
K16N6T5	9768	2	2	0	12033	13	7836.2	57.6%	-23.2%
K16N6T10	12238	2	237.2	0	19167	4	8536	60.1%	-56.6%
K16N12T5	9141	2	41.8	0	16221	7	7634	58.8%	-77.4%
K16N12T10	11636	3	1154.2	0	17595	4	8754.2	65.5%	-51.2%
K24N6T5	11533	3	651.1	0	20489	6	7223	75.2%	-77.7%
K24N6T10	17942	2	1910	0	27562	3	8510	82.9%	-53.6%
K24N12T5	18531	4	753.3	0	32772	8	7486.5	85.6%	-76.8%
K24N12T10	16635	5	3454.8	0	28366	16	8689.7	90.2%	-70.5%
Average	10901	2	685.2	0	17097	8	7663.2	61.3%	-51.1%

We generate a set of demand data of size 100 from a normal distribution with a mean of 15 and a standard deviation of 5. Then, set $v = 0.9$ to identify the uncertainty set (3) by SVC. For ease of calculation, we set $V = 10$. The model parameters are set according to the method described in Section 5.1. We obtain the relevant performance metrics of the CCG method and Benders decomposition algorithm on 12 groups of instances, as shown in Table 3.

As shown in Table 3, all instances were solved to optimality by the CCG algorithm in the restricted time, achieving $GAP = 0$. The average running time of the CCG algorithm is 685.2 s, which is significantly lower than that of the Benders decomposition method. In contrast, the average optimal UB for the Benders decomposition method is higher than the CCG method, and the $diff^*$ is 51.1%. These results demonstrate the effectiveness of the CCG algorithm in solving the SND problem, as also observed in prior studies (Wang & Qi, 2020; Zhang et al., 2024).

In addition, we analyze the impact of the number of extreme points $|S|$ on the CCG algorithm. We perform further experiments on the instance K8N6T5. Specifically, we set $|S| \in \{5, 10, 50, 100, 500, 1000\}$, and the average time to solve 30 times is as an evaluation metric. The results are shown in Fig. 5. For larger $|S|$ values (e.g., $|S| = 500$ and $|S| = 1000$), the computational time increases significantly. This is because as the scenario set grows larger, the master problem may require observing more scenarios to achieve convergence. In this context, more scenarios imply a larger and more complex master problem (with more resource variables and constraints), indicating longer computation times. This indicates that larger $|S|$ sizes require higher computational resources. Therefore, if decision-makers can obtain a representative set of scenarios, it helps save computational time. The number of extreme points in the demand set is much smaller than that in all scenarios, which also demonstrates to some extent the computational advantage of the SVC method for constructing uncertainty sets.

It is worth noting that, as stated in Zhang et al. (2024), the performance of CCG may vary in different computing environments. In general, better computer configuration means better algorithmic performance. Therefore, our numerical study basically reflects the performance comparison of the two algorithms.

5.3. Sensitivity analysis

In this section, the focus is on the exploration of the influence mechanism of key model parameters on solutions. The parameters in the model are as follows: the maximum number of trucks V , the unit outsourcing cost λ , and the uncertainty v (where v represents the maximum proportion of abnormal samples). The parameter value ranges were chosen to encompass a variety of realistic operational scenarios, informed by common observations in logistics practice and supported by existing literature (see Sun et al., 2025; Wang & Qi, 2019). Specifically, when one

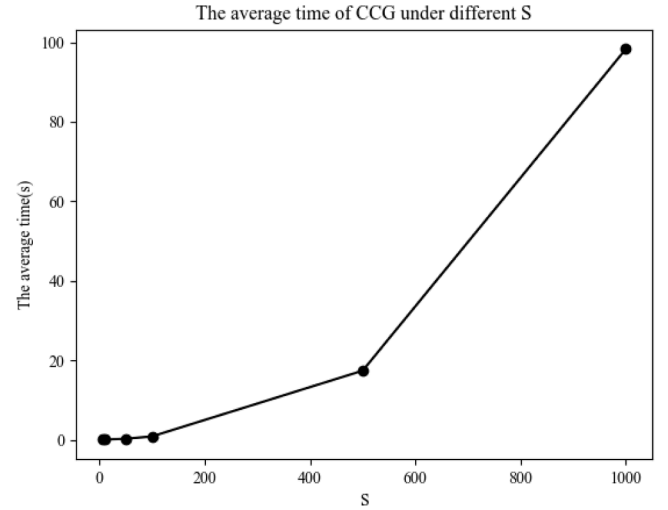


Fig. 5. The average time of CCG under different $|S|$.

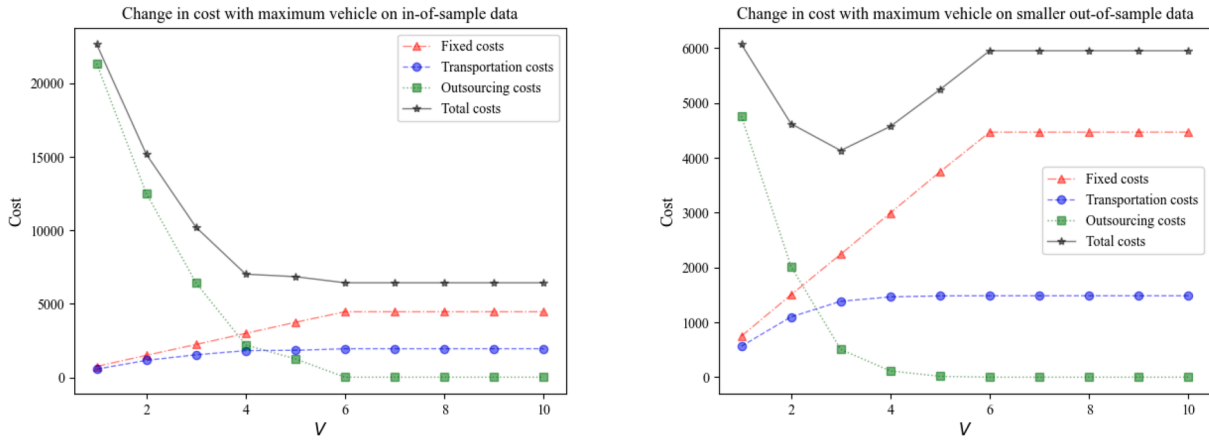
parameter is changed and other parameters are fixed, the optimal decision x is then determined. Then its efficacy is assessed by evaluating its performance with both in-of-sample data and an out-of-sample test set. We use the expected value of fixed costs, transportation costs, and outsourcing costs as evaluation metrics. The detailed numerical results are as follows.

5.3.1. The effect of maximum vehicle

We analyze the performance of the model for different values of the maximum vehicle V , ranging from 1 to 10. In case of a change in V , $u = 20$, $v = 0.7$ and $\lambda = 150$. All experiments are conducted on instance K8N6T5. The results are shown in Fig. 6.

In Fig. 6(a), it is noteworthy that the total costs do not escalate with the number of assets when the characteristics of the out-of-sample demands are similar to those of the in-sample demands. This is because, when the maximum number of trucks, denoted as V , is low. The service network is unable to handle all demands, and the volume exceeding network capacity must rely on outsourcing for fulfillment, resulting in expensive outsourcing costs. Conversely, as the number of trucks improves, the dependence on outsourcing diminishes, leading to a reduction in outsourcing costs to nearly zero.

Furthermore, we investigate the difference in total costs between $V = 1$ and $V = 4$ while varying the parameters v and λ . Specifically, we derive decision variables x for both $V = 1$ and $V = 4$ across different values of v and λ , utilizing all available in-sample data to assess the total costs associated with the implementation of the decisions x . The total costs for $V = 1$ are then subtracted from those for $V = 4$. As illustrated



(a) Change in cost with maximum vehicle V on in-of-sample data (b) Change in cost with maximum vehicle V on smaller out-of-sample data

Fig. 6. Change in cost with maximum vehicle V .

in Fig. 7, it has been observed that when unit outsourcing costs are low, the total costs for $V = 1$ and $V = 4$ are identical. In this case, all freight is fulfilled through outsourcing, and the logistics service provider does not need to utilize trucks to design the service network. Additionally, it is evident that as v increases (resulting in the less conservative solution) and λ rises, the difference in total costs has increased. This indicates that when $v \rightarrow 1$ or when $\lambda \rightarrow +\infty$, adding more trucks is more conducive to lowering total costs. This also partly explains the necessity of constructing a network with sufficient capacity. Note that the conclusions obtained from out-of-sample data are consistent.

Conversely, if the out-of-sample data has a smaller mean compared to the in-of-sample data, the conclusion differs slightly. As shown in Fig. 6(b), the total cost exhibits a tendency to decrease first and then increase. In the initial phase, companies can mitigate high outsourcing costs by increasing the number of trucks, resulting in a decrease in total costs. However, the allocation of an excessive number of trucks leads to resource redundancy, which means fixed costs cannot be sufficiently offset by the reduction in outsourcing expenses. Ultimately leading to an increase in total costs. This tendency also highlights the negative impact of overly conservative solutions on company profitability. Consequently, our objective is to achieve a balance by formulating decisions that are robust yet not excessively conservative.

5.3.2. The effect of unit outsourcing cost

In this section we study the impact of the unit outsourcing cost of freight λ on every cost. We set $V = 5$, $v = 0.7$ and $u = 20$, and $\lambda \in \{10, 25, 50, 75, 100, 125, 150\}$. All experiments are conducted on instance K8N6T5.

Fig. 8 illustrates the significant impact of unit outsourcing cost λ on the performance of solutions that are made by the DDRO model. Notably, with the unit outsourcing cost increasing, total costs exhibit an upward trend, while the outsourcing costs, represented by the green dotted line in Fig. 8, initially rise before subsequently declining. This is because companies tend to meet demand through outsourcing when costs are relatively low. Conversely, as outsourcing costs escalate, companies are prompted to invest in the development of their own service network to mitigate reliance on external outsourcing. In this context, the network structure stabilizes due to the fixed parameter $V = 5$, resulting in an increase in total costs as the value of λ rises, given that the volume of outsourced freight remains constant.

Then, we give a further explanation numerically. Since we generate $f_{ij} \sim U(100, 200)$ and $c_{ij} \sim U(10, 30)$. Assuming $u = 15$ and $|\mathcal{T}| = 5$. Then

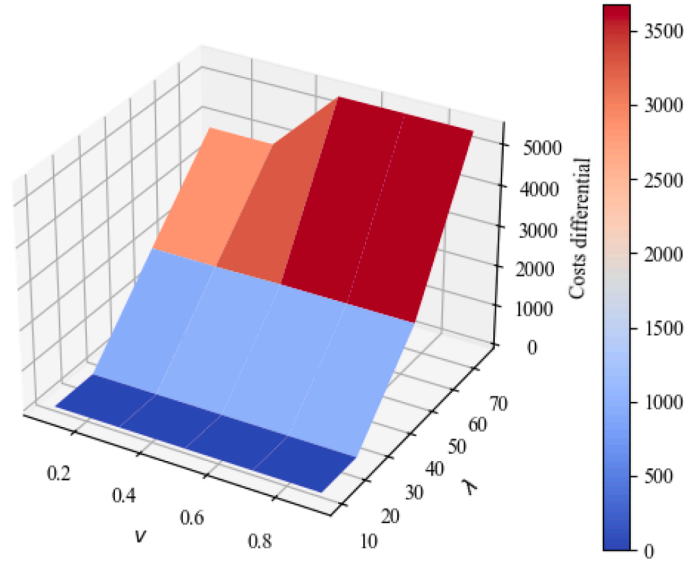
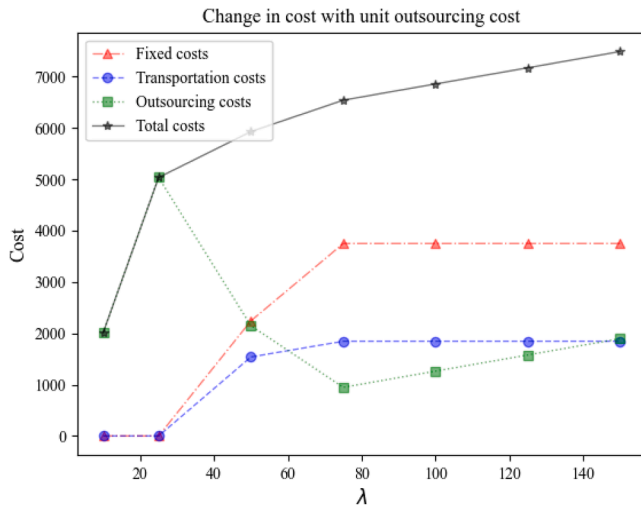
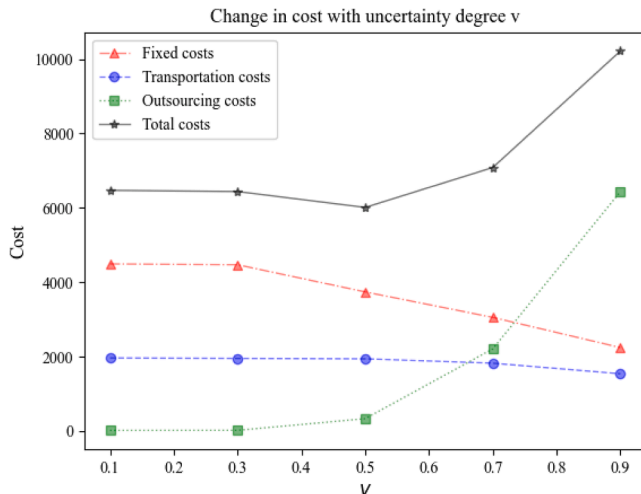
when a vehicle is transporting only one type of freight, the demands are 15 in a route. So the unit cost of transporting freight over the network uc and $uc \in [37, 77]$, approximately. *i.e.*, when $\lambda \in [37, 77]$, logistics service providers may choose to deliver some freight through outsourcing while utilizing trucks for the remainder. However, in practice, the emergence of consolidation causes the above intervals to shift to the left (e.g., in the event that the truck in question has a capacity of 20 and the demand for freight is 15, the truck is able to utilize the remaining capacity to transport other freight). Therefore, as mentioned by Lo et al. (2013), it is advantageous for companies to evaluate λ and uc to determine whether outsourcing should be considered as a viable option, as it may lead to reduced costs.

5.3.3. The effect of uncertainty degree

This section examines the impact of the uncertainty control parameter v on the DDRO model's solutions. First, we set $V = 5$, $\lambda = 150$ and $u = 20$. The uncertainty set is identified using the SVC method at $v = 0.1, 0.3, 0.5, 0.7, 0.9$ based on historical data of size (100,5), generated from a normal distribution with a mean of 15 and a standard deviation of 5. We use a simpler instance, K5N6T5, and observe the performance of the model. The results are shown in Fig. 9.

As shown in Fig. 9, we observe that the fixed cost decreases as v increases. This is because increasing v means more data is classified as outliers, reducing the size of the uncertainty set and resulting in a less robust solution. It can be observed that when $v \rightarrow 1$, outsourcing costs rise rapidly, it indicates that network capacity is too low to deliver freight effectively. Furthermore, the total cost decreases and then increases with v . When the parameter v approaches 1, the solution loses its robustness. Hence, when faced with out-of-sample data, the constructed network can barely hedge against risk, incurring expensive outsourcing costs. Whereas $v = 0.1 \rightarrow 0.5$, networks still have a strong ability to hedge risk. Notably, we find that decision performance is best when $v = 0.5$. This suggests that neither an overly conservative nor an unconservative condition is an optimal decision for the company.

The alteration in fixed costs means that variations in the parameter v lead to changes to the service network structure. Fig. 10 illustrates the optimal configurations derived from the DDRO model, with the uncertainty degree v ranging from 0.3 to 0.9. According to Liium et al. (2009) and Wang and Qi (2020), we define that a cycle is a path starting from a node at the beginning of the time period 0 and arriving at the same node at the end of the time period $T - 1$. It is evident that as v increases, both the number of cycles and the number of vehicles employed decrease, indicating that the DDRO model progressively generates solutions that

The difference of total costs between $V = 1$ and $V = 4$ as varying v and λ .Fig. 7. The difference of total costs between $V = 1$ and $V = 4$ as varying v and λ .Fig. 8. Change in cost with different λ .Fig. 9. Change in cost with uncertainty degree v .

are less robust with higher values of v . Specifically, at $v = 0.3$, the DDRO model opts for a conservative approach, utilizing six vehicles to complete three cycles (refer to Fig. 10(a)). Conversely, at $v = 0.9$, only three vehicles are utilized to accomplish two cycles (see Fig. 10(d)), resulting in a reduction of fixed costs. However, the decrease in cycles and the number of vehicles (as depicted in Fig. 10) significantly diminishes the likelihood of consolidation, resulting in expensive outsourcing costs. To further clearly indicate this result, we employ the load factor as a metric for evaluation, monitoring its variations in both in-of-sample and out-of-sample data across different values of v . The load factor is computed as follows:

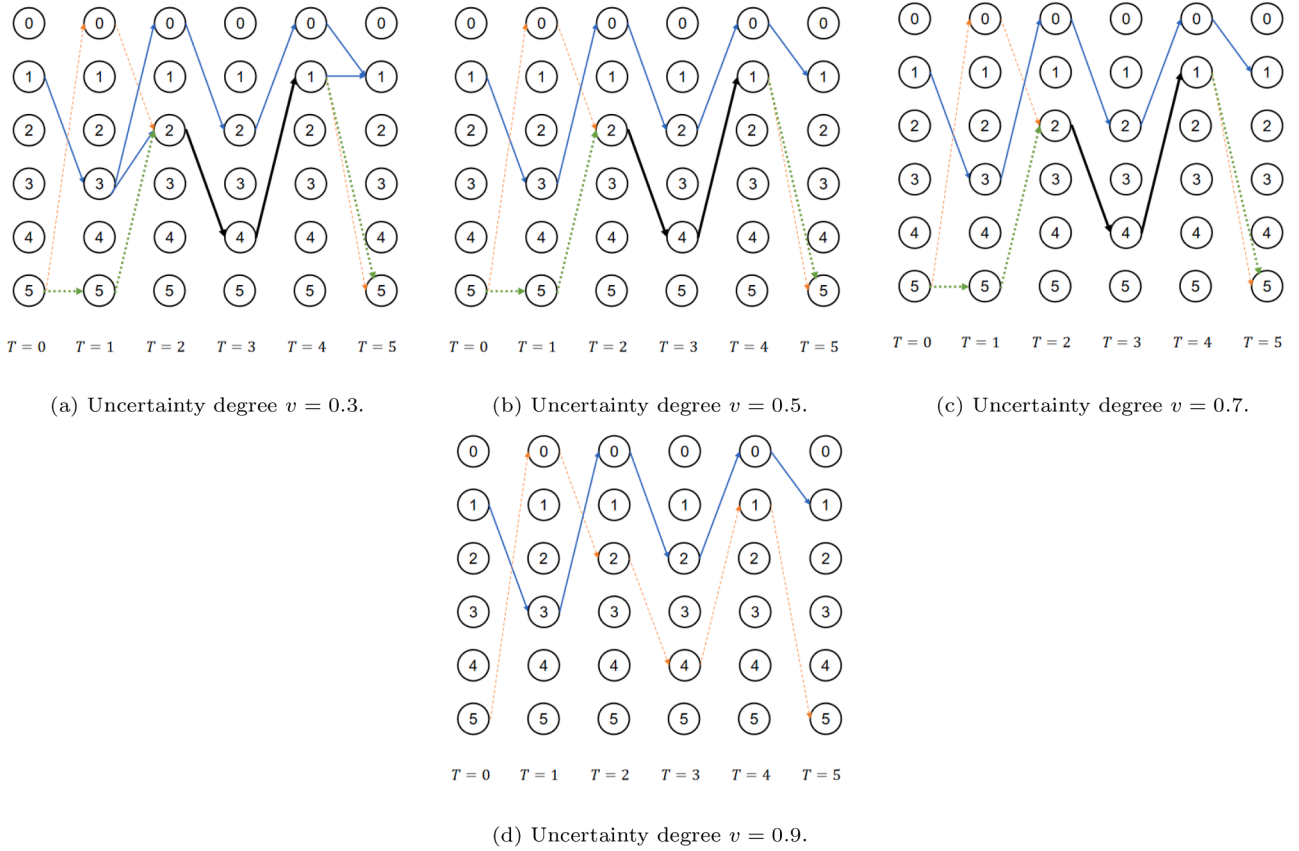
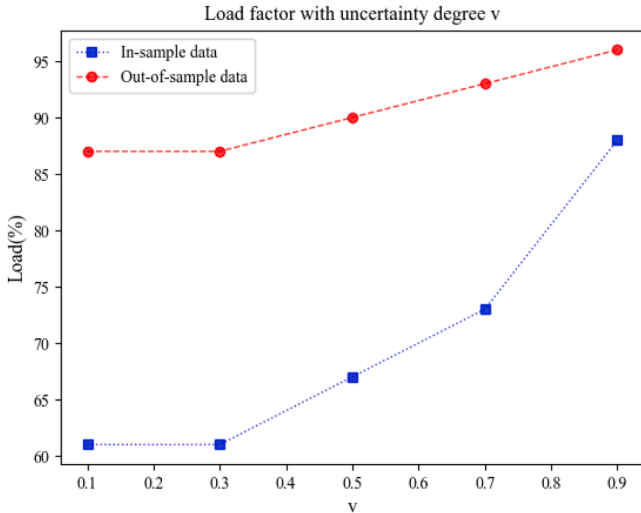
$$\text{Load Factor} = \frac{\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \sum_{k \in \mathcal{K}} \sum_{t \in \mathcal{T}} y_{ij}^{kt}}{\sum_{i \in \mathcal{N}} \sum_{j \in \mathcal{N}} \sum_{t \in \mathcal{T}} x_{ij}^t \cdot u}, \quad \text{if } y_{ij}^{kt} > 0.$$

From Fig. 11, we see that $v = 0.1 \rightarrow 0.9$, the gap between in-of-sample and out-of-sample data load factors narrows. This trend illustrates that DDRO becomes more conservative when the value of v is smaller, effectively hedging against risks by reserving additional capacity. Conversely, higher v values lead to less conservative decisions.

In summary, the decision-maker possesses the ability to influence the level of conservatism by modifying the parameter v . A decrease in the value of v leads to a more resilient decision-making process. This indicates that companies ought to adapt the value of v in accordance with their risk tolerance during the planning phase, thereby achieving a balance between economic considerations and robustness, ultimately facilitating more economically efficient decision-making.

5.4. The value of DDRO

In this section, to validate the advantages of the proposed DDRO model, we evaluated the performance of various model outputs on out-of-sample data. These models include the deterministic model, the two-stage stochastic optimization model, the traditional robust optimization model, and the DDRO model proposed in this paper. For the deterministic model and the two-stage stochastic optimization model, please refer to Bai et al. (2014). Specifically, we use 500 out-of-sample data to evaluate and compare solutions made by these models, using average total cost as the evaluation metric. For the DDRO model, after generating the demands data, the uncertainty set is identified using the SVC method with specified parameters v value. Then the DDRO

Fig. 10. Results obtained by different values of v .Fig. 11. Load factor with uncertainty degree v .

model is solved to obtain the decision $\mathbf{x}_{DDRO}$ based on set 10. The RO model is solved directly using the nominal demand and maximum deviation. The calculation method is as follows: $\bar{q}_k$ is $\frac{1}{|D|} \cdot \sum_n D_k^n, \forall k \in \mathcal{K}$ and $\hat{q}_k = \max_{0 \leq n \leq |D|} D_k^n - \bar{q}_k, \forall k \in \mathcal{K}$. Then, the RO set is as follows:

$$\mathcal{U}_1 = \left\{ \mathbf{q} \mid q_k = \bar{q}_k + p_k \cdot \delta \hat{q}_k, \sum_{k \in \mathcal{K}} p_k \leq \Gamma, 0 \leq p_k \leq 1, \forall k \in \mathcal{K} \right\}. \quad (16)$$

To evaluate robustness, we control the parameters v and δ to ensure an almost comparable number of outliers in the two uncertainty sets. Additionally, the deterministic model is based on nominal demand resolution, with the solution denoted as $\mathbf{x}_D$. To ensure fairness in evaluation, the stochastic optimization model employs a *sample average approximation* method. The scenario set is defined as S , the convex hull vertex set identified by SVC, with each scenario occurring with probability $\frac{1}{|S|}$. Then, we obtain the decisions $\mathbf{x}_D, \mathbf{x}_{SO}, \mathbf{x}_{RO}$ and $\mathbf{x}_{DDRO}$. We fix $\mathbf{x}$ and generate demands from different distributions acting on MP to re-evaluate the cost of the model. The results are shown in Fig. 12.

As illustrated in Fig. 12, the cost of solutions based on deterministic models is significantly higher than other solutions. This is because the fluctuation of demand is ignored during the planning phase, resulting in an oversimplified service network structure that struggles to handle excessive demand. Therefore, it is essential to consider demand uncertainty in advance. For the other three models, the solutions output by the DDRO model proposed in this paper demonstrate superior performance. Specifically, compared to the solution of the two-stage stochastic optimization model, the solution of the DDRO model consistently achieves lower costs. This shows that minimizing the average cost across all demand scenarios cannot avoid the severe negative impact of the worst-case scenario. Although the probability is low, if it occurs, it will significantly damage the profitability of logistics service providers. Furthermore, it is observed that the DDRO approach incurs lower costs compared to the traditional RO method. This cost efficiency can be attributed to DDRO's utilization of the convex hull of historical data as the uncertain set, which presents tighter bounds and fewer blank areas in comparison to RO, thereby rendering DDRO less conservative. It is noteworthy that when $v = 0.7$, the cost of the DDRO solution is slightly higher than that of the RO model. As mentioned above, the solution exhibits limited robustness at this point. Since the uncertainty

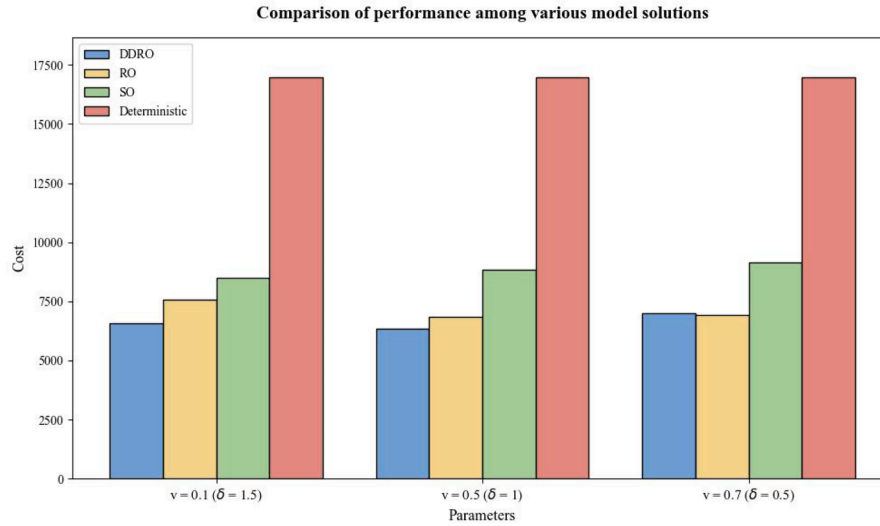


Fig. 12. Comparison of performance among various model solutions.

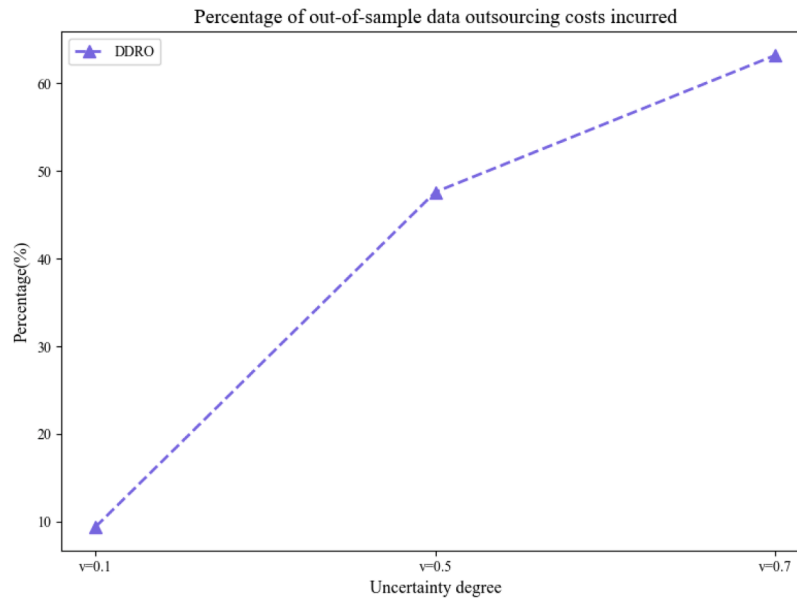


Fig. 13. Percentage of out-of-sample data outsourcing costs incurred.

set identified by SVC becomes more compact, the ability to hedge risks is diminished. This reflects that when data is insufficient, utilizing the RO model for decision-making may be a better choice for logistics service providers. Besides, DDRO offers probabilistic assurances that decisions will fulfill performance criteria, particularly when there is minimal fluctuation in the characteristics of the customer demand data. As shown in Fig. 13, DDRO has the capacity to ascertain that $(1-v)\%$ of the system's data meets expectations, thereby rendering decisions more targeted and adaptable. In contrast, the RO method can hardly provide such a probabilistic guarantee.

Based on these comparative results of the RO model and DDRO model, we draw the following conclusions:

(i) Unlike the RO model, the DDRO model utilizes an adaptive methodology to formulate the uncertainty set based on historical data. This approach facilitates decision-making that is both robust and cost-effective while maintaining validity. Note, in scenarios characterized by limited data availability, solutions derived from the RO model generally exhibit increased conservatism, which may lead to enhanced economic advantages.

(ii) By adjusting the parameters v of the SVC approach, we provide a DDRO model that allows decision-makers to design more reliable service networks, informed by their experiential knowledge or risk preferences.

6. Managerial insights and discussion

This paper explores the service network design problem under demand uncertainty, simultaneously incorporating data-driven approaches and robust optimization to improve service network performance. Based on a clear analysis and identification of the advantages of designing service networks utilizing a data-driven robust optimization framework, in this section, we provide some managerial insights for logistics service providers in practical implementation.

First, logistics service providers should scientifically determine the budget for their owned fleet size based on their risk tolerance and cost structure. When outsourcing costs are low, they may appropriately rely on outsourcing to help absorb fluctuations; conversely, they should moderately increase their owned freight capacity to develop a more reliable service network. Moreover, logistics service providers are advised

to avoid decision-making approaches that are excessively optimistic or overly cautious, as both extremes may impede profitability. On the one hand, decision-making approaches that depend solely on average demand fail to account for demand variability, thereby incurring substantial outsourcing costs. On the other hand, excessively allocating capacity to accommodate events with very low probability leads to consistently underutilized resources. In situations where the true distribution of demand data cannot be accurately determined, adopting a compromise risk tolerance in the decision-making process may represent the most effective approach. Nonetheless, this issue warrants comprehensive and ongoing investigation within the context of the SND problem.

A limitation of this study is that the uncertainty sets are derived exclusively from historical data at the beginning of the planning horizon, which may constrain the solution's effectiveness in mitigating risks. In practice, once the schedule is implemented, real-time demand data can be continuously collected throughout operations. When integrating such demand data into the analysis, it is essential to iteratively identify the uncertainty set to develop robust solutions. This approach has the potential to reduce costs and enhance profitability while necessitating only minimal modifications to the current service network configuration. However, this requires carefully weighing the scheduling costs associated with making the modification against the profits that can be gained by implementing it. Furthermore, even over multiple quarters, it's possible for the logistics service providers to consistently identify the uncertainty set and reallocate resources accordingly. This process evolves into a problem of strategic significance. Finally, it is noteworthy that the demand scenarios obtained based on the SVC method may not be uniformly distributed across the learned hypersphere. This may fail to capture portions of the uncertainty space implicitly learned by SVC, potentially leaving certain boundary directions unprotected. Although numerical experiments demonstrate that the SVC-based approach can identify critical demand scenarios and generate high-quality solutions, further consideration is still required for certain important freight contexts. There are two approaches that may resolve this issue: (i) Utilizing more advanced forecasting techniques and incorporating these forecasted demand scenarios into historical data; (ii) Generating additional demand scenarios based on existing data according to specific rules (refer to Sarayloo et al., 2020). However, introducing these new scenarios requires careful trade-offs: on the one hand, increased costs due to solution redundancy; on the other hand, enhanced capability to hedge against demand fluctuations.

In summary, employing a DDRO method can enhance the performance of the freight service network and bring benefits for the logistics service providers.

7. Conclusion

To mitigate the adverse effects of demand fluctuations on companies, it is imperative to design a service network with adequate capacity to manage risks associated with uncertainty in demand. To our best knowledge, this paper presents the first study on the data-driven robust service network design problem, aiming to ensure a company's profitability under fluctuating demand. We propose a method that constructs uncertainty sets via support vector clustering; based on this, a two-stage robust optimization model is developed. To efficiently solve this model, a customized column-and-constraint generation algorithm was designed. Numerical experiments validate the efficacy of the CCG algorithm, demonstrating its superiority over the Benders decomposition method in terms of computational efficiency and solution quality. Additionally, the advantages of the DDRO framework are highlighted, as it facilitates robust decision-making while exhibiting less conservativeness compared to traditional RO methods. Furthermore, we provide managerial insights indicating that the adoption of either excessively unconservative or overly conservative strategies may lead to increased transportation costs, thereby negatively impacting profitability. The decisions derived from the methodology proposed in this paper are char-

acterized by robustness and cost-effectiveness, potentially yielding enhanced benefits in the context of uncertain demand.

In this paper, we have carried out a preliminary exploration of DDRO SND. In the future, we will further develop innovative data-driven methods to address a service network design problem that balances robustness and conservatism through adaptive optimization and dynamic decision-making mechanisms to meet the challenges of practical applications under complex demand fluctuations. For example, when historical demand data displays multiple peaks that form distinct clusters, it may be essential to conduct a thorough classification of the data prior to conducting uncertainty sets. In this case, the model presented in this study is expanded into a data-driven robust optimization model that incorporates multiple uncertainty sets. The appropriate classification and construction of uncertainty sets, along with the development of methodologies to handle more complex models, represent promising future research. Additionally, incorporating carbon emissions into the model to facilitate the attainment of sustainability objectives represents a promising avenue for further research. Moreover, we intend to explore the development of an efficient algorithm for addressing large-scale robust service network design problems.

CRediT authorship contribution statement

Lingpu Zhang: Conceptualization, Methodology, Software, Formal analysis, Writing – original draft, Writing – review & editing; **Min Huang:** Methodology, Writing – review & editing, Supervision, Project administration, Funding acquisition; **Yuxin Zhang:** Software, Formal analysis, Writing – review & editing; **Ziteng Wang:** Formal analysis, Writing – review & editing; **Miao Li:** Formal analysis, Writing – review & editing; **Xingwei Wang:** Formal analysis, Writing – review & editing.

Data availability

Data will be made available on request.

Declaration of competing interest

None.

Acknowledgments

This work is supported by the NSFC Key Supported Project of the Major Research Plan Grant No. 92567302; the NSFC Grant No. U25A20431 62432003; and the Liaoning Revitalizing Talent Program No. XLYC2202045.

References

- Alexandre, V., Alexandre, S., David, P., Jos'e M, A., & C, N. G. (2020). Two-stage robust unit commitment for co-optimized electricity markets: An adaptive data-driven approach for scenario-based uncertainty sets. *IEEE Transactions on Sustainable Energy*, 11(2), 958–969.
- An, Y., & Zeng, B. (2015). Exploring the modeling capacity of two-stage robust optimization: Variants of robust unit commitment model. *IEEE Transactions on Power Systems*, 30(1), 109–122.
- Andersen, J., Christiansen, M., & Crainic, T. G. (2011). Branch and price for service network design with asset management constraints. *Transportation Science*, 45(1), 33–49.
- Andersen, J., Crainic, T. G., & Christiansen, M. (2009a). Service network design with asset management: Formulations and comparative analyses. *Transportation Research Part C: Emerging Technology*, 17(2), 197–207.
- Andersen, J., Crainic, T. G., & Christiansen, M. (2009b). Service network design with management and coordination of multiple fleets. *European Journal of Operational Research*, 193, 377–389.
- Bai, R., Wallace, W. S., & Li, J. (2014). Stochastic service network design with rerouting. *Transportation Research Part B: Methodological*, 60(1), 50–65.
- Baron, O., Berman, O., Fazel-Zarandi, M. M., & Roshanaei, V. (2019). Almost robust discrete optimization. *European Journal of Operational Research*, 276(2), 451–465.
- Ben-Tal, A., & Goryashko, A. (2004). Adjustable robust solutions of uncertain linear programs. *Mathematical Programming*, 99(2), 351–376.
- Bertsimas, D., Gupta, V., & Kallus, N. (2018). Data-driven robust optimization. *Mathematical Programming*, 167(2), 235–292.

- Bertsimas, D., & Sim, M. (2004). The price of robustness. *Operations Research*, 52(1), 35–53.
- Boland, N., Hewitt, M., Marshall, L., & Savelsbergh, M. (2017). The continuous-time service network design problem. *Operations Research*, 65(5), 1303–1321.
- Cortes, C., & Vapnik, V. (1995). Support-vector networks. *Machine Learning*, 20, 273–297.
- Crainic, T. G. (2000). Service network design in freight transportation. *European Journal of Operational Research*, 122(2), 272–288.
- Crainic, T. G., & Rei, W. (2025). 50 years of operations research for the tactical planning of consolidation-based freight transportation. *EURO Journal on Transportation and Logistics*, 14, 100157.
- Crainic, T. G., & Rousseau, J. M. (1986). Multicommodity, multimode freight transportation: A general modeling and algorithmic framework for the service network design problem. *Transportation Research Part B: Methodological*, 20(3), 225–242.
- Ghamlouche, I., Crainic, T. G., & Gendreau, M. (2003). Cycle-based neighbourhoods for fixed-charge capacitated multicommodity network design. *Operations Research*, 51(4), 655–667.
- Ghamlouche, I., Crainic, T. G., & Gendreau, M. (2004). Path relinking, cycle-based neighbourhoods and capacitated multicommodity network design. *Operations Research*, 131(1), 109–133.
- Hewitt, M. (2022). The flexible scheduled service network design problem. *Transportation Science*, 56(4), 1000–1021.
- Hewitt, M., Crainic, T. G., Nowak, M., & Rei, W. (2019). Scheduled service network design with resource acquisition and management under uncertainty. *Transportation Research Part B: Methodological*, 128, 324–343.
- Kris, B., An, C., & Gerrit, K. (2013). Optimal shipping routes and vessel size for intermodal barge transport with empty container repositioning. *Computers in Industry*, 64(2), 155–164.
- Liu, C. J., Zhang, J. L., Lin, S., & Shen, Z.-J. M. (2023). Service network design with consistent multiple trips. *Transportation Research Part E: Logistics and Transportation Review*, 171, 103028.
- Lium, A.-G., Crainic, T. G., & Wallace, S. W. (2009). A study of demand stochasticity in service network design. *Transportation Science*, 43(2), 144–157.
- Lo, H. K., An, K., & Lin, W. (2013). Ferry service network design under demand uncertainty. *Transportation Research Part E: Logistics and Transportation Review*, 59, 48–70.
- Long, J., Zhu, J., Zhai, J., Xu, T., Liang, C., & Zhao, L. (2025). Data-driven robust operating optimization of energy-material coupled system in refineries under uncertainty. *Expert Systems with Applications*, 267, 126184.
- Marshall, L., Boland, N., Savelsbergh, M., & Hewitt, M. (2020). Interval-based dynamic discretization discovery for solving the continuous-time service network design problem. *Transportation Science*, 55(1), 29–51.
- Mirzaei, E., Nikbakhsh, E., & Kashan, H. A. (2025). Accelerated benders decomposition algorithm for the reliable service network design problem under disruption. *Computers & Industrial Engineering*, 207, 111234.
- Mohsen, R. G., Donya, R., & Mohammad, M. (2024). Data-driven robust optimization for a multi-trip truck-drone routing problem. *Expert Systems with Applications*, 241, 122485.
- Pedersen, M. B., Crainic, T. G., & Madsen, O. B. G. (2009). Models and tabu search meta-heuristics for service network design with asset balance requirements. *Transportation Science*, 43(2), 158–177.
- Ruchika, S., & Patten, J. (2023). Data-driven robust portfolio optimization with semi mean absolute deviation via support vector clustering. *Expert Systems with Applications*, 224, 120000.
- Sarayloo, F., Crainic, T. G., & Rei, W. (2020). A learning-based matheuristic for stochastic multicommodity network design. *INFORMS Journal on Computing*, 33(2), 643–656.
- Scherr, Y. O., Hewitt, M., & Mattfeld, D. C. (2022). Stochastic service network design for a platooning service provider. *Transportation Research Part C: Emerging Technologies*, 144, 103912.
- Scherr, Y. O., Neumann-Saavedra, B. A., Hewitt, M., & Mattfeld, D. C. (2018). Service network design for same day delivery with mixed autonomous fleets. *Transportation Research Procedia*, 30, 23–32.
- Shang, C., Huang, X., & You, F. (2017). Data-driven robust optimization based on kernel learning. *Computers & Chemical Engineering*, 106, 404–479.
- Sun, Y., Qiu, R. Z., Shao, S. S., Sun, M. H., & Fan, Z. P. (2025). Omnichannel strategies and data-driven robust inventory policies with demand uncertainties. *Computers & Operations Research*, 173, 106830.
- Teypez, N., Schrenk, S., & Cung, V. D. (2010). A decomposition scheme for large-scale service network design with asset management. *Transportation Research Part E: Logistics and Transportation Review*, 46(1), 156–170.
- Truden, C., & Hewitt, M. (2024). The service network design problem with fleet and emissions management. *Transportation Research Part C: Emerging Technologies*, 166, 104769.
- Wang, Z. J., & Qi, M. Y. (2019). Service network design considering multiple types of services. *Transportation Research Part E: Logistics and Transportation Review*, 126, 1–14.
- Wang, Z. J., & Qi, M. Y. (2020). Robust service network design under demand uncertainty. *Transportation Science*, 54(3), 565–853 + C2.
- Wu, L., & Hifi, M. (2021). Data-driven robust optimization for the itinerary planning via large-scale GPS data. *Knowledge-Based Systems*, 231, 107437.
- Xiang, X., Fang, T., Liu, C. C., & Pei, Z. (2022). Robust service network design problem under uncertain demand. *Computers & Industrial Engineering*, 172(PA), 108615.
- Yaghini, M., Karimi, M., & Rahbar, M. (2015). A cutting-plane neighborhood structure for fixed-charge capacitated multicommodity network design problem. *INFORMS Journal on Computing*, 27, 48–58.
- Zeng, B., & Zhao, L. (2013). Solving two-stage robust optimization problems using a column-and-constraint generation method. *Operations Research Letters*, 41(5), 457–461.
- Zhang, Q. H., Huang, M., & Wang, H. H. (2024). Collaborative service network design for multiple logistics carriers considering demand uncertainty. *Symmetry*, 16 (8), 1083.