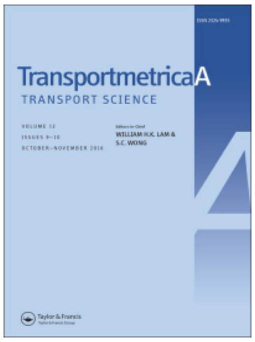

Transportmetrica A: Transport Science

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/ttra21

Distributionally robust optimisation model for multi-objective hub location problem via considering ambiguousness

Fanghao Yin, Yankui Liu & Yanju Chen

TRANSPORTMETRICA A: TRANSPORT SCIENCE
https://doi.org/10.1080/23249935.2022.2094494

Distributionally robust optimisation model for multi-objective hub location problem via considering ambiguousness

Fanghao Yin[a], Yankui Liu[b] and Yanju Chen[a]

[a]Risk Management and Financial Engineering Lab, College of Mathematics and Information Science, Hebei University, Baoding, People's Republic of China; [b]Hebei Key Laboratory of Machine Learning and Computational Intelligence, College of Mathematics and Information Science, Hebei University, Baoding, People's Republic of China

ABSTRACT
We formulate a new multi-objective hub location problem by considering economic, customer satisfaction and environmental objectives. The economic objective includes loss cost in addition to transportation cost and hub construction cost. The environmental objective covers noise pollution cost and carbon emissions trading cost. The customer satisfaction objective is innovatively defined as the sum of the transportation time satisfaction and transportation quality satisfaction. Transportation costs, noise levels and carbon emissions are assumed as uncertain parameters. In practice, the probability distributions of uncertain parameters are often ambiguous. To characterise this ambiguity, we first construct an ambiguity set to propose a distributionally robust multi-objective hub location model and derive its safe approximation under mean and dispersion ambiguity set. Then, from the perspective of the government, a goal programming model is established. Finally, we apply the proposed model to design China's super logistics hub network to verify the model's validity and better performance.

ARTICLE HISTORY
Received 14 December 2021
Accepted 11 June 2022

KEYWORDS
Multi-objective hub location problem; distributionally robust optimisation; ambiguity set; safe approximation; goal programming

1. Introduction

Hub location problems are the important network optimisation problems which can greatly save costs and resources, and they play an extremely important role in today's economic globalisation. The hub location problems are concerned with locating hub facilities and allocating non-hub nodes to hub facilities in order to route the traffic between origin-destination pairs (Alumur and Kara 2008). Regarding the hub capacity, there are two main types of hub location problems: uncapacitated hub location problem (e.g. Contreras, Cordeau, and Laporte 2011) and capacitated hub location problem (e.g. Correia, Nickel, and Saldanha-da-Gama 2018). They differ in whether the capacity of each hub is finite. In addition, single allocation hub location problem considers each non-hub node can send and

CONTACT Yanju Chen chen-yanju@163.com; yanjuchen@hbu.edu.cn Risk Management and Financial Engineering Lab, College of Mathematics and Information Science, Hebei University, Baoding 071002, Hebei, People's Republic of China

receive traffic through a single hub facility (e.g. Azizi and Salhi 2022), while multiple allocation hub location problem permits non-hub node to be allocated to more than one hub facilities (e.g. Ghaffarinasab 2018).

The seminal research in hub location problems began with the pioneering work of O'Kelly (1987). The original quadratic integer programming formulation with a non-convex objective function was formulated for the single allocation *p*-hub median problem by O'Kelly (1987). Subsequently, the integer programming formulations for four types of discrete hub location problems were presented. These are the *p*-hub median problem, the uncapacitated hub location problem, *p*-hub center problem and hub covering problem (see, Campbell 1994). Because of the known computational difficulty, Campbell (1996) proposed two new heuristic algorithms for the single allocation *p*-hub median problem. In the same time, another paper (Ernst and Krishnamoorthy 1996) presented efficient algorithms for the uncapacitated single allocation *p*-hub median problem. Some recent researches have addressed the hub location problems as well as their variants (e.g. Huang et al. 2018; Wang and Qin 2021; Korani and Eydi 2021; Kayışoğlu and Akgün 2021).

With respect to multi-objective nature of the hub location problems, some literatures have considered different objective functions (see Section 1.2). It was found that time and cost are major concerns in most of the literature for hub location problems. However, environmental factors in transportation networks' design have recently received more attention. We analyse the decision-making environment of hub location problem and find that economic cost and environmental factors can be regarded as objective functions under two different decision-making bodies of enterprises and government, respectively. Moreover, a complete hub location problem should include the decision-making environment, in which there exist the government, enterprises and consumers. To this end, it is necessary to construct a model that includes the three decision-making bodies. Based on the above analysis, this study models the hub location problem as a multi-objective formulation with economic, environmental and customer satisfaction objectives. Thereinto, we innovatively define the customer satisfaction objective including the transportation time satisfaction and transportation quality satisfaction.

In recent literatures, the hub location problems focus on location and network design decisions in an uncertain environment (see Section 1.2). There are three main methods to deal with uncertainty: stochastic optimisation, robust optimisation and fuzzy optimisation. Traditionally, these three methods do not consider the ambiguity in the model parameters' distribution information. However, the ambiguity of probability distribution in the uncertain hub location problem is getting more and more attention, which motivates us to have an in-depth discussion. The distributionally robust optimisation method can make full use of partial distribution information to provide the optimal decisions that can resist the ambiguity of distribution. Motivated by the above considerations, it is meaningful and feasible to introduce the distributionally robust optimisation method under the ambiguity set including mean and dispersion information in the hub location problem.

1.1. Main contributions

From a new viewpoint, we study the multi-objective hub location problem, which extends the hub location problem in previous literature by introducing three objective functions. In addition, for the multi-objective hub location problem, we introduce a new ambiguity set including mean and dispersion information, and propose a distributionally robust multi-objective hub location model. More specifically, the major contributions of this paper are as follows:

- We develop a new multi-objective hub location problem that includes economic, customer satisfaction and environmental objectives. In the economic objective, we first put forward the loss cost of accidents in the transportation, which makes the problem more realistic. In addition, we integrate carbon emissions and noise pollution as environmental objective. Importantly, we define a new customer satisfaction objective that includes both transportation time satisfaction and transportation quality satisfaction.
- We propose a distributionally robust hub location model, in which transportation costs, noise levels and carbon emissions are described as random variables with ambiguous distribution information. Moreover, we derive the safe approximation of the proposed distributionally robust optimisation model under mean and dispersion ambiguity set. After that, we build a goal programming model based on the safe approximation. The resulting framework is a single objective mixed integer second-order cone programming (MISOCP) model, which can be solved efficiently by commercial optimisation solver.
- We demonstrate the effectiveness of the proposed model by the case study about China's super logistics hub network. In order to illustrate the advantages of our new approach, we compare it with the classical robust optimisation method. The results show that our new approach not only is effective when the distributions of random parameters are ambiguous, but also reduces the cost by considering the probability distribution information of the random parameters.

The remainders of this paper are organised as follows. The next subsection reviews the relevant literature of hub location problems. In Section 2, we propose a new distributionally robust multi-objective hub location problem. Safe approximation under mean and dispersion ambiguity set is derived in Section 3. The goal programming model is presented for multi-objective hub location problem in Section 4. In Section 5, we report numerical results about a practical problem and obtain some interesting management implications. Section 6 gives the conclusions of the paper. The proofs of the main results are provided in the Appendix.

1.2. Related literature

Our research is related to two streams of literature about hub location problems: the first stream is multi-objective hub location problems; another related literature stream is uncertain hub location problems. Therefore, this subsection reviews these two streams in the literatures.

1.2.1. Multi-objective hub location problems

Multi-objective hub location problems have been widely studied in recent years. Its earlier theoretical study was found by da Graça Costa, Captivo, and Clímaco (2008), which only included a bi-objective model (cost and time). After that, Alumur and Kara (2008) predicted that the multi-objective hub location problems may be a research direction in the future. Ghodratnama, Tavakkoli-Moghaddam, and Azaron (2013) presented a bi-objective model including total costs (covering cost, transportation cost, opening cost and reopening cost of facilities, activating cost facilities in hubs, and vehicle using cost) and total times. In contrast, the total costs and maximum transportation time could also be used as two objective functions (see, Rahimi et al. 2016). Ghezavati and Hosseinifar (2018) regarded the value at risk and the costs of establishing a hub and material transportation as two objective functions.

On the 25th anniversary of the hub location problems, Campbell and O'Kelly (2012) pointed out that it would be a meaningful exploration to combine the environmental problems with the hub location problems. Since then, the research of Mohammadi, Torabi, and Tavakkoli-Moghaddam (2014) is the case in point. In addition to the total transportation cost objective, they also considered the total cost of noise pollution and the consumed energy of vehicles as the other two objectives. Similarly, Zhalechian et al. (2017) proposed a three-objective model including the noise pollution. Moreover, the carbon emissions were considered as objective function by Musavi and Bozorgi-Amiri (2017). Parsa et al. (2019) not only considered the transport cost, but also studied the greenhouse gas emissions, fuel consumption and noise in the design of airline hub-and-spoke network.

Note that our work on multi-objective research is different from the previous literatures about multi-objective hub location problems (see, Mohammadi, Torabi, and Tavakkoli-Moghaddam (2014, 2019) and Musavi and Bozorgi-Amiri (2017). Twofold differences made by this paper relative to Mohammadi, Torabi, and Tavakkoli-Moghaddam (2014, 2019) and Musavi and Bozorgi-Amiri (2017) are emphasised here. Firstly, we establish a new multi-objective optimisation model including economy, customer satisfaction and environment objectives. In our model, economic objective includes not only the transportation costs and hub construction costs, but also the loss cost due to some accidents, while Mohammadi, Torabi, and Tavakkoli-Moghaddam (2014, 2019) and Musavi and Bozorgi-Amiri (2017) only considered the transportation costs and hub construction costs. Carbon emissions and noise pollution are constructed in environmental objective. Secondly, we restructure a new customer satisfaction objective via including both transportation time satisfaction and transportation quality satisfaction, which is distinctly different from the problem proposed by Mohammadi, Torabi, and Tavakkoli-Moghaddam (2014, 2019) and Musavi and Bozorgi-Amiri (2017).

1.2.2. Uncertain hub location problems

The hub location problems in uncertain environment had become a hot spot and research focus. According to method classification of dealing with uncertainty, the literatures about uncertain hub location problems can be roughly divided into stochastic hub location problems, fuzzy hub location problems and robust hub location problems.

When the uncertain parameters are described as random variables and the exact probability distribution is known, many researchers applied stochastic optimisation method to model the uncertain hub location problems. Contreras, Cordeau, and Laporte (2011) assumed demands and transportation costs are random variables and designed a

Monte-Carlo simulation-based algorithm to solve their problems. In many cases, the uncertainty was expressed by a finite set of scenarios (see, Alumur, Nickel, and Saldanha-da-Gama 2012). Sadeghi et al. (2015) assumed the road capacity as random variable and regarded road capacity reliability as a probability in *p*-hub covering location problem. In addition, the two-stage stochastic programming model for real-world air network was established by some researchers (e.g. Ahmadi et al. 2015). Recently, Azizi, Vidyarthi, and Chauhan (2018) proposed a hub-and-spoke network model under random demand and congestion. Correia, Nickel, and Saldanha-da-Gama (2018) regarded demand as random variable and studied a multi-period stochastic capacitated multiple allocation hub location problem. Shang et al. (2021a) presented the expected value and the chance-constrained programming model for multi-modal hub location problem, and they derived the equivalent integer second-order cone programming form of the proposed model. Similarly, Hu et al. (2021) also studied a joint chance-constrained programming model under the premise of considering the balanced utilisation of hub capacities, and carried out related transformations.

The fuzzy optimisation method is often used to model the decision system in which the data exhibit subjective uncertainty, and it depends on accurate possibility distribution. In hub location problem, Chou (2010) proposed a fuzzy hub location selection model and applied it to hub locations in Southeastern Asia. Yang, Liu, and Yang (2013, 2014) studied a new risk aversion *p*-hub center problem with fuzzy travel times. Yang, Yang, and Gao (2017) developed a two-phase approach for fuzzy hub-and-spoke network design problem. Wang et al. (2018) considered the expected value criterion and the critical value criterion for the fuzzy hub-and-spoke based road-rail intermodal transportation network problem.

The robust optimisation method describes uncertainty by an uncertainty set (Ben-Tal, Ghaoui, and Nemirovski 2009). In recent years, the robust optimisation method has been developed for hub location problems. For instance, in the uncapacitated multiple allocation *p*-hub median problem, Talbi and Todosijević (2017) applied robust optimisation method to study the uncertainty of flows. In different setting, by using robust optimisation method, Yang and Yang (2017) studied the uncertainty of discount factor in the *p*-hub median problem. When demand was considered as an uncertain parameter, it may be depicted as a polyhedral uncertainty set (see, Meraklı and Yaman 2016, Ghaffarinasab 2018). Cheng et al. (2018) studied a two-stage robust optimisation approach for the reliable network design problem.

In these three methods, stochastic optimisation and fuzzy optimisation methods rely on accurate distribution information, while classical robust optimisation method does not require distribution information at all. However, in most practical cases, the distribution information is ambiguous, or partially known. In these cases, the distributionally robust optimisation method is an effective research tool. Recently, this method has also been applied to the hub location problems. For example, Yin et al. (2019) considered *p*-hub median problem by addressing the uncertain carbon emissions from the transportation and discussed two types of ambiguity sets including the probability distributions under the bounded perturbations with zero means and Gaussian perturbations with partial knowledge of expectations and variances. Under a similar ambiguity set, Shang et al. (2021b) investigated a cluster-based hierarchical hub location problem through a distributionally robust optimisation method. Wang, Chen, and Liu (2020) equivalently transformed the adaptive distributionally robust hub location problem into a non-adaptive classical robust model under the assumption of independence, and its second-stage routing decision followed the optimal static policy. Yin and Zhao (2021) considered the mean and conditional value-at-risk criteria for a hub interdiction median problem to implement a data-driven distributionally robust optimisation approach. Although the above literatures studied the distributionally robust hub location problems, it is obvious that they only considered a single objective function, and did not study the model under the ambiguity set including mean and dispersion. In order to make readers more clearly understand the contribution of this paper, we list the differences between this paper and its closely related literature in Table 1. Inspired by this, we propose a multi-objective hub location model that includes customer satisfaction objective, and analyse the characteristics of the constructed model under the ambiguity set including mean and dispersion information.

2. Distributionally robust multi-objective hub location problem

In this section, we focus on a multi-objective hub location problem including economic, customer satisfaction, and environmental objectives. On the basis of the existing research (Mohammadi, Torabi, and Tavakkoli-Moghaddam 2014; Zhalechian et al. 2017; Jiang et al. 2020), we ameliorate the hub location problem to make it more realistic by considering the loss cost of accidents, the transportation mode between hubs and customer satisfaction. Thereafter, constituting the ambiguity sets based on the pieces of distribution information, we establish a new distributionally robust multi-objective hub location model.

2.1. Problem description

The classical hub location problem aims to find the locations of p hubs and the allocations of non-hub nodes from the known nodes set N so that the total transportation cost or time is minimised. In this paper, we extend the classic p-hub location problem by considering the capacity level c of hubs and the transport mode m between hubs. The capacity of the hub is subdivided into several levels according to the size of the hub capacity. For example, it is divided into three levels (high, medium and low), in which the capacity level corresponding to high is the largest and the capacity level corresponding to low is the smallest. The set of capacity levels for hubs is denoted as $C, c \in C$. In the same consideration as Shang et al. (2021a), we consider multi-modal transportation between hubs. The set of transport modes between hubs is abbreviated as M. Different from previous studies, we consider three objective functions: economic goal, customer satisfaction goal and environmental goal. Therefore, our main research problem is to design the network structure with p hubs based on the known nodes set and relevant data, determine the capacity levels of hubs, and select the transport mode between hubs to minimise the economic, customer satisfaction and environmental objectives of the whole hub network. The simplified hub network structure is exhibited in Figure 1, where i, j denote non-hub nodes and k, l denote hub nodes.

In order to present the mathematical formulation for p-hub location problem, we give some main assumptions. The model in this paper depends on the following assumptions: (1) A non-hub node can only be assigned to a single hub; (2) The number of hubs is predefined as p; (3) All hubs are interconnected, and there is no connection between non-hub

Table 1. Related literature review.

Recent research	Number of objectives		Meaning of objectives			Optimisation method			
	Single objective	Multi-objective	Economic	Environmental	Customer satisfaction	Stochastic	Fuzzy	Robust	Distributionally robust
Mohammadi, Torabi, and Tavakkoli-Moghaddam (2014)		✓	✓	✓		✓	✓		
Musavi and Bozorgi-Amiri (2017)		✓	✓	✓					
Mohammadi, Jula, and Tavakkoli-Moghaddam (2019)		✓	✓			✓			
Shang et al. (2021a)	✓		✓			✓			
Talbi and Todosijević (2017), Ghaffarinasab (2018)	✓		✓					✓	
Wang et al. (2018)		✓	✓				✓		
Yin et al. (2019)	✓		✓	✓					✓
Wang, Chen, and Liu (2020), Shang et al. (2021b), Yin and Zhao (2021)	✓		✓						✓
This paper		✓	✓	✓	✓				✓

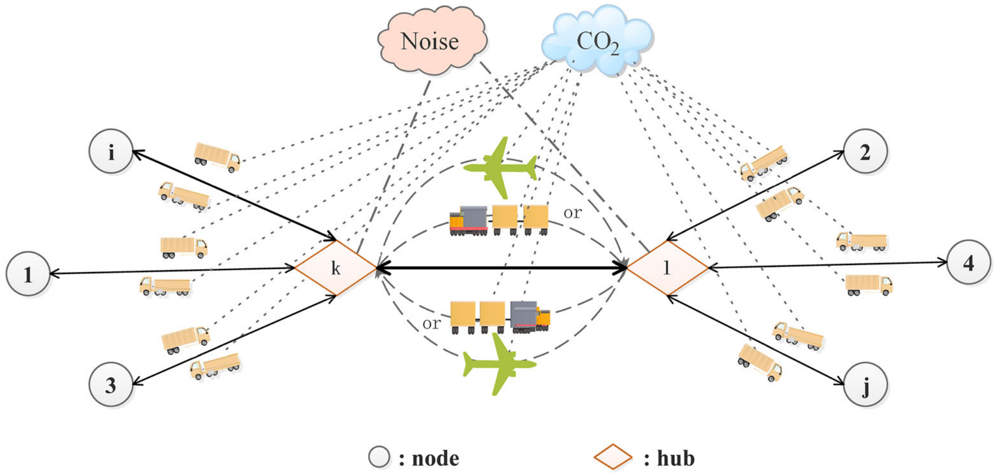

Figure 1. Simplified hub network structure.

nodes; (4) The capacity of the hub is restricted; (5) It is known that the probability of accidents on planes and trains is so small that they can be neglected. So we assume that the probability of accidents between hubs is zero. Based on the above assumptions, we describe the modelling process of our hub location problem and the final model in detail in the following contents.

2.2. Objective functions

For the practicality of hub location problem, we present a multi-objective optimisation problem involving economic, customer satisfaction and environmental objectives. In the economic objective, we take into account the occurrence of accidents at the transport, which is of practical significance. In the customer satisfaction objective, we define the novel transportation time satisfaction and transportation quality satisfaction. In the environmental objective, we not only consider the noise pollution of hub but also take into account the carbon emissions produced in transportation. The specific forms of the three objectives are elaborated below.

2.2.1. Economic objective

This objective minimises the total economic costs including total transport cost, accident loss cost and hub construction cost. Let's start with a few notations. $\eth_{ik}$,w_{ik} and δ_{ik} denote transportation distance, flow and unit transportation cost from non-hub node i to hub k, respectively. Therefore, the transportation cost from non-hub node i to hub k can be expressed as $\delta_{ik}w_{ik}\eth_{ik}$. As we all know, due to the sudden occurrence of bad weather, road congestion, traffic accidents and various natural disasters, there is loss cost in the process of transportation. In our design, the accident loss cost is calculated by increasing the proportion of transportation cost. Hence the actual transportation cost (including loss cost) from i to k is $(1 + p_{ik})\delta_{ik}w_{ik}\eth_{ik}$, where p_{ik} indicates loss ratio caused by accidents from non-hub node i to hub k. Because of the impact of scale economy, the transportation cost between hubs is reduced with the discount coefficient α_m. Therefore, the transportation cost between hubs is $\alpha_m \delta_{kl}^m w_{kl} \eth_{kl}$. In the case of the two-hub location problem ($i \to k \to l \to j$), the transportation cost from i to j is calculated as $((1+p_{ik})\delta_{ik} w_{ik} \eth_{ik} + \alpha_m \delta_{kl}^m w_{kl} \eth_{kl} + (1+p_{lj})\delta_{lj} w_{lj} \eth_{lj}) X_{iklj}^m$, in which X_{iklj}^m is 1 when transport mode m is used between hubs k and l, and the route (i,k,l,j) is a path, otherwise, X_{iklj}^m is 0. Moreover, F_k^c indicates fixed construction cost for hub k with capacity c. So we have the following economic objective:

$$\sum_{i,j,k,l \in N} \sum_{m \in M} [(1+p_{ik})\delta_{ik} w_{ik} \eth_{ik} + \alpha_m \delta_{kl}^m w_{kl} \eth_{kl} + (1+p_{lj})\delta_{lj} w_{lj} \eth_{lj}] X_{iklj}^m + \sum_{k \in N} \sum_{c \in C} F_k^c Y_{kk}^c, \tag{1}$$

in which $Y_{kk}^c = 1$ if node k is established as a hub with capacity level c, otherwise, $Y_{kk}^c = 0$.

What is noteworthiness is that δ exhibit the random characteristics based on historical data, hence they can be considered as random variables. In our model, we assume that $\delta = \bar{\delta} + z\hat{\delta}$, where $\bar{\delta}$ are nominal values, $\hat{\delta}$ are basic shifts, z are random variables. When the precise probability distribution P of random variables z is known, we propose a value at risk model for economic objective as follows:

$$\begin{aligned} &\min\ \kappa \\ &\text{s.t.}\ \Pr_{z\sim P}\Bigg\{ \sum_{i,j,k,l\in N}\sum_{m\in M} [(1+p_{ik})(\bar{\delta}_{ik} + z_i^1 \hat{\delta}_{ik}) w_{ik} \eth_{ik} + \alpha_m (\bar{\delta}_{kl}^m + z_k^2 \hat{\delta}_{kl}^m) w_{kl} \eth_{kl} \\ &\qquad + (1+p_{lj})(\bar{\delta}_{lj} + z_l^3 \hat{\delta}_{lj}) w_{lj} \eth_{lj}] X_{iklj}^m + \sum_{k\in N}\sum_{c\in C} F_k^c Y_{kk}^c \le \kappa \Bigg\} \ge 1-\epsilon, \end{aligned} \tag{2}$$

where ϵ represents the tolerance level for probability constraint violations, κ represents the maximum budget economic cost. In most cases, the precise probability distributions of random variables are unknown, which are only characterised by their partial distribution information. The distribution P of random variables z is ambiguous, that is, P belongs to an ambiguity set $\mathcal{P}^1$. The corresponding chance constraint becomes an ambiguous chance constraint. Hence, the value at risk objective with ambiguous chance constraint can be represented as follows:

$$\begin{aligned} &\min\ \kappa \\ &\text{s.t.}\ \inf_{P\in\mathcal{P}^1} \Pr_{z\sim P}\Bigg\{ \sum_{i,j,k,l\in N}\sum_{m\in M} [(1+p_{ik})(\bar{\delta}_{ik} + z_i^1 \hat{\delta}_{ik}) w_{ik} \eth_{ik} + \alpha_m (\bar{\delta}_{kl}^m + z_k^2 \hat{\delta}_{kl}^m) w_{kl} \eth_{kl} \\ &\qquad + (1+p_{lj})(\bar{\delta}_{lj} + z_l^3 \hat{\delta}_{lj}) w_{lj} \eth_{lj}] X_{iklj}^m + \sum_{k\in N}\sum_{c\in C} F_k^c Y_{kk}^c \le \kappa \Bigg\} \ge 1-\epsilon. \end{aligned} \tag{3}$$

2.2.2. *Customer satisfaction objective*

This objective minimises the total customer satisfaction index including transportation time satisfaction and transportation quality satisfaction. In past research, the soft time windows can be divided into many types according to the penalties calculation method (see, Niknamfar and Niaki 2016). In general, the penalties are calculated for the outside of the limited time interval, both early and late. There is a type of soft time window that is more like a hard time window. It has a certain degree of allowance, and also has to calculate the corresponding penalties. Both the part above a certain upper bound and the part below the lower boundary are regarded as unacceptable parts, the penalties will be set to infinity. In this paper, according to the actual logistics and transport services, the semi-time window is presented, which has one side time window. In this semi-time window, the upper bound of transportation time from origin i to destination j is denoted as L_{ij}. If $t_{ij} > L_{ij}$, the penalty cost is infinite.

Moreover, in the hub location problem, $t_{ik} + t_{kl}^m + t_{lj}$ is the transportation time from node i to node j. Due to the influence of various uncontrollable factors, the actual transportation time may increase some delay time. The actual transportation time is $t_{ij} = (1 + p_{ik})t_{ik} + t_{kl}^m + (1 + p_{lj})t_{lj}$ by the same way of calculating the loss cost. Hence, according to semi-time window, we define the transportation time satisfaction of customer as

$$\frac{L_{ij} - ((1 + p_{ik})t_{ik} + t_{kl}^m + (1 + p_{lj})t_{lj})}{L_{ij}}, \tag{4}$$

where the molecule represents the difference between the maximum permissible transportation time and the actual transportation time. The meaning of the defined transportation time satisfaction (4) is the ratio of the time interval between the maximum permissible transportation time and the actual transportation time. The larger the ratio, the shorter the actual transportation time. That is, consumers receive the goods earlier. Obviously, consumers want to receive their goods as soon as possible, so the result is that the greater the transportation time satisfaction, the more favourable for customers. In addition, when $L_{ij} = ((1 + p_{ik})t_{ik} + t_{kl}^m + (1 + p_{lj})t_{lj})$, the actual transportation time is stuck on the maximum permissible of the transportation time. At this point, the transportation time satisfaction is 0, which expresses the impartial attitude of consumers, that is, consumers do not get good satisfaction, but the delivery is not delayed.

In addition, the customer requires not only the delivery time of the goods as fast as possible but also the loss of the goods as small as possible. Therefore, we define transportation quality satisfaction as $\frac{p_{ik}+p_{lj}}{2}$. The goal of customer is to obtain a smaller loss rate of goods. As a result, combining transportation time satisfaction and transportation quality satisfaction, we construct a new customer satisfaction objective function as

$$\max \sum_{i,j,k,l \in N} \sum_{m \in M} \left(\frac{L_{ij} - ((1 + p_{ik})t_{ik} + t_{kl}^m + (1 + p_{lj})t_{lj})}{L_{ij}} - \frac{p_{ik} + p_{lj}}{2} \right) X_{iklj}^m. \tag{5}$$

It is found by analysis that this objective function successfully achieves dimension elimination and unit unification through the fractional structure. And the first item and the second item in the parentheses of the generated objective function (5) are both unitless real numbers between 0 and 1.

2.2.3. Environmental objective

This objective minimises the total environmental costs including noise pollution cost and carbon emissions cost. More specifically, according to the model described by Mohammadi, Torabi, and Tavakkoli-Moghaddam (2014), the cost of noise pollution at hub k is presented as $\phi[\exp(\xi L_{10,k} - \xi L_{\max,k}) - 1]$, where $L_{10,k}$ is the total noise level calculated at a reference distance of 10 metres from the nearside carriage way edge at hub k (hourly); $L_{\max,k}$ is the threshold measured in dB(A) at hub k; ϕ, ξ are constant coefficients. For more information, we refer interested readers to Mohammadi, Torabi, and Tavakkoli-Moghaddam (2014). Beyond that, low-carbon logistics has also received a lot of attention recently (see, Jiang, Zhang, and Meng 2021). One of the low-carbon logistics policies is called the carbon cap-and-trade policy (see, Mohammed et al. 2017), which requires each firm to be assigned a free amount of initial carbon emissions. If a firm emits less amount of carbon than its prescribed carbon cap (the deficiency is recorded as E^-), then it can sell the unused amount of carbon emissions. On the other hand, if a firm emits more amount of carbon than its prescribed carbon cap (the excess is marked as E^+), it may purchase additional carbon emission credit. However, in the actual decision-making environment, decision-maker of the enterprise does not know whether to purchase or sell carbon emissions in advance. The proposed model is to make a decision about this and output the enterprise's trading volume E^+ or E^-. Hence, we define $\delta_e(E^+ - E^-)$ as the cost through carbon emissions trading in a single cycle of the entire network, where δ_e is unit carbon emission price. Therefore, environmental objective of hub network is expressed as

$$\sum_{k\in N}\sum_{c\in C}\phi[\exp(\xi L_{10,k} - \xi L_{\max,k}) - 1]Y^c_{kk} + \delta_e(E^+ - E^-). \tag{6}$$

Noteworthy, noise level $L_{10,k}, \forall k \in N$ are affected by many factors including the number of heavy vehicles and traffic speed. The speed of transportation and the number of heavy vehicles exhibit random characteristics based on historical data, so $L_{10,k}, \forall k \in N$ are considered as random variables in this paper. We assume that $L_{10,k} = \bar{L}_{10,k} + z^4_k \hat{L}_{10,k}, \forall k \in N$, where $\bar{L}_{10,k}$ is nominal value, $\hat{L}_{10,k}$ is basic shifts, z^4_k is random variable. For a given precise distribution P, the expectation of environmental objective can be represented as follows:

$$\min\ \mathbf{E}_{z\sim P}\left\{\sum_{k\in N}\sum_{c\in C}\phi[\exp(\xi(\bar{L}_{10,k} + z^4_k\hat{L}_{10,k}) - \xi L_{\max,k}) - 1]Y^c_{kk} + \delta_e(E^+ - E^-)\right\}. \tag{7}$$

However, the probability distribution of random variable z^4_k is usually ambiguous. In view of this, we introduce an ambiguity set $\mathcal{P}^2$, which contains a family of distributions with the same characteristics as the true probability distribution. Hence, the worst-case expected environmental objective with ambiguity set can be represented as follows:

$$\min \sup_{P\in\mathcal{P}^2} \mathbf{E}_{z\sim P}\left\{\sum_{k\in N}\sum_{c\in C}\phi[\exp(\xi(\bar{L}_{10,k} + z^4_k\hat{L}_{10,k}) - \xi L_{\max,k}) - 1]Y^c_{kk} + \delta_e(E^+ - E^-)\right\}. \tag{8}$$

Although E^+ and E^- exist at the same time and both of them are non-negative real numbers in our model, the optimal decision with either $E^+ = 0$ or $E^- = 0$ will be finally obtained through the constructed model.

2.3. Constraints

2.3.1. Basic constraints

Constraint (9) indicates the number of hubs. Constraint (10) requires that a non-hub node can only be connected to one hub. Constraint (11) indicates that as long as node k is selected as hub, non-hub nodes i can connect to it. Constraint (12) indicates that at most one capacity level can be selected for a hub node. Constraint (13) selects single hub pair (k,l) for each origin-destination pair (i, j) and a specific transportation mode m. Constraint (14) ensures that a route (i, k, l, j) employs one transportation mode at most. Constraint (15) obliges that if node k is located as hub, there is the only one transport mode between hubs

k and l. Constraint (16) is the binary constraint.

$$\sum_{c\in C}\sum_{k\in N} Y_{kk}^{c} = p. \tag{9}$$

$$\sum_{c\in C}\sum_{k\in N} Y_{ik}^{c} = 1, \forall i \in N. \tag{10}$$

$$Y_{ik}^{c} \le Y_{kk}^{c}, \forall i,k \in N, c \in C. \tag{11}$$

$$\sum_{c\in C} Y_{kk}^{c} \le 1, \forall k \in N. \tag{12}$$

$$\sum_{k,l\in N}\sum_{m\in M} X_{iklj}^{m} = 1, \forall i,j \in N. \tag{13}$$

$$\sum_{m\in M} X_{iklj}^{m} \le 1, \forall i,j,k,l \in N. \tag{14}$$

$$\sum_{l\in N}\sum_{m\in M} X_{iklj}^{m} \le \sum_{c\in C} Y_{kk}^{c}, \forall i,k,j \in N. \tag{15}$$

$$X_{iklj}^{m}, Y_{kk}^{c}, Y_{ik}^{c} \in \{0,1\}, \forall i,j,k,l \in N, m \in M, c \in C. \tag{16}$$

2.3.2. Capacity constraints

Constraint (17) ensures that the flow passing through hub k does not exceed the capacity of hub, where w_{ik} is the flow from non-hub node i to hub k, $\hbar_k^c$ is the capacity of node k as a hub with level c.

$$\sum_{c\in C}\sum_{i\in N} (w_{ik} + w_{ki}) Y_{ik}^{c} \le \sum_{c\in C} \hbar_k^c Y_{kk}^{c}, \forall k \in N. \tag{17}$$

Constraint (18) indicates that the actual transportation time cannot exceed L_{ij} for each origin-destination pair (i,j), where L_{ij} is the upper bound of transportation time from origin i to destination j.

$$\sum_{m\in M} [(1+p_{ik})t_{ik} + t_{kl}^{m} + (1+p_{lj})t_{lj}] X_{iklj}^{m} \le L_{ij}, \forall i,j,k,l \in N. \tag{18}$$

Considering the sustainability of logistics transport, carbon emissions from vehicles are taken into account during our modelling process. We use e_{ik} to represent the carbon emissions generated per unit distance between non-hub and hub. Discount coefficient of carbon emission cost between hubs is marked as β_m. Therefore, we can obtain that the carbon emissions produced in a complete route (i,k,l,j) is $e_{ik}\eth_{ik} + \beta_m e_{kl}^{m}\eth_{kl} + e_{lj}\eth_{lj}$. In view of the carbon policy, the difference between the total carbon emissions generated in the hub network and the given free initial carbon quota E^{cap} can be traded as a carbon credit. In the course of trading, the resulting capacity constraint of carbon emissions is obtained as follows:

$$\sum_{i,j,k,l\in N}\sum_{m\in M} (e_{ik}\eth_{ik} + \beta_m e_{kl}^{m}\eth_{kl} + e_{lj}\eth_{lj}) X_{iklj}^{m} + E^{-} \le E^{cap} + E^{+}.$$

According to historical data, the carbon emissions from transportation exhibit random characteristics. So e are considered as random variables in this paper. We assume that $e =$

$\bar{e} + z\hat{e}$, where $\bar{e}$ are nominal values, $\hat{e}$ are basic shifts, z are random variables. For tolerance level ϵ, the following chance constraint is obtained.

$$\Pr_{z\sim P}\left\{\sum_{i,j\in N}\sum_{k,l\in N}\sum_{m\in M}((\bar{e}_{ik}+z_i^5\hat{e}_{ik})\eth_{ik}+\beta_m(\bar{e}_{kl}^m+z_k^6\hat{e}_{kl}^m)\eth_{kl}+(\bar{e}_{lj}+z_l^7\hat{e}_{lj})\eth_{lj})X_{iklj}^m+E^-\right.$$
$$\left.\le E^{cap}+E^+\right\}\ge 1-\epsilon.$$

In particular, the probability distribution P of random variables z is ambiguous and belongs to ambiguity set $\mathcal{P}^3$. Hence, the following ambiguous chance constraint about carbon emissions is obtained.

$$\inf_{P\in\mathcal{P}^3}\Pr_{z\sim P}\{\sum_{i,j,k,l\in N}\sum_{m\in M}((\bar{e}_{ik}+z_i^5\hat{e}_{ik})\eth_{ik}+\beta_m(\bar{e}_{kl}^m+z_k^6\hat{e}_{kl}^m)\eth_{kl}+(\bar{e}_{lj}+z_l^7\hat{e}_{lj})\eth_{lj})X_{iklj}^m+E^-$$
$$\le E^{cap}+E^+\}\ge 1-\epsilon. \tag{19}$$

2.4. Distributionally robust model

Under the premise of the proposed objective functions and constraints, we obtain the following distributionally robust multi-objective hub location model.

$$\begin{cases}
\min & \kappa\\
\max & \sum_{i,j,k,l\in N}\sum_{m\in M}\left(\dfrac{L_{ij}-((1+p_{ik})t_{ik}+t_{kl}^m+(1+p_{lj})t_{lj})}{L_{ij}}-\dfrac{p_{ik}+p_{lj}}{2}\right)X_{iklj}^m\\
\min & \sup_{P\in\mathcal{P}^2}\mathbf{E}_{z\sim P}\left\{\sum_{k\in N}\sum_{c\in C}\phi[\exp(\xi(\bar{L}_{10,k}+z_k^4\hat{L}_{10,k})-\xi L_{\max,k})-1]Y_{kk}^c+\delta_e(E^+-E^-)\right\}\\
\text{s.t.} & \inf_{P\in\mathcal{P}^1}\Pr_{z\sim P}\left\{\sum_{i,j,k,l\in N}\sum_{m\in M}[(1+p_{ik})(\bar{\delta}_{ik}+z_i^1\hat{\delta}_{ik})w_{ik}\eth_{ik}+\alpha_m(\bar{\delta}_{kl}^m+z_k^2\hat{\delta}_{kl}^m)w_{kl}\eth_{kl}\right.\\
& \left.+(1+p_{lj})(\bar{\delta}_{lj}+z_l^3\hat{\delta}_{lj})w_{lj}\eth_{lj}]X_{iklj}^m+\sum_{k\in N}\sum_{c\in C}F_k^cY_{kk}^c\le\kappa\right\}\ge 1-\epsilon\\
& \text{and constraints (9)}-(19).
\end{cases} \tag{20}$$

For the above formulated distributionally robust multi-objective hub location model, a problem that arises in practice is the need to commit to a distribution family $\mathcal{P}$ with only partial information about the probability distribution. Therefore, the proposed model is unfortunately a semi-infinite programming and it results in a severely computationally intractable character. A more challenging difficulty is dealing with the expectation of environmental objective (8) and ambiguous chance constraints (3) and (19). Hence, in an effort to address these issues, we discuss how to deal with them in the next section, so as to transform the distributionally robust multi-objective hub location model into computationally solvable form.

3. Safe approximation under mean and dispersion information

The distribution P is ambiguous in many practical problems. Therefore, when we specify partial distributional information of the model's uncertain parameters, an ambiguity set $\mathcal{P}$

of distribution is effective, which contains the true nominal distribution P. As we are discovering, the structure of $\mathcal{P}$ completely determines the difficulty degree of solving this problem. In this section, we assume that ambiguity set $\mathcal{P}$ contains only mean and dispersion information (μ, d), where d is the mean absolute deviations from the means μ. For more information about (μ, d), we refer interested readers to Postek et al. (2018). Postek et al. (2018) showed that for the ambiguity set $\mathcal{P}_{\mu,d}$, no equivalent closed-form result exists. Hence, in the following sections, we will show how to compute an upper bound of expected environmental objective (20) and seek the safe approximations of ambiguous chance constraints (3) and (19). The mentioned 'safe' in this paper means that the optimal solution of the derived model is still in the feasible domain of the original model. So the derived model is the safe approximation model of the original model. The reason is that the feasible domain of the derived approximation model is a subset of the feasible domain of the original problem, that is to say, the optimal solution of the approximation model must be feasible for the original model.

3.1. The upper bound of expected value

In the expected environmental objective, we observe that only z_k^4 is random variable and its distribution P belongs to an ambiguity set $\mathcal{P}^2$. For simplicity, the expected environmental objective can be rewritten as follows:

$$\min \sup_{P\in\mathcal{P}^2} \sum_{k\in N}\sum_{c\in C} \phi[\exp(\xi\,\bar{L}_{10,k} - \xi L_{\max,k})\mathbf{E}_{z\sim P}[\exp(\xi z_k^4\,\hat{L}_{10,k})] - 1]Y_{kk}^c + \delta_e(E^+ - E^-). \quad (21)$$

To address the above objective, a challenging difficulty is dealing with $\mathbf{E}_{z\sim P}\exp(\xi z_k^4\,\hat{L}_{10,k})$, $\forall P$
$\in \mathcal{P}^2$. Under the mean and dispersion information, as Postek et al. (2018), we also assume that z_k^4 has a support contained in $[-1, 1]$ and mean 0, the ambiguity set $\mathcal{P}_{\mu,d}^2$ is defined as follows:

$$\mathcal{P}_{\mu,d}^2 = \{P : \mathrm{supp}(z_k^4) \subseteq [-1, 1], E_P[z_k^4] = 0, E_P[|z_k^4|] = d_k^4, k \in N\}. \quad (22)$$

Observing that the ambiguity set $\mathcal{P}_{\mu,d}^2$ has only support, mean and dispersion constraints, due to the ambiguity of distribution, we cannot obtain an exact form of the worst-case expectation. To overcome this problem, in this paper, we provide a good upper bound of $\mathbf{E}_{z\sim P}\exp(\xi z_k^4\,\hat{L}_{10,k})$ as shown in the following theorem.

Theorem 3.1: *Let $z_k^4, k \in N$ be mutually independent random variables. Then, for the given parameters ξ and $\hat{L}_{10,k}$, we have*

$$\sup_{P\in\mathcal{P}_{\mu,d}^2} \mathbf{E}_{z\sim P}\exp(\xi z_k^4\,\hat{L}_{10,k}) \le \frac{d_k^4}{2}\exp(-\xi\,\hat{L}_{10,k}) + \frac{d_k^4}{2}\exp(\xi\,\hat{L}_{10,k}) + (1 - d_k^4)\exp(0)$$
$$= d_k^4\cosh(\xi\,\hat{L}_{10,k}) + 1 - d_k^4, \quad (23)$$

where $\cosh(\cdot)$ *is hyperbolic cosine function. That is,* $d_k^4\cosh(\xi\,\hat{L}_{10,k}) + 1 - d_k^4$ *is an upper bound of* $\mathbf{E}_{z\sim P}\exp(\xi z_k^4\,\hat{L}_{10,k})$.

Based on the derivation of the above theorem, we get the upper bound of the expected value as a safe approximation of the original expectation. So the worst-case expected environmental objective is converted into the following form:

$$\min \sum_{k\in N}\sum_{c\in C} \phi[\exp(\xi\,\bar{L}_{10,k} - \xi L_{\max,k})(d_k^4\cosh(\xi\,\hat{L}_{10,k}) + 1 - d_k^4) - 1]Y_{kk}^c + \delta_e(E^+ - E^-). \tag{24}$$

3.2. Safe approximation of ambiguous chance constraints

The distributionally robust hub location model is a semi-infinite programming since there are ambiguous chance constraints, which usually cannot be solved directly. Meanwhile, we can observe that ambiguous chance constraints (3) and (19) have the same structure, the difference between them is that the ambiguity sets are different. In order to express more intuitively, we first study the safe approximation of the ambiguous chance constraint (3) under mean and dispersion information. The results of (19) can be obtained similarly. We rewrite (3) as the following ambiguous chance constraint:

$$\begin{aligned}\inf_{P\in\mathcal{P}^1} \Pr_{z\sim P}\Bigg\{&\sum_{i,j,k,l\in N}\sum_{m\in M}[(1+p_{ik})\,\bar{\delta}_{ik}\,w_{ik}\eth_{ik} + \alpha_m\,\bar{\delta}_{kl}^m\,w_{kl}\eth_{kl} + (1+p_{lj})\,\bar{\delta}_{lj}\,w_{lj}\eth_{lj}]X_{iklj}^m \\ &+ \sum_{i\in N} z_i^1 \sum_{j,k,l\in N}\sum_{m\in M}(1+p_{ik})\,\hat{\delta}_{ik}\,w_{ik}\eth_{ik}X_{iklj}^m + \sum_{k\in N} z_k^2 \sum_{i,j,l\in N}\sum_{m\in M}\alpha_m\,\hat{\delta}_{kl}^m\,w_{kl}\eth_{kl}X_{iklj}^m \\ &+ \sum_{l\in N} z_l^3 \sum_{i,j,k\in N}\sum_{m\in M}(1+p_{lj})\,\hat{\delta}_{lj}\,w_{lj}\eth_{lj}X_{iklj}^m + \sum_{k\in N}\sum_{c\in C}F_k^cY_{kk}^c \le \kappa\Bigg\} \ge 1-\epsilon.\end{aligned} \tag{25}$$

For brevity, adding the additional variable y_0, y_i^1, y_k^2 and y_l^3 to the chance constraint, let

$$\begin{aligned} y_0 &= \sum_{i,j,k,l\in N}\sum_{m\in M}[(1+p_{ik})\bar{\delta}_{ik}w_{ik}\eth_{ik} + \alpha_m\bar{\delta}_{kl}^m w_{kl}\eth_{kl} \\ &\quad + (1+p_{lj})\bar{\delta}_{lj}w_{lj}\eth_{lj}]X_{iklj}^m + \sum_{k\in N}\sum_{c\in C}F_{kk}^cY_k^c - \kappa, \\ y_i^1 &= \sum_{j,k,l\in N}\sum_{m\in M}(1+p_{ik})\,\hat{\delta}_{ik}\,w_{ik}\eth_{ik}X_{iklj}^m, \\ y_k^2 &= \sum_{i,j,l\in N}\sum_{m\in M}\alpha_m\,\hat{\delta}_{kl}^m\,w_{kl}\eth_{kl}X_{iklj}^m, \\ y_l^3 &= \sum_{i,j,k\in N}\sum_{m\in M}(1+p_{lj})\,\hat{\delta}_{lj}\,w_{lj}\eth_{lj}X_{iklj}^m. \end{aligned} \tag{26}$$

Then, by the above substitution, the ambiguous chance constraint (3) is converted into the following form

$$\inf_{P\in\mathcal{P}^1} \Pr_{z\sim P}\{y_0 + \sum_{i\in N} z_i^1y_i^1 + \sum_{k\in N} z_k^2y_k^2 + \sum_{l\in N} z_l^3y_l^3 \le 0\} \ge 1-\epsilon. \tag{27}$$

Now we deduce the safe approximation of the ambiguous chance constraint (27). Under the mean and dispersion information, we assume that the components z_i^1, z_k^2, z_l^3 have support contained in [−1, 1], the ambiguity set $\mathcal{P}^1_{\mu,d}$ is defined as follows:

$$
\begin{aligned}
\mathcal{P}^1_{\mu,d} = \{&P : \mathrm{supp}(z_i^1) \subseteq [-1,1], \mathrm{supp}(z_k^2) \subseteq [-1,1], \mathrm{supp}(z_l^3) \subseteq [-1,1], \\
&E_P z_i^1 = 0, E_P z_k^2 = 0, E_P z_l^3 = 0, \\
&E_P|z_i^1| = d_i^1, E_P|z_k^2| = d_k^2, E_P|z_l^3| = d_l^3, i,k,l \in N\}.
\end{aligned} \tag{28}
$$

Due to the above ambiguity set, the equivalent forms of (27) cannot be obtained, thus, we provide its safe approximation.

Theorem 3.2: *Given $y_0, y_i^1, y_k^2, y_l^3, P \in \mathcal{P}^1_{\mu,d}$, if there exist auxiliary variables γ, τ that satisfy the second-order cone constraint system*

$$
\begin{cases}
y_0 = \gamma_0 + \tau_0, \\
y_i^1 = \gamma_i^1 + \tau_i^1, \forall i \in N, \\
y_k^2 = \gamma_k^2 + \tau_k^2, \forall k \in N, \\
y_l^3 = \gamma_l^3 + \tau_l^3, \forall l \in N, \\
\gamma_0 + \sum\limits_{i \in N} |\gamma_i^1| + \sum\limits_{k \in N} |\gamma_k^2| + \sum\limits_{l \in N} |\gamma_l^3| \le 0, \\
2\ln(1/\epsilon)\left(\sum\limits_{i \in N} (\sigma_i^1 \tau_i^1)^2 + \sum\limits_{k \in N} (\sigma_k^2 \tau_k^2)^2 + \sum\limits_{l \in N} (\sigma_l^3 \tau_l^3)^2\right) \le \tau_0^2,
\end{cases} \tag{29}
$$

where

$$
\begin{aligned}
\sigma_i^1 &= \sup_{t \in R} \sqrt{\frac{2\ln(d_i^1 \cosh(t) + 1 - d_i^1)}{t^2}}, \\
\sigma_k^2 &= \sup_{t \in R} \sqrt{\frac{2\ln(d_k^2 \cosh(t) + 1 - d_k^2)}{t^2}}, \\
\sigma_l^3 &= \sup_{t \in R} \sqrt{\frac{2\ln(d_l^3 \cosh(t) + 1 - d_l^3)}{t^2}},
\end{aligned} \tag{30}
$$

then the constraint system (29) *is a safe approximation of* (27), *and all feasible solution in the constraint system is feasible in* (27).

In combination with the conclusion of Theorem 3.2, if y_0, y_i^1, y_k^2, y_l^3 are substituted, we can obtain the safe approximation of (3) as follows:

$$\begin{cases} \sum_{i,j,k,l \in N} \sum_{m \in M} [(1+p_{ik})\, \bar{\delta}_{ik}\, w_{ik} \eth_{ik} + \alpha_m\, \bar{\delta}_{kl}^m\, w_{kl} \eth_{kl} + (1+p_{lj})\, \bar{\delta}_{lj}\, w_{lj} \eth_{lj}] X_{iklj}^m \\ + \sum_{k \in N} \sum_{c \in C} F_k^c Y_{kk}^c - \kappa = \gamma_0 + \tau_0, \\ \sum_{j,k,l \in N} \sum_{m \in M} (1+p_{ik})\, \hat{\delta}_{ik}\, w_{ik} \eth_{ik} X_{iklj}^m = \gamma_i^1 + \tau_i^1, \forall i \in N, \\ \sum_{i,j,l \in N} \sum_{m \in M} \alpha_m\, \hat{\delta}_{kl}^m\, w_{kl} \eth_{kl} X_{iklj}^m = \gamma_k^2 + \tau_k^2, \forall k \in N, \\ \sum_{i,j,k \in N} \sum_{m \in M} (1+p_{lj})\, \hat{\delta}_{lj}\, w_{lj} \eth_{lj} X_{iklj}^m = \gamma_l^3 + \tau_l^3, \forall l \in N, \\ \gamma_0 + \sum_{i \in N} |\gamma_i^1| + \sum_{k \in N} |\gamma_k^2| + \sum_{l \in N} |\gamma_l^3| \le 0, \\ 2\ln(1/\epsilon) \left(\sum_{i \in N} (\sigma_i^1 \tau_i^1)^2 + \sum_{k \in N} (\sigma_k^2 \tau_k^2)^2 + \sum_{l \in N} (\sigma_l^3 \tau_l^3)^2 \right) \le \tau_0^2, \end{cases} \tag{31}$$

in which $\sigma_i^1, \sigma_k^2, \sigma_l^3$ are given by formula (30).

Similarly, we derive the safe approximation of the ambiguous chance constraint (19). Under the mean ($E_P z_i^5 = E_P z_k^6 = E_P z_l^7 = 0$) and dispersion information ($E_P|z_i^5| = d_i^5, E_P|z_k^6| = d_k^6, E_P|z_l^7| = d_l^7$), the safe approximation result of (19) is

$$\begin{cases} \sum_{i,j,k,l \in N} \sum_{m \in M} (\bar{e}_{ik}\, \eth_{ik} + \beta_m\, \bar{e}_{kl}^m\, \eth_{kl} + \bar{e}_{lj}\, \eth_{lj}) X_{iklj}^m + E^- - E^{cap} - E^+ = \eta_0 + \theta_0, \\ \sum_{j,k,l \in N} \sum_{m \in M} \hat{e}_{ik}\, \eth_{ik} X_{iklj}^m = \eta_i^5 + \theta_i^5, \forall i \in N, \\ \sum_{i,j,l \in N} \sum_{m \in M} \hat{e}_{kl}^m\, \beta_m \eth_{kl} X_{iklj}^m = \eta_k^6 + \theta_k^6, \forall k \in N, \\ \sum_{i,j,k \in N} \sum_{m \in M} \hat{e}_{lj}\, \eth_{lj} X_{iklj}^m = \eta_l^7 + \theta_l^7, \forall l \in N, \\ \eta_0 + \sum_{i \in N} |\eta_i^5| + \sum_{k \in N} |\eta_k^6| + \sum_{l \in N} |\eta_l^7| \le 0, \\ 2\ln(1/\epsilon) \left(\sum_{i \in N} (\varsigma_i^5 \theta_i^5)^2 + \sum_{k \in N} (\varsigma_k^6 \theta_k^6)^2 + \sum_{l \in N} (\varsigma_l^7 \theta_l^7)^2 \right) \le \theta_0^2, \end{cases} \tag{32}$$

where η, θ are auxiliary variables, and

$$\varsigma_i^5 = \sup_{t \in R} \sqrt{\frac{2\ln(d_i^5 \cosh(t) + 1 - d_i^5)}{t^2}},$$

$$\varsigma_k^6 = \sup_{t \in R} \sqrt{\frac{2\ln(d_k^6 \cosh(t) + 1 - d_k^6)}{t^2}},$$

$$\varsigma_l^7 = \sup_{t \in R} \sqrt{\frac{2\ln(d_l^7 \cosh(t) + 1 - d_l^7)}{t^2}}. \tag{33}$$

In summary, we derive the safe approximations of ambiguous chance constraints (3) and (19). So the distributionally robust multi-objective hub location model is transformed into a computationally solvable form, which is a deterministic second-order cone programming model. However, this form is still a multi-objective optimisation problem. In the next section, to solve this model, by using the goal programming method, we transform it into a single objective optimisation model.

4. The goal programming formulation for multi-objective hub location problem

Depending on the specific ambiguity set with the mean and dispersion information in Section 3, the proposed model is approximately transformed into the following mixed integer second-order cone optimisation model

$$\begin{cases} \min & \kappa \\ \max & \sum_{i,j,k,l\in N}\sum_{m\in M}\left(\frac{L_{ij}-((1+p_{ik})t_{ik}+t_{kl}^{m}+(1+p_{lj})t_{lj})}{L_{ij}}-\frac{p_{ik}+p_{lj}}{2}\right)X_{iklj}^{m} \\ \min & \sum_{k\in N}\sum_{c\in C}\phi[\exp(\xi\,\bar{L}_{10,k}-\xi L_{\max,k})(d_k^4\cosh(\xi\,\hat{L}_{10,k})+1-d_k^4)-1]Y_{kk}^{c}+\delta_e(E^+-E^-) \\ \text{s.t.} & \text{constraints (9)}-(18),\ (24),\ (31)\text{ and }(32). \end{cases} \tag{34}$$

Based on the mixed integer second-order cone optimisation model (34) with three objective functions, we establish a goal programming model under resource constraints for multi-objective hub location problem. The concept of goal programming model originated from Charnes and Cooper (1961). After that, the goal programming is well-known by more and more researchers (Aouni and Kettani 2001) because it can be readily solved through general-purpose optimisation software.

We assume that the aspiration levels of the three objective functions are $\mathcal{F}_1, \mathcal{F}_2$ and $\mathcal{F}_3$, respectively. Meanwhile, the positive and negative deviations of economic objective, customer satisfaction objective and environmental objective are denoted as (π_1^+,π_1^-), (π_2^+,π_2^-) and (π_3^+,π_3^-), respectively. From the government's point of view, based on the idea of sustainable development, the environmental objective is the first priority. Then the customer satisfaction objective is the second priority. In order to ensure the normal operation of the company, as far as possible to reduce the cost, the economic objective is the third priority. Based on the above description, we establish the following goal programming model:

$$\begin{cases} \min & \mathbb{P}_1\pi_3^+ + \mathbb{P}_2\pi_2^- + \mathbb{P}_3\pi_1^+ \\ \text{s.t.} & \kappa+\pi_1^- - \pi_1^+ = \mathcal{F}_1 \\ & \sum_{i,j,k,l\in N}\sum_{m\in M}\left(\frac{L_{ij}-((1+p_{ik})t_{ik}+t_{kl}^{m}+(1+p_{lj})t_{lj})}{L_{ij}}-\frac{p_{ik}+p_{lj}}{2}\right)X_{iklj}^{m} \\ & \quad +\pi_2^- - \pi_2^+ = \mathcal{F}_2 \\ & \sum_{k\in N}\sum_{c\in C}\phi[\exp(\xi\,\bar{L}_{10,k}-\xi L_{\max,k})(d_k^4\cosh(\xi\,\hat{L}_{10,k})+1-d_k^4)-1]Y_{kk}^{c}+\delta_e(E^+-E^-) \\ & +\pi_3^- - \pi_3^+ = \mathcal{F}_3 \\ & \pi_1^+,\pi_1^-,\pi_2^+,\pi_2^-,\pi_3^+,\pi_3^- \geq 0 \\ & \text{constraints (9)}-(18),\ (24),\ (31)\text{ and }(32). \end{cases} \tag{35}$$

where $\mathbb{P}_1, \mathbb{P}_2, \mathbb{P}_3$ are weight coefficients, and $\mathbb{P}_1 \gg \mathbb{P}_2 \gg \mathbb{P}_3$.

According to the goal programming method, the multi-objective optimisation model is converted into a single-objective optimisation model by considering the priority order of the objective functions, so the model (35) is different from the original model (34). This transformation is beneficial for solvability of our model in practical settings. The proposed goal programming model (35) is still an MISOCP model, which can be solved directly by commercial optimisation software.

Remark 4.1: Note that the adopted goal programming method is not directly applied to the original multi-objective distributionally robust optimisation model, but is applied to the safe approximation model transformed in Section 3. The ambiguousness in the original model cannot be handled directly. Hence, our indirect processing method can effectively handle the ambiguousness and transform the multi-objective optimisation model into a single-objective optimisation model, which is readily solved by general commercial optimisation solver.

5. A case study about China's super logistics hub network

In this section, we apply the proposed model to the actual case study of China's super logistics hub network. In addition, the performance of the proposed model in practical cases is analysed to verify the model's validity.

In Section 5.1, the problem background of designing China's super logistics hub network and the test data required are described. The computational results are presented in Section 5.2. In order to present the advantages of our model, we do some contrastive study in Section 5.3. Finally, the management significances of the proposed method are summarised in Section 5.4.

5.1. Application case and test data

As we all know, Memphis can be called the 'U. S. hub', because it is not only the most important transportation hub in the United States, but also the world's largest air logistics hub. Analogously, for China, which city has this potential? Over the past few years, many central cities in China have been competing for the status of 'China Memphis', such as Xi'an, Wuhan, etc. Which city is the most likely to succeed? In this paper, we select 12 potential cities in China as shown in Figure 2, to abstract China's super logistics hub network design into a hub location problem. From the point of view of a hub location problem, we try to choose several cities that are most likely to become China's super logistics hub by the proposed distributionally robust hub location model.

The data about the 12 potential cities ($N = 12$) are from searching relevant websites. We obtain Euclidean distances $\mathfrak{d}$ between any two cities, which are measured in Google Maps. According to 'the traffic safety law of the PRC', the speed of trucks on expressways shall not exceed 100 km/h, and shall not be less than 60 km/h. Thus we assume that the speed of a truck is a random value between 60 and 100 km/h. Then the transportation time $\mathbf{t}$ between non-hub node and hub is equal to the quotient of the distance and the speed of truck. There are two transport modes between hubs ($M = \{1, 2\}$, aircraft and train), $m = 1$ for aircraft mode, $m = 2$ for train mode. When the transport mode between hubs is aircraft, we get the transportation time $\mathbf{t}^1$ from Google Maps. Since the speed of express cargo trains in China is

Figure 2. The locations of 12 potential cities in China.

120 km/h, we assume that the speed of ordinary freight trains is a random value between 80 and 120 km/h. Thus, when the transport mode between hubs is train, the transportation time $\mathbf{t}^2$ is equal to the quotient of the distance and the speed of train. We get traffic flow $\mathbf{w}$ between cities by seeking each city's 'Statistical Yearbook 2017' (Since this is a small hub network, the magnitude of traffic flow is reduced by 10^{-7} times.).

The capacity of the hub is divided into three levels: high level, medium level, and low level, i.e. $C = 3$. We assume that the capacity $\hbar$ of hub corresponding to the three levels is equal to the flow of the highway, railway, and air traffic in the 'Statistical Yearbook 2017', respectively (The order of magnitude of the flow is reduced by 10^{-5} times). We take the transaction price of Beijing carbon trading market on 3 September 2018 as the unit carbon emission price δ_e, which is 67.51CNY (http://www.tanpaifang.com/tanhangqing/). Based on 'National Road Freight Transport Price Index', $\overline{\delta}$ is a random value between 0.233 and 0.424 CNY (http://www.crtm.cn/ezine/15132.html), $\overline{\delta^1}$ is a random value between 2 and 4 CNY, and $\overline{\delta^2}$ is a random value between 0.08 and 0.14 CNY. In addition, the basic shift of uncertain parameters is equal to 5% of the nominal value. We assume that the mean absolute deviations $\mathbf{d}$ are random values between 0 and 0.5. The parameters ς and σ are calculated based on equations (30) and (33). The right side of each equation is an optimisation problem for univariate nonlinear function with respect to t. For a given value of parameter $\mathbf{d}$, with the help of Matlab software, we can draw the image of this function and calculate its supremum value. That is, the values of the parameters ς and σ are obtained. Since

Table 2. The remaining parameters values.

Parameter value	λ	$L_{\max,k}$ (dB)	E^{cap} (kg)	ϕ
	200	55	10,000	1
Parameter value	ξ	α_m	β_m	p_{ik}
	0.25	[0.2, 0.2]	[0.2, 0.2]	**U**(0, 0.04)
Parameter value	F_k^1 (million)	F_k^2 (million)	F_k^3 (million)	$L_{ij}(h)$
	U(0.45, 0.50)	**U**(0.40, 0.45)	**U**(0.35, 0.40)	**U**(25, 35)
Parameter value	$\bar{L}_{10,k}$ (dB)	$\bar{e}_{ik}$ (kg)	$\bar{e}_{kl}^1$ (kg)	$\bar{e}_{kl}^2$ (kg)
	U(75, 100)	**U**(0.3, 0.5)	**U**(0.7, 0.9)	**U**(0.5, 0.7)

Table 3. The optimal values of the proposed model.

The number of hubs	Positive deviation	Negative deviation	The optimal values (CNY)
$p = 1$	$\pi_1^+ = 841460$	$\pi_1^- = 0$	$Ec = 3,441,460$
	$\pi_2^+ = 0$	$\pi_2^- = 1.3717$	$Cs = 93.6283$
	$\pi_3^+ = 0$	$\pi_3^- = 1437.8$	$En = 6562.2$
$p = 2$	$\pi_1^+ = 278,330$	$\pi_1^- = 0$	$Ec = 2,878,330$
	$\pi_2^+ = 5.3465$	$\pi_2^- = 0$	$Cs = 100.3465$
	$\pi_3^+ = 0$	$\pi_3^- = 1143.2$	$En = 6856.8$
$p = 3$	$\pi_1^+ = 30,768$	$\pi_1^- = 0$	$Ec = 2,630,768$
	$\pi_2^+ = 8.1472$	$\pi_2^- = 0$	$Cs = 103.1472$
	$\pi_3^+ = 0$	$\pi_3^- = 1026.8$	$En = 6973.2$

$\mathbb{P}_1 \gg \mathbb{P}_2 \gg \mathbb{P}_3$, we assume that $\mathbb{P}_1 = 10^4, \mathbb{P}_2 = 10^0, \mathbb{P}_3 = 10^{-4}$ and $\epsilon = 0.02$. The values of the other parameters are shown in Table 2.

5.2. Case results

Considering this problem's size, we only study the cases when $p = 1, 2$ and 3, that is, only one, two or three nodes are chosen as hubs. Depending on the data obtained, the proposed model is solved by using the optimisation solver CPLEX 12.8. All computational results are solved on an Inter(R) Core(TM) i5-7200U 2.50 GHz personal computer with 12 GB RAM operating under Windows 10.

In order to be more reasonable, by referring to the optimal values of the corresponding single-objective models, the aspiration levels of three goals are $\mathcal{F}_1 = 2,600,000, \mathcal{F}_2 = 95$, and $\mathcal{F}_3 = 800$, respectively. Solving the proposed goal programming model, the optimal values are shown in Table 3.

In Table 3, *Ec*, *Cs* and *En* represent the optimal values of economic, customer satisfaction and environmental objectives, respectively. From Table 3 horizontally, when $p = 1$, we can find that the positive deviations of the customer satisfaction and environmental objectives are 0, which indicate that the environmental objective is achieved. That is, the customer satisfaction and environmental cost do not exceed the given goal value. By contrast, it is found that the economic objective exceeds the given goal value. When $p = 2$ and 3, we find that, unlike $p = 1$, the positive deviations relative to environmental objective is 0. This means that the environmental objective in the goal programming model has been achieved no more than the given goal value. But the economic objective is not achieved. This may be due to the fact that economic objective is regarded as the lowest in the previous target priorities. By longitudinal observation of Table 3, we observe that when the

Table 4. The optimal solutions of the proposed model.

The number of hubs	E^{cap}	E^{+}	E^{-}	The actual carbon emissions	The optimal hubs	The capacity level of hub
$p = 1$	14,000	0	42,493	56,493	11	Medium
$p = 2$	12,000	0	38,918	50,918	9,11	Low,Medium
$p = 3$	10,000	0	37,521	47,521	9,10,11	Low,Medium, Medium

value of p increases, the optimal value of environmental objective increases accordingly. The reason is that as the number of hubs increases, the corresponding noise pollution cost increases, which increases the environmental objective. As for the economic objective, with the increase of the number of hubs and the impact of economies of scale, the incurred costs decreases as shown in Table 3.

The optimal solutions obtained by our model are shown in Table 4. When the number of hubs is 3, the optimal hub nodes are 9, 10 and 11. The corresponding cities are Dalian, Xiamen and Ezhou. The optimal hub nodes are 9 and 11 if the number of hubs is 2. The optimal hub node is 11 if only one hub is needed. From this observation, it is inferred that Ezhou is most likely to become 'China Memphis' by using our model. Of course, this result, which is based on our established hub location model, is only a reference for the decision maker. Table 4 not only gives the locations of the hubs, but also gives the corresponding capacity levels of the hubs. The transport mode between hubs, not given in the table, obtained by the proposed model is air transportation. What's more, as we can see from the fifth column of the table, with the increase of the number of hubs, the actual amount of carbon emissions is reduced due to the impact of economies of scale.

In a hub network, the scale effect between hubs is an important factor for decision-makers to achieve their goals. In our proposed model, the scale effect is measured by the discount coefficient. There are two discount coefficients in our model, one is α_m, which can affect the saving proportion of transportation cost in economic objective, and the other is β_m, which can affect the saving proportion of inter-hub carbon emission in capacity constraint. We will study the influence of these two discount coefficients on the optimal results of the model, respectively. In general, the value of discount coefficient is between 0.2 and 0.8. Therefore, we take the discount coefficient as (0.2: 0.1: 0.8), which means that 0.2 is the initial value point, 0.8 is the final value point, and 0.1 is the step length. On this basis, for the value of α_m, 7×7 parameter values are obtained by the cross combination. Under different parameters, the proposed goal programming model is solved separately, and the optimal values of the three objective functions are obtained as shown in Figure 3.

From Figure 3, we can observe the change trend of the three objective values in α_m. In Figure 3(a), under the same α_2, it is obvious that the optimal value of economic objective increases with the increase of α_1. By contrast, the influence of α_2 is not obvious, but it also slightly affects the optimal value of economic objective. For example, as shown in Figure 3(a), when $\alpha_1 = 0.7$ and α_2 changes from 0.2 to 0.3, the optimal value of economic objective in the model increases from 2,674,455 to 2,674,747. Since α_m exists only in the mathematical expression of economic objective, it intuitively does not affect customer satisfaction objective and environmental objective. Even so, we find that α_m has an impact on customer satisfaction objective and environmental objective at some angles. This impact is shown in Figure 3(b) and (c). It is because there is

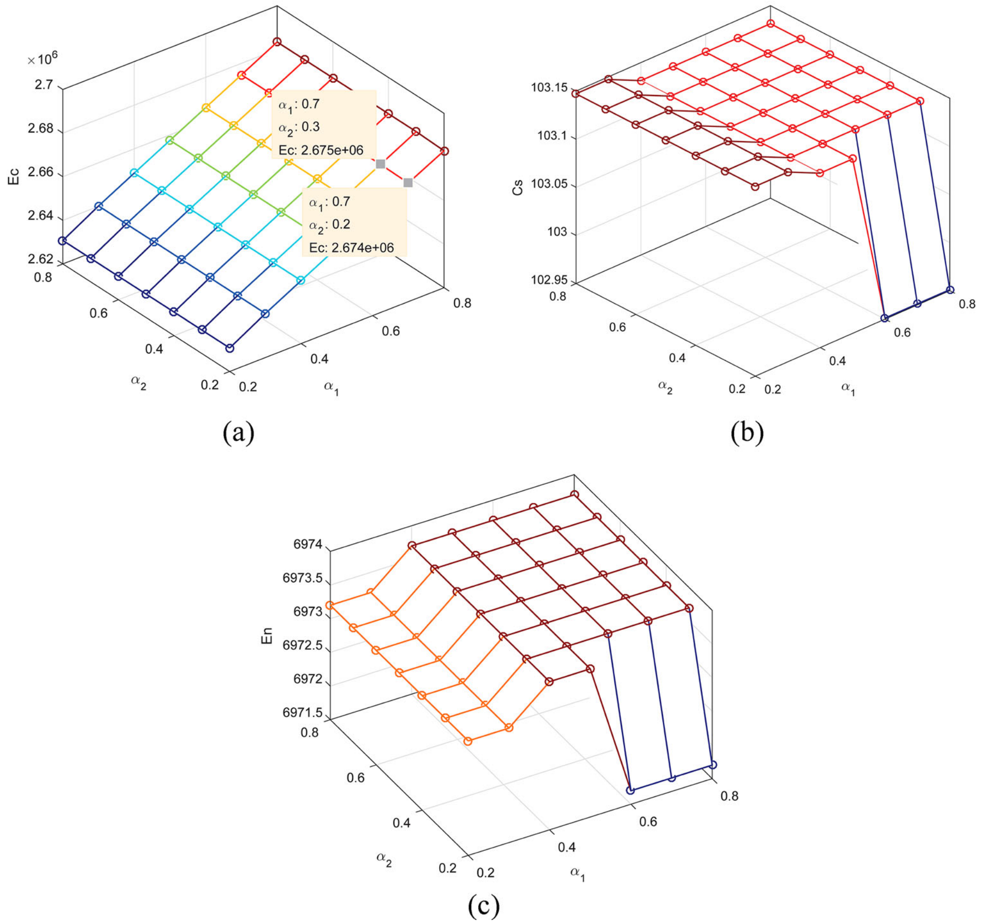

Figure 3. The optimal values corresponding to the value of α_m. (a) *Ec* corresponding to the value of α_m. (b) *Cs* corresponding to the value of α_m. (c) *En* corresponding to the value of α_m

a close relationship and conflict between the three objectives of our proposed model. α_m indirectly affects the customer satisfaction objective and environmental objective by changing the economic objective. Similar to the sensitivity analysis of α_m, the sensitivity study of discount coefficient β_m can also be obtained, which is omitted here for brevity.

5.3. Contrastive study with robust optimisation model

For the proposed DRO model (denoted as DRO-(μ, d)), when the uncertain parameters only have support contained in $[-1, 1]$, that is, the distribution information of uncertain parameters is unknown, the proposed model degenerates to a classical robust optimisation model. To illustrate the effectiveness of the proposed model, we compare the proposed model with the classical robust optimisation model. Under the support information, the classical robust optimisation model still needs to be transformed into a deterministic robust counterpart problem that can be solved via commercial solver. For details regarding the transformation

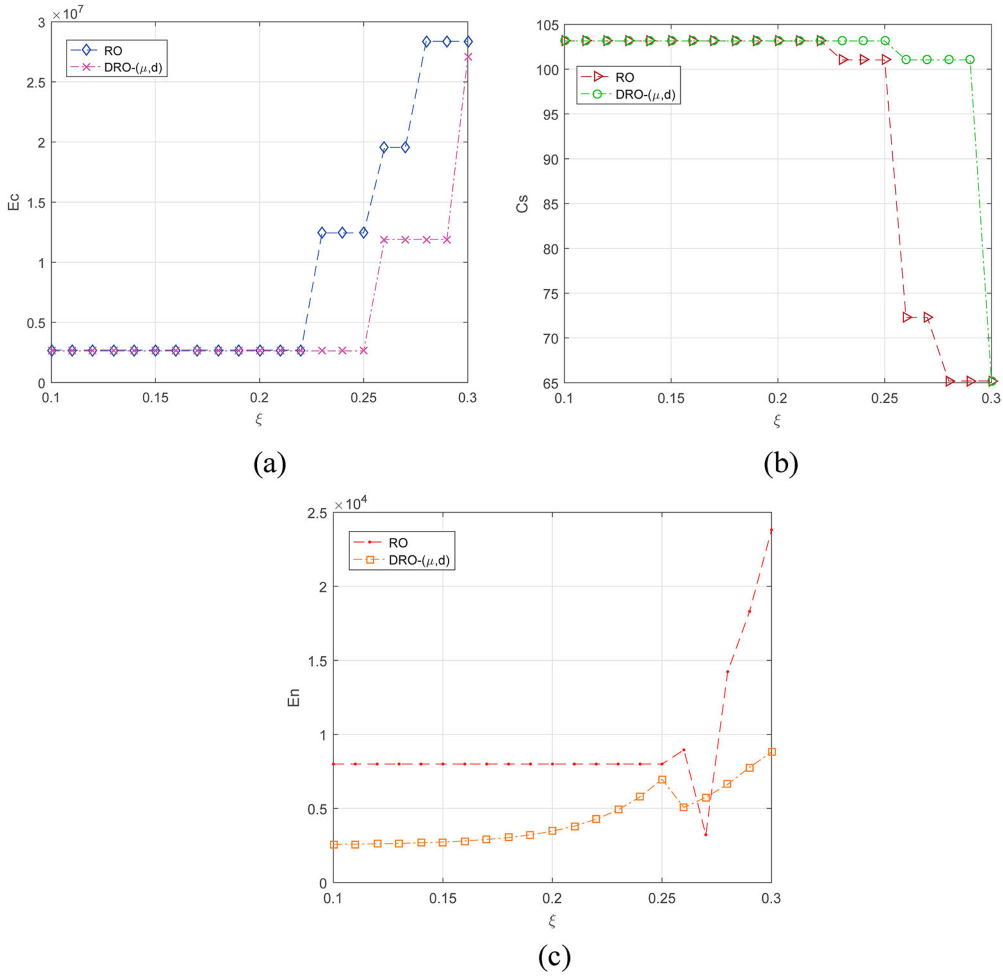

Figure 4. The influence of parameter ξ on the optimal values. (a) The influence of parameter ξ on *Ec*. (b) The influence of parameter ξ on *Cs*. (c) The influence of parameter ξ on *En*.

process, please refer to the study by Ben-Tal, Ghaoui, and Nemirovski (2009). We select 21 values of parameter ξ in [0.1, 0.3]. When ξ takes different values, it is worth noting that the model we solve is the goal programming model under the safety approximation of the DRO model, which corresponds to the goal programming model under the transformation of the classical robust optimisation model. For the proposed model and the classical robust optimisation model, the influences of parameter ξ on their optimal results are shown in Figure 4 and Table 5.

From the performance of the optimal values in Figure 4, we observe that, under the same value of ξ, our model gets better optimal objectives. This shows that the proposed model yields the lower costs and higher customer satisfaction, which is better than the classical robust optimisation model. In addition, with the increase of parameter ξ, Figures 4(a) and (b) show that the optimal values of economic objective and customer satisfaction objective of our model both remain unchanged first and then present an increasing and decreasing

Table 5. The influence of parameter ξ on the optimal solution.

	The optimum hub location		The capacity level of hub	
ξ	RO	DRO-(μ, d)	RO	DRO-(μ, d)
0.10	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.11	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.12	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.13	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.14	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.15	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.16	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.17	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.18	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.19	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.20	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.21	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.22	9, 10, 11	9, 10, 11	Low, Medium, Medium	Low, Medium, Medium
0.23	7, 9, 10	9, 10, 11	Medium, Low, Medium	Low, Medium, Medium
0.24	7, 9, 10	9, 10, 11	Medium, Low, Medium	Low, Medium, Medium
0.25	7, 9, 10	9, 10, 11	Medium, Low, Medium	Low, Medium, Medium
0.26	7, 9, 10	7, 9, 10	Medium, Low, Medium	Medium, Low, Medium
0.27	7, 9, 10	7, 9, 10	Medium, Low, Medium	Medium, Low, Medium
0.28	4, 9, 10	7, 9, 10	Medium, Low, Medium	Medium, Low, Medium
0.29	4, 9, 10	7, 9, 10	Medium, Low, Medium	Medium, Low, Medium
0.30	4, 9, 10	4, 9, 10	Medium, Low, Medium	Medium, Low, Medium

trend, respectively. The same observation is obtained for the classical robust optimisation model.

Table 5 lists the optimal hub locations and hub capacity levels of the DRO model and the classical robust optimisation model. From the table, we can find that, no matter how ξ changes, the optimal solutions of the two models do not change when ξ is relatively small. But when ξ is large, the optimal solutions of the two models change. The optimal solution of classical robust optimisation model begins to change after $\xi = 0.22$, while our model starts to change after $\xi = 0.25$. This shows that our proposed model is less sensitive to parameter ξ.

5.4. *Management insights*

In the previous subsection, we represent the computational results of our model and analyse the effects of the key parameters' changes on the optimal decision, by which we can provide the reference range of parameters for decision makers in practice. In addition, our method is compared with classical robust optimisation method. According to the results of numerical experiments, we give some managerial insights for decision-makers:

- The proposed distributionally robust multi-objective hub location model can help decision-makers to obtain the optimal solution by incorporating the ambiguity of distribution information. Therefore, decision-makers may adopt our model when the distributions of random parameters are ambiguous. Moreover, if decision-makers know the mean and dispersion information of random variables, they may directly use the conclusion on safe approximation of our model in Section 3.
- The sensitivity analysis of discount coefficient shows that the discount coefficient affects all the three objective values of our model. Moreover, the discount coefficient reflects the

degree of scale economy in practice. Thus, decision-makers may set the value of discount coefficient according to the real size of economy.

- Compared with the classical robust optimisation model, our new model produces the lower costs and higher customer satisfaction. This shows an advantage of the proposed method. Therefore, when the distributions of random parameters are ambiguous, to save costs and improve customer satisfaction, decision-makers may apply our model to design a better hub network structure by taking advantage of as much distribution information as possible.
- From Figure 4 and Table 5, we find that the parameter ξ affects the optimal results of the proposed model. But when parameter ξ takes value within the range of [0.1, 0.25], the optimal hub network provided by our model remains unchanged. Therefore, when the proposed model is used to design a hub network in practice, decision-makers should analyse the specific conditions to determine the appropriate support set of parameter ξ.

6. Conclusions

In this paper, we studied a new multi-objective hub location problem, in which economic, customer satisfaction and environmental objectives were optimised. In particular, we proposed a new customer satisfaction objective by defining transportation time satisfaction and transportation quality satisfaction. Moreover, the unit transportation costs, carbon emissions and noise levels exhibited random characteristics, and only partial distribution information could be obtained. The adopted detailed technique approach and the obtained meaningful results are shown below:

We first proposed a distributionally robust multi-objective hub location model when the ambiguous probability distribution was characterised via general ambiguity set. The ambiguity set characterised the possible range of the real probability distribution. In addition, when the ambiguity set was characterised by the mean and dispersion of the random variable, we derived the safe approximation of the proposed model. This resulting safe approximation was a multi-objective MISOCP model.

Then, we applied goal programming method to deal with the multi-objective MISOCP model. From the government's point of view, the environmental objective was taken as the first priority, then the customer satisfaction and economic objectives were ranked accordingly as the second and third priorities. The resulted model was a single objective MISOCP model, which could be solved directly using a commercial optimisation software.

Finally, we applied the proposed model to design China's super logistics network. The computational results verified the effectiveness of our new method in the case the probability distributions of random parameters were ambiguous. We compared the proposed method with the classical robust optimisation method. The results showed that our method had a better performance by taking advantage of the probability distribution information, instead of just using the support information of random parameters. Moreover, according to the computational results, we provided some management implications that might be helpful to decision-makers in hub network design and researchers in other fields.

While our proposed model has been transformed into a computationally solvable model by employing safe approximation and goal programming, the method of solving the transformed model directly with the optimisation solver has the limitation that it can only deal with small-scale cases. For large-scale cases, it is necessary to develop heuristic algorithms to obtain their Pareto fronts. This is one of our future research directions. Additionally, in our future research, we will study the application of distributionally robust optimisation methods to our hub location problem under other ambiguity sets.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the National Natural Science Foundation of China under Grant Nos. 61773150 and 71801077.

References

Ahmadi, T., H. Karimi, H. Davoudpour, and S. A. Hosseinijou. 2015. "A Robust Decision-Making Approach for *p*-hub Median Location Problems Based on Two-stage Stochastic Programming and Mean-Variance Theory: A Real Case Study." *International Journal of Advanced Manufacturing Technology* 77 (9-12): 1943–1953.

Alumur, S., and B. Y. Kara. 2008. "Network Hub Location Problems: The State of the Art." *European Journal of Operational Research* 190 (1): 1–21.

Alumur, S. A., S. Nickel, and F. Saldanha-da-Gama. 2012. "Hub Location Under Uncertainty." *Transportation Research Part B: Methodological* 46 (4): 529–543.

Aouni, B., and O. Kettani. 2001. "Goal Programming Model: A Glorious History and a Promising Future." *European Journal of Operational Research* 133 (2): 225–231.

Azizi, N., and S. Salhi. 2022. "Reliable Hub-and-Spoke Systems with Multiple Capacity Levels and Flow Dependent Discount Factor." *European Journal of Operational Research* 298 (3): 834–854.

Azizi, N., N. Vidyarthi, and S. S. Chauhan. 2018. "Modelling and Analysis of Hub-and-Spoke Networks Under Stochastic Demand and Congestion." *Annals of Operations Research* 264 (1-2): 1–40.

Ben-Tal, A., L. E. Ghaoui, and A. Nemirovski. 2009. *Robust Optimization*. Princeton: Princeton University Press.

Campbell, J. F. 1994. "Integer Programming Formulations of Discrete Hub Location Problems." *European Journal of Operational Research* 72 (2): 387–405.

Campbell, J. F. 1996. "Hub Location and the *p*-hub Median Problem." *Operations Research* 44 (6): 923–935.

Campbell, J. F., and M. E. O'Kelly. 2012. "Twenty-five Years of Hub Location Research." *Transportation Science* 46 (2): 153–169.

Charnes, A., and W. W. Cooper. 1961. *Management Models and Industrial Applications of Linear Programming*. New York: Wiley.

Cheng, C., M. Qi, Y. Zhang, and L. M. Rousseau. 2018. "A Two-Stage Robust Approach for the Reliable Logistics Network Design Problem." *Transportation Research Part B: Methodological* 111: 185–202.

Chou, C. C. 2010. "Application of FMCDM Model to Selecting the Hub Location in the Marine Transportation: A Case Study in Southeastern Asia." *Mathematical and Computer Modelling* 51 (5-6): 791–801.

Contreras, I., J. F. Cordeau, and G. Laporte. 2011. "Stochastic Uncapacitated Hub Location." *European Journal of Operational Research* 212 (3): 518–528.

Correia, I., S. Nickel, and F. Saldanha-da-Gama. 2018. "A Stochastic Multi-Period Capacitated Multiple Allocation Hub Location Problem: Formulation and Inequalities." *Omega* 74: 122–134.

da Graça Costa, M., M. E. Captivo, and J. Clímaco. 2008. "Capacitated Single Allocation Hub Location Problem—A bi-Criteria Approach." *Computers & Operations Research* 35 (11): 3671–3695.

Ernst, A. T., and M. Krishnamoorthy. 1996. "Efficient Algorithms for the Uncapacitated Single Allocation *p*-hub Median Problem." *Location Science* 4 (3): 139–154.

Ghaffarinasab, N. 2018. “An Efficient Matheuristic for the Robust Multiple Allocation *p*-hub Median Problem Under Polyhedral Demand Uncertainty.” *Computers & Operations Research* 97: 31–47.
Ghezavati, V., and P. Hosseinifar. 2018. “Application of Efficient Metaheuristics to Solve a New bi-Objective Optimization Model for Hub Facility Location Problem Considering Value at Risk Criterion.” *Soft Computing* 22 (1): 195–212.
Ghodratnama, A., R. Tavakkoli-Moghaddam, and A. Azaron. 2013. “A Fuzzy Possibilistic bi-Objective Hub Covering Problem Considering Production Facilities, Time Horizons and Transporter Vehicles.” *International Journal of Advanced Manufacturing Technology* 66 (1-4): 187–206.
Hu, Q. M., S. Hu, J. Wang, and X. Li. 2021. “Stochastic Single Allocation Hub Location Problems with Balanced Utilization of Hub Capacities.” *Transportation Research Part B: Methodological* 153: 204–227.
Huang, D., Z. Liu, X. Fu, and P. T. Blythe. 2018. “Multimodal Transit Network Design in a Hub-and-Spoke Network Framework.” *Transportmetrica A: Transport Science* 14 (8): 706–735.
Jiang, J., D. Zhang, and Q. Meng. 2021. “Impact Analysis of Investment Coordination Mechanisms in Regional Low-Carbon Logistics Network Design.” *Transportation Research Part D: Transport and Environment* 92. doi:10.1016/j.trd.2021.102735.
Jiang, J., D. Zhang, Q. Meng, and Y. Liu. 2020. “Regional Multimodal Logistics Network Design Considering Demand Uncertainty and CO_2 Emission Reduction Target: A System-Optimization Approach.” *Journal of Cleaner Production* 248, doi:10.1016/j.jclepro.2019.119304.
Kayışoğlu, B., and İ Akgün. 2021. “Multiple Allocation Tree of Hubs Location Problem for non-Complete Networks.” *Computers & Operations Research* 136, doi:10.1016/j.cor.2021.105478.
Korani, E., and A. Eydi. 2021. “Bi-level Programming Model and KKT Penalty Function Solution Approach for Reliable Hub Location Problem.” *Expert Systems with Applications* 184. doi:10.1016/j.eswa.2021.115505.
Meraklı, M., and H. Yaman. 2016. “Robust Intermodal Hub Location Under Polyhedral Demand Uncertainty.” *Transportation Research Part B: Methodological* 86: 66–85.
Mohammadi, M., P. Jula, and R. Tavakkoli-Moghaddam. 2019. “Reliable Single-Allocation Hub Location Problem with Disruptions.” *Transportation Research Part E: Logistics and Transportation Review* 123: 90–120.
Mohammadi, M., S. A. Torabi, and R. Tavakkoli-Moghaddam. 2014. “Sustainable Hub Location Under Mixed Uncertainty.” *Transportation Research Part E: Logistics and Transportation Review* 62: 89–115.
Mohammed, F., S. Z. Selim, A. Hassan, and M. N. Syed. 2017. “Multi-Period Planning of Closed-Loop Supply Chain with Carbon Policies Under Uncertainty.” *Transportation Research Part D: Transport and Environment* 51: 146–172.
Musavi, M., and A. Bozorgi-Amiri. 2017. “A Multi-Objective Sustainable Hub Location-Scheduling Problem for Perishable Food Supply Chain.” *Computers & Industrial Engineering* 113: 766–778.
Niknamfar, A. H., and S. T. A. Niaki. 2016. “Fair Profit Contract for a Carrier Collaboration Framework in a Green hub Network Under Soft Time-Windows: Dual Lexicographic max-min Approach.” *Transportation Research Part E: Logistics and Transportation Review* 91: 129–151.
O’Kelly, M. E. 1987. “A Quadratic Integer Program for the Location of Interacting Hub Facilities.” *European Journal of Operational Research* 32 (3): 393–404.
Parsa, M., A. S. Nookabadi, S. D. Flapper, and Z. Atan. 2019. “Green Hub-and-Spoke Network Design for Aviation Industry.” *Journal of Cleaner Production* 229: 1377–1396.
Postek, K., A. Ben-Tal, D. Den Hertog, and B. Melenberg. 2018. “Robust Optimization with Ambiguous Stochastic Constraints Under Mean and Dispersion Information.” *Operations Research* 66 (3): 814–833.
Rahimi, Y., R. Tavakkoli-Moghaddam, M. Mohammadi, and M. Sadeghi. 2016. “Multi-objective Hub Network Design Under Uncertainty Considering Congestion: An M/M/c/K Queue System.” *Applied Mathematical Modelling* 40 (5-6): 4179–4198.
Sadeghi, M., F. Jolai, R. Tavakkoli-Moghaddam, and Y. Rahimi. 2015. “A New Stochastic Approach for a Reliable *p*-hub Covering Location Problem.” *Computers & Industrial Engineering* 90: 371–380.
Shang, X., K. Yang, B. Jia, and Z. Gao. 2021a. “The Stochastic Multi-Modal Hub Location Problem with Direct Link Strategy and Multiple Capacity Levels for Cargo Delivery Systems.” *Transportmetrica A: Transport Science* 17 (4): 380–410.

Shang, X., K. Yang, B. Jia, and Z. Gao. 2021b. "Distributionally Robust Cluster-Based Hierarchical Hub Location Problem for the Integration of Urban and Rural Public Transport System." *Computers & Industrial Engineering* 155, doi:10.1016/j.cie.2021.107181.
Talbi, E. G., and R. Todosijević. 2017. "The Robust Uncapacitated Multiple Allocation *p*-hub Median Problem." *Computers & Industrial Engineering* 110: 322–332.
Wang, S., Z. Chen, and T. Liu. 2020. "Distributionally Robust Hub Location." *Transportation Science* 54 (5): 1189–1210.
Wang, J., and Z. Qin. 2021. "Modelling and Analysis of Uncertain Hub Maximal Covering Location Problem in the Presence of Partial Coverage." *Journal of Intelligent & Fuzzy Systems* 40 (5): 9987–10002.
Wang, R., K. Yang, L. Yang, and Z. Gao. 2018. "Modeling and Optimization of a Road-Rail Intermodal Transport System Under Uncertain Information." *Engineering Applications of Artificial Intelligence* 72: 423–436.
Yang, K., Y. Liu, and G. Yang. 2013. "Solving Fuzzy *p*-hub Center Problem by Genetic Algorithm Incorporating Local Search." *Applied Soft Computing* 13 (5): 2624–2632.
Yang, K., Y. Liu, and G. Yang. 2014. "Optimizing Fuzzy *p*-hub Center Problem with Generalized Value-at-Risk Criterion." *Applied Mathematical Modelling* 38 (15-16): 3987–4005.
Yang, M., and G. Yang. 2017. "Robust Optimization for the Single Allocation *p*-hub Median Problem Under Discount Factor Uncertainty." *Journal of Uncertain Systems* 11 (3): 230–240.
Yang, K., L. Yang, and Z. Gao. 2017. "Hub-and-Spoke Network Design Problem Under Uncertainty Considering Financial and Service Issues: A Two-Phase Approach." *Information Sciences* 402: 15–34.
Yin, F., Y. Chen, F. Song, and Y. Liu. 2019. "A New Distributionally Robust *p*-hub Median Problem with Uncertain Carbon Emissions and its Tractable Approximation Method." *Applied Mathematical Modelling* 74: 668–693.
Yin, F., and Y. Zhao. 2021. "Modelling Data-Driven Distributionally Robust Risk-Averse Hub Interdiction Median Problem Under Hypothesis Test." *Computers & Industrial Engineering* 157. doi:10.1016/j.cie.2021.107323.
Zhalechian, M., R. Tavakkoli-Moghaddam, Y. Rahimi, and F. Jolai. 2017. "An Interactive Possibilistic Programming Approach for a Multi-Objective Hub Location Problem: Economic and Environmental Design." *Applied Soft Computing* 52: 699–713.

Appendix A. Proofs of Theorems 3.1 and 3.2

Proof of Theorem 3.1: When the support, expectation and deviation of the random variable z_k^4 are known, the distribution of the random variable belongs to the following ambiguity set.

$$\mathcal{P}^2_{\mu,d_k^4} = \{\mathrm{P} : \mathrm{supp}(z_k^4) \subseteq [a,b], \mathrm{E_P}[z_k^4] = \mu, \mathrm{E_P}[|z_k^4 - \mu|] = d_k^4\}.$$

In order to obtain the upper bound of the expected value of a convex function $\exp(\xi z_k^4 \hat{L}_{10,k})$ containing a random variable z_k^4, when the true probability distribution of z_k^4 belongs to an ambiguity set $\mathcal{P}_{\mu,d_k^4}$, the following upper bound is given by Postek et al. (2018):

$$\max_{\mathrm{P}\in\mathcal{P}^2_{\mu,d_k^4}} \mathrm{E_P}[\exp(\xi z_k^4 \hat{L}_{10,k})] = p_1 \exp(\xi a \hat{L}_{10,k}) + p_2 \exp(\xi \mu \hat{L}_{10,k}) + p_3 \exp(\xi b \hat{L}_{10,k}),$$

where $p_1 = \frac{d_k^4}{2(\mu-a)}$, $p_2 = 1 - \frac{d_k^4}{2(\mu-a)} - \frac{d_k^4}{2(b-\mu)}$, $p_3 = \frac{d_k^4}{2(b-\mu)}$. Due to $\mu = 0$, $a = -1$ and $b = 1$, we obtain: $p_1 = \frac{d_k^4}{2}, p_2 = 1 - d_k^4, p_3 = \frac{d_k^4}{2}$. Therefore, the following conclusions are obtained.

$$\sup_{P\in\mathcal{P}^2_{\mu,d}} E_{z\sim P}[\exp(\xi z_k^4 \hat{L}_{10,k})]$$

$$\leq \frac{d_k^4}{2}\exp(-\xi\,\hat{L}_{10,k}) + \frac{d_k^4}{2}\exp(\xi\,\hat{L}_{10,k}) + (1-d_k^4)\exp(0)$$
$$= d_k^4\cosh(\xi\,\hat{L}_{10,k}) + 1 - d_k^4.$$

The proof of theorem is complete. ■

Proof of Theorem 3.2: The proof of Theorem 3.2 is based on the following conclusions.

Theorem 3 (Ben-Tal, Ghaoui, and Nemirovski (2009))**.** *If the random variables z_i^1, z_k^2, z_l^3 obey the following properties P.1-2 and $a^- \leq z \leq a^+$, $a^- \leq \mu^- \leq \mu^+ \leq a^+$, for all i, k, l.*

P.1. For any $i,k,l \in N$, z_i^1, z_k^2 and z_l^3 are independent random variables such that $\text{supp}(z_i^1) \subseteq [-1,1]$, $\text{supp}(z_k^2) \subseteq [-1,1]$, $\text{supp}(z_l^3) \subseteq [-1,1]$,

P.2. The distributions P_i, P_k, P_l of the components z_i^1, z_k^2, z_l^3 are such that

$$\int \exp(ts)dP_i(s) \leq \exp(\max\{(\mu_i^1)^+ t, (\mu_i^1)^- t\} + \frac{1}{2}(\sigma_i^1)^2 t^2), \forall t \in R,$$
$$\int \exp(ts)dP_k(s) \leq \exp(\max\{(\mu_k^2)^+ t, (\mu_k^2)^- t\} + \frac{1}{2}(\sigma_k^2)^2 t^2), \forall t \in R,$$
$$\int \exp(ts)dP_l(s) \leq \exp(\max\{(\mu_l^3)^+ t, (\mu_l^3)^- t\} + \frac{1}{2}(\sigma_l^3)^2 t^2), \forall t \in R$$

with known constants $(\mu_i^1)^- \leq (\mu_i^1)^+$, $(\mu_k^2)^- \leq (\mu_k^2)^+$ and $(\mu_l^3)^- \leq (\mu_l^3)^+$. Then, the following robust counterpart

$$\begin{cases} y_0 = \gamma_0 + \tau_0, \\ y_i^1 = \gamma_i^1 + \tau_i^1, \forall i \in N, \\ y_k^2 = \gamma_k^2 + \tau_k^2, \forall k \in N, \\ y_l^3 = \gamma_l^3 + \tau_l^3, \forall l \in N, \\ \gamma_0 + \sum\limits_{i\in N} \max[(a_i^1)^-\gamma_i^1, (a_i^1)^+\gamma_i^1] + \sum\limits_{k\in N} \max[(a_k^2)^-\gamma_k^2, (a_k^2)^+\gamma_k^2] + \sum\limits_{l\in N} \max[(a_l^3)^-\gamma_l^3, (a_l^3)^+\gamma_l^3] \leq 0, \\ \tau_0 + \sum\limits_{i\in N} \max[(\mu_i^1)^-\tau_i^1, (\mu_i^1)^+\tau_i^1] + \sum\limits_{k\in N} \max[(\mu_k^2)^-\tau_k^2, (\mu_k^2)^+\tau_k^2] + \sum\limits_{l\in N} \max[(\mu_l^3)^-\tau_l^3, (\mu_l^3)^+\tau_l^3] \\ +\sqrt{2\ln(1/\epsilon)}\sqrt{\sum\limits_{i\in N}(\sigma_i^1\tau_i^1)^2 + \sum\limits_{k\in N}(\sigma_k^2\tau_k^2)^2 + \sum\limits_{l\in N}(\sigma_l^3\tau_l^3)^2} \leq 0 \end{cases}$$

of $\{y_0 + \sum_{i\in N} z_i^1 y_i^1 + \sum_{k\in N} z_k^2 y_k^2 + \sum_{l\in N} z_l^3 y_l^3 \leq 0\}$ corresponding to the perturbation set

$$\mathcal{U} = \left\{ \eta \in R^{3N} : \exists \gamma \in R^{3N} : \begin{array}{c} \mu^- \leq \eta - \gamma \leq \mu^+ \\ \sqrt{\sum \gamma^2/\sigma^2} \leq \sqrt{2\ln(1/\epsilon)} \\ a^- \leq \eta \leq a^+ \end{array} \right\}$$

is a safe approximation of (27)*. Moreover, every decision variable x that can be extended to a feasible solution (x,γ,τ) to the ambiguous chance constraint* (3) *is feasible for the chance constraint* (27).

In Theorem 3, if $(\mu_i^1)^- = (\mu_i^1)^+ = (\mu_k^2)^- = (\mu_k^2)^+ = (\mu_l^3)^- = (\mu_l^3)^+ = 0$, $(a_i^1)^- = (a_k^2)^- = (a_l^3)^- = -1$, $(a_i^1)^+ = (a_k^2)^+ = (a_l^3)^+ = 1$, we can get the conclusion of Theorem 3.2, but we also need to get the values of σ_i^1, σ_k^2 and σ_l^3. Firstly, we substitute z_i^1, z_k^2 and z_l^3 into P.2 and get the following results:

$$\begin{cases} \int_{-1}^{1} \exp(tz_i^1)dP_i(z_i^1) \leq \exp\left(\max\{(\mu_i^1)^+ t, (\mu_i^1)^- t\} + \frac{1}{2}(\sigma_i^1)^2 t^2\right), \forall t \in R, \forall P \in \mathcal{P}_{\mu,d}^1, \\ \int_{-1}^{1} \exp(tz_k^2)dP_k(z_k^2) \leq \exp\left(\max\{(\mu_k^2)^+ t, (\mu_k^2)^- t\} + \frac{1}{2}(\sigma_k^2)^2 t^2\right), \forall t \in R, \forall P \in \mathcal{P}_{\mu,d}^1, \\ \int_{-1}^{1} \exp(tz_l^3)dP_l(z_l^3) \leq \exp\left(\max\{(\mu_l^3)^+ t, (\mu_l^3)^- t\} + \frac{1}{2}(\sigma_l^3)^2 t^2\right), \forall t \in R, \forall P \in \mathcal{P}_{\mu,d}^1. \end{cases} \quad (36)$$

According to Postek et al. (2018), we have

$$\begin{cases} \sup_{P\in\mathcal{P}^1_{\mu,d}} \left\{ \int_{-1}^{1} \exp(tz_i^1) dP_i(z_i^1) \right\} = d_i^1 \cosh(t) + 1 - d_i^1, \\ \sup_{P\in\mathcal{P}^1_{\mu,d}} \left\{ \int_{-1}^{1} \exp(tz_k^2) dP_k(z_k^2) \right\} = d_k^2 \cosh(t) + 1 - d_k^2, \\ \sup_{P\in\mathcal{P}^1_{\mu,d}} \left\{ \int_{-1}^{1} \exp(tz_l^3) dP_l(z_l^3) \right\} = d_l^2 \cosh(t) + 1 - d_l^2. \end{cases} \tag{37}$$

Thus, if we substitute (37) into (36), we get

$$\begin{cases} d_i^1 \cosh(t) + 1 - d_i^1 \le \exp\left(\max\{(\mu_i^1)^+ t, (\mu_i^1)^- t\} + \frac{1}{2}(\sigma_i^1)^2 t^2 \right), \forall t \in R, \\ d_k^2 \cosh(t) + 1 - d_k^2 \le \exp\left(\max\{(\mu_k^2)^+ t, (\mu_k^2)^- t\} + \frac{1}{2}(\sigma_k^2)^2 t^2 \right), \forall t \in R, \\ d_l^2 \cosh(t) + 1 - d_l^2 \le \exp\left(\max\{(\mu_l^3)^+ t, (\mu_l^3)^- t\} + \frac{1}{2}(\sigma_l^3)^2 t^2 \right), \forall t \in R. \end{cases}$$

Setting $(\mu_i^1)^+ = (\mu_i^1)^- = 0$, $(\mu_k^2)^+ = (\mu_k^2)^- = 0$, $(\mu_l^3)^+ = (\mu_l^3)^- = 0$, then we obtain the following conclusion

$$\begin{cases} d_i^1 \cosh(t) + 1 - d_i^1 \le \exp\left(\frac{1}{2}(\sigma_i^1)^2 t^2 \right), \forall t \in R, \\ d_k^2 \cosh(t) + 1 - d_k^2 \le \exp\left(\frac{1}{2}(\sigma_k^2)^2 t^2 \right), \forall t \in R, \\ d_l^2 \cosh(t) + 1 - d_l^2 \le \exp\left(\frac{1}{2}(\sigma_l^3)^2 t^2 \right), \forall t \in R \end{cases}$$

$$\Leftrightarrow \begin{cases} \sigma_i^1 = \sup_{t\in R} \sqrt{\frac{2\ln(d_i^1 \cosh(t) + 1 - d_i^1)}{t^2}}, \\ \sigma_k^2 = \sup_{t\in R} \sqrt{\frac{2\ln(d_k^2 \cosh(t) + 1 - d_k^2)}{t^2}}, \\ \sigma_l^3 = \sup_{t\in R} \sqrt{\frac{2\ln(d_l^3 \cosh(t) + 1 - d_l^3)}{t^2}}. \end{cases}$$

The proof is complete. ■