CIRRELT

CIRRELT-2024-20

Handling Ambiguity in Stochastic Humanitarian Supply Chain Network Design

Mohammad Daneshvar
Sanjay Dominik Jena
Walter Rei

June 2024

Bureau de Montréal
Université de Montréal
C.P. 6128, succ. Centre-Ville
Montréal (Québec) H3C 3J7
Tél : 1-514-343-7575
Télécopie : 1-514-343-7121

Bureau de Québec
Université Laval,
2325, rue de la Terrasse
Pavillon Palasis-Prince, local 2415
Québec (Québec) G1V 0A6
Tél : 1-418-656-2073
Télécopie : 1-418-656-2624

Handling Ambiguity in Stochastic Humanitarian Supply Chain Network Design

Mohammad Daneshvar*, Sanjay Dominik Jena, Walter Rei

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT) and Analytics, Operations, and Information Technologies Department, School of Management, Université du Québec à Montréal

Abstract. The design and operation of Humanitarian Supply Chain Networks after a natural disaster are among the most complex activities conducted by humanitarian organizations, involving different sources of uncertainty. Typically, the assessments of damage and the resulting demand for resources in the affected region are estimated using different data sources (e.g. surveys and satellite imagery). However, inconsistent estimates of uncertain parameters obtained from the use of multiple data sources may result in ambiguity, posing difficulties to define the planning problem. We here aim at mitigating such ambiguity by developing four mathematical models that deal with ambiguity with varying degrees of conservatism regarding the obtained estimations of the uncertainty. The performance of each proposed model is evaluated, considering two data sources across four ambiguity patterns and 20 problem instances generated using real-world data from the 2018 Indonesia earthquake. The results highlight the benefits of the Minimization of Maximum Data-Source Penalty when the ambiguity pattern is unknown and the decision-maker has equally high confidence in all data sources.

Keywords: Humanitarian Supply Chain; Humanitarian Relief Network; Stochastic Programming; Ambiguity; post-disaster; Aid Planning

Acknowledgement: This research was enabled in part by support provided by Calcul Québec (calculquebec.ca) and the Digital Research Alliance of Canada (alliance.can.ca).

---

* Corresponding author: daneshvar.mohammad@courrier.uqam.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2024

1 Introduction

In 2013, the total amount of funding requested by humanitarian organizations worldwide was 12.8 billion US dollars (OCHA, 2022). However, the total amount of donations received by humanitarian organizations in the same year was 8.3 billion US dollars (OCHA, 2022), accounting for only 65% of the requested funding. In all other years of the previous decade, the satisfaction rate of requested annual funding requested by humanitarian organizations was even lower. Since then, the total amount of funding requests of humanitarian organizations has increased to 51.6 billion US dollars (OCHA, 2022) in 2022, from which only 29.7 billion US dollars was provided, covering 57.5% of the requested amount. These numbers indicate severe budget shortages for humanitarian organizations. Three-quarters of the expenses of humanitarian organizations are related to the logistics of humanitarian relief operations (Besiou and Van Wassenhove, 2020; Van Wassenhove, 2006; HELP Logistics AG, 2018). An efficient management of available resources for humanitarian relief operations is thus paramount for efficiently providing relief to the affected population. This includes allocating the available budget over time, the effective use of the available staff to perform and support the humanitarian operations, and the planning of the transportation operations to distribute critical supplies.

Effective distribution planning of critical supplies among vulnerable people after a natural disaster is crucial, given its direct impact on the population's health. The required planning is complex and mandates coordination among several stakeholders, while a series of crucial decisions need to be made to ensure operational success. A major complexity of the planning process stems from the high level of uncertainty in the informational environment in which tactical decisions, such as those related to infrastructure and logistics services, are made. Furthermore, a considerable multitude of stakeholders, including the governments, military, donors, and humanitarian organizations, require coordination at multiple levels. For example, when a disaster happens, the affected region is oftentimes divided into subregions where different humanitarian organizations will operate. This enables a better coverage of the affected region. For security reasons, military personnel are often called upon to protect humanitarian organizations, their staff, and volunteers when deployed in the field to distribute aid. Communication and coordination between the military and humanitarian organizations is thus a pivotal part of the distribution of critical supplies. Lastly, a coordinated effort between humanitarian organizations and the media is also required to bring attention to the crisis that occurred, which, in turn, helps to fundraise and collect the required budget for the necessary operations.

In this paper, we are specifically interested in the design and operation of a Humanitarian Supply Chain Network (HSCN) after the occurrence of a natural disaster over a defined planning horizon. An HSCN is defined as a physical network of hubs that are used to receive, store, and distribute critical supplies among the vulnerable population, where transportation services are planned to move the critical supplies. The design of an HSCN requires a comprehensive understanding of various parameters, including geographic and demographic data, financial constraints (e.g. the budget), and the availability of essential resources. While the values of some of these parameters are known at the design time, the values of others (e.g., demand, transportation, and storage capacities) are uncertain and estimated during the assessments of damages and needs. Such assessment of damages and needs starts immediately after the disaster, evaluating the damaged infrastructures and demand of affected populations (Balcik, 2017; Balcik and Yanıkoğlu, 2020). The assessment process must operate quickly to provide the required information to decision-makers. Given that on-site assessments of all impacted locations within a constrained timeframe are not feasible, supplementary data sources are employed to expedite the procurement of essential information. A data source (e.g. surveys, satellite imagery, governmental reports, and media) is a database from which data is collected or retrieved to help define probabilistic models accounting for uncertain components involved in the HSCN design problem. Hence, a finite set of probabilistic models is available to estimate the uncertain parameters in the HSCN design problem. Although a high level of confidence is observed for such obtained probabilistic models, they may have discrepancies, directly causing ambiguity in the informational context in which the planning process of humanitarian relief operations occurs. As such, ambiguity in the HSCN design problem is a problem setting in which inconsistent probability distributions are associated with the uncertain parameters (Langewisch and Choobineh, 1996). For instance, Grass et al. (2021) discuss a real-world humanitarian relief problem from Syria where two data sources provide estimations on the level of demand and in some demand points, the estimations barely overlap.

As a result, we are here interested in solving an HSCN design problem with discrepancies in the estimations of the uncertainty in demand and in both transportation and storage capacity obtained from various data sources. We propose four distinct optimization models with varying levels of conservatism, incorporating the inherent ambiguity in the HSCN design problem. This contribution advances the understanding and set of available tools for HSCN design by accounting for ambiguity in the studied problem. We are specifically interested in identifying the circumstances in which the proposed models are a superior choice to the commonly used two-stage stochastic model in the literature that does not explicitly account for such ambiguity. To this end, we conduct a comprehensive empirical evaluation, assessing the performance of the proposed optimization models. These experiments are performed with two different data sources and encompass four unique ambiguity patterns. Moreover, the utilized instances are generated from a 2018 Indonesian earthquake dataset, ensuring that the findings have practical relevance.

The remainder of the paper is structured as follows. Section 2 covers the literature on handling uncertainty in HSCN design problems and the modeling of uncertainty and ambiguity in general. Section 3 is dedicated to the problem definition. The proposed models are introduced in Section 4. The experiments and results are discussed in Section 5. Finally, we conclude in Section 6.

2 Literature review

In this section, we position our study within the existing literature. We review the literature on both the humanitarian relief problems under uncertainty and the optimization methods proposed to model and solve them. Subsection 2.1 presents a literature review on uncertainty in humanitarian relief studies, discussing the sources of uncertainty, the assessment process to obtain the required information after a disaster, and the ambiguity in estimating the uncertain parameters involved in humanitarian relief planning. Subsection 2.2 then reviews the relevant Operations Research literature on different approaches to model uncertainty, including Stochastic Programming and Robust Optimization.

2.1 Uncertainty in humanitarian relief

The design of an HSCN necessitates access to geographical and demographical data, as well as information regarding the availability of critical supplies and other essential resources (e.g. budget) for humanitarian relief distribution operations. While some information is known at the design phase (e.g. the available routes connecting hubs), some crucial information (e.g. the capacity of each route for transportation) becomes available over time. It is crucial to consider such uncertainty in the design process, providing adaptability to the changing circumstances (e.g. varying realizations of demand), allocating resources more effectively, and enhancing disaster response, ultimately saving lives and reducing the impact of the disaster on the affected populations.

Demand is the most common and often the most impactful uncertain component in humanitarian relief studies (Balcik and Beamon, 2008; Dönmez et al., 2021; Anaya-Arenas et al., 2014). Additional sources of uncertainty in humanitarian relief problems include a lack of information on the affected population, the urban or rural structure of the affected region, and the intensity of the natural disaster and its secondary impacts (e.g. landslide or aftershock). Further uncertain components in humanitarian relief studies are travel time, supply, network reliability, shipping cost, and shipping capacity (Anaya-Arenas et al., 2014; Tofighi et al., 2016; Daneshvar et al., 2023).

Damage and demand assessments are conducted after the natural disaster, providing probabilistic models that represent the uncertain components of the planning problem. Considering the limited time and a lack of resources available for the assessments, the affected region is divided into smaller subregions, and the assessments are taken by sampling sites in each subregion (Balcik and Yanıkoğlu, 2020; Balcik, 2017). In addition to the data obtained from on-site visits, other data sources are also available for the assessments. Data sources in humanitarian relief problems either need experts' interpretation (e.g. satellite imagery, media) or belong to previous natural disasters in the region (e.g. historical data) (Yáñez-Sandivari et al., 2021). Previous disasters' data sources are mostly used in pre-disaster studies (Balcik et al., 2019). However, since each natural disaster has unique characteristics (Chen et al., 2011), post-disaster studies rely more on the experts' interpreted data sources from the current natural disaster.

Benini et al. (2017) explain different types of responses provided by experts in assessments during humanitarian operations, including probability, continuous scale, and scalar quantity estimations. Probability estimation is a single-value estimation often used to estimate the likelihood of occurrences of an event. In continuous scale estimation, the expert indicates a range or a single value over a defined scale, estimating the uncertain components. Finally, scalar quantity is a three-point estimation method, including the minimum, maximum, and most probable values. The obtained data points form a triangular distribution representing the uncertain components (Hakimifar et al., 2021), providing higher accuracy than the former estimations. In this paper, we use the triangular distribution estimation method. A set of discrete scenarios is then generated from distributions estimating the value of uncertain components (Grass and Fischer, 2016; Gutjahr and Nolz, 2016; Grass et al., 2021). Although the obtained estimations can be made with a high level of confidence, they might have discrepancies, resulting in ambiguity within the HSCN design problem. The common approach in the literature is to use stochastic programming to model the uncertainties. However, such an approach does not reflect the discrepancies between different data sources and could result in a sub-optimal solution based on each individual data source (Grass et al., 2021; Benini, 2017). In Grass et al. (2021), the authors proposed a machine learning approach leveraging graph clustering and stochastic optimization techniques to address the challenges encountered by humanitarian decision-makers in a shelter location problem with ambiguity in demand. While the approach proposed in Grass et al. (2021) enables the ambiguity to be analyzed, our work focuses on optimization models for the HSCN design in a problem setting that involves ambiguity. We provide mathematical models that explicitly account for the ambiguity in the here-studied HSCN design problem using combinations of stochastic programming, robust optimization, and goal programming.

2.2 Uncertainty and ambiguity in operations research

We first review stochastic programming and robust optimization, both approaches are used to model and solve optimization problems with uncertainties. We then discuss goal programming, a mathematical optimization technique that can be used in optimization problems with multiple conflicting objectives.

Stochastic programming is the paradigm of choice for problems with uncertain components that can be formulated, with a high level of confidence, by probability distributions. In contrast, robust optimization is employed when the statistical information regarding the uncertain components of the problem is limited (e.g., only the upper bound and lower bounds of the probability distribution are available).

In stochastic programming problems (Birge and Louveaux, 2011), a set of scenarios (i.e., random realizations) generated from the probability distributions of the uncertain components represent the probabilistic outcomes of the uncertain parameters. However, probability distributions are not always available. Ben-Tal and Nemirovski (1998) introduce the concept of robust optimization, considering the possible realizations of uncertain parameters regardless of their probability distribution. To control the level of conservatism of the obtained solutions, the authors defined an uncertainty set (i.e., a predefined range of potential variations of uncertain parameters) that prevents all uncertain components from taking their worst-case values simultaneously, reducing the level of conservatism of the original max-min model introduced by Wald (1945). Further uncertainty sets are used in the literature, including polyhedral, norm-bounded, interval and chance-constrained uncertainty sets (Pluymers et al., 2005; Ben-Tal and Nemirovski, 2002; Abedor et al., 1995). The choice of uncertainty set depends on the nature of the problem and the level of conservatism or robustness desired in the optimization process. The polyhedral uncertainty set is defined by linear constraints, restricting the potential values of uncertain parameters (Pluymers et al., 2005). The norm-bounded uncertainty set is used where the magnitude of deviation is known but not the direction (Abedor et al., 1995). When only bounds on uncertain parameter values are known, the interval uncertainty set is used (Ben-Tal and Nemirovski, 2002). Using a chance-constrained probability set involves defining a probability distribution for uncertain parameters and setting constraints limiting the probability of violating the constraints below a specified threshold.

Finally, our work is also related to goal programming (Charnes and Cooper, 1957), a technique used to balance multiple conflicting objectives. Variations of goal programming have been used in the literature, including its combination with stochastic programming (Aouni et al., 2012) and robust goal programming (Ghahtarani and Najafi, 2013). In this study, we consider an HSCN design problem with the estimation of uncertain parameters obtained from multiple data sources, and optimizing the model based on each data source could be seen as a different goal. Hence, we use goal programming to propose a model that explicitly accounts for uncertainty and ambiguity in the HSCN design problem; see Minimization of expected opportunity loss approach in Section 4.2.

3 Problem description

In this section, we present the here-considered HSCN design problem. Specifically, this section introduces the general characteristics of the HSCN design problem under uncertainty, including the network structure, both deterministic and uncertain parameters, and the decisions involved.

An HSCN is a physical network of hubs connected by transportation services (Daneshvar et al., 2023). The designed HSCN receives, stores, transports, and distributes critical supplies to beneficiary groups over a defined planning horizon, aiming to minimize the harm to people's health caused by unmet demand. Here, we introduce the terminology used in the rest of the paper. The target population, called beneficiary groups, is the relocated people who live in temporary shelters such as camps, schools, and sports centers. We divide the planning horizon into operational time frames, referred to as time periods. A time period indicates the required amount of time during which a shipment is received, stored, and distributed in the affected region, plus the time that beneficiaries consume them.

Each beneficiary group needs a set of supplies that are called critical supplies. Some critical supplies are provided only once during the first period (e.g. tent and blanket), and others at every period (e.g. food). The beneficiary groups pick up the critical supplies from physical locations called Distribution Centers (DC). Failure to satisfy the demand for each critical supply causes a penalty. The required quantity of each critical supply for each individual or household is known (IFRC, 2021), but the number of individuals in each beneficiary group is uncertain. The source of this uncertainty is due to both a lack of information and the possibility of change in the number of individuals in each group over time caused by secondary impacts (e.g. aftershocks following an earthquake, landslide following a flood). Therefore, the exact level of demand may never be available and is considered uncertain.

A high level of demand and limited resources typically prevent the HSCN from fully satisfying the demand of the beneficiary groups. The portion of demand that is not satisfied is denoted as unmet demand. Each unit of unmet demand negatively affects the level of demand in the next period, the degree of which can be accounted for by using the notion of a spread factor (Daneshvar et al., 2023). For instance, in the natural disasters that happen during pandemics, lack of access to face masks and alcohol-based disinfectants results in the spread of the epidemy and increases the demand for test kits and related medication (Sakamoto et al., 2020). The spread factor indicates the impact of one unit of unmet demand of a critical supply on the demand level for critical supplies in the next period. The demand in each period is therefore defined as the sum of a base demand calculated based on available estimations and a residual demand, which is the effect of unmet demand in the previous period. Equation (1) computes the total demand $\hat{d}_l^{kt}$ of beneficiary group l for critical supply k at period t. In this equation, $\tilde{d}_l^{kt}$ represents the base demand of beneficiary group l for critical supply k at period t, $s^{k'k}$ represents the spread factor of critical supply k' on supply k, and $\bar{a}_{il}^{k't-1}$ represents the allocated critical supply k' to beneficiary group l from DC i at period $t-1$.

$$\underbrace{\hat{d}_l^{kt}}_{\text{total demand}} = \underbrace{\tilde{d}_l^{kt}}_{\text{base demand}} + \underbrace{\sum_{k' \in K} s^{k'k}(\hat{d}_l^{k't-1} - \sum_{i \in V_{DC}} \bar{a}_{il}^{k't-1})}_{\text{residual demand}} \tag{1}$$

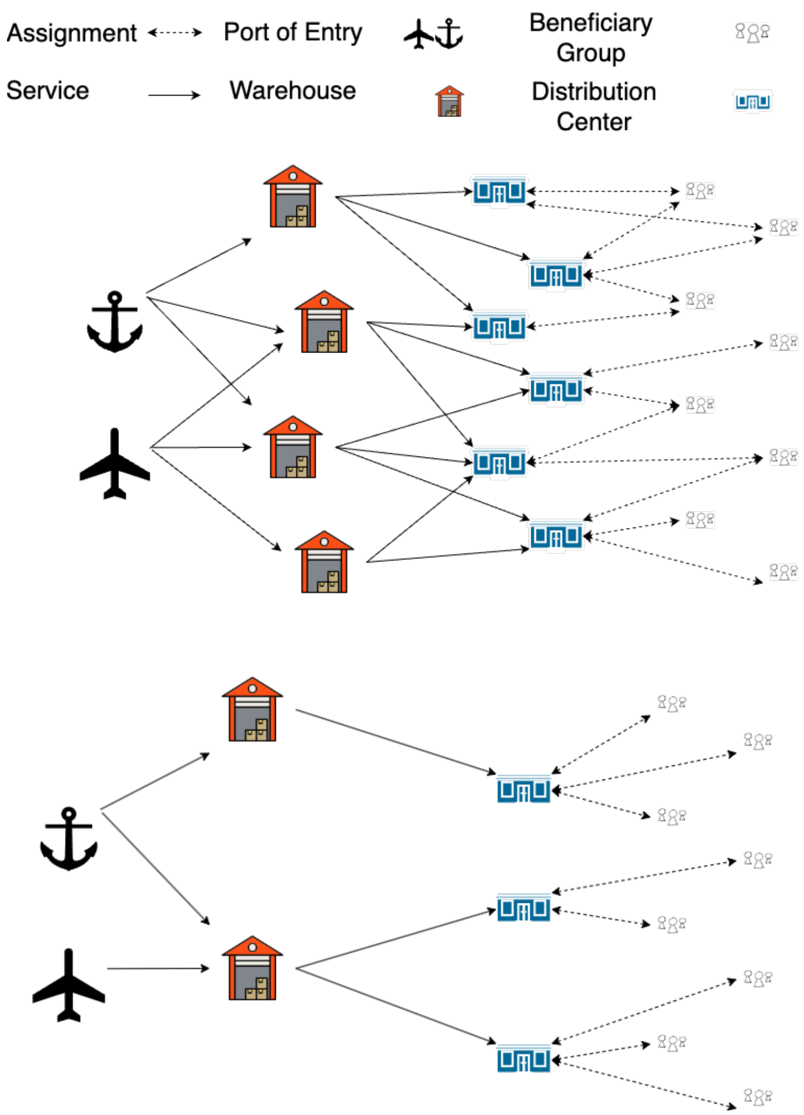

Figure 1: An HSCN illustration. Top: all available hubs, services, and assignments. Bottom: a designed HSCN, including the selected hubs, services, and assignments in an example HSCN planning solutions.

We consider a three-layer structure HSCN, as exemplified in Figure 1. The first layer of hubs consists of ports of entry that receive critical supplies from international humanitarian organizations. The second layer of hubs includes warehouses. A warehouse is a hub that receives critical supplies from ports of entry, stores them and sends them to the third layer of the HSCN, which consists of DCs. There is a fixed cost to use each selected hub for the considered planning horizon. In addition to the fixed cost of selecting a hub, there is an additional fixed cost for reserving the inventory resources available at selected warehouses. Inventory resources are only available at the warehouses, with the possibility to store critical supplies over the considered planning period. A unit of inventory resources could be a classroom in a school or a container located in a field used as a temporary warehouse. The fixed cost associated with utilizing inventory resources within warehouses is proportional to the requested capacity allocation at each respective warehouse, subject to the maximum available storage capacity of the warehouse. Each beneficiary group is assigned to a DC where it can pick up its allocated critical supplies at each time period. Transportation services move the critical supplies between selected hubs. We assume that each transportation resource commutes only between its origin and destination hubs, returning to its origin hub after delivering the critical supplies. The total transportation capacity between two hubs is given by the sum of the transportation services between these hubs. A unit of transportation resources could be a truck, a boat, a train wagon, or a helicopter. There is a fixed cost (e.g. for drivers, staff, and security escorts) for selecting each unit of transportation resources, as well as a variable flow cost (e.g. fuel) proportional to the travel distance of the transportation resources. The total costs incurred by the design decisions are limited by the initial budget available at the time of the design. Over the subsequent periods, the humanitarian organization receives donations for operational expenses (e.g. flow cost). Unused budget at any time period is carried over to the next period.

In the aftermath of a natural disaster, the extent of the impact of the event on the affected population (e.g. demand) and state of the region (e.g. transportation and inventory capacities) is uncertain. Over time, additional information may become available, reducing the contextual uncertainty (e.g. conducting direct observations in the affected region might enable a more accurate quantification of the needs of the affected population). However, humanitarian organizations cannot afford to wait for such information to plan and deploy the aid. Rapid responses are crucial to minimize harm in the affected region. Furthermore, the local resources necessary for the HSCN design might be notably scarce, and delaying the procurement of such resources could lead to an inflationary spiral (Holguín-Veras et al., 2012). Therefore, critical decisions regarding the structure and capacities of the HSCN need to be made amidst uncertainty, while other decisions concerning the allocation of the available resources can be made once additional information is obtained and uncertainty levels are diminished. We here consider a two-stage setting where the first decision stage occurs at the beginning of the planning horizon when the HSCN is designed under a rather high level of uncertainty. For the second stage (when the operational decisions are made), we assume that all stochastic parameters (e.g. demand and transportation capacity) become known.

The scenarios (i.e., realizations of the uncertain parameters) are generated using the probability distributions obtained from assessments conducted in the region. However, as multiple data sources (e.g. satellite imagery and governmental reports) are involved in the assessments, the probability distributions obtained might have inconsistencies, resulting in ambiguity. Specifically, when different assessments are performed using different data sources to quantify the same demands, they may yield different random distributions. Recalling that the same level of confidence is assigned to all assessments, the ambiguity arises from the uncertainty about which probabilities should be used during the planning process. In the next section, we present optimization models that explicitly account for such sources of ambiguity when solving the here-considered problem.

4 Optimization model

In this section, we propose a variety of optimization models that explicitly deal with the ambiguity that stems from inconsistent estimations of the uncertain components obtained from various data sources. Subsection 4.1 recalls how the HSCN problem is formulated as a two-stage stochastic optimization model under the general assumption that a single data source is used to generate a single scenario set Ψ. Then, in subsection 4.2, we introduce a series of optimization models that explicitly consider the ambiguity faced in the HSCN design problem under study.

4.1 HSCN design model

We model the HSCN design problem as a two-stage stochastic model, introduced in Daneshvar et al. (2023). In this model, the hubs and transportation services are selected from available hubs, represented by the set V, and services, represented by the set A. The set of hubs contains three subsets, including the port of entry hubs, V_I, the warehouse hubs, V_W, and the DC hubs, V_{DC}. The selected hubs and services will be part of the HSCN network over the entire planning horizon. The planning horizon consists of a sequence of time periods represented by the set T. The designed HSCN is used to distribute a set of critical supplies, represented by the set K, among the beneficiary groups which are represented by the set L, over the planning horizon. In order to model uncertain parameters, we use scenarios generated from estimations provided by data sources. In the HSCN design model, a set of scenarios are generated from a single data source, represented by Ψ.

There are some costs related to the design and some others for the operation of the HSCN. The former includes the cost of selecting hubs, represented by the parameter f_i, $i \in V$, the selection cost of the inventory resources assigned to warehouses, represented by the parameter $\hat{f}_i$, $i \in V_W$, and the cost of selecting transportation resources for services, represented by the parameter $\hat{f}_{ij}$, $(i, j) \in A$. We model the operational cost by the parameter c^k_{ij}, which represents the cost of transporting one unit of critical supply $k \in K$ by service $(i, j) \in A$. The parameter u_{ij} illustrates the capacity of one unit of transportation resource of the service $(i, j) \in A$, and the parameter u_i, $i \in V_W$ expresses the capacity of one unit of inventory resources. The parameter m_{ij} defines the maximum number of transportation resources available for the service $(i, j) \in A$, and the maximum number of inventory resources available for the warehouse $i \in V_W$ is indicated by the parameter m_i.

The parameter z^0 represents the initial budget, and the parameters z^t indicates the received donations at each period $t \in T$. The parameter n^{kt}_i demonstrates the maximum quantity of each critical supply $k \in K$ that can be made available at a point of entry hub $i \in V_I$ at period $t \in T$. The parameter $g^t_{i\psi}$ represents the percentage of available inventory resources of hub $i \in V_W$ at period $t \in T$, in scenario $\psi \in \Psi$. Furthermore, the parameter $g^t_{ij\psi}$ indicates the percentage of available transportation resources of service $(i, j) \in A$ at period $t \in T$. The base demand for critical supply $k \in K$ of a beneficiary group $l \in L$ at period $t \in T$ in scenario $\psi \in \Psi$ is represented by the parameter $d^{kt}_{l\psi}$. The parameter b^k specifies the penalty of unmet demand for the critical supply $k \in K$. The parameter $s^{k'k}$ represents the spread factor for one unit of the critical supply $k' \in K$ over the critical supply $k \in K$. Finally, the parameter $\hat{d}^{kt}_{l\psi}$ illustrates the total demand of the beneficiary group $l \in L$ for the critical supply $k \in K$ at period $t \in T$ in scenario $\psi \in \Psi$.

Table 1: Decision variables of the two-stage stochastic model.

First-stage	
$x_{ij} \in \{0,1\}$	1 if service $(i,j) \in A$, is selected to be part of the HSCN; 0 otherwise.
$y_i \in \{0,1\}$	1 if hub $i \in V$, is selected to be part of the HSCN; 0 otherwise.
$\hat{x}_{ij} \in \mathbb{N}^0$	Number of units of transport resources selected for service $(i,j) \in A$.
$\hat{y}_i \in \mathbb{N}^0$	Number of units of inventory resources selected for hub $i \in V_W$.
$a_{il} \in \{0,1\}$	1 if beneficiary group $l \in L$, is assigned to DC $i \in V_{DC}$; 0 otherwise.
Second-stage	
$\bar{x}^{kt}_{ij_\psi} \geq 0$	Quantity of critical supply $k \in K$ transferred through service $(i,j) \in A$ at period $t \in T$ in scenario $\psi \in \Psi$.
$\bar{a}^{kt}_{il_\psi} \geq 0$	Quantity of critical supply $k \in K$ at period $t \in T$ allocated to beneficiary group $l \in L$ from DC $i \in V_{DC}$ in scenario $\psi \in \Psi$.
$r^{kt}_{i_\psi} \geq 0$	Inventory level (in number of units) of critical supply $k \in K$ at warehouse $i \in V_W$ at the end of period $t \in T$ in scenario $\psi \in \Psi$.
$\hat{r}^{kt}_{i_\psi} \geq 0$	Inventory level (in number of units) of critical supply $k \in K$ at warehouse $i \in V_W$ at the beginning of period $t \in T$ in scenario $\psi \in \Psi$.

Table 1 defines the decision variables of our model. Starting from the decision variables that are made in the first stage, the decision variable x_{ij} indicates if the service $(i, j) \in A$ is included in the HSCN, and the decision variable $\hat{x}_{ij}$ denotes the number of transport resources that are assigned to the service. Likewise, the decision variable y_i takes value 1 if the hub $i \in V$ is part of the network, and 0 otherwise. Decision variable $\hat{y}_i$ indicates the number of inventory resources assigned to the warehouse $i \in V_W$. Finally, decision variable a_{il} takes value 1 if the beneficiary group $l \in L$ is assigned to the DC $i \in V_{DC}$, and 0 otherwise.

The following are the second-stage decision variables of the model. The decision variable $\bar{x}^{kt}_{ij_\psi}$ indicates the quantity of the critical supply $k \in K$ transferred at period $t \in T$ on service $(i, j) \in A$ in scenario $\psi \in \Psi$. Furthermore, the decision variable $\bar{a}^{kt}_{il_\psi}$ indicates the quantity of critical supply $k \in K$ allocated to beneficiary group $l \in L$ at period $t \in T$ from DC $i \in V_{DC}$ in scenario $\psi \in \Psi$. $\hat{r}^{kt}_{i_\psi}$ represents the level of inventory of warehouse $i \in V_W$ for critical supply $k \in K$ at the beginning of period $t \in T$ in scenario $\psi \in \Psi$, and the decision variable $r_{i_\psi}^{kt}$ represents the inventory level at the end of that time period.

$$\min \sum_{\psi\in\Psi} p_\psi Q_\psi(\hat{x},\hat{y},a) \tag{2}$$

$$s.t.\quad 2x_{ij} \le y_i + y_j \quad \forall (i,j)\in A, \tag{3}$$

$$\hat{y}_i \le m_i y_i \quad \forall i \in V_W, \tag{4}$$

$$\hat{x}_{ij} \le m_{ij} x_{ij} \quad \forall (i,j)\in A, \tag{5}$$

$$\sum_{i\in V} f_i y_i + \sum_{i\in W} \hat{f}_i \hat{y}_i + \sum_{(i,j)\in A} \hat{f}_{ij}\hat{x}_{ij} \le z^0, \tag{6}$$

$$\sum_{i\in V_{DC}} a_{il} = 1 \quad \forall l\in L, \tag{7}$$

$$a_{il} \le y_i \ \ \forall i\in V_{DC},\ \forall l\in L, \tag{8}$$

$$\hat{x}_{ij}\in\mathbb{N}^0,\ \hat{y}_i\in\mathbb{N}^0,\ x_{ij}\in\{0,1\},\ y_i\in\{0,1\}, a_{il}\in\{0,1\},\ \forall i\in V,\ \forall (i,j)\in A. \tag{9}$$

Where $Q_\psi(\hat{x},\hat{y},a)$ calculates the minimum penalty over the defined periods for the scenario $\psi\in\Psi$ with first stage decision variables values being fixed to $\hat{x},\hat{y},a$. $Q_\psi(\hat{x},\hat{y},a)$ is defined as follows:

$$Q_\psi(\hat{x},\hat{y},a) := min \sum_{t\in T}\sum_{k\in K} b^k \sum_{l\in L}(\hat{d}_{l\psi}^{kt} - \sum_{i\in V_{DC}} \bar{a}_{il_\psi}^{kt}), \tag{10}$$

$$s.t.\quad \sum_{k\in K} \bar{x}_{ij_\psi}^{kt} \le u_{ij} g_{ij_\psi}^t \hat{x}_{ij} \quad \forall (i,j)\in A,\ \forall t\in T, \tag{11}$$

$$\bar{a}_{il_\psi}^{kt} \le \sum_{(j,i)\in A} u_{ji} g_{ji_\psi}^t m_{ji} a_{il}, \quad \forall i\in V_{DC},\ \forall l\in L,\ \forall k\in K,\ \forall t\in T, \tag{12}$$

$$\bar{a}_{il_\psi}^{kt} \le \hat{d}_{l\psi}^{kt}, \quad \forall i\in V_{DC},\ \forall l\in L,\ \forall k\in K,\ \forall t\in T, \tag{13}$$

$$\sum_{l\in L} \bar{a}_{il_\psi}^{kt} = \sum_{j\in W} \bar{x}_{ji_\psi}^{kt}, \quad \forall i\in V_{DC}, \forall k\in K,\ \forall t\in T, \tag{14}$$

$$\hat{d}_{l\psi}^{kt} = d_{l\psi}^{kt} + \sum_{k'\in K} s^{k'k}(\hat{d}_{l\psi}^{k't-1} - \sum_{i\in V_{DC}} \bar{a}_{il_\psi}^{k't-1}), \quad \forall l\in L,\ \forall k\in K,\ \forall t\in T, \tag{15}$$

$$\sum_{i\in V} f_i y_i + \sum_{i\in W}\hat{f}_i\hat{y}_i + \sum_{(i,j)\in A}\hat{f}_{ij}\hat{x}_{ij} + \sum_{t'=1}^{t}\sum_{(i,j)\in A}\sum_{k\in K} c_{ij}^k \bar{x}_{ij_\psi}^{kt'} \le z^0 + \sum_{t'=1}^{t} z^{t'}, \quad \forall t\in T, \tag{16}$$

$$\hat{r}_{j_\psi}^{kt} \le r_{j_\psi}^{kt-1} \quad \forall j\in V_W,\ \forall k\in K,\ \forall t\in T, \tag{17}$$

$$\sum_{k\in K} \hat{r}_{j_\psi}^{kt} \le u_j g_{j_\psi}^t \hat{y}_j \quad \forall j\in V_W,\ \forall t\in T, \tag{18}$$

$$\sum_{k\in K} r_{j_\psi}^{kt} \le u_j g_{j_\psi}^t \hat{y}_j \quad \forall j\in V_W,\ \forall t\in T, \tag{19}$$

$$r_{j_\psi}^{kt} = \hat{r}_{j_\psi}^{kt} + \sum_{(i,j)\in A} \bar{x}_{ij_\psi}^{kt} - \sum_{(j,i)\in A} \bar{x}_{ji_\psi}^{kt}, \quad \forall j\in V_W,\ \forall k\in K,\ \forall t\in T, \tag{20}$$

$$\sum_{(i,j)\in A} \bar{x}^{kt}_{ij_\psi} \leq n^{kt}_i \quad \forall i \in V_I,\ \ \forall k \in K,\ \forall t \in T, \tag{21}$$

$$\bar{x}^{kt}_{ij_\psi} \geq 0,\ \bar{a}^{kt}_{il_\psi} \geq 0,\ r^{kt}_{i_\psi} \geq 0,\ \hat{r}^{kt}_{i_\psi} \geq 0,\ \forall (i,j) \in A,\ \forall i \in V,\ \forall k \in K,\ \forall t \in T. \tag{22}$$

As a two-stage HSCN design model, the objective function (2) minimizes the expected penalty of unmet demand over the set of scenarios Ψ. Constraints (3) indicate that only services with selected hubs at their origin and destination are permitted for selection. Constraints (4) limit the number of inventory resources at each warehouse to the maximum available inventory resources at that warehouse. Similarly, constraints (5) limit the number of transportation resources for each transportation service to the maximum number of available transportation resources for that transportation service. Constraints (6) limit the total cost of selecting hubs and assigning resources to warehouses and transportation services to the initial budget. Constraints (7) ensure that each beneficiary group is assigned to exactly one DC. Constraints (8) limit the assignment of beneficiary groups to DCs that are selected to be part of the HSCN. Constraints (9) indicate the bounds of the decision variables.

The objective function (10) minimizes the total penalty of unmet demand over the planning horizon for a given scenario ψ. Constraints (11) limit the quantity of transported critical supplies over services to the available capacity of services at each time period. Constraints (12) ensure that the allocated amount of critical supplies to each beneficiary group from each DC are limited by the maximum amounts of the critical supplies received by the DC at each period. Constraints (13) limit the allocated critical supplies to each beneficiary group to the total demand of that beneficiary group at each period. Constraints (14) ensure that the total quantity of critical supplies that are delivered to each DC is equal to the total quantity of critical supplies allocated to beneficiary groups at each period. Constraints (15) formulate the total demand of each beneficiary group for each critical supply at each period as the summation of the base demand and the residual demand of that critical supply. Constraints (16) ensure that the amount for the overall costs of the first stage and first t periods that is paid, is limited by the summation of the initial budget and the received donation up to that period.

Constraints (17) restrict the inventory level of each critical supply at each warehouse at each period by its level at the end of the last period. Constraints (18) limit the inventory level of critical supplies at the beginning of each period to the available inventory capacity of that warehouse in that period. Similarly, Constraints (19) limit the inventory level of critical supplies at the end of each period to the available inventory capacity of that warehouse in that period. Constraints (20) indicate the inventory level of critical supplies at the end of each period as the sum of the inventory level at the beginning of that period and the received quantity of critical supplies at that period minus the quantity of shipped critical supplies to DCs at that period. Constraints (21) ensure that the quantity of shipped critical supplies from each port of entry is limited by the capacity of each port of entry at that period. Constraint (22) are the non-negativity requirements imposed on the all the second-stage decision variables.

4.2 Proposed HSCN design models

The HSCN design model ignores the ambiguity in the obtained estimates from multiple data sources. This subsection introduces the HSCN design models that explicitly handle the discussed ambiguity in the problem and provide alternative optimization methods for the HSCN design model.

Assume the decision-makers should make a series of decisions, represented here by $\boldsymbol{x} \in X$, X indicating the feasible set for the decisions while facing uncertainty represented here by parameter vector ξ. We further assume $F(\boldsymbol{x}, \xi)$ represents the function the decision makers seek to optimize and computes the penalty obtained by using the decision vector x when the uncertain parameters get the value ξ. Let there be e data sources with $\Psi_1, \Psi_2, ..., \Psi_e$ being their corresponding scenario sets. The set containing all scenario sets obtained from available data sources is called the ambiguity set (Bayraksan and Love, 2015), represented by $\mathbb{P} := \{\Psi_1, \Psi_2, ..., \Psi_e\}$. We then define solution $\boldsymbol{x}^*_i$, $i \in \{1, 2, ..., e\}$ as the solution that obtains the minimum expected value of function $F(\boldsymbol{x}, \xi)$ for all possible values of $\xi \in \Psi_i$:

$$\boldsymbol{x}^*_i \in arg \min_{\boldsymbol{x}\in X} \mathbb{E}_{\xi\in\Psi_i}[F(\boldsymbol{x}, \xi)],\ \ \forall\ i \in \{1, 2, ..., e\} \tag{23}$$

Assuming $e = 1$, then (23) delivers a single solution $\boldsymbol{x}_1^*$. However, we have e data sources available, and we assume that:

$$\boldsymbol{x}_i^* \neq \boldsymbol{x}_j^* \quad \forall i \neq j \quad \text{and} \quad i, j \in \{1, 2, ..., e\}.$$

We define the opportunity loss of i'th data source when using solution $\boldsymbol{x}$, represented by $\epsilon_i(\boldsymbol{x})$, as the disparity between minimum expected value derived from the optimal solution $\boldsymbol{x}_i^*$ and the attained value when using solution $\boldsymbol{x}$, mathematically formulated as:

$$\epsilon_i(\boldsymbol{x}) := \mathbb{E}_{\xi \in \Psi_i}[F(\boldsymbol{x}, \xi)] - \mathbb{E}_{\xi \in \Psi_i}[F(\boldsymbol{x}_i^*, \xi)].$$

Then, we assume, employing the optimal solution $\boldsymbol{x}_j^*$, derived from considering scenario set Ψ_j, within the optimization function associated with the scenario set Ψ_i, results in a significant opportunity loss:

$$\epsilon_i(\boldsymbol{x}_j^*) >> \epsilon_i(\boldsymbol{x}_i^*) = 0 \quad \forall i, j \in \{1, 2, ..., e\}.$$

One then seeks to find a single solution $\boldsymbol{x}^*$ such that:

$$\boldsymbol{x}^* \in X \quad and \quad \epsilon_i(\boldsymbol{x}^*) \approx 0 \quad \forall i \in \{1, 2, ..., e\}.$$

Minimization of expected opportunity loss

We first present a goal programming (Charnes and Cooper, 1957) approach where minimizing data source-specific expected penalties are treated as distinct goals to reach. We refert this model to one of *Minimization of expected opportunity loss (MIN-OppLoss).* Its objective (24) is defined as the total expected opportunity loss by associating the same weight to the deviations from each data source-specific target value.

$$\min_{\boldsymbol{x}} \sum_{i=1}^{e} \mathbb{E}_{\xi \in \Psi_i}[F(\boldsymbol{x}, \xi)] - \mathbb{E}_{\xi \in \Psi_i}[F(x_i^*, \xi)] \tag{24}$$

The second part of the objective, $\mathbb{E}_{\xi \in \Psi_i}[F(x_i^*, \xi)]$, being a constant, can be removed from the formulation, making objectives (24) and (25) equivalent.

$$\min_{\boldsymbol{x}} \sum_{i=1}^{e} \mathbb{E}_{\xi \in \Psi_i}[F(\boldsymbol{x}, \xi)] \tag{25}$$

This model is equivalent to the commonly used model in the literature (Daneshvar et al., 2023) that does not specifically consider ambiguity in the HSCN design problem. In other words, its considered distribution is defined as the union of the individual distributions of the various data sources.

In order to define the MIN-OppLoss model, one can replace the objective function (2) with the objective function (26). In an alternative view, Minimization of expected opportunity loss approach could be interpreted as the equivalent of the HSCN design model where Ψ is replaced by Ψ_{Total}. In (26), p_ξ represents the probability of scenario ξ if data source Ψ is providing the accurate estimation of uncertain parameters.

$$\min \sum_{i=1}^{e} \left[\sum_{\psi \in \Psi_i} p_\psi Q_\psi(\hat{x}, \hat{y}, a) \right] \tag{26}$$

subject to Constraints (3), (4), (5), (6), (7), (8), (9).

Minimization of Maximum Scenario Penalty

Using the classical robust optimization approach (Soyster, 1973), minimizing the worst-case scenario outcome using the scenario set $\Psi_{total} := \bigcup_{i=1}^{e} \Psi_i$. In other words, it provides a robust solution against uncertainties by considering the most adverse outcome while maintaining feasibility:

$$\min_{\boldsymbol{x}} \max_{\xi \in \Psi_{total}} F(\boldsymbol{x}, \xi). \tag{27}$$

We refer to this model as the *MIN-MaxScenPen* model by scenario. The obtained solution is expected to perform worse compared to other presented models on most of the realization of uncertain parameters while resulting in less harm in extreme scenarios.

Developing the Minimization of Maximum Scenario Penalty model is achieved by introducing an auxiliary decision variable, Θ, in the objective function, replacing the original objective function (2). Additionally, the constraint (29) is added to the first stage of the model, limiting the objective function value based on the penalties associated with each scenario in the scenario set Ψ_{Total}. Furthermore, constraint (30) indicates the bounds of the decision variable Θ.

$$\min \quad \Theta \tag{28}$$

subject to constraints (3), (4), (5), (6), (7), (8), (9), (29), and (30).

$$\Theta \geq Q_{\psi}(\hat{x}, \hat{y}, a), \quad \forall \psi \in \Psi_{total} \tag{29}$$

$$\Theta \geq 0. \tag{30}$$

Minimization of Expected Data-Source Penalty

The MIN-MaxScenPen model focuses on extreme cases, leading to overly cautious decisions and not capturing the full range of possible scenarios. To expand the number of scenarios involved in the solution using the concept of data source, we propose a model *Minimization of Expected Data-Source Penalty (MIN-ExpDSPen)*, which is based on robust optimization and aims to minimize data source level expected penalty. The MIN-ExpDSPen model defines an objective that minimizes the maximum expected penalty of data source-specific scenarios within the ambiguity set, as presented by objective (31):

$$\min_{\boldsymbol{x}} \max_{\Psi_i \in \mathbb{P}} \mathbb{E}_{\xi \in \Psi_i}[F(\boldsymbol{x}, \xi)] \tag{31}$$

The aim of this objective is to find a solution x that minimizes the highest level of expected penalty among the scenario sets in $\mathbb{P}$.

In this model, the objective function (32) replaces the objective function (2). Furthermore, we add the constraint (30) to the first stage of the model. Constraints (33) ensure that the objective function value of the first stage is more than the expected penalty of the designed HSCN over the scenario set generated from estimations obtained from each data source.

$$\min \quad \Theta \tag{32}$$

subject to constraints (3), (4), (5), (6), (7), (8), (9), (30) and:

$$\Theta \geq \sum_{\psi \in \Psi_i} p_{\psi} Q_{\psi}(\hat{x}, \hat{y}, a), \quad \forall i = 1, 2, \dots e. \tag{33}$$

Minimization of Maximum Data-Source Penalty

While MIN-OppLoss approach minimizes the expected penalty over scenarios from all data sources, it does not consider the variance of the opportunity loss. In other words, the obtained HSCN may perform inadequately over some data sources and very well on others. In contrast, the here proposed model *Minimization of Maximum Data-Source Penalty (MIN-MaxDSPen)*, grounded in robust optimization, aims to address the variability in opportunity losses among scenario sets in $\mathbb{P}$. Consequently, this method minimizes the maximum opportunity loss within each data source's scenario set, as depicted in objective (34).

$$\min_{\boldsymbol{x}} \max_{\Psi_i \in \mathbb{P}} \mathbb{E}_{\xi \in \Psi_i}[F(\boldsymbol{x}, \xi)] - \mathbb{E}_{\xi \in \Psi_i}[F(\boldsymbol{x}_i^*, \xi)]. \tag{34}$$

Objective (34) provides a solution x with the minimum opportunity loss among all data sources.

To model this approach, the objective function (35) replaces the objective function (2). The range of the expected penalty in the HSCN design problem defines the domain of Θ. Therefore, the domain of the decision variable Θ defined by constraint (30) is in the range of positive real numbers. Constraints (36) in the first stage ensure that the value of auxiliary decision variable Θ is always more than the gap between the expected penalty of designed HSCN in this model and HSCNs designed by HSCN design model using Ψ_i, Where i could point to any data source from 1 to e.

$$\min \quad \Theta \tag{35}$$

subject to constraints (3), (4), (5), (6), (7), (8), (9) and:

$$\Theta \geq \sum_{\psi \in \Psi_i} p_\psi(Q_\psi(\hat{x}, \hat{y}, a) - Q_\psi(\hat{x}_i^*, \hat{y}_i^*, a_i^*)), \quad \forall i = 1, 2, ..., e. \tag{36}$$

5 Experimental results

In this section, we design and apply a set of experiments to study the performance of the proposed models for the HSCN design problem. Subsection 5.1 introduces the data set, including the characteristics of the natural disaster, the affected region, the sources used in the data preparation, and the scenario generation. Subsection 5.2 presents the experimental results, including a Pareto frontier analysis, a ranked-based analysis, and a comparative performance analysis over the solutions obtained by executing the proposed models on the introduced instances. Finally, managerial insights are presented.

5.1 Data set

We use a data set (Daneshvar et al., 2023) from the 2018 earthquake in Lombok island at Indonesia. More than 1500 aftershocks have been recorded in the region but most of them were weak shakes. The most important quakes in the region are presented in Table 2. The earthquake forced 445,343 individuals to relocate into 2,700 camps on Lombok and the neighboring islands. The Indonesian government announced a state of emergency from July 29th to August 26th, which is here considered as the planning horizon. We divide the planning horizon into 4 time periods, each presenting a week during the state of emergency. The International Organization for Migration (IOM) has published the list of all camps including the location and number of individuals in each camp (IOM, 2019). In this study, we consider 349 beneficiary groups on the island with a total population of 52,128 individuals in 15,993 households. Considering that clean water was distributed among beneficiary groups from local resources using 21 water trucks (IFRC, 2019), we only consider shelter, food, and hygiene packs as the critical supplies that are brought in from outside the affected region and to be transported and distributed using the designed HSCN. We use the standard required quantity of each critical supply (IFRC, 2021) per individual or per household that is calculated and published by the International Federation of Red Cross and Red Crescent Societies (IFRC).

earthquake	date	strength
main earthquake	2018/07/29	6.4 Richter magnitude scale
first strong aftershock	2018/08/05	7.0 Richter magnitude scale
second strong aftershock	2018/08/09	5.9 Richter magnitude scale
third strong aftershock	2018/08/26	6.4 Richter magnitude scale

Table 2: Most important earthquakes on Lombok island in 2018.

Palang Merah Indonesia (PMI) is the local branch of IFRC in Indonesia which was responsible for the distribution of critical supplies in the affected region. We used the published reports of IFRC and PMI to complete our data set (IFRC, 2019). According to these reports, PMI used four ports of entry and six warehouses in their HSCN. Furthermore, PMI signed contracts with third-party companies to transport critical supplies among the hubs. However, since the details of the contracts are not included within the reports of the IFRC and PMI, we consulted the local transportation companies' websites for the cost and capacity of their services.

Since the locations of the DCs are not provided in the IFRC reports, we use the DBSCAN algorithm (Ester et al., 1996) to generate DCs using the beneficiary groups as the candidate locations. The DBSCAN algorithm uses two parameters: the epsilon parameter that denotes the neighborhood radius of the DCs in the same cluster, and the minimum number of neighbors to cluster the beneficiary groups based on distance and density. The value of these parameters is set by a domain expert, leading to the most appropriate cluster for the study problem (Mendes and Cardoso, 2006).

The transportation costs are calculated based on the driving distance between hubs. The walking distances between beneficiary groups for the DBSCAN algorithm are obtained from an online routing engine (Luxen and Vetter, 2011), which operates on the OpenStreetMap.

In addition to the data associated with the deterministic parameters, the demand and damage assessments provide the necessary information to estimate the uncertain parameters, including demand, transportation, and inventory capacities. Since the assessments are time-consuming processes, the affected region is divided into smaller sub-regions (e.g., 81 sub-regions in this case study) to speed up the process, and the damage and demand assessments are performed on a set of locations sampled from each sub-region (Balcik and Yanıkoğlu, 2020; Balcik, 2017). In the following experiments, we assume two data sources are available, providing estimations on the value of uncertain parameters. The estimations derived from these two data sources are inconsistent, with the first data source always yielding more pessimistic estimates than the second data source.

5.1.1 Ambiguity Patterns

We here focus on four different ambiguity patterns, illustrated in Figure 2, which characterize the different relationships that two different distributions can have to each other. Ambiguity pattern (a) represents the estimations provided by two data sources, each with a high level of uncertainty and no overlap, causing a high level of ambiguity. Ambiguity pattern (b) contains the same level of uncertainty as ambiguity pattern (a), as they both have the same range of estimation for uncertain parameters. However, as the estimations provided by the two data sources overlap, the level of ambiguity in (b) is lower than (a). The mode of the distributions in the ambiguity pattern (c) is the same as in the ambiguity pattern (a). However, the range of the distributions in (c) is less than (a), reducing both uncertainty and ambiguity levels. Finally, the probability distributions presented in ambiguity pattern (d) have the same range as ambiguity pattern (c), but there is no gap between the two distributions, reducing the ambiguity in (d) compared to (c). We can consider each of these four distinct ambiguity patterns for each problem instance.

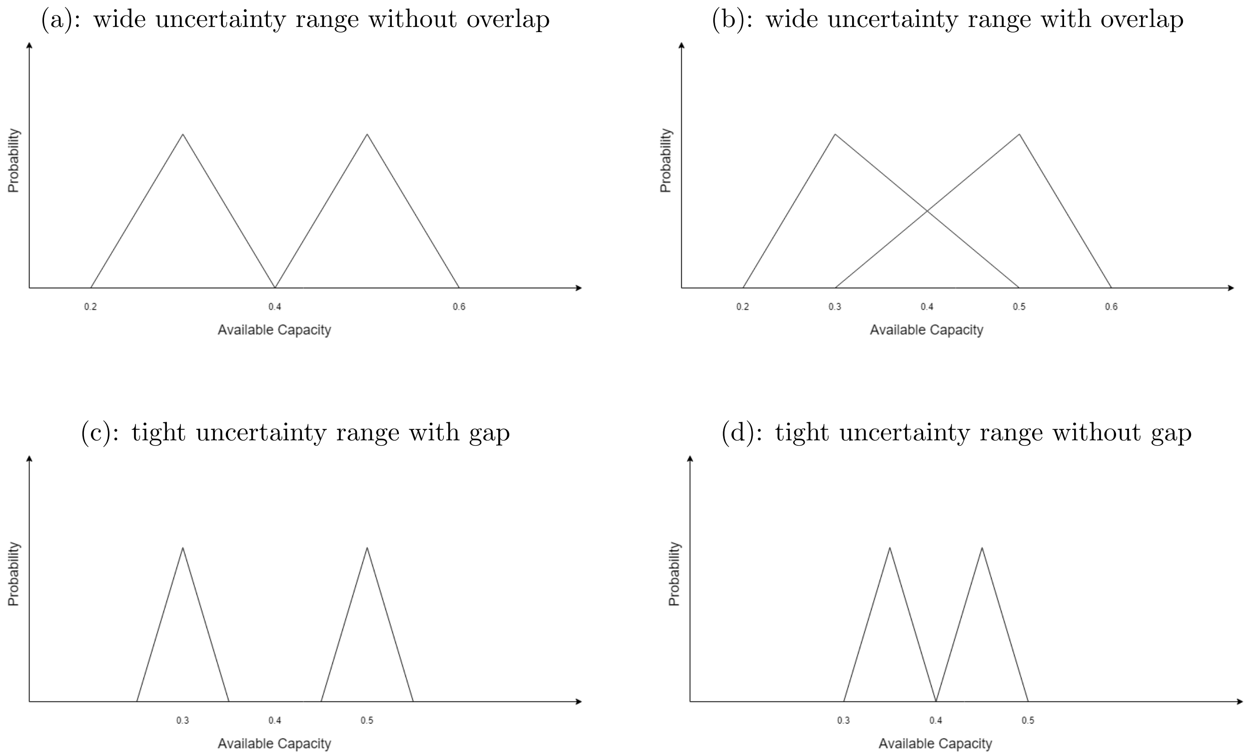

Figure 2: The four considered ambiguity patterns for two data sources.

5.1.2 Scenario Sampling

As we do not have access to the raw assessment data of the earthquake on Lombok Island, we simulated triangular distributions (Hakimifar et al., 2021; Benini et al., 2017). The minimum and maximum values of the triangular distributions are set within the range of the data gathered from humanitarian organizations' websites IFRC (2019). To approximate the proposed models, as well as to evaluate the performance of the obtained solutions, a set of scenarios is required that effectively captures the different variations of the uncertain parameters. Assuming the same confidence level for all data sources used to obtain the probability distributions, an equal number of scenarios are generated from each probability distribution. Furthermore, we consider equal probability for all scenarios generated from each triangular distribution. A total of 3000 scenarios are generated from each data source (6000 per problem instance), which we here assume to represent the ground truth (i.e., an accurate estimation of the possible realization of the uncertain parameters). However, since solving such a problem would be computationally intractable, using the Sample Average Approximation method (Kleywegt et al., 2002), we generate smaller scenario sets (i.e., sample scenario sets) with 300 scenarios per data source and solve the models for such smaller scenario sets. Each instance is then composed of two ground truths (one per data source) and two sample scenario sets.

5.1.3 Instance Generation

We generate multiple problem instances, each including one ground truth and one sample scenario set. Each instance is generated using the data from the 2018 earthquake in Indonesia. First, a subset of the beneficiary groups containing at least 80 percent of the 349 beneficiary groups is randomly selected for each instance to enhance the variability among instances. The DBSCAN algorithm then generates the candidate DCs. Two or three candidate warehouses and points of entry are selected at random, and a set of candidate services is added between hubs in different layers. Finally, the available budget depends on a budget ratio parameter relative to the *population* size. Equation (37) defines how the budget ratio is formulated.

$$budget\ \ ratio = \frac{z^0 + \sum_{t \in T} z^t}{population}. \tag{37}$$

The first ten instances are generated with a budget ratio of 640, as used in (Daneshvar et al., 2023). Ten additional instances are generated with a budget ratio of 512, computed by considering only 80 percent of the former budget ratio. This amounts to a total number of 20 instances.

5.2 Computational Results

This section presents the experimental results to determine the most suitable model for decision-makers to adopt under each ambiguity pattern, and based on their preference for either optimism or conservatism. To this end, we first analyze the Pareto frontier to evaluate the dominance of the obtained solutions in Section 5.2.1. Then, a ranking analysis is carried out in Section 5.2.2 to identify the best performing models under different ambiguity patterns. Finally, a comparative performance analysis complements the previous studies in Section 5.2.3, identifying average performance of the models and their relative performance to the competing models. Managerial insights are then summarized in Section 5.2.4.

The data for the above-mentioned analyses is prepared as follows. For each instance-ambiguity pattern, the sample scenario sets are used to obtain two solutions, each using one of the data sources. The solutions are then evaluated using the corresponding instance's ground-truth scenario set. In some instances, resources and budgets are sufficiently high, causing all models' solutions to have negligible differences in performance and to satisfy almost all the demand. In the following experiments, we exclude instances with a percentage difference of 2% (i.e., the best and worst solution evaluation gap is less than two percent of the best solution). The reason is that such a small percentage gap provides limited insights into the relative efficacy of the alternative models. With equal confidence levels attributed to both data sources in each instance, the following analysis presents findings outlined according to the evaluation results of the studied instances. The implementation employs the Pyomo software package (Hart et al., 2011, 2017), executed on the Calcul Québec servers with a computational infrastructure featuring 6 CPU cores and 256 GB of memory.

5.2.1 Pareto Frontier Analysis

The Pareto frontier represents the set of optimal solutions where enhancing one criterion comes at the expense of another, highlighting the inherent trade-offs in multi-objective decision-making problems. In the here-studied problem, we have two ground-truth scenario sets, with the expected penalty of each solution for each ground-truth serving as one criterion. To better understand the models' relative performance in different ambiguity patterns, we calculated the number of instances each model locates on the Pareto frontier. The Pareto frontier consists of all feasible solutions that are not dominated by any other feasible solution. A solution is dominated if another solution is better for at least one objective and no worse for the others. Figure 3 outlines the proportion of instances within each model wherein a solution is attained on the Pareto frontier.

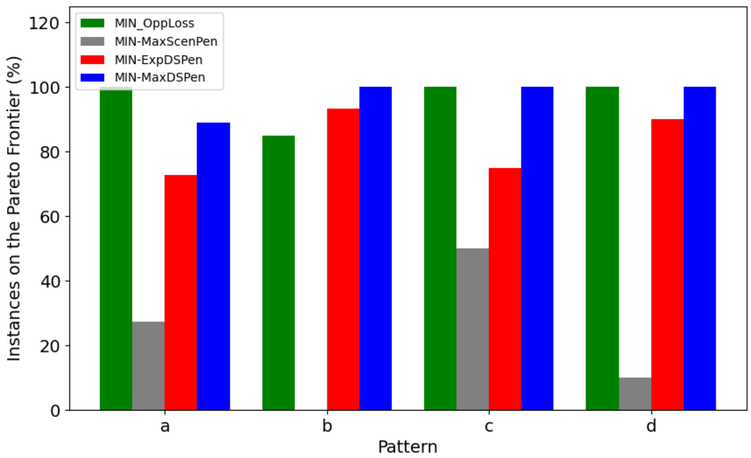

Figure 3: The percentage of instances each model's solution was on the Pareto frontier.

None of the models consistently attains Pareto optimality. Considering the ambiguity patterns (c) and (d), it is noteworthy that the MIN-OppLoss and MIN-MaxDSPen models consistently reside on the Pareto frontier (i.e., the solution is not dominated by any other solution for any criteria). However, for ambiguity pattern (a), only the MIN-OppLoss model consistently lies on the Pareto frontier, while MIN-MaxDSPen lies on the frontier most of the time. Meanwhile, for ambiguity pattern (b), the MIN-MaxDSPen model consistently lies on the Pareto frontier.

While no model obtains nondominated solutions for all ambiguity patterns, MIN-OppLoss and MIN-MaxDSPen solutions are always nondominated when only considering ambiguity patterns (c) and (d). While such an attribute indicates the value of MIN-OppLoss and MIN-MaxDSPen models, in many real-world HSCN problems, indicating the ambiguity pattern is a complex task. We hence perform additional experiments to gain more insights into the proposed models that are valid for all ambiguity patterns.

5.2.2 Ranking Analysis

Our interest lies in tracing the individual performance of models concerning each data source. To this end, we now analyze the ranking of solutions obtained by the models across various instances. This approach affords an understanding of how effectively the models address each element of the multi-objective optimization problem.

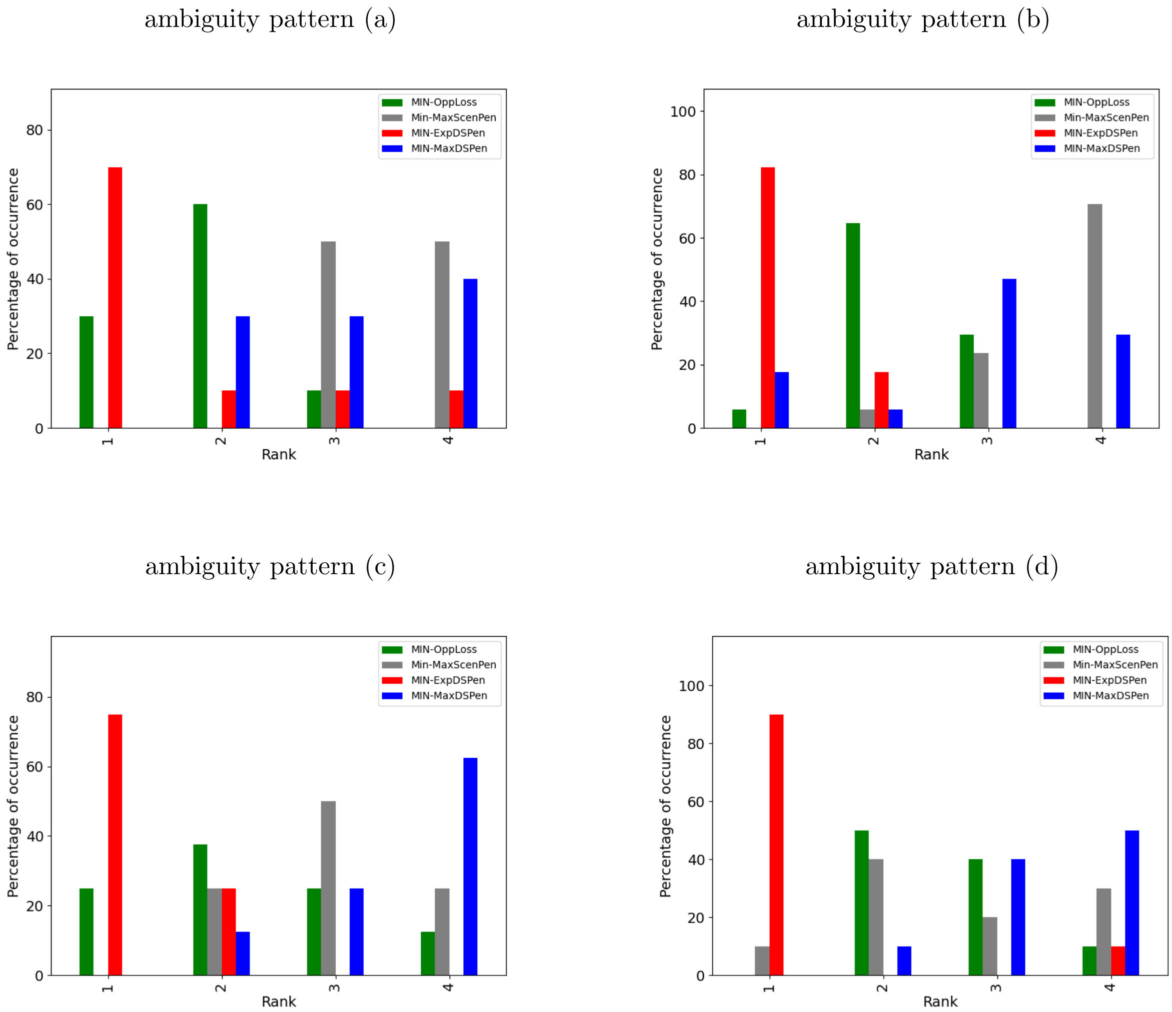

Figure 4: The ranking distribution of modes over the 20 instances on the first data source.

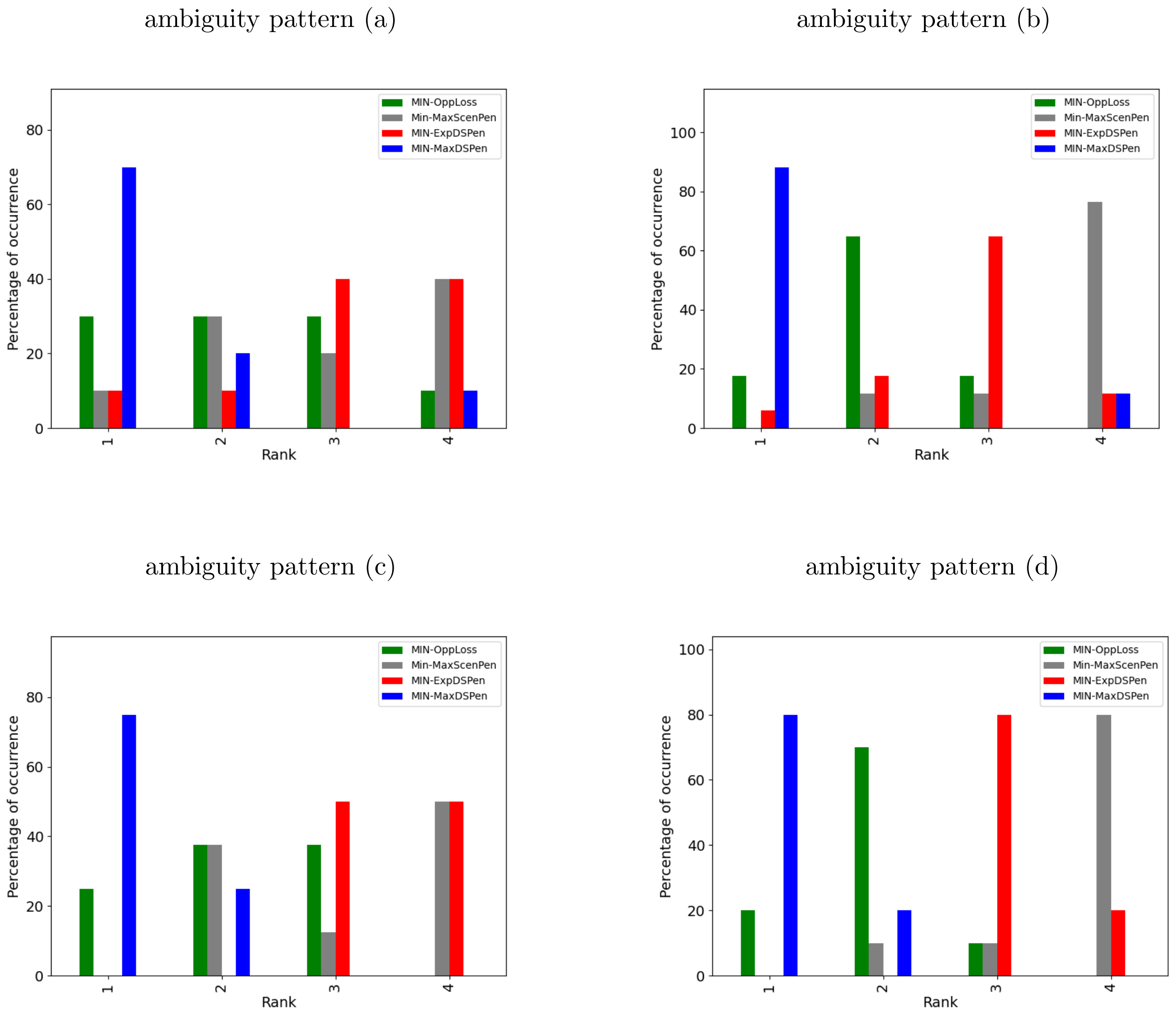

Figure 5: The ranking distribution of modes over the 20 instances on the second data source.

We evaluate the proposed models' solutions by ranking them according to their performance on the GTs, and then computing the frequency with which each model's solution obtains each rank. This procedure is carried out independently for each data source. Figure 4 and Figure 5 represent the ranking distributions of models for the first and second data sources, respectively. According to Figure 4, considering the first data source, MIN-ExpDSPen has the highest probability of providing the best-performing solution. Furthermore, Figure 5 indicates that MIN-MaxDSPen has the highest probability of providing the best solution when considering the second data source. Therefore, in a problem setting where the decision-makers are biased toward one of the data sources, regardless of the ambiguity pattern, the best choise is MIN-ExpDSPen when biased towards the first data source, and MIN-MaxDSPen when biased toward the second data source.

5.2.3 Comparative Performance Analysis

Within the context of the ranked-based analysis, it is noteworthy that none of the models consistently attain the highest ranking. For instance, when considering the first data source in the ranked-based analysis of ambiguity pattern (a), for 30% of instances, the MIN-ExpDSPen model fails to secure the first rank. To provide a clearer picture of which optimization models yield the most efficient results overall, we conduct a series of comparative analyses that directly assess the results obtained using each proposed model relative to the top-ranked model in each case. The motivation for these analyses is to offer decision-makers insights into the potential risks associated with selecting a particular optimization approach based on the observed trends of the ranking results. Moreover, in instances with equal confidence levels associated with the available data sources, the utility of the ranking analysis diminishes. We, therefore, use the performance gap (p-gap) and absolute performance gap (abs-p-gap) to evaluate the performance of the introduced models in all instances and across the considered ambiguity patterns. The evaluation of each solution represents the expected penalty of the solution over the ground truth. Therefore, the evaluations are converted to the p-gap, enabling comparisons to be conducted across the instances. For each instance, we identify the Best Evaluation Value (BEV) as the lowest penalty evaluated on the ground-truth among all four models. Then, equation (38) calculates the p-gap of each model.

$$\text{p-gap} = [(\text{model}'\text{s evaluation} - \text{BEV})/\text{BEV}] * 100 \tag{38}$$

Furthermore, the absolute gap between solutions obtained from different models, abs-p-gap is calculated as the gap between the model's evaluation and BEV, see equation (39).

$$\text{abs-p-gap} = [(\text{model}'\text{s evaluation} - \text{BEV})] \tag{39}$$

Table 3 presents the results of this experiment, including the average abs-p-gap and penalty over considered instances. The table is structured to show results for each data source separately, followed by the total mean across both data sources.

For each model and metric, results are presented separately for the first and second data sources. Each row represents the average absolute performance gap (abs-p-gap) and expected penalty for the specified model and ambiguity pattern. The "mean" columns provide the average values obtained for all ambiguity patterns for each data source.

The "Total mean" section shows the overall average values for both data sources combined, providing a comprehensive view of the models' performance across all data sources.

To increase the readability of the table, the total *penalty* values in Table 3 have been scaled down by a factor of one million.

When considering the first data source, the MIN-ExpDSPen model has the lowest *abs-p-gap* mean values, outperforming other models when using the uncertainty estimations provided by the first data source in the evaluation. MIN-ExpDSPen seeks to find a solution that minimizes the expected penalty of the data source with the highest expected penalty, hence outperforming other models when evaluated on the first (pessimistic) data source. Similarly, considering the second data source, the MIN-MaxDSPen outperforms other models. In particular, for the second data source, the conservative models, including MIN-ExpDSPen and MIN-MaxScenPen, have relatively high mean values. Furthermore, the MIN-OppLoss model does not explicitly consider data-source ambiguity. Therefore, MIN-MaxDSPen outperforms other models when the evaluation is performed on the optimistic data source.

Finally, the "Total mean" section presents the overall mean across both data sources. In other words, the values in the "Total mean" section indicate the opportunity loss over the instances considered. MIN-MaxDSPen has the least *abs-p-gap* value, making it an attractive option under data-source ambiguity.

Figure 6 presents the performance profile of the here studied models. The top figures represent the performance profile over one data source, and the figure at the bottom shows the performance profile when considering both data sources. In these figures, the x-axis represents the threshold of the *p-gap*, and the y-axis indicates the percentage of instances with a lower *p-gap* than the value indicated on the x-axis. The performance profile of the first data source indicates that the MIN-ExpDSPen model has the best *p-gap* for about 90 percent of the instances. For the remaining instances, its *p-gap* becomes rather high when compared to the other models. The second data source performance profile indicates that the MIN-MaxDSPen model outperforms other models by a considerable margin for almost 90% of the instances. Finally, the performance profile over both data sources suggests that MIN-MaxDSPen is superior to all other models, followed by MIN-OppLoss. Both the MIN-ExpDSPen and the MIN-MaxScenPen models underperform the other models.Overall, MIN-MaxDSPen shows clear benefits under data source ambiguity, being the best performing model for most instances. For instances where it is not the best-performing model, it underperforms other models less than its competitors do.

5.2.4 Managerial Insights

The following managerial insights can be summarized from the conducted experiments, providing a valuable understanding of the presented models accounting for the here-studied ambiguity in the HSCN design problem.

Model	Metric	pattern (a)	pattern (b)	pattern (c)	pattern (d)	mean
FIRST DATA SOURCE						
MIN-OppLoss	abs-p-gap	5.38	44.4	0.01	47.9	**24.42**
	penalty	248.38	158.40	301.47	136.80	**211.26**
MIN-MaxScenPen	abs-p-gap	9.29	52.6	0.00	29.1	**22.74**
	penalty	257.84	174.22	306.48	141.22	**219.94**
MIN-ExpDSPen	abs-p-gap	8.54	10.20	1.50	36.7	**14.23**
	penalty	247.18	156.64	300.22	134.74	**209.69**
MIN-MaxDSPen	abs-p-gap	4.67	26.7	0.00	46.5	**19.46**
	penalty	257.06	166.76	318.20	146.44	**222.11**
SECOND DATA SOURCE						
MIN-OppLoss	abs-p-gap	75.9	59.9	87.4	42.5	**66.42**
	penalty	174.84	31.17	34.15	45.44	**71.4**
MIN-MaxScenPen	abs-p-gap	14.7	1.80	37.4	49.5	**25.85**
	penalty	167.41	42.11	39.93	53.17	**75.65**
MIN-ExpDSPen	abs-p-gap	26.3	1.55	57.9	58.8	**36.13**
	penalty	177.40	36.96	39.77	50.48	**76.15**
MIN-MaxDSPen	abs-p-gap	8.51	0.0	23.8	37.4	**17.42**
	penalty	171.28	27.96	29.73	42.56	**67.88**
BOTH DATA SOURCES						
MIN-OppLoss	abs-p-gap	40.64	52.15	43.71	45.2	**45.42**
	penalty	211.61	94.78	167.81	91.12	**141.33**
MIN-MaxScenPen	abs-p-gap	11.99	27.2	18.70	39.3	**24.29**
	penalty	212.63	108.17	173.21	97.20	**147.79**
MIN-ExpDSPen	abs-p-gap	17.42	5.88	29.7	47.75	**25.18**
	penalty	212.29	96.8	170.00	92.61	**142.92**
MIN-MaxDSPen	abs-p-gap	6.59	13.35	11.90	41.95	**18.44**
	penalty	214.17	97.36	173.97	94.5	**144.99**

Table 3: The average abs-p-gap and expected penalty values (in millions) over the studied problem instances

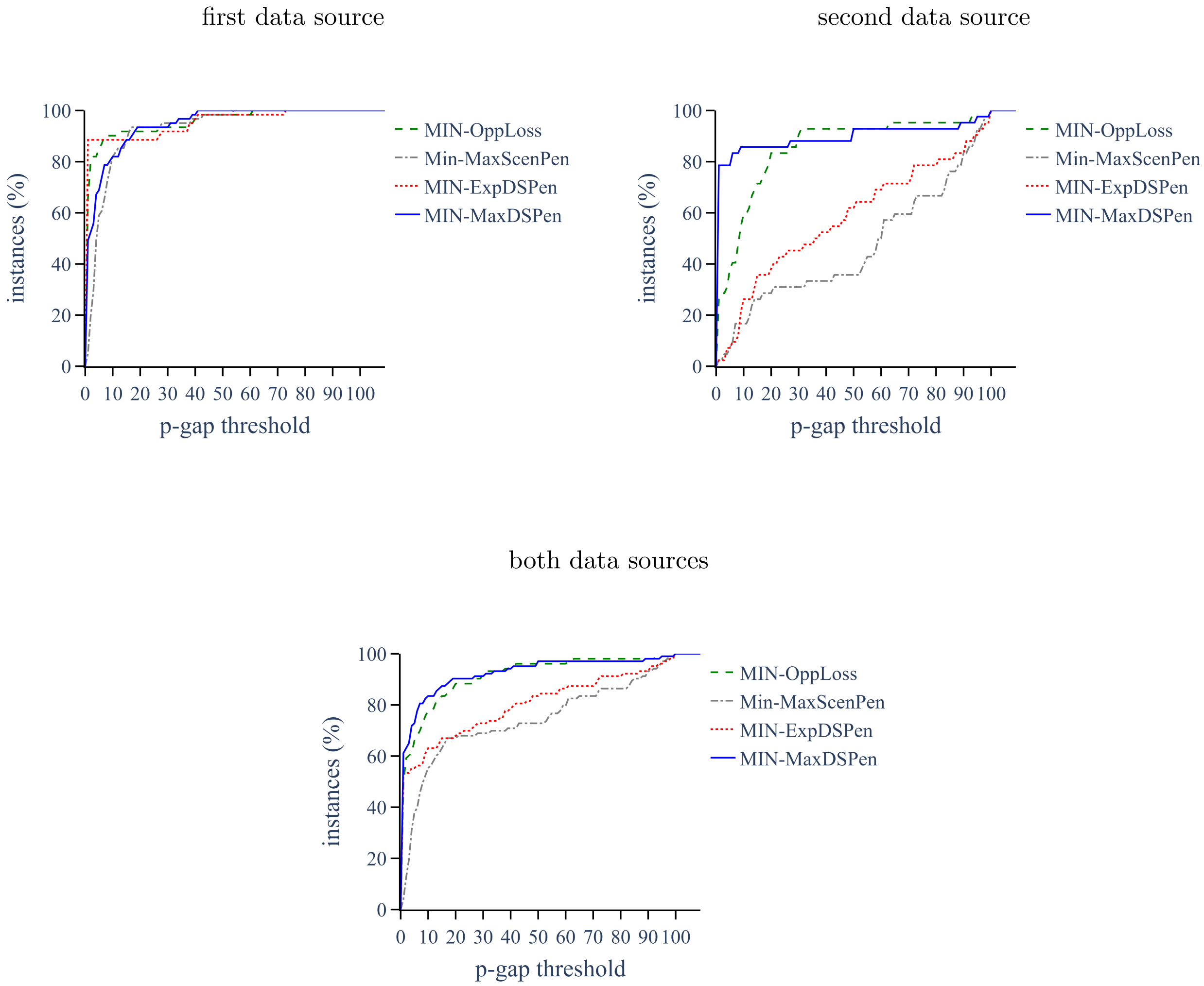

Figure 6: Performance profile of the studied models considering either one or both data sources.

1. Regardless of the level of risk aversion, under data source ambiguity, explicitly accounting for such ambiguity in the HSCN design problem has economic advantages on average and mitigates risk, alleviating the worst-case outcome (see, e.g., Table 3 and Figure 6).

2. In problem settings with a low level of uncertainty and ambiguity (i.e., ambiguity patterns (c) and (d)), which translate into a planning context where there is a higher level of confidence regarding the estimations obtained, MIN-OppLoss (the most popular approach in the literature) and MIN-MaxDSPen solutions are always on the Pareto frontier, representing the best trade-offs between the estimations obtained from the data sources (see Figure 3).

3. The MIN-ExpDSPen model, regardless of the ambiguity pattern, has the highest probability of obtaining the best solution when assessed based on the information obtained from the pessimistic (first) data source (see Figure 4, Table 3 and Figure 6). As such, if a decision-maker is interested in hedging the risk based on the most pessimistic assessments, this is the approach of choice.

4. The MIN-MaxDSPen, regardless of the ambiguity pattern, has the highest probability of obtaining the best solution when assessed based on the information obtained from the optimistic (second) data source (see Figure 5, Table 3 and Figure 6). This conclusion aligns with the fact that MIN-OppLoss does not explicitly consider ambiguity, whereas MIN-ExpDSPen and MIN-MaxScenPen are conservative models prioritizing the pessimistic data source.

5. According to the data presented in Table 3 and Figure 6, in situations where no bias towards a specific data source is evident and the ambiguity pattern remains indistinct, the MIN-MaxDSPen model emerges as the most favourable choice.

6 Conclusion

In this paper, we have proposed four optimization methods, including Minimization of expected opportunity loss (MIN-OppLoss), Minimization of Maximum Data-Source Penalty (MIN-MaxDSPen), Minimization of Expected Data-Source Penalty (MIN-ExpDSPen), and Minimization of Maximum Scenario Penalty, that explicitly account for ambiguity caused by inconsistent estimates of uncertain parameters obtained from multiple data sources involved in the demand and damage assessments in the context of solving the HSCN design problem. We then compare the performance of the proposed models over four different ambiguity patterns with two data sources on 20 instances extracted from a real-world data set on the 2018 Indonesia earthquake.

The results obtained and analysis performed led us to the following conclusion. In an HSCN design problem with narrow uncertainty and ambiguity (e.g. ambiguity patterns (c) and (d)), then MIN-MaxDSPen and MIN-OppLoss models could be used as their solutions always fit on the Pareto frontier, indicating a solution that is not dominated by solutions obtained from other models. Furthermore, if the ambiguity pattern is unknown and the decision-makers are slightly biased toward one of the two data sources, then the following applies: if they are biased toward the pessimistic data source, they should use MIN-ExpDSPen, and if biased toward the optimistic data source, they should use MIN-MaxDSPen. Finally, if no information is available regarding the ambiguity pattern with the same confidence level toward the data sources, then the decision-makers should use the MIN-MaxDSPen model.

This paper opens several research directions on the impact of ambiguity in HSCN design problems. Of particular interest is the question of how to design and adjust the HSCN in response to evolving patterns of ambiguity as new information becomes available throughout the planning horizon. Another research line would involve considering a varying level of confidence in the data sources while also considering a higher number of data sources directly leading to higher levels of complexity in the ambiguity patterns.

Acknowledgement

This research was enabled in part by support provided by Calcul Québec (calculquebec.ca) and the Digital Research Alliance of Canada (alliance.can.ca).

References

Abedor, J., Nagpal, K., Khargonekar, P. P., and Poolla, K. (1995). Robust regulation in the presence of norm-bounded uncertainty. *IEEE Transactions on Automatic Control*, 40(1):147–153.

Anaya-Arenas, A. M., Renaud, J., and Ruiz, A. (2014). Relief distribution networks: a systematic review. *Annals of Operations Research*, 223(1):53–79.

Aouni, B., Ben Abdelaziz, F., and La Torre, D. (2012). The stochastic goal programming model: theory and applications. *Journal of Multi-Criteria Decision Analysis*, 19(5-6):185–200.

Balcik, B. (2017). Site selection and vehicle routing for post-disaster rapid needs assessment. *Transportation research part E: logistics and transportation review*, 101:30–58.

Balcik, B. and Beamon, B. M. (2008). Facility location in humanitarian relief. *International Journal of logistics*, 11(2):101–121.

Balcik, B., Silvestri, S., Rancourt, M.-È., and Laporte, G. (2019). Collaborative prepositioning network design for regional disaster response. *Production and Operations Management*, 28(10):2431–2455.

Balcik, B. and Yanıkoğlu, İ. (2020). A robust optimization approach for humanitarian needs assessment planning under travel time uncertainty. *European Journal of Operational Research*, 282(1):40–57.

Bayraksan, G. and Love, D. K. (2015). Data-driven stochastic programming using phi-divergences. In *The operations research revolution*, pages 1–19. INFORMS.

Ben-Tal, A. and Nemirovski, A. (1998). Robust convex optimization. *Mathematics of operations research*, 23(4):769–805.

Ben-Tal, A. and Nemirovski, A. (2002). On tractable approximations of uncertain linear matrix inequalities affected by interval uncertainty. *SIAM Journal on Optimization*, 12(3):811–833.

Benini, A., Chataigner, P., Noumri, N., Parham, N., Sweeney, J., and Tax, L. (2017). The use of expert judgment in humanitarian analysis theory, methods and applications. *Geneva Assessment Capacities Project*, (August):1–218.

Benini, A., C. P. N. N. P. N. S. J. T. L. (2017). The use of expert judgment in humanitarian analysis—theory, methods, applications. *Assessment Capacities Project - ACAPS*.

Besiou, M. and Van Wassenhove, L. N. (2020). Humanitarian operations: A world of opportunity for relevant and impactful research. *Manufacturing & Service Operations Management*, 22(1):135–145.

Birge, J. R. and Louveaux, F. (2011). *Introduction to stochastic programming*. Springer Science & Business Media.

Charnes, A. and Cooper, W. W. (1957). Management models and industrial applications of linear programming. *Management science*, 4(1):38–91.

Chen, A. Y., Peña-Mora, F., and Ouyang, Y. (2011). A collaborative gis framework to support equipment distribution for civil engineering disaster response operations. *Automation in Construction*, 20(5):637–648.

Daneshvar, M., Jena, S. D., and Rei, W. (2023). A two-stage stochastic post-disaster humanitarian supply chain network design problem. *Computers & Industrial Engineering*, 183:109459.

Dönmez, Z., Kara, B. Y., Karsu, Ö., and Saldanha-da Gama, F. (2021). Humanitarian facility location under uncertainty: Critical review and future prospects. *Omega*, page 102393.

Ester, M., Kriegel, H.-P., Sander, J., Xu, X., et al. (1996). A density-based algorithm for discovering clusters in large spatial databases with noise. In *kdd*, volume 96, pages 226–231.

Ghahtarani, A. and Najafi, A. A. (2013). Robust goal programming for multi-objective portfolio selection problem. *Economic Modelling*, 33:588–592.

Grass, E. and Fischer, K. (2016). Prepositioning of relief items under uncertainty: A classification of modeling and solution approaches for disaster management. *Logistics Management*, pages 189–202.

Grass, E., Ortmann, J., Balcik, B., and Rei, W. (2021). A machine learning approach to deal with ambiguity in the humanitarian decision making.

Gutjahr, W. J. and Nolz, P. C. (2016). Multicriteria optimization in humanitarian aid. *European Journal of Operational Research*, 252(2):351–366.

Hakimifar, M., Balcik, B., Fikar, C., Hemmelmayr, V., and Wakolbinger, T. (2021). Evaluation of field visit planning heuristics during rapid needs assessment in an uncertain post-disaster environment. *Annals of Operations Research*, pages 1–42.

Hart, W. E., Laird, C. D., Watson, J.-P., Woodruff, D. L., Hackebeil, G. A., Nicholson, B. L., Siirola, J. D., et al. (2017). *Pyomo-optimization modeling in python*, volume 67. Springer.

Hart, W. E., Watson, J.-P., and Woodruff, D. L. (2011). Pyomo: modeling and solving mathematical programs in python. *Mathematical Programming Computation*, 3(3):219–260.

HELP Logistics AG (2018). Supply chain expenditure and preparedness investment opportunities. report, international federation of red cross and red crescent societies, switzerland. *The International Federation of Red Cross and Red Crescent Societies, the Philippine Red Cross.*

Holguín-Veras, J., Jaller, M., Van Wassenhove, L. N., Pérez, N., and Wachtendorf, T. (2012). On the unique features of post-disaster humanitarian logistics. *Journal of Operations Management*, 30(7-8):494–506.

IFRC (2019). Emergency appeal mdrid013.

IFRC (2021). The standard products catalogue.

IOM (2019). Indonesia displacement data - lombok earthquake site assessment, international organization for migration (iom), displacement tracking matrix (dtm).

Kleywegt, A. J., Shapiro, A., and Homem-de Mello, T. (2002). The sample average approximation method for stochastic discrete optimization. *SIAM Journal on Optimization*, 12(2):479–502.

Langewisch, A. and Choobineh, F. (1996). Stochastic dominance tests for ranking alternatives under ambiguity. *European Journal of Operational Research*, 95(1):139–154.

Luxen, D. and Vetter, C. (2011). Real-time routing with openstreetmap data. In *Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems*, GIS '11, pages 513–516, New York, NY, USA. ACM.

Mendes, A. B. and Cardoso, M. G. (2006). Clustering supermarkets: the role of experts. *Journal of Retailing and Consumer Services*, 13(4):231–247.

OCHA (2022). Trends in response plan/appeal requirements.

Pluymers, B., Rossiter, J. A., Suykens, J. A., and De Moor, B. (2005). The efficient computation of polyhedral invariant sets for linear systems with polytopic uncertainty. In *Proceedings of the 2005, American control conference, 2005.*, pages 804–809. IEEE.

Sakamoto, M., Sasaki, D., Ono, Y., Makino, Y., and Kodama, E. N. (2020). Implementation of evacuation measures during natural disasters under conditions of the novel coronavirus (covid-19) pandemic based on a review of previous responses to complex disasters in japan. *Progress in disaster science*, 8:100127.

Soyster, A. L. (1973). Convex programming with set-inclusive constraints and applications to inexact linear programming. *Operations research*, 21(5):1154–1157.

Tofighi, S., Torabi, S. A., and Mansouri, S. A. (2016). Humanitarian logistics network design under mixed uncertainty. *European Journal of Operational Research*, 250(1):239–250.

Van Wassenhove, L. N. (2006). Humanitarian aid logistics: supply chain management in high gear. *Journal of the Operational research Society*, 57(5):475–489.

Wald, A. (1945). Statistical decision functions which minimize the maximum risk. *Annals of Mathematics*, pages 265–280.

Yáñez-Sandivari, L., Cortés, C. E., and Rey, P. A. (2021). Humanitarian logistics and emergencies management: New perspectives to a sociotechnical problem and its optimization approach management. *International Journal of Disaster Risk Reduction*, 52:101952.