Nested logic-based Benders decomposition for disaster preparedness planning with horizontal coordination

Penghui Guo[a], Zhijie Sasha Dong[b,*], Jianjun Zhu[a]

[a]*College of Economics and Management, Nanjing University of Aeronautics and Astronautics, Nanjing 211106, China*
[b]*School of Transportation, Southeast University, Nanjing 211189, China*

Abstract

Disaster preparedness planning includes capacity building, delivery planning, and requirements estimation. Uncertainties of disaster make it challenging to place the proper amount of relief resources at the right locations. This paper develops a two-stage stochastic preparedness planning problem for stockpiling and distributing relief items. The model allows horizontal coordination to rebalance inventory from relief facilities in surplus to where in shortage by a new vehicle routing problem variant with inter-facility routes. This paper also proposes the concept of service level, by defining it as the lowest proportion of satisfied demand for all the affected communities, which not only helps with humanitarian equity but also makes the model a tool for comprehending the value of budgets. Under certain service levels, cost requirements are estimated by the objective function. A nested logic-based Benders decomposition algorithm with optimality and feasibility cuts is designed to solve the problem exactly. In contrast with the classical implementation that builds a new branch-and-bound tree in each iteration, the branch-and-check, which searches on a single branch-and-bound tree, is implemented as an alternative. Based on the numerical study instances, extensive experiments validate the efficiency of the algorithm, the value of horizontal coordination, and the impact of integrating vehicle routing.

Keywords: `humanitarian logistics, prepositioning, horizontal coordination, two-stage stochastic programming, logic-based Benders decomposition`

1. Introduction

Over the past twenty years, 7,348 natural disasters caused 1.23 million deaths and 4.03 billion people affected worldwide (UNDRR, 2020). Humanitarian logistics has played a crucial role in combating disasters (Kara and Savaşer, 2017). For instance, in the 2021 Haiti Earthquake, humanitarian assistance was delivered to about 457,800 people with more than 200,000 ready-to-eat food, 2854,000 Gallons of water, and 628,000 non-food item kits from August to November 2021, according to reports by the United Nations Office for the Coordination of Humanitarian Affairs (OCHA) (OCHA, 2021). In four phases of disaster operations management (i.e., mitigation, preparedness, response, and recovery) (Özdamar and Ertem, 2015), preparedness planning is vital in terms of time-saving and cost-saving among humanitarian logistics activities. Effective planning not only accelerates the response by more than a week, resulting in the salvation of numerous lives but also yields substantial cost savings, doubling the value of future response-related expenditures (UNICEF and WFP, 2015). However, factors like the uncertainty of disaster impact and the limitation of relief resources make designing an effective and efficient preparedness plan challenging.

Preparedness planning includes locating facilities and preparing for transportation and relief supplies (Kara and Savaşer, 2017, Sabbaghtorkan et al., 2020). For instance, the United Nations Humanitarian Response Depot (UNHRD) and the International Federation of Red Cross and Red Crescent Societies (IFRC) operate several facilities around the world with emergency stocks to provide immediate assistance to the affected areas (Balcik et al., 2019). One major challenge of planning is that it is difficult to predict disasters, including the location, intensity, and timing (Sabbaghtorkan et al., 2020). As a result, the demands of affected communities can be either overestimated or underestimated, causing a mismatch between the supplies of relief facilities and the demands of affected communities. When there is a lack of coordination among relief facilities, to avoid potential stockout, reliable preparedness planning tends to be conservative through purchasing and holding additional relief items at each facility. Horizontal coordination which refers to the relationships and interactions among different actors operating within the same level of the relief chain (Balcik et al., 2010), is promising in this situation through peak shaving and valley filling. In the commercial context, horizontal coordination has been applied and widely addressed in literature for expensive low-demand items (e.g., spare parts) (van Wijk et al., 2019, Topan and van der Heijden, 2020). The similarity between expensive low-demand items in the commercial context and relief items in the humanitarian context is the motivating factor for this study. Relief items are expensive since it is estimated that the average cost of providing relief to victims can go up to

*Corresponding author
Email addresses: `guopenghui@outlook.com` (Penghui Guo), `sasha@seu.edu.cn` (Zhijie Sasha Dong)

USD 93.19 per beneficiary (Webster et al., 2008). As a reference, the gross national income (GNI) per capita of Haiti in 2020 was USD 1,363 (UN, 2022), so USD 93.19 is nearly the GNI per capita per month. Relief items are in low-demand because they are not consumed routinely but may be stored for a long period before it is consumed. This also implies that relief items are expensive in terms of inventory holding. Similar characteristics between relief items and expensive low-demand items imply the potential for horizontal coordination in humanitarian operations management. In recent years, horizontal coordination has also been applied in the humanitarian context. For example, the Caribbean Disaster and Emergency Management Agency (CDEMA) maintains a relief network with shared inventory to ensure coordination among 18 participating states (Balcik et al., 2019, Rodríguez-Pereira et al., 2021). The UNHRD network which is managed by World Food Program (WFP) allows participating members to borrow and loan relief stock from warehouses around the world (Toyasaki et al., 2017). However, horizontal coordination has not been well addressed in the humanitarian operations management literature. Horizontal coordination adds flexibility to the relief network through inventory rebalancing. When warehousing resources or relief items are limited, without horizontal coordination, there may even be no feasible response plan. Nevertheless, the improper allocation of relief resources to facilities may cause delayed delivery, since additional time may be consumed for inter-facility transportation. Incorporating horizontal coordination into the planning problem rather than using it as a remedy in the post-disaster response phase is a feasible way to improve the effectiveness and efficiency of planning.

Though alleviating human suffering is always the priority of humanitarian logistics, the cost of humanitarian activities is also essential from the perspective of responders, which can be humanitarian organizations or public sectors. For humanitarian organizations that rely on donations (which are always insufficient) to operate (Keshvari Fard et al., 2022), financial transparency puts pressure on cost effectiveness (Baker and Salway, 2016). Especially in the era of COVID-19, driven by the pandemic and weakened economies, those in need are growing while spending by public donors dropped (DI, 2021). For public sectors, justifying the use of funds to taxpayers asks for an appropriate costing methodology. Methods based on unit-based costing have been suggested by the OCHA and the literature, which estimates the overall cost for a humanitarian response using the cost per delivery or person served (OCHA, 2018, Sabbaghtorkan et al., 2020). These methods are usually allocation models or transportation models, which use the average unit cost per delivery or person as input. However, as the OCHA pointed out, one major drawback of unit-based costing is that it is difficult to estimate the average unit cost. This highlights the gap between the preparedness planning literature and practice. The cause for the average unit cost estimation difficulty is twofold. Firstly, planning decisions are made before disasters while the unit cost may vary with the actual disaster situation. Secondly, even for a certain disaster situation, the unit cost for delivering relief items to a beneficiary is also difficult to obtain, since multiple affected communities can be visited by the same vehicle. We close the gap by integrating the routing decisions into the preparedness planning problem. With the vehicle routing problem (VRP), the aforementioned average unit cost per delivery or person is replaced with the average unit routing cost (per mile) that is constant in disaster situations. The cost per mile can be obtained much more accurately and simply from transportation service providers. As a result, the overall cost can be estimated more accurately. To examine how the sufficiency of relief items influences the monetary cost or vice versa, we introduce a concept namely service level. The service level is defined as the lowest ratio of satisfied demand for all the affected communities in all the cases. Since the motivation of this work is to support preparedness decisions, decision-makers can adjust the service level to multiple values to figure out the marginal cost of the relief item sufficiency and the marginal contribution of the budget. By leveraging our costing method with the service level parameter, the decision-makers can comprehend the value of an investment, and the ability of budget planning can be enhanced.

The goal of this study is a more effective, efficient, and flexible disaster preparedness planning method, and we will determine the location and quantity of relief items such that they can be delivered to affected communities at the minimum cost once a disaster occurs. The preparedness planning problem is formulated using two-stage stochastic programming. The first-stage decisions are facility location and relief item preparedness. In the second stage, to reduce the impact of uncertainty, horizontal coordination is taken into account to rebalance inventory. The second-stage problem dispatches relief items through a vehicle routing problem rather than the transportation problem or allocation problem that has been widely addressed in the literature (Section 2.1). The vehicle routing decisions allow decision-makers to respond more efficiently and estimate costs more accurately. The objective is to minimize the sum of planning costs (location, procurement) and expected response cost (routing) under a particular service level. It is worth noting that the service level parameter makes our model a tool for decision-makers to evaluate the marginal cost of supply sufficiency. Additionally, we consider the affected community heterogeneity. For example, it is pretty common that some communities are isolated due to road damage. In this case, these communities require special vehicles (e.g., assault boats and all-terrain vehicles) to be reached. In other words, all the affected communities are classified into two categories. We ensure that the demands of both categories can be satisfied. Along with a service level guarantee, humanitarian equity can be achieved. Horizontal coordination is the major among the aforementioned characteristics in our model. Technically, the problem with horizontal coordination has a significantly larger decision space than that without horizontal coordination, and the decision space of the former contains the decision space of the latter, so integrating horizontal coordination always results in a preparedness plan that is not worse than the other one. However, the horizontal coordination problem requires more computational efforts, despite the fact that the two-stage stochastic programming and vehicle routing problem themselves are known to be computationally challenging.

The stochastic programming problem can be formulated with an *extensive form* (or deterministic equivalent) that can be directly fed to the off-the-shelf solvers. In the extensive form, the number of variables and constraints can be huge. Hence solving large-scale problems in this manner is usually prohibitive. The hierarchical nature of our problem inspires us to solve it with the Benders decomposition (BD) (Benders, 1962). The original BD is incapable of solving our problem as it requires subproblems to be linear programming (LP) while our second-stage problem (subproblem in BD) contains integer variables. Fortunately, the latest advance in the class of Benders methods, the logic-based Benders decomposition (LBBD), can be applied to any class of optimization problems (Hooker and Ottosson, 2003). In this work, the subproblem in the standard LBBD is further decomposed and there are three echelons of problems corresponding with location, allocation, and routing, respectively. Hence, we call it nested logic-based Benders decomposition (NLBBD). Since subproblem solving can be a bottleneck for successfully implementing LBBD when the subproblem is complex, the advantage of our NLBBD method is the subproblem is further decomposed. Before convergence, the NLBBD exchanges information between upper-echelon problems and lower-echelon problems through trial solutions and logic-based Benders cuts. The logic-based Benders cuts cannot be generated follow a standard way like the original BD does, but need to be tailored for a specific problem. Furthermore, as our second-stage problem is not *relatively complete* and contains an objective (rather than a feasibility checking problem), not only *feasibility cuts* but also *optimality cuts* are needed to guide searching for the optimum. Moreover, some *valid inequalities*, which are either adapted from the literature or specialized for our problem, are introduced to reinforce the mathematical model. These valid inequalities are also compatible with the decomposed problems. A variant of LBBD, namely *branch-and-check* (Thorsteinsson, 2001, Hooker, 2019), is also implemented. The branch-and-check adds cutting planes on a single branch-and-bound tree to avoid solving the master problem from scratch each time a cutting plane is added. Finally, based on the NLBBD and branch-and-check method, extensive numerical experiments are conducted on specially designed instances. In summary, this work contributes to both the modeling and solution methodology as follows.

- Both studies in commercial operations management and the practice of humanitarian organizations imply the potential of coordination in humanitarian operations management. However, coordination is still in its infancy in the literature on humanitarian operations management. We study a preparedness planning problem with horizontal coordination that allows inventory rebalance to alleviate the impact of inaccurately estimated demand. Since providing relief supplies to affected communities sufficiently is the main concern of this study, we further involve the heterogeneity of the community in our model to help achieve humanitarian equity. Numerical results show that horizontal coordination is especially valuable for situations where facility capacity is limited, i.e., horizontal coordination makes infeasible problems feasible. When the facility capacity is sufficient, horizontal coordination can result in more than 10% cost savings in our instances.
- Costing is also one of the crucial activities in the preparedness phase. Most of the existing literature involves costing based on average unit cost, which can be difficult to estimate. This may impact both the accuracy of costing and the quality of a preparedness plan. We fill this gap by introducing the routing decision to the preparedness planning problem. Numerical results demonstrate that the average-unit-cost-based method may overestimate or underestimate the overall cost and usually yields solutions worse than the routing-based method. Additionally, a service level parameter is introduced to help decision-makers figure out the trade-off between the cost and the sufficiency of supplies.
- We develop an efficient NLBBD algorithm with a certificate of optimality. LBBD is a framework that needs to be customized for specific classes of problems, and this work is one of the very few applications of LBBD on stochastic programming (Section 2.3). Equipped with the specialized logic-based Benders feasibility cuts and optimality cuts, the NLBBD algorithm allows incomplete recourse and subproblem objectives. To the best of our knowledge, we are the first that develop both feasibility cut and optimality cut in the literature on LBBD for stochastic programming problems. Experiments using the extensive form, standard NLBBD, and branch-and-check are conducted. The results demonstrate that our algorithm outperforms the extensive form significantly. Furthermore, since our model captures activities of humanitarian operations management in common (location, allocation, and routing) and our NLBBD is general enough (incomplete recourse, subproblem objective), we believe the NLBBD can be extended to other classes of humanitarian operations management problems.

The remainder of this paper is organized as follows. Section 2 reviews the existing literature related to this paper. Section 3 presents the preparedness planning problem formally. To solve the mathematical model efficiently, in Section 4, the NLBBD method is introduced. Finally, extensive numerical results are reported in Section 5 and Section 6 to validate our contributions, followed by conclusions in Section 7.

2. Literature review

2.1. Disaster preparedness planning

Disaster preparedness planning includes capacity requirements estimation, capacity building, and capacity delivery planning (DHS, 2011). The term "capacity" can refer to intangible (e.g., financial resources) and physical (e.g., inventory) capacity (Van Wassenhove, 2006, Kunz et al., 2014). In this study, the preparedness planning problems of interest are mainly monetary requirements estimation, physical inventory building, and inventory delivery planning. The physical inventory building and inventory delivery planning in preparedness planning are also known as *prepositioning*. Prepositioning can be either tactical or strategical, depending on the length of time horizon from prepositioning to specific disaster occurrence (Sabbaghtorkan et al., 2020). The tactical prepositioning problems are usually formulated with deterministic models (An et al., 2013, Arslan et al., 2021). When prepositioning decisions are made a relatively long period before disasters, the uncertainty of disasters has to be involved. Consequently, robust models (Paul and Wang, 2019, Velasquez et al., 2020, Wang and Chen, 2020, Wang et al., 2021b) and stochastic models (Table 1) are in the majority of strategical prepositioning studies. Readers may refer to Sabbaghtorkan et al. (2020) for a comprehensive literature review on prepositioning. In the following, we focus on stochastic prepositioning for relief facility location and relief item distribution, i.e., physical capacity building and delivery planning, and discuss how the monetary requirements estimation is included in their models.

Table 1: Stochastic prepositioning problems

Paper	1st	2nd	1st objective	2nd objective
Rawls and Turnquist (2010)	LI	C	Facility + procurement	Commodity flow + holding + penalty$^{\ominus}$
Mete and Zabinsky (2010)	LI	T	Operating cost	Transportation time + penalty$^{\ominus}$
Hong et al. (2015)	LI	C	Facility + procurement	Transportation + penalty$^{\ominus}$ + penalty$^{\oplus}$
Alem et al. (2016)	I	C	Procurement + vehicle	Holding + penalty$^{\ominus}$
Pacheco and Batta (2016)	L	T	Procurement	Transportation + penalty$^{\ominus}$ + penalty$^{\oplus}$
Rezaei-Malek et al. (2016)	LI	T	Facility + procurement	Transportation + penalty$^{\oplus}$, utility, difference of utility
Tofighi et al. (2016)	LI	T	Facility + holding	Distribution time, travel time, penalty$^{\ominus}$ + penalty$^{\oplus}$
Morrice et al. (2016)	LI	T	Transportation	Transportation + penalty$^{\ominus}$
Caunhye et al. (2016)	LI	R	Facility	Transportation time + transshipment time + penalty$^{\ominus}$
Manopiniwes and Irohara (2017)	LI	R	Facility + procurement	Shipping
Noham and Tzur (2018)	L	A	-	Allocated quantity / response time
Torabi et al. (2018)	I	T	Procurement + holding	Procurement + transportation + penalty$^{\ominus}$
Ni et al. (2018)	LI	C	Facility + procurement	Transportation + penalty$^{\ominus}$ + penalty$^{\oplus}$
Elçi and Noyan (2018)	LI	T	Facility + procurement	Transportation + penalty$^{\ominus}$
Sanci and Daskin (2019)	L	C	Facility	Transportation + penalty$^{\ominus}$
Paul and Zhang (2019)	LI	T	Facility + procurement + shipping capacity	Shipping capacity + shipping + deprivation
Hu and Dong (2019)	LI	T	Facility + procurement + transportation	Procurement + transportation + holding + penalty$^{\ominus}$
Aslan and Çelik (2019)	LI	R	Facility + procurement	Transportation + outsourcing + repair
Erbeyoğlu and Bilge (2020)	LI	T	Facility + procurement	-
Wang et al. (2021a)	LI	T	Facility + procurement	Transportation + penalty$^{\ominus}$
Sanci and Daskin (2021)	L	T	Facility	Transportation + penalty$^{\ominus}$
This paper	LI	R	Facility + procurement + vehicle	Routing

[1] `1st`, the first-stage decisions; `2nd`, the second-stage decisions; `L`, location; `I`, inventory; `A`, allocation; `T`, transportation; `C`, commodity flow; `R`, routing; `penalty`$^{\ominus}$, unsatisfied demand penalty; `penalty`$^{\oplus}$, unused inventory penalty; `-`, not applicable.

The prepositioning problem was first introduced by Rawls and Turnquist (2010). They proposed a two-stage stochastic prepositioning problem, in which the first-stage decisions include storage facility locations and sizes, as well as inventory for various types of supplies. The second stage is formulated as a multi-commodity network flow problem to distribute relief items. More recently, for example, Alem et al. (2016) developed a two-stage stochastic network flow model to make first-stage emergency aid prepositioning, vehicle preparation, and second-stage resource distribution decisions in disaster relief. Noham and Tzur (2018) studied a disaster preparedness and response problem, which addresses network designing and resource allocation with the actual decision-making process. Sanci and Daskin (2019) integrated location and network restoration decisions for emergency relief. The second stage makes distribution and network restoration decisions jointly, both of which are formulated as commodity flow problems. Erbeyoğlu and Bilge (2020) developed a two-stage stochastic model for achieving robust response. Their first-stage decisions include facility selection and inventory prepositioning. The second stage determines relief item distribution quantity. We summarize the stochastic prepositioning problems in Table 1 with their first-stage decisions, second-stage decisions, first-stage objectives, and second-stage objectives. In terms of the first-stage decisions, it is clear that location and inventory (LI) are the majority (14 out of 20). As a result, the first-stage objectives are mainly facility costs plus procurement costs. Most of the existing studies (17 out of 20) formulate the second stage under the framework of allocation problem (A), transportation problem (T), or commodity flow problem (C), which require a prior estimate of the "unit" cost. As discussed in Section 1, obtaining the unit cost can be difficult, if not impossible. Apart from that, in the second-stage objective, it is pretty common (15 out of 20) to include penalty costs, the coefficients of which are even more difficult to obtain. Including penalty costs results in a mixture of monetary cost and humanitarian costs, which may cause difficulties in understanding the trade-off between monetary cost and humanitarian cost. Though decision-makers can modify the penalty factor to obtain different solutions, all these solutions cannot guarantee the service level for a certain community, as our model separates the monetary cost and humanitarian cost (service level) in the objective and constraints, respectively. Also, the mixed second-stage objective makes the models incapable of estimating monetary requirements, which is one of the main aspects of preparedness planning.

Replacing the simpler second-stage problems (allocation, transportation, and commodity flow) with vehicle routing problems can be significant for monetary requirements estimation. The only works that integrate the vehicle routing problem into the second stage of the two-stage stochastic prepositioning problem are ones by Manopiniwes and Irohara (2017), Aslan and Çelik (2019), Caunhye et al. (2016). The mathematical models of Manopiniwes and Irohara (2017), Aslan and Çelik (2019) rely on predefined routes as parameters. However, the number of variables associated with all the possible routes can be huge (Boccia et al., 2011). Thus the predefined routes have to be limited within an artificially designed set. The study by Caunhye et al. (2016) is mostly closely related to our work and it is the only one that includes typical "routing constraints" (e.g., subtour elimination) in the model. They developed a location-routing problem for emergency preparedness and response planning with two-stage stochastic programming. In the second stage, vehicle routing and uncertainty (transportation time, external supply, and demand) are taken into consideration. The problem is solved by transforming the two-stage model into a single-stage counterpart (deterministic equivalent). Due to computationally intractability, a small-size illustrative example with four warehouses, four affected communities, and three scenarios is presented in their paper. And they pointed out that a specifically designed algorithm is required for more realistic disaster situations with more potential warehouses and affected communities. Aside from the specifically designed algorithm, the main difference between Caunhye et al. (2016) and ours is that their model allows unsatisfied demand. Consequently, penalty cost, which is hard to estimate in practice, needs to be predetermined. A poorly designed penalty cost may lead to a less robust solution. In addition, the heterogeneity of affected communities is not considered in their work. Our work does not need any parameters that are hard to obtain and ensures robustness for all disaster cases. Along with the consideration for affected community heterogeneity and service level, humanitarian equity can be achieved.

We summarize our efforts to address the limitations of the literature. Firstly, to avoid the unit cost parameter that is difficult to obtain, the second-stage problem is modeled with the vehicle routing problem. Secondly, to help decision-makers understand the trade-off between monetary costs and humanitarian costs, our objective function consists of only monetary cost and the service level constraints measure humanitarian cost (unlike the penalty-based model in the literature). Thirdly, community heterogeneity is considered for humanitarian equity. More generally, it is worth noting that the location-routing problem with stochastic demand (SLRP) (Albareda-Sambola et al., 2007) has characteristics in common with our stochastic prepositioning-routing problem, i.e., location and routing. However, the *recourse* decisions (second-stage decisions) in the stochastic programming models are distinct. The SLRP recourse decisions are replenishment routes that return to the depot based on the *priori routes* (first-stage decisions), while ours are complete vehicle routing problems.

2.2. Horizontal coordination

Humanitarian *coordination* was defined by Wankmüller and Reiner (2019) as *"process of organizing, aligning and differentiating of participating NGOs' actions based on regional knowledge, know-how, specialization and resource availability to reach a shared goal in the context of disasters"*. It can be *vertical* (between upstream and downstream) or *horizontal* (among peers) (Balcik et al., 2010) for managing information, risks, and resources (Wankmüller and Reiner, 2019, Kumar and Singh, 2022). This study focuses on *horizontal coordination* of resources (relief items) among peers through *lateral transshipment*, which originated in commercial operations management (Lee, 1987) and has shown potential for reducing costs of holding, shipping, and waiting (Grahovac and Chakravarty, 2001). Readers may refer to Balcik et al. (2010), Wankmüller and Reiner (2019), Adsanver et al. (2024) for literature reviews on humanitarian coordination.

In the following, we focus on the literature in recent years that studied the horizontal coordination of relief resources from the operational perspective. Note that works that studied the coordination between humanitarian organizations based on game theory are excluded. Kamyabniya et al. (2019) proposed a capacity-sharing coordination mechanism for platelets provision in humanitarian relief operations. The proposed mechanism avoids the wastage of platelets significantly. Davis et al. (2013) studied the coordination of supply position and distribution within a cooperative relief network. In their network, shipment can be either among suppliers or from a supplier to a consumer. The value of coordination on cost saving is validated by a case study. Setiawan et al. (2019) dealt with the resource location and distribution problem for sudden-onset disasters. Their numerical results show that resource sharing by breaking down barriers of administrative boundaries is beneficial. Rottkemper et al. (2012) developed a multi-period inventory relocation and redistribution model in humanitarian operations management. In each period, transshipment between regional depots is allowed. Baskaya et al. (2017) developed deterministic prepositioning models that incorporate lateral transshipment of relief items in humanitarian logistics. The case study shows that the percentage of lateral transshipment increases when beneficiaries require higher quality service. Doodman et al. (2019) developed a multi-period stochastic pre-positioning model considering lateral transshipment among distribution centers. The value of lateral transshipment on cost saving is validated by numerical experiments. Wang et al. (2021c) developed a two-stage stochastic model with stochastic demand for the distribution of disaster supplies with lateral transshipment opportunity. All these results suggest the potential of horizontal coordination in humanitarian operations.

Horizontal coordination is still in its infancy in the literature on humanitarian operations management, especially with uncertainty. To the best of our knowledge, there are only seven papers on humanitarian horizontal coordination, and among them, three (Davis et al., 2013, Doodman et al., 2019, Wang et al., 2021c) incorporated uncertainty. We demonstrate the distinction between the literature and this study as follows. Firstly, most of the studies exploited the deterministic (and static) model (Baskaya et al., 2017) and dynamic model (Kamyabniya et al., 2019, Setiawan et al., 2019, Rottkemper et al., 2012), and showing a tendency to tactical decisions (Kamyabniya et al., 2019). Our problem is more strategic than tactical since facility location decisions that can have a long-term impact are considered. Secondly, the tactical tendency makes horizontal coordination be used in a reactive manner in the response phase. We analyze the value of integrating horizontal coordination into the preparedness problem (Section 6.1), rather than using it as a remedy. Thirdly, incorporating horizontal coordination with vehicle routing (as described in Section 1 and Section 2.1) adds computational challenges to our model, and asks for an efficient specialized algorithm. The second-stage vehicle routing problem with horizontal coordination is a new variant of the multi-depot vehicle routing problem with inter-depot routes (Schiffer et al., 2019).

2.3. Logic-based Benders decomposition

The hierarchical nature of two-stage stochastic programming makes stage-wise decomposition promising for it. The BD (Benders, 1962) and the L-shaped method (Van Slyke and Wets, 1969) (a type of BD that was specialized for stochastic programming) are well-known stage-wise decomposition methods that use linear hyperplanes to outer approximate the value function of the second-stage problem. The BD and the L-shaped method only allow linear subproblems. However, the subproblem may contain integer (including 0-1 integer) variables for the two-stage stochastic programming. The integer L-shaped (Laporte and Louveaux, 1993) extends the Benders method by allowing mixed-integer subproblems, though the master problem can only be 0-1 integer. The LBBD (Hooker and Ottosson, 2003) generalizes the existing BD method, and it can be applied to any class of optimization problems, e.g., mixed integer programming, non-linear programming, and feasibility-checking problem (Rahmaniani et al., 2017). For the comprehensive literature review of BD, readers can refer to Rahmaniani et al. (2017).

Similar to the BD, the LBBD decomposes an original mathematical programming problem into one relaxed *master problem* and some *subproblems*. The master problem provides a lower bound (in the case of the minimization problem) for the original problem and it feeds a temporary solution to the subproblem. The subproblems add *feasibility cuts* to the master problem when there is an infeasible subproblem. When all the subproblems are feasible, an upper bound for the original problem is obtained. *Optimality cuts* are generated and added to the master problem when it turns out the recourse cost is underestimated by the master problem. The algorithm terminates when the lower bound is close enough to the upper bound. Compared with the BD, the LBBD relies on the *inference dual*, which is a generalization for the linear programming dual in BD. Unlike the BD which has a standard procedure to generate Benders cuts, logic-based Benders cuts for LBBD are problem-specific. The LBBD has been applied to facility location (e.g., Fazel-Zarandi and Beck, 2012, Fazel-Zarandi et al., 2013), routing (e.g., Fachini and Armentano, 2020), and scheduling (e.g., Lombardi et al., 2010, Riise et al., 2016, Guo et al., 2021, Elçi and Hooker, 2022) problems in various contexts. In the context of humanitarian operations management, Erbeyoğlu and Bilge (2020) is the only one that applied LBBD. We recommend Hooker (2019) for a comprehensive review of LBBD. Most of the works using LBBD formulate only one type of Benders cut, either optimality cuts or feasibility cuts. In our decomposition framework, both feasibility cuts and optimality cuts are formulated, so more complicated characteristics (incomplete recourse, response cost) of disaster response are allowed to be captured. Without a feasibility cut, the stochastic programming problem is required to be relatively complete, i.e., for any first-stage solution, the second-stage problem is always feasible. Without an optimality cut, the second-stage response cost cannot be involved. Furthermore, our work extends the existing five works (Lombardi et al., 2010, Fazel-Zarandi et al., 2013, Erbeyoğlu and Bilge, 2020, Guo et al., 2021, Elçi and Hooker, 2022) that applied LBBD to stochastic programming. To the best of our knowledge, we are the first that develop both feasibility cuts and optimality cuts in the literature on LBBD for stochastic programming problems.

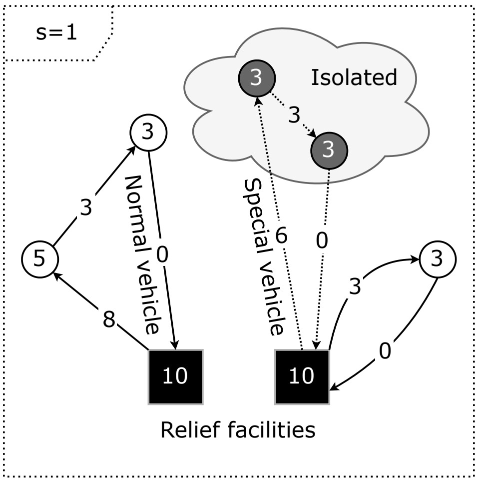
The NLBBD, which is also known as three-level (Fazel-Zarandi et al., 2013), recursive (Riise et al., 2016), or three-stage (Guo et al., 2021) LBBD, is a variant of the standard LBBD. The NLBBD further decomposes each subproblem in LBBD into one master problem and some subproblems. NLBBD is suitable for problems that are inherently multi-stage, or static but contain harder subproblems in the standard LBBD. The NLBBD shares information and generates cutting planes recursively. Fazel-Zarandi et al. (2013) used three-level LBBD to solve facility location and fleet management problems with random travel time. (Riise et al., 2016) used recursive LBBD to solve the deterministic multi-mode outpatient appointment scheduling problem. Guo et al. (2021) used three-stage LBBD to solve the distributed operating room scheduling problem with uncertain surgery durations. Our work is the first that uses NLBBD in the context of humanitarian operational management. Since our model captures activities of humanitarian operations management in common (location, allocation, and routing) and the NLBBD framework allows involving incomplete recourse and response cost, our NLBBD framework can be extended to other classes of preparedness planning problems. It is also worth noting that for large-scale stochastic programming, the NLBBD makes parallel computing possible. Though this is out of the scope of this study. In summary, this study not only contributes to the LBBD literature on stochastic programming but also contributes to the NLBBD literature, and we are the first that present an exact solution method with optimality certificates for preparedness planning with horizontal coordination.

3. Stochastic prepositioning-routing problem

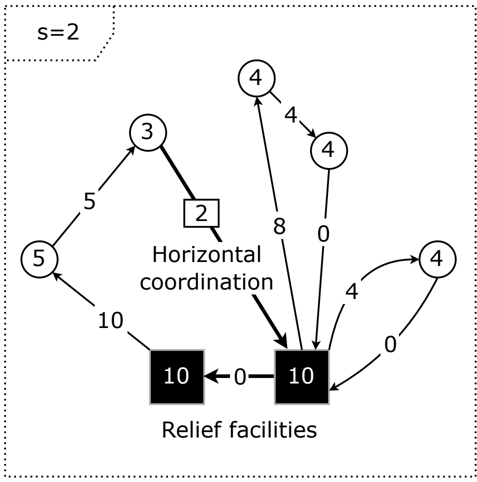
3.1. Problem description

Humanitarian logistics involves two phases for distributing relief supplies: global transportation and local transportation. For example, the UNHRD operates six global depots to procure, store, and transport emergency supplies. These depots have been chosen for their transport connections and proximity to disaster-prone areas. Inventory in these depots allows relief items to be dispatched within 48 hours upon requests of partners. The process of transporting items from global depots to partners' facilities is known as global transportation. Partner organizations of UNHRD distribute relief items to affected people, after the deliveries in bulk have arrived from UNHRD (WFP, 2019, UNICEF, 2023). Our study focuses on distributing relief items from local facilities to affected people, i.e., local transportation. The main differences between global transportation and local transportation are the quantity of items and the time required. Transporting goods from global depots to local facilities is typically done in bulk. The time required, using UNHRD as an example, is in days. Whereas, transporting from local facilities to affected people takes hours and the demand of each affected community can be less than a truckload. Our modeling choice is made based on these differences. Firstly, since it takes hours for local facilities to reach affected people, we do not consider the variations in visiting time between these communities. Secondly, since the demand of each community can be less than a truckload, we model the second-stage decisions based on the vehicle routing problem, instead of the transportation problem, which can be more suitable for global transportation.

A relief network containing nodes (relief facilities and affected communities) and links (roads) is considered in this study. Decisions associated with the proposed relief network consist of both pre-disaster planning decisions and post-disaster response decisions. The planning decisions have to be made with uncertain information, that is the condition of affected communities and roads is only revealed after a disaster strikes. These uncertainties are characterized by discrete scenarios that are generated from historical data or predictions for a future event, depending on whether the plan is made as a long-term strategy or for an oncoming predictable disaster, e.g., a typhoon. Each scenario has a specific probability and corresponds to a realization of uncertain parameters, e.g., the demand of affected communities, and whether a community can be reached by normal vehicles. To figure out the trade-off between the sufficiency of relief items and the monetary cost, we introduce a service level parameter, which is defined as the minimum ratio of demand that can be satisfied for all the affected communities in all the scenarios. The goal of the preparedness planning problem is (i) to make planning and response decisions such that the demand of all the affected communities can be met under particular service-level constraints, and (ii) to estimate the overall cost, which is the sum of the planning cost and expected response cost. To be more specific, we use Figure 1 as an illustrative example of the proposed relief network.



(a) Without horizontal coordination



(b) With horizontal coordination

Figure 1: The second-stage routing decisions for two different scenarios $s = 1$ and $s = 2$, with the same first-stage location and prepositioning decisions. Figure 1(a) shows the case with lower demands but isolated affected communities, requiring no horizontal coordination but a special vehicle. Figure 1(b) involves higher demands without isolated community, necessitating horizontal coordination.

In Figure 1, the circles represent affected communities, the black boxes represent relief facilities, and the arrows represent vehicle routes. The numbers on circles, boxes, and arrows represent the quantity of relief item demand, inventory, and flow, respectively. As presented in Figure 1, we assume that in the planning phase: (i) two relief facilities are opened among all the candidates, with each having a capacity of ten; (ii) the prepositioning inventory of each relief facility is its maximum capacity; and (iii) two and three potentially affected communities are assigned to the left relief facility and the right relief facility, respectively. Due to the concern of consistency (Stavropoulou et al., 2019) (personnel can become more competent by familiarizing themselves with their assigned region), geographical restrictions (Uichanco, 2021) (a physical boundary may exist between communities affected by a disaster), and administrative boundaries (Setiawan et al., 2019) (the government is responsible only for its own administration area), we assume that each affected community is assigned to one relief facility. This assumption is aligned with the humanitarian relief practice and is common in the humanitarian logistic literature with horizontal coordination. From the practical perspective, assigning affected communities to facilities before a disaster aligns with the pre-scripted mission assignments principle of the National Incident Management System (NIMS) in the U.S. (DHS, 2008), which helps to identify resources and capabilities required, and also helps to organize and train deployable personnel. Assigning before a disaster is also a counterbalance for duplication of effort that facing humanitarian organizations (UNDP, 2022, TOI, 2011, IPC-IG, 2022). Duplication occurs when a secondary agency provides the aid that was the primary responsibility of another agency, and then the primary agency also provides aid later (eCFR, 2009). This can be prevented by ensuring each affected community one-time delivery from a single source, as we have assumed. In the literature, assigning communities to facilities is common, particularly in studies with horizontal coordination (please refer to the supplementary material for a brief literature review). Completing assignments before a disaster enhances the reliability of relief systems by enabling more predictable post-disaster actions. This is because each facility knows which region it is responsible for. Horizontal coordination, exemplified later by Figure 1, primarily aims to increase flexibility. Our modeling choices (assignment + horizontal coordination) are a compromise between flexibility and reliability. Nevertheless, this assumption can be relaxed by simply removing the assignment decisions or moving them to the second stage.

Since assignment decisions are made before the actual demand of affected communities is known, unexpected relief item shortages or surplus for opened relief facilities may be incurred. Therefore, horizontal coordination from one facility to another is introduced to rebalance inventory in the response stage. Figure 1(a) and Figure 1(b) illustrate the response-stage decisions with two scenarios. In Figure 1(a), the total supply of the left relief facility is ten, and the total demand of affected communities assigned to it is eight. The total supply of the right relief facility is ten, and the total demand of affected communities assigned to it is nine. That is for each relief facility, the total supply can meet the total demand. Thus, horizontal coordination is not necessary. Note that in Figure 1(a), two of the affected communities are isolated, thus special vehicle is used. In Figure 1(b), the total demand of the right relief facility is twelve, while its supply is ten. Hence, two units of relief items are transported from the left relief facility that has a surplus (ten units of supplies, eight units of demand). Without horizontal coordination, there is no feasible solution for scenario $s = 2$. One may argue that the assumption of assignment decisions in the planning stage is too strong. However, even when we relax this constraint, there is still no feasible solution for $s = 2$, since it is impossible to pack items of size $\{5, 3, 4, 4, 4\}$ into two bins each has a capacity of ten. Using vehicle routing problems with split deliveries can result in feasible solutions. On the other hand, split deliveries imply that each affected community may be visited multiple times. This should be avoided considering the hazard of disaster-affected areas. In summary, there are four types of decisions that need to be made in the preparedness planning problem: location, inventory, allocation, and routing. The allocation decision includes the number of relief items to be distributed from facilities to affected communities and from one facility to another. The objective is to distribute relief items to affected communities with the minimum cost, i.e., facility, procurement, and expected routing costs.

3.2. Two-stage stochastic programming model

We formulate the disaster preparedness planning problem as a stochastic prepositioning-routing problem (SPRP) using two-stage stochastic programming. The first-stage decisions need to be decided before the disaster occurs, and the second-stage decisions are for the response actions, namely recourse decisions. The goal of the two-stage stochastic programming is to minimize the sum of the first-stage cost and expected recourse cost. Notations for sets, parameters, and decision variables used are summarized as follows.

Sets:

T	Set of relief facility candidates.
D	Set of potentially affected communities.
N	Set of both relief facility candidates and potentially affected communities, $N = T \cup D$.
N_i	Set of nodes excluding i, $N_i = N \setminus \{i\}$.
A	Set of arcs between nodes in N, $A = \{(p, q) \mid p \in N, q \in N\}$.
K	Set of available facility types.
H	Set of all available vehicle types.
H'	Set of available special vehicle types, $H' \subset H$.
S	Set of scenarios. The probability of $s \in S$ is denoted as P_s.
${D'}^s$	Set of isolated affected communities in scenario $s \in S$.

Parameters:

M	A sufficiently large number.
r	Service level that restricts the minimal ratio of satisfied demands, $r \geq 0$.
Q^k	Capacity of relief facility of type $k \in K$.
c^k	Opening cost of relief facility of type $k \in K$.
c	Unit procurement cost of relief item.
n^h	Number of available vehicles of type $h \in H$.
Q	Capacity of vehicles.
c_1^h	Fixed cost of preparing a type $h \in H$ vehicle at a relief facility.

c_{pq}^h	Unit cost of vehicle of type $h \in H$ to travel on arc $(p,q) \in A$.
m_j^s	Demand of affected community $j \in D$ in scenario $s \in S$.
P_s	Probability of scenario $s \in S$.
e_j^s	Binary parameter that indicates whether $j \in D$ is isolated in scenario $s \in S$. $e_j^s = 1$ if affected community j is isolated in scenario s, $e_j^s = 0$ otherwise. An isolated affected community can only be visited by special vehicles of type $h \in H'$.

Decision variables:

y_{ik}	Binary variable that indicates whether a type $k \in K$ facility is opened at $i \in T$. $y_{ik} = 1$ if relief facility i of type k is opened, $y_{ik} = 0$ otherwise.
l_i	Quantity of initial inventory of relief facility $i \in T$.
w_i^h	Number of vehicle of type $h \in H$ prepared at relief facility $i \in T$.
x_{ij}	Binary variable that indicates whether community $j \in D$ is assigned to $i \in T$. $x_{ij} = 1$ if community j is assigned to relief facility i, $x_{ij} = 0$ otherwise.
t_{ij}^s	Quantity of relief item transported from $i \in T$ to $j \in D$ in scenario $s \in S$.
$t_{ii'j}^s$	Quantity of relief item transported from $i \in T$ to $j \in D$ through $i' \in T$ in scenario $s \in S$.
$z_{pq}^{(s,i,h)}$	Binary variable that indicates whether vehicle of type $h \in H$ that dispatched from $i \in T$ travels on arc $(p,q) \in A$ in scenario $s \in S$. $z_{pq}^{(s,i,h)} = 1$ if vehicle of type h from relief facility i travels on arc (p,q) in scenario s, $z_{pq}^{(s,i,h)} = 0$ otherwise.
$f_{pq}^{(s,i,h)}$	Quantity of relief item that flows from relief facility $i \in T$ on vehicle of type $h \in H$ on arc $(p,q) \in A$ in scenario $s \in S$.

Auxiliary decision variables:

x_{pq}^s	Binary variable that indicates whether $q \in D \cup T$ accept relief items from $p \in T$ in scenario $s \in S$. In each scenario $s \in S$, for relief facility $q \in T$, $x_{pq}^s = 1$ if and only if $\sum_{j \in D} t_{pqj}^s > 0$, while for community $q \in D$, $x_{pq}^s = x_{pq}$.
$Q_p^{(s,i)}$	Quantity of relief item transported from relief facility $i \in T$ to affected community or transfer point (relief facility) $p \in D \cup T$ in scenario $s \in S$.

In the planning phase, a subset of relief facilities is chosen from the candidate set T. Each facility can be established with a specific capacity option within K. The opening cost for facility $i \in T$ of capacity $Q^k \in K$ is c^k. Relief items and vehicles of type $H \setminus H'$ and H' are prepared at each opened relief facility. The relief items in a quantity of l_i, purchased with price c and prepared at each opened relief facility, cannot exceed facility capacity. The goal of the planning phase is to determine the location and capacity (y_{ik}) of relief facilities, determine the initial inventory of relief items (l_i) and vehicles (w_i^h) prepared at opened facilities, and assign (x_{ij}) affected communities to opened facilities. The planning-phase decisions are formulated as follows.

$$
\begin{aligned}
\text{[SPRP] minimize } & \sum_{i \in T} \sum_{k \in K} c^k y_{ik} + \sum_{i \in T} c\, l_i + \sum_{i \in T} \sum_{h \in H} c_1^h w_i^h + \mathcal{Q}\left(y_{ik}, l_i, w_i^h, x_{ij}\right) && (1.1)\\
\text{subject to } & \sum_{k \in K} y_{ik} \le 1, && \forall i \in T \quad (1.2)\\
& l_i \le \sum_{k \in K} y_{ik} Q^k, && \forall i \in T \quad (1.3)\\
& \sum_{i \in T} x_{ij} = 1, && \forall j \in D \quad (1.4)\\
& x_{ij} \le \sum_{k \in K} y_{ik}, && \forall i \in T, j \in D \quad (1.5)\\
& \sum_{i \in T} w_i^h \le n^h, && \forall h \in H \quad (1.6)\\
& y_{ik}, x_{ij} \in \{0,1\}, && \forall i \in T, j \in D, k \in K \quad (1.7)\\
& l_i, w_i^h \in \mathbb{Z}_0^+, && \forall i \in T, h \in H \quad (1.8)
\end{aligned}
$$

(1.1) minimizes the total cost, which includes both first-stage and expected second-stage costs. The first-stage costs comprise facility opening, relief item procurement, and vehicle preparation expenses. The second-stage *recourse* cost — the expected cost of taking first-stage decisions — is denoted as $\mathcal{Q}(\cdot)$, and will be discussed in detail later. (1.2)–(1.6) are the first-stage prepositioning constraints. (1.2) ensures that at most one type of relief facility can be opened at each candidate location. (1.3) ensures that the number of relief items cannot exceed the facility's capacity. (1.4) is a consistency constraint that ensures every affected community is always served by the same relief facilities whatever the outcoming scenario is. However, the "=" can be replaced with "$\ge$" to allow each affected community to be served by multiple facilities. (1.5) ensures affected communities can be assigned to a relief facility only if the relief facility is opened. (1.6) limits the number of vehicles that can be used. (1.7) and (1.8) are domains of decision variables where "$\in \mathbb{Z}_0^+$" is equivalent to "≥ 0 and integer".

In the response phase, due to uncertainty, unexpected shortages or surpluses may be incurred. To rebalance inventory, facilities in shortage can be replenished by others in surplus. Thus, there are two types of transportation methods: direct shipment t^s_{ij}, i.e., transportation from i to j; and horizontal coordination $t^s_{ii'j}$, i.e., transportation from i to j through i'. For all the scenarios $s \in S$, the quantity of relief items allocated to an affected community at least meets the demand of the affected community m^s_j multiplied by the service level r. The response-phase decisions aim to allocate relief items to affected communities through both direct shipment (t^s_{ij}) and horizontal coordination ($t^s_{ii'j}$). For each scenario $s \in S$ with realized affected communities demand quantity m^s_j and accessibility e^s_j, each affected community $j \in D$ must be visited by one vehicle with the number of relief items defined by allocation variables t^s_{ij} and $t^s_{ii'j}$. In $s \in S$, some affected communities may be isolated, making them inaccessible to normal vehicles. These communities, denoted as $D'^s = \{j \in D | e^s_j = 1\}$, can only be reached by special vehicles of type $h \in H'$. Let the binary variable $z^{(s,i,h)}_{pq} = 1$ if a vehicle of type h that is dispatched by facility i in scenario s travel on arc $(p,q) \in A$, the quantity of commodity flow on this arc is denoted as $f^{(s,i,h)}_{pq}$, and the cost for type h vehicle to travel on (p,q) is c^h_{pq}. The flow variable is restricted by vehicle capacity Q. The response-stage problem is formulated to an allocation and vehicle routing problem based on the single-commodity flow formulation (Baldacci et al., 2008, Koç et al., 2016). In the model, we use subscript $(p,q) \in A$ for variables $t^s_{..}$, $t^s_{...}$, $x^s_{..}$, $z^{s,i,h}_{..}$, and $f^{(s,i,h)}_{..}$ in the routing constraints (2.9)–(2.20) to distinguish them from the superscript i in $z^{(s,i,h)}_{..}$ and $f^{(s,i,h)}_{..}$.

$$
\begin{aligned}
&[\text{SPRP}]\ \mathcal{Q}\left(y_{ik}, l_i, w^h_i, x_{ij}\right) = \\
&\text{minimize} \sum_{s\in S} \mathsf{P}_s \sum_{i\in T}\sum_{h\in H}\sum_{(p,q)\in A} c^h_{pq} \cdot z^{(s,i,h)}_{pq} && (2.1)\\
&\text{subject to } t^s_{ij} \le M x_{ij}, && \forall i \in T, j \in D, s \in S \quad (2.2)\\
&t^s_{ii'j} \le M x_{i'j}, && \forall i \in T, i' \in T, j \in D, s \in S \quad (2.3)\\
&\sum_{j\in D} t^s_{ii'j} \le M \sum_{k\in K} y_{ik}, && \forall i \in T, i' \in T, s \in S \quad (2.4)\\
&\sum_{j\in D} t^s_{ii'j} \le M \sum_{k\in K} y_{i'k}, && \forall i \in T, i' \in T, s \in S \quad (2.5)\\
&\sum_{j\in D} t^s_{ij} + \sum_{i'\in T}\sum_{j\in D} t^s_{ii'j} \le l_i, && \forall i \in T, s \in S \quad (2.6)\\
&\sum_{i\in T} t^s_{ij} + \sum_{i\in T}\sum_{i'\in T} t^s_{ii'j} \le m^s_j, && \forall j \in D, s \in S \quad (2.7)\\
&\sum_{i\in T} t^s_{ij} + \sum_{i\in T}\sum_{i'\in T} t^s_{ii'j} \ge \lceil r \cdot m^s_j \rceil, && \forall j \in D, s \in S \quad (2.8)\\
&\sum_{q\in N_i} z^{(s,i,h)}_{iq} \le w^h_i, && \forall i \in T, h \in H, s \in S \quad (2.9)\\
&x^s_{pq} = x_{pq}, && \forall p \in T, q \in D, s \in S \quad (2.10)\\
&\frac{1}{M}\sum_{j\in D} t^s_{pqj} \le x^s_{pq} \le \sum_{j\in D} t^s_{pqj}, && \forall p, q \in T, s \in S \quad (2.11)\\
&x^s_{pp} = 1, && \forall p \in T, s \in S \quad (2.12)\\
&\sum_{h\in H}\sum_{p\in N} z^{(s,i,h)}_{pq} = x^s_{iq}, && \forall i \in T, q \in N_i, s \in S \quad (2.13)\\
&\sum_{h\in H'}\sum_{p\in N} z^{(s,i,h)}_{pq} = x^s_{iq}, && \forall i \in T, q \in D'^s, s \in S \quad (2.14)\\
&\sum_{p\in N} z^{(s,i,h)}_{pq} - \sum_{p\in N} z^{(s,i,h)}_{qp} = 0, && \forall i \in T, q \in N_i, h \in H, s \in S \quad (2.15)\\
&f^{(s,i,h)}_{pq} \le M x^s_{ip}, && \forall i \in T, p \in N, q \in N, h \in H, s \in S \quad (2.16)\\
&f^{(s,i,h)}_{pq} \le M x^s_{iq}, && \forall i \in T, p \in N, q \in N, h \in H, s \in S \quad (2.17)\\
&\sum_{h\in H}\sum_{q\in N} f^{(s,i,h)}_{qp} - \sum_{h\in H}\sum_{q\in N} f^{(s,i,h)}_{pq} = Q^{(s,i)}_p, && \forall i \in T, p \in N_i, s \in S \quad (2.18)\\
&Q^{(s,i)}_q z^{(s,i,h)}_{pq} \le f^{(s,i,h)}_{pq} \le (Q - Q^{(s,i)}_p) z^{(s,i,h)}_{pq}, && \forall i \in T, (p,q) \in A, h \in H, s \in S \quad (2.19)\\
&Q^{(s,i)}_p = \begin{cases} t^s_{ip} + \sum_{q\in T} t^s_{qip} & \forall i \in T, p \in D, s \in S \\ \sum_{q\in D} t^s_{ipq} & \forall i \in T, p \in T, p \ne i, s \in S \\ 0 & \forall i \in T, p \in T, p = i, s \in S \end{cases} && (2.20)\\
&t^s_{ij}, t^s_{ii'j} \in \mathbb{Z}^+_0, && \forall i \in T, i' \in T, j \in D, s \in S \quad (2.21)\\
&x^s_{pq} \in \{0,1\}, && \forall p \in T, q \in N, s \in S \quad (2.22)\\
&z^{(s,i,h)}_{pq} \in \{0,1\}, && \forall i \in T, (p,q) \in A, h \in H, s \in S \quad (2.23)\\
&f^{(s,i,h)}_{pq}, Q^{(s,i)}_p \in \mathbb{Z}^+_0, && \forall i \in T, (p,q) \in A, h \in H, s \in S \quad (2.24)
\end{aligned}
$$

(2.1) minimizes the expected second-stage routing costs, i.e., the last term in (1.1). (2.2)–(2.8) are allocation constraints. (2.2) ensures every community only accepts relief items from the facility it is assigned to. (2.3) ensures relief facility i can transport relief items through facility i' to affected community j only if affected community j was assigned to relief facility i'. (2.4) and (2.5) ensure relief items can be transported from relief facility i to affected community j through relief facility i' only if both i and i' is opened. (2.6) ensures the total quantity of relief item transport from relief facility i is no more than its stock. (2.7) and (2.8) ensure that all the affected communities in all the scenarios obtain relief items no more than their demand but satisfy the service level r requirement. The service level parameter, denoted by r, serves as a measure of reliability for a plan. It acts as a factor of safety and enables decision-makers to estimate the monetary requirements for different levels of reliability. By analyzing the marginal cost associated with increasing reliability, they can determine the optimal service level that balances cost and performance. (2.9)–(2.20) are routing constraints. (2.9) ensures the number of vehicles does not exceed the number that has been prepared at a relief facility. (2.10) and (2.11) introduce auxiliary variables x^s_{pq}, which indicate the assignment in a specific scenario. Specifically, except for the affected communities assigned to relief facility p, transfer facilities $\{q \in T | \exists_{j\in D} t^s_{pqj} \geq 1\}$ are also considered assigned to p. (2.12) permits all relief facilities to dispatch vehicles in every scenario. (2.13)–(2.15) ensure that an arc (p, q) can be visited by a vehicle from i only if both p and q are assigned to i. (2.16) and (2.17) link binary assignment variables and continuous variables. (2.18)–(2.20) define the commodity flow. In (2.20), $Q^{(s,i)}_p$ is the number of relief items that need to be directly delivered from i to p in scenario s, and p can be either the affected community or relief facility. Due to (2.19), unlike the standard single-commodity flow formulation that is mixed-integer linear programming (MIP), our model becomes a mixed-integer quadratically constrained programming (MIQCP). Though (2.19) is non-convex, binary variables in quadratic terms allow off-the-shelf solvers to handle it by linearization. Furthermore, our nested decomposition solution method tackles this quadratic constraint automatically, because variables in the quadratic terms are decomposed into problems of the different echelons. (2.21)–(2.24) are domains of decision variables.

3.3. *Valid inequalities*

The proposed SPRP model can be further enhanced in the following ways. Firstly, it is apparent that some variables must be zero in the optimal solution, hence they are set to zeros by (3.1) and (3.3) to reduce the searching space. Secondly, valid inequalities (3.2) and (3.4)–(3.7) are introduced to strengthen the continuous relaxation of SPRP. These reinforcements can be beneficial for finding better feasible solutions and providing a tightener lower bound. (3.1) is adapted from the strengthened inequalities used by Salhi et al. (2014) for the deterministic multi-depot heterogeneous fleet vehicle routing problem. (3.2) is adapted from the valid inequalities used by Albareda-Sambola et al. (2009) and Fazel-Zarandi and Beck (2012) for the deterministic plant location problem. (3.1) and (3.2) are stochastic counterparts for those of deterministic problem. Additionally, we introduce (3.3)–(3.7) to reinforce the formulation.

$$[\mathrm{VI}]\ f^{(s,i,h)}_{pi} = 0, \qquad \forall i \in T, p \in N, h \in H, s \in S \tag{3.1}$$

$$\sum_{k\in K} y_{ik} \leq \sum_{s\in S}\sum_{q\in N_i} x^s_{iq}, \qquad \forall i \in T \tag{3.2}$$

$$t^s_{iij} = 0, \qquad \forall i \in T, j \in D, s \in S \tag{3.3}$$

$$w^h_i \leq \sum_{s\in S}\sum_{q\in N_i} x^s_{iq}, \qquad \forall i \in T, h \in H \tag{3.4}$$

$$\sum_{h\in H} w^h_i Q \geq \sum_{j\in D} t^s_{ij} + \sum_{i'\in T}\sum_{j\in D} t^s_{i'ij} + \sum_{i'\in T}\sum_{j\in D} t^s_{ii'j}, \qquad \forall i \in T, s \in S \tag{3.5}$$

$$\sum_{h\in H'} w^h_i Q \geq \sum_{j\in D'^s} t^s_{ij} + \sum_{i'\in T}\sum_{j\in D'^s} t^s_{i'ij}, \qquad \forall i \in T, s \in S \tag{3.6}$$

$$\sum_{i\in T} l_i \geq \max_{s\in S}\left(\sum_{j\in D} \lceil r \cdot m^s_j \rceil\right) \tag{3.7}$$

(3.1) imposes a vehicle must dispatch all the loaded relief items before it completes the trip and returns to a facility. (3.2) makes sure a candidate facility can be opened only if at least one affected community is assigned to it. (3.3) avoids relief facilities to transport items to themselves. Similar to (3.2), (3.4) make sure that vehicles can be prepared at a facility only if at least one point is assigned to this facility. (3.5) and (3.6) impose the minimum vehicle capacity that is prepared. The right-hand side of (3.6) is different from that of (3.5) because it is assumed that relief facilities are located at less vulnerable places such that a relief facility will not be isolated after the disaster occurs, consequently, the situation to dispatch special vehicles to another facility not necessarily occur. Though in some cases letting special vehicles deliver resources to another relief facility may reduce the overall cost, (3.6) is valid since it is a lower bound on the special vehicle capacity. (3.5) and (3.6) imply the bounds for normal vehicles. (3.7) imposes that the resources prepositioned at relief facilities at least meet the maximum total demand of affected communities among all the scenarios.

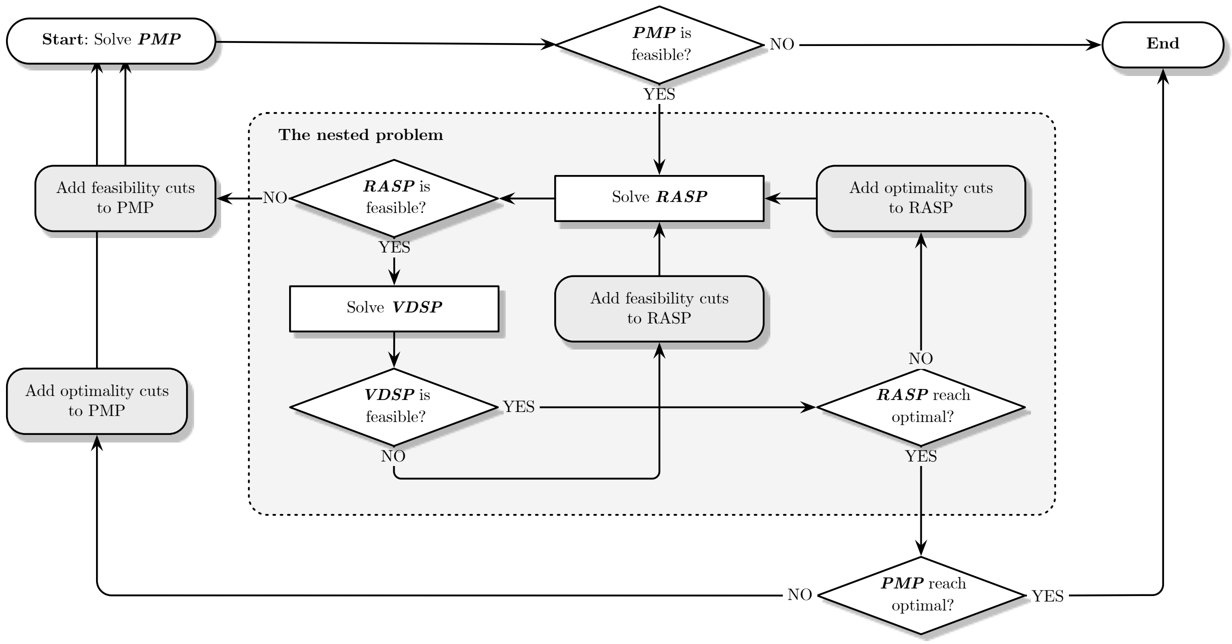


Figure 2: The NLBBD flowchart. The original SPRP (Section 3) is decomposed into a prepositioning master problem (PMP) (Section 4.1.1), a resource allocation subproblem (RASP) (Section 4.1.2), and a set of vehicle dispatching subproblems (VDSP) (Section 4.1.3). Logic-based Benders optimality cuts and feasibility cuts are proposed for the PMP (Section 4.2.1) and the RASP (Section 4.2.2).

4. Solution method

The solution method based on the NLBBD is developed in this section. Section 4.1 decomposes the original SPRP model. Section 4.2 introduces four sets of logic-based Benders cuts for the decomposed problems, two of which are feasibility cuts and the other two are optimality cuts. Section 4.3 presents the enhancements to the NLBBD method. Notations used in this section, proofs of cut validity, pseudocode, and implementation details for the algorithm are provided in the supplementary material.

4.1. Nested logic-based Benders Decomposition

The number of constraints and decision variables associated with SPRP grows with the scenario number, facility number, and affected community number. Therefore, the problem with a large number of scenarios, facilities, or affected communities cannot be solved efficiently using off-the-shelf solvers. To solve the problem efficiently, we develop a solution method based on the LBBD (Hooker and Ottosson, 2003, Hooker, 2019). The LBBD method, which is similar to BD, decomposes the original model into one master problem and some subproblems, but LBBD can handle any class of optimization problems. A set of problem-specific logic-based Benders cuts (*feasibility cuts* and/or *optimality cuts*) are used to link the master problem and subproblems. The master problem in LBBD is a relaxation of the original problem. In the objective function of the master problem, the objective of subproblems is replaced with some *estimators*. These estimators usually underestimate the subproblem objectives, when the master problem does not obtain sufficient information (Benders cuts) from the subproblems. The value functions of these estimators with respect to the master problem decision variables are usually nonlinear and noncontinuous for stochastic programming problems. However, with given master problem variable values (temporary solution), the true value of estimators can be computed by subproblems relatively easily. The difference between estimator values and objective values (or infeasibility proof) of subproblems implies the optimality of the solution in the current iteration for the original problem. The Benders optimality cuts exclude the temporary master problem solution when gaps between estimator values and subproblem objective values are detected. The Benders feasibility cuts exclude the temporary master problem solution when the subproblem is infeasible with the master problem solution. The LBBD converges to the optimum in a finite number of iterations when estimators being close enough to the subproblem objectives.

The hierarchical nature of two-stage stochastic programming, as well as the particular structure of the proposed SPRP, motivate us to extend the LBBD to NLBBD. The first-stage prepositioning decisions have a global impact, while the second-stage response (allocation and routing) decisions only have a relatively local impact. Fixing the first-stage decisions y_{ik}, w_i^h, and l_i results in $|S|$ independent multi-depot vehicle routing problems. Furthermore, once assignment decisions x_{ij} and allocation decisions t_{ij}^s and $t_{ii'j}^s$ are fixed, at most $|S| \cdot |T|$ independent, and single-depot, vehicle routing problem are obtained. The term "nested" in the NLBBD is used to express that the subproblems in the standard LBBD are further decomposed. As shown in Figure 2, the original SPRP can be decomposed into a prepositioning master problem (PMP) (Section 4.1.1), a resource allocation subproblem (RASP) (Section 4.1.2), and a set of vehicle dispatching subproblems (VDSP) (Section 4.1.3). The upper echelon problems feed the trial solution to lower echelon problems, while the lower echelon problems generate either Benders feasibility cuts or Benders optimality cuts to guide searching for the optimum (Section 4.2). Two versions of NLBBD, namely classical NLBBD (NLBBD-C) and modern branch-and-check (NLBBD-B&C) (Thorsteinsson, 2001, Hooker, 2019), are implemented. The only difference between NLBBD and NLBBD-B&C is that NLBBD builds a PMP branch-and-bound tree from scratch each time a new set of Benders cuts is generated, while NLBBD-B&C generates Benders cut and adds it to the branch-and-bound tree once an incumbent (integer feasible) solution is found. Thus, NLBBD-C builds up multiple branch-and-bound trees during iterations, while LBBD-B&C searches for the optimum in a single branch-and-bound tree, which implies the behavior differences between them. Branch-and-check is known to be efficient especially when the master problem is harder compared with the subproblem (Beck, 2010, Hooker, 2019). The pseudocode of LBBD-C and LBBD-B&C, and implementation details can be found in the supplementary material. The following subsections introduce the resulting decomposed problems.

4.1.1. Prepositioning master problem

Let $\mathcal{R}_{is}$ be an estimator for the second-stage recourse cost of specific $i \in T, s \in S$, we have the following PMP which copies constraints that contain only first-stage decisions (y_{ik}, l_i, x_{ij}, and w_i^h) from the SPRP. The constraints that contain both first-stage decisions and second-stage decisions are left for RASP and VDSPs. The proposed PMP is a relaxation of the original SPRP since it contains only a subset of the constraints in the original SPRP.

$$
\begin{aligned}
\text{[PMP] minimize } & \sum_{i\in T}\sum_{k\in K} c^k y_{ik} + \sum_{i\in T} c\, l_i + \sum_{i\in T}\sum_{h\in H} c_1^h w_i^h + \mathbb{E}_{s\in S}\left(\sum_{i\in T}\mathcal{R}_{is}\right) && && (4.1)\\
\text{subject to } & (1.2)\text{–}(1.8), (3.7) && && (4.2)\\
& \sum_{k\in K} y_{ik} \le \sum_{j\in D} x_{ij}, && \forall i\in T && (4.3)\\
& w_i^h \le \sum_{j\in D} x_{ij}, && \forall i\in T, h\in H && (4.4)\\
& \sum_{h\in H} w_i^h Q \ge l_i, && \forall i\in T && (4.5)\\
& \sum_{h\in H} w_i^h Q \ge \sum_{j\in D} \lceil r\cdot m_j^s\rceil \cdot x_{ij}, && \forall i\in T, s\in S && (4.6)\\
& \sum_{h\in H'} w_i^h Q \ge \sum_{j\in D'^s} \lceil r\cdot m_j^s\rceil \cdot x_{ij}, && \forall i\in T, s\in S && (4.7)\\
& (7.1), (7.2), && \text{[feasibility cuts]} && (4.8)\\
& (8), && \text{[optimality cuts]} && (4.9)\\
& \mathcal{R}_{is} \in \mathbb{R}_0^+, && \forall i\in T, s\in S && (4.10)
\end{aligned}
$$

(4.1) minimizes the sum of the first-stage cost and the expected second-stage cost. (4.2) replicates constraints (1.2)–(1.8) from the model SPRP. It also includes the valid inequality (3.7). (4.3)–(4.7) strengthen the PMP formulation as suggested by Hooker (2019) to contain relaxation of subproblems in the master problem. The relaxation is expressed in terms of master problem variables. (4.3) is similar to the valid inequality (3.2) that ensures a facility can be opened only if at least one affected community is assigned to it. (4.4) is similar to (3.4), ensuring vehicles can be prepared at a facility that has at least one affected community assigned to it. (4.5) ensures the total capacity of prepared vehicles can at least meet the total supply. (4.6) ensures the total capacity of vehicles prepared at a facility at least meets the demand of affected communities (adjusted by r) assigned to this facility. (4.7) is similar to (4.6) about the total capacity of special vehicles. (4.8) and (4.9) are Benders cuts that are discussed in detail in Section 4.2.1. (4.10) is domains of decision variable where "$\in \mathbb{R}_0^+$" is equivalent to "≥ 0".

4.1.2. Resource allocation subproblem

The RASP is an intermediary between the PMP and the VDSPs. For one thing, the RASP determines the quantity of relief items for VDSPs. For another, the RASP estimates the expected routing cost for the PMP. Note that the RASP is scenario-separable, such that we can formulate the RASP with $|S|$ independent subproblems. In this case, we do not have to solve the whole (relatively large-scale) RASP, but solve $|S|$ independent RASP. Though this decomposition for the RASP may help improve computational efficiency further, we keep the RASP undivided mainly because our experiment shows that the computational time for the RASP is less significant than that for the VDSPs. In the following subsections, we let $\bar{\cdot}, \tilde{\cdot}$, and $\hat{\cdot}$ denote variables values obtained by the PMP, the RASP, and the VDSPs, respectively.

$$
\begin{aligned}
\text{[RASP] } & \mathbb{E}_{s\in S}\left(\sum_{i\in T}\tilde{\mathcal{R}}_{is}\right) = \\
\text{minimize } & \sum_{s\in S} \mathsf{P}_s \sum_{i\in T} \mathcal{R}'_{is} && && (5.1)\\
\text{subject to } & (2.2)\text{–}(2.8), (2.21), (3.3) && && (5.2)\\
& \sum_{h\in H} w_i^{sh} Q \ge \sum_{j\in D} t_{ij}^s + \sum_{i'\in T}\sum_{j\in D} t_{i'ij}^s + \sum_{i'\in T}\sum_{j\in D} t_{ii'j}^s, && \forall i\in T, s\in S && (5.3)
\end{aligned}
$$

$$\sum_{h \in H'} w_i^{sh} Q \geq \sum_{j \in D'^s} t_{ij}^s + \sum_{i' \in T} \sum_{j \in D'^s} t_{i'ij}^s, \qquad \forall i \in T, s \in S \tag{5.4}$$

$$w_i^{sh} = \bar{w}_i^h, \qquad \forall i \in T, h \in H, s \in S \tag{5.5}$$

$$(10.1)\text{–}(10.4), \qquad \text{[feasibility cuts]} \tag{5.6}$$

$$(11.1)\text{–}(11.3), \qquad \text{[optimality cuts]} \tag{5.7}$$

$$w_i^{sh} \in \mathbb{Z}_0^+, \qquad \forall i \in T, h \in H, s \in S \tag{5.8}$$

$$\mathcal{R}'_{is} \in \mathbb{R}_0^+, \qquad \forall i \in T, s \in S \tag{5.9}$$

(5.1) minimizes the expected routing cost with the given PMP solution. In (5.1), $\tilde{\mathcal{R}}_{is}$ denotes the value of routing cost obtained by RASP, and $\mathcal{R}'_{is}$ is the estimator for routing cost in the RASP for the VDSPs. $\mathcal{R}'_{is}$ is different from that in the PMP, which is denoted as $\mathcal{R}_{is}$. Constraints (5.2) remains the same as (2.2)–(2.8) except in (5.2) y_{ik}, l_i, x_{ij} are replaced with $\bar{y}_{ik}, \bar{l}_i, \bar{x}_{ij}$ yielded by the PMP, respectively. (5.3) provides a lower bound on the total capacity of vehicles prepared at a facility. (5.4) is the counterpart of (5.3) for special vehicles. (5.5) introduces variables w_i^{sh}, which indicates the number of type h vehicles at facility i in scenario s. (5.6) and (5.7) are Benders feasibility cuts and optimality cuts that are discussed in detail in Section 4.2.2. (5.8) and (5.9) are domains of decision variables. The purpose of RASP is to check whether the first stage location decisions $\bar{x}_{ij}$, assignment decisions $\bar{y}_{ik}$, and relief item storage decisions $\bar{l}_i$ provide a feasible resource allocation scheme given the demand and service level r. Once a RASP is infeasible with the given first-stage decision variables, feasibility cuts are generated for the PMP (see Section 4.2 for details). When the RASP is feasible we proceed to solve the VDSPs in Section 4.1.3.

4.1.3. Vehicle dispatching subproblems

Given the decisions yielded by PMP and RASP, there are at most $|S| \cdot |T|$ independent VDSPs considering that some of the relief facilities may not be opened. Suppose that $x_{ij}, y_{ik}, t_{ij}^s, t_{ii'j}^s$ are temporarily fixed to $\bar{x}_{ij}, \bar{y}_{ik}, \tilde{t}_{ij}^s, \tilde{t}_{ii'j}^s$, $\forall i, i' \in T, j \in D, s \in S$ by the PMP and the RASP. For each $i \in T$ that $\sum_{k \in K} \bar{y}_{ik} = 1$ (opened facility): let $\tilde{D}_i^s \subset D$, where $e_j^s = 0$ and $\bar{x}_{ij} = 1 \ \forall j \in D$, denote non-isolated affected communities assigned to i; let $\tilde{D'}_i^s = D \setminus \tilde{D}_i^s$, where $e_j^s = 1$ and $\bar{x}_{ij} = 1 \ \forall j \in D$, denote isolated affected communities assigned to i; let $\tilde{T}_i^s \subset T$, where $\sum_{j \in D} \tilde{t}_{ii'j}^s \geq 1 \ \forall i' \in T$, denote relief facilities that accept relief items from i. For each relief facility i in scenario s, $\tilde{\mathcal{G}}^{(s,i)} = (\tilde{\mathcal{N}}^{(s,i)}, \tilde{\mathcal{A}}^{(s,i)})$ denotes a symmetric complete graph with node set $\tilde{\mathcal{N}}^{(s,i)} = \{i\} \cup \tilde{D'}_i^s \cup \tilde{D}_i^s \cup \tilde{T}_i^s$ and arc set $\tilde{\mathcal{A}}^{(s,i)} = \{(p,q) | p, q \in \tilde{\mathcal{N}}^{(s,i)}\}$. Let facility $\{i\}$ be the depot, the customer set for $\{i\}$ can be denoted as $\tilde{\mathcal{N}}_i^{(s,i)} = \tilde{\mathcal{N}}^{(s,i)} \setminus \{i\} = \tilde{D'}_i^s \cup \tilde{D}_i^s \cup \tilde{T}_i^s$. The quantity of acquired relief item of demand point in $\tilde{\mathcal{N}}_i^{(s,i)}$ is denoted as $\tilde{Q}_j^{(s,i)} = \tilde{t}_{ij}^s, \forall j \in \tilde{D'}_i^s \cup \tilde{D}_i^s$, $\tilde{Q}_{i'}^{(s,i)} = \sum_{j \in D} \tilde{t}_{ii'j}^s, \forall i' \in T_i^s$. Then, let integer variables $f_{pq}^{(s,i,h)}$ indicate the quantity of relief item transported on arc $(p,q) \in \tilde{\mathcal{A}}^{(s,i)}$ in scenario s by vehicle of type h that dispatched from relief facility i, let integer variables $x_{pq}^{(s,i,h)} = 1$ indicate the vehicle h dispatched from relief facility i travels on arc $(p,q) \in \tilde{\mathcal{A}}^{(s,i)}$ in scenario s, $x_{pq}^{(s,i,h)} = 0$ otherwise. Note that the isolated affected communities in $\tilde{D'}_i^s$ can only be visited by special vehicle $h \in H'$. For each $(i,s) \in T \times S$ pair, we have a VDSP as follows.

$$[\text{VDSP}] \ \hat{\mathcal{R}'}_{is} =$$

$$\text{minimize} \sum_{h \in H} \sum_{(p,q) \in \tilde{\mathcal{A}}^{(s,i)}} c_{pq}^h \cdot z_{pq}^{(s,i,h)} \tag{6.1}$$

$$\text{subject to} \sum_{q \in \tilde{\mathcal{N}}_i^{(s,i)}} z_{iq}^{(s,i,h)} \leq \tilde{w}_i^{sh}, \qquad \forall h \in H \tag{6.2}$$

$$\sum_{h \in H} \sum_{p \in \tilde{\mathcal{N}}^{(s,i)}} z_{pq}^{(s,i,h)} = 1, \qquad \forall q \in \tilde{\mathcal{N}}_i^{(s,i)} \tag{6.3}$$

$$\sum_{h \in H'} \sum_{p \in \tilde{\mathcal{N}}^{(s,i)}} z_{pq}^{(s,i,h)} = 1, \qquad \forall q \in \tilde{D'}_i^s \tag{6.4}$$

$$\sum_{p \in \tilde{\mathcal{N}}^{(s,i)}} z_{pq}^{(s,i,h)} - \sum_{p \in \tilde{\mathcal{N}}^{(s,i)}} z_{qp}^{(s,i,h)} = 0, \qquad \forall q \in \tilde{\mathcal{N}}_i^{(s,i)}, h \in H \tag{6.5}$$

$$\sum_{h \in H} \sum_{q \in \tilde{\mathcal{N}}^{(s,i)}} f_{qp}^{(s,i,h)} - \sum_{h \in H} \sum_{q \in \tilde{\mathcal{N}}^{(s,i)}} f_{pq}^{(s,i,h)} = \tilde{Q}_p^{(s,i)}, \qquad \forall p \in \tilde{\mathcal{N}}_i^{(s,i)} \tag{6.6}$$

$$\tilde{Q}_q^{(s,i)} z_{pq}^{(s,i,h)} \leq f_{pq}^{(s,i,h)} \leq (Q - \tilde{Q}_p^{(s,i)}) z_{pq}^{(s,i,h)}, \qquad \forall (p,q) \in \tilde{\mathcal{A}}^{(s,i)}, h \in H \tag{6.7}$$

$$f_{pi}^{(s,i,h)} = 0, \qquad \forall p \in \tilde{\mathcal{N}}^{(s,i)}, h \in H \tag{6.8}$$

$$z_{pq}^{(s,i,h)} \in \{0,1\}, \qquad \forall (p,q) \in \tilde{\mathcal{A}}^{(s,i)}, h \in H \tag{6.9}$$

$$f_{pq}^{(s,i,h)} \in \mathbb{Z}_0^+, \qquad \forall (p,q) \in \tilde{\mathcal{A}}^{(s,i)}, h \in H \tag{6.10}$$

(6.1) minimizes the routing cost. (6.2) ensures that the vehicle of type h used in relief facility i in every scenario s cannot exceed the number of vehicles dispatched to i in the prepositioning phase. (6.3) ensures that every demand node in the VDSP is visited exactly once. (6.4) ensures that the isolated demand nodes are visited by the special vehicle exactly once. (6.3)–(6.5) ensure that the isolated demand nodes cannot be visited by normal vehicles. (6.6) ensures the relief items preserved at the affected community or transfer facility are exactly equal to its demand. (6.7) eliminates the sub-tours. As a counterpart of the quadratic constraint (2.19), (6.7) is linear since the transportation decisions $\tilde{t}^s_{ij}$ and $\tilde{t}^s_{ii'j}$ are fixed to specific values. (6.8) is a valid inequality that is similar to (3.1), imposing that a vehicle must return to a facility without a remainder item after a trip is finished. (6.9) and (6.10) are domains of decision variables.

4.2. Logic-based Benders cuts

Two types of Benders cuts, i.e., Benders feasibility cuts and Benders optimality cuts, are proposed for the PMP and the RASP. Once the lower echelon problem is infeasible with the decisions provided by the upper echelon problem, feasibility cuts are generated by the lower echelon problem for the upper echelon problem, i.e., the RASP to the PMP and the VDSP to the RASP. Once the lower echelon problem is optimized but its solution indicates that the upper echelon problem underestimates the objective value, optimality cuts are generated by the lower echelon problem for the upper echelon problem. A valid Benders feasibility cut must (i) excludes the current master problem solution if it is not globally feasible, and (ii) not remove any globally feasible solution (Chu and Xia, 2004). A valid Benders optimality cut must (iii) excludes the current master problem solution if it is not globally optimal, and (iv) not remove any globally optimal solution (Guo et al., 2021). These properties hold for our logic-based Benders cuts. (i) and (iii) guarantee finite convergence if the master problem variables have a finite domain. (ii) and (iv) guarantee optimality because the cuts never remove globally optimal solutions. Hence, our NLBBD algorithm converges to the optimum in a finite number of iterations.

4.2.1. Benders cuts for PMP

Feasibility cuts: We feed vehicle variable values ($\bar{w}^h_i$), location variable values ($\bar{y}_{ik}$), assignment variable values ($\bar{x}_{ij}$), and inventory variable values ($\bar{l}_i$) yielded by the PMP to the RASP, to check the feasibility of vehicle variables w^{sh}_i. When the RASP turns out to be infeasible after some Benders cuts are added by the VDSPs, it is implied that the PMP solution is also not globally feasible. This kind of infeasibility is caused by the conflict between (5.5) and (5.6). The reason why other variables ($\bar{y}_{ik}$, $\bar{x}_{ij}$, and $\bar{l}_i$) are neglected to check the feasibility is that (3.7) (in (4.2)) ensures sufficient supplies, so that there is at least one feasible way to allocate relief items.

In an iteration u of PMP, let $\bar{w}^{h(u)}_i$ be the vehicle variable values and $\bar{x}^{(u)}_{ij}$ be the assignment variable values. Given the service level r and affected community demands m^s_j, the minimum number of vehicle requirements of facility i can be obtained by solving a series of bin packing problems $\mathbb{B}(L, Q)$ where L is a list of items with certain sizes and Q is the capacity of the bin (see details in the supplementary material). For each scenario, let $\mathbb{B}^{s(u)}_i\left(\lceil r \cdot m^s_j \rceil \forall j \in \bar{D}^{(u)}_i, Q\right)$ be the lower bound of the vehicle number, then the minimum number of vehicles required to be prepared at facility i is $\max_{s\in S}\left\{\mathbb{B}^{s(u)}_i\left(\lceil r \cdot m^s_j \rceil \forall j \in \bar{D}^{(u)}_i, Q\right)\right\}$, which is sufficient for the worst scenario. Considering that in a future iteration, customers in $\bar{D}^{(u)}_i = \{j \in D | \bar{x}^{(u)}_{ij} = 1\}$ may be assigned to other facilities rather than i, we present the Benders feasibility cuts generated by the RASP for the PMP, inspired by Fazel-Zarandi and Beck (2012) for deterministic facility location problem with homogeneous customer, as

$$\sum_{h\in H} w^h_i \geq \max_{s\in S}\left\{\mathbb{B}^{s(u)}_i\left(\lceil r \cdot m^s_j \rceil \forall j \in \bar{D}^{(u)}_i, Q\right)\right\} - \sum_{j\in \bar{D}^{(u)}_i}(1 - x_{ij}), \qquad \forall i \in \mathcal{I}^{(u)} \tag{7.1}$$

$$\sum_{h\in H'} w^h_i \geq \mathbb{B}^{s(u)}_i\left(\lceil r \cdot m^s_j \rceil \forall j \in \bar{D'}^{s(u)}_i, Q\right) - \sum_{j\in \bar{D'}^{s(u)}_i}(1 - x_{ij}), \qquad \forall i \in \mathcal{I}^{(u)}, s \in S \tag{7.2}$$

where $\mathcal{I}^{(u)}$ is the set of facilities that do not have a sufficient number of vehicles in iteration u. The right-hand side of (7.1) is the minimum number of vehicles that are required to be prepared at facility i considering that some affected community j may be reassigned to other facilities. (7.2) is the counterpart of (7.1) for special vehicles, where $\bar{D'}^{s(u)}_i = \{j \in D | e^s_j = 1, \bar{x}^{(u)}_{ij} = 1\}$ denotes the set of isolated communities.

Optimality cuts: When the RASP is optimized with a given PMP solution but there is a gap between the RASP objective and the PMP recourse objective, it implies that the PMP underestimates the recourse cost, thus optimality cuts are generated and added to the PMP. The idea of our optimality cuts for the PMP is, that when the assignment decisions x_{ij} are modified, the performance of the PMP may be further improved. In other words, the PMP can try other assignments, and the new assignments may result in RASP with a lower objective. If no better objective is obtained from the RASP, the recourse estimators in the PMP are increased in turn. Let $\bar{D}^{(u)}_i \cup \{i\}$ denote the cluster of points that are assigned to facility i in iteration u, the PMP solution may be improved by reassigning some of the points in this cluster to other facilities in the future iterations. We present optimality cuts as

$$\mathcal{R}_{is} \geq \tilde{\mathcal{R}}^{(u)}_{is} - \sum_{j\in \bar{D}^{(u)}_i}\left[\max_{p\in \bar{D}^{(u)}_i \cup \{i\}}\left\{2 \cdot \max_{h\in H} c^h_{pj}\right\}\right](1 - x_{ij}). \tag{8}$$

The first term of the right-hand side of (8) is the routing cost obtained by solving the RASP. For each affected community j assigned to facility i in iteration u, if j is assigned to another facility rather than i, the maximum routing cost reduction is the term in the square bracket, namely cutting coefficient. If j is removed from the cluster $\bar{D}_i^{(u)} \cup \{i\}$, it reduces in maximum by twice the maximum travel cost between j and another point in $\bar{D}_i^{(u)} \cup \{i\}$. (8) is inspired by the optimality cut by Fachini and Armentano (2020) for the deterministic single-depot heterogeneous fleet vehicle routing problem with time windows. (8) has a single-cut counterpart that can be formulated as

$$\sum_{s\in S}\mathsf{P}_s\sum_{i\in T}\mathcal{R}_{is} \geq \sum_{s\in S}\mathsf{P}_s\sum_{i\in T}\tilde{\mathcal{R}}_{is}^{(u)} - \sum_{i\in T}\sum_{j\in\bar{D}_i^{(u)}}\left[\max_{p\in\bar{D}_i^{(u)}\cup\{i\}}\left\{2\cdot\max_{h\in H} c_{pj}^h\right\}\right](1-x_{ij}). \tag{9}$$

For the problem with larger $|T|$ and $|S|$, (8) may result in a larger master problem with more constraints. By introducing (9), the problem size can be reduced significantly. But on the other hand, (9) provides poorer recourse cost approximation than (8). The trade-off between (8) and (9) is the number of iterations and runtime for solving the master problem in each iteration. In our instances, (8) is used since it is more efficient than (9).

4.2.2. Benders cuts for RASP

Feasibility cuts: Given a particular transportation solution $(\tilde{w}_i^{sh}, \tilde{t}_{i,j}^s, \tilde{t}_{ii'j}^s)$ yielded by the RASP, if the resulting (s,i)-specific VDSP solution is infeasible, feasibility cuts are generated to strengthen the RASP. When the demands of affected communities are separable, i.e., VDSP can be formulated as a VRP with separable demands, the RASP solution will never be infeasible for the VDSP since (4.5), (5.3), and (5.4) guarantee sufficient vehicle capacity, furthermore (3.7) (in (4.2)) and horizontal coordination guarantee sufficient supply. In our settings, the demand is inseparable, i.e., all the demands of an affected community must be fulfilled by a single vehicle. In this case, the infeasibility is caused by the fact that the truckload is not fully used.

In a particular iteration u, let $\tilde{D}_i^{s(u)} = \{j \in D | \tilde{t}_{ij}^{s(u)} \geq 1\}$ and $\tilde{D'}_i^{s(u)} = \{j \in \tilde{D}_i^{s(u)} | e_j^s = 1\}$ be all the customers and isolated customers that obtain relief items directly from facility i, $\tilde{T}_i^{s(u)} = \{i' \in T | \sum_{j\in D}\tilde{t}_{ii'j}^{s(u)} \geq 1\}$ be the facilities that obtain relief items from facility i such that these items can be transported to other customers, $\tilde{t}_{ii'j}^{s(u)}$ and $\tilde{t}_{ij}^{s(u)}$ denote the transshipment decisions that fed by the RASP. Let $\mathcal{I}^{s(u)}$ denote a set of i for which the (s,i)-specific VDSPs are proved to be infeasible in iteration u, we have the Benders feasibility cuts as

$$\sum_{h\in H} w_i^{sh} \geq \mathbb{B}_i^{s(u)}\left(\tilde{t}_{ij}^{s(u)}, \tilde{t}_{ii'j}^{s(u)}, Q\right) - \sum_{j\in\tilde{D}_i^{s(u)}} W_{ij}^{s(u)} - \sum_{i'\in\tilde{T}_i^{s(u)}} W_{ii'}^{s(u)}, \qquad \forall i\in\mathcal{I}^{s(u)}, s\in S \tag{10.1}$$

$$\sum_{h\in H'} w_i^{sh} \geq \mathbb{B}_i^{s(u)}\left(\tilde{t}_{ij}^{s(u)}, Q\right) - \sum_{j\in\tilde{D'}_i^{s(u)}} W_{ij}^{s(u)}, \qquad \forall i\in\mathcal{I}^{s(u)}, s\in S \tag{10.2}$$

$$W_{ij}^{s(u)} = \begin{cases} 1, & \text{if } t_{ij}^s \leq \tilde{t}_{ij}^{s(u)} - 1 \\ 0, & \text{otherwise} \end{cases}, \qquad \forall i\in\mathcal{I}^{s(u)}, j\in\tilde{D}_i^{s(u)}, s\in S \tag{10.3}$$

$$W_{ii'}^{s(u)} = \begin{cases} 1, & \text{if } \sum_{j\in D} t_{ii'j}^s \leq \sum_{j\in D}\tilde{t}_{ii'j}^{s(u)} - 1 \\ 0, & \text{otherwise} \end{cases}, \qquad \forall i\in\mathcal{I}^{s(u)}, i'\in\tilde{T}_i^{s(u)}, s\in S. \tag{10.4}$$

There are two possible ways to cut off the current infeasible RASP solution: (i) adding vehicles to facility i; (ii) decreasing the quantity of relief items transported to other facilities and communities. (10.1)–(10.4) are based on this idea. The first term of the right-hand side of (10.1), $\mathbb{B}_i^{s(u)}\left(\tilde{t}_{ij}^s, \tilde{t}_{ii'j}^s, Q\right)$, is the minimum number of vehicles obtained by solving a bin packing problem. The capacity of bin is Q, item set is $\tilde{D}_i^{s(u)} \cup \tilde{T}_i^{s(u)}$, and sizes of items are $\tilde{t}_{ij}^{s(u)}$ and $\sum_{j\in D}\tilde{t}_{ii'j}^{s(u)}$. The second term of the right-hand side of (10.1), $\sum_{j\in\tilde{D}_i^{s(u)}} W_{ij}^s$, is the maximum reduction of vehicle quantity considering that the direct transportation quantity may be decreased by the RASP. Similarly, the third term of the right-hand side of (10.1), $\sum_{i'\in\tilde{T}_i^{s(u)}} W_{ii'}^s$, is the maximum reduction of vehicle quantity considering that the horizontal coordination quantity may be decreased. For a facility $j \in \tilde{D}_i^{s(u)}$, when the relief items that are transported to it horizontally are reduced, the maximum vehicle quantity reduction of $i \in \mathcal{I}^{s(u)}$ is one. When the item transportation quantity is increased, the minimum vehicle quantity requirement cannot be lower than $\mathbb{B}_i^{s(u)}(\tilde{t}_{ij}^s, \tilde{t}_{ii'j}^s, Q)$. In summary, (10.1) is the lower bound on the number of all types of vehicles given the resource allocation decisions; (10.2) is the lower bound on the number of special vehicles given the resource allocation decisions.

Optimality cuts: In an iteration u, when the RASP solution is feasible for a (s,i)-specific VDSP, but there is a gap between VDSP objective value $\hat{\mathcal{R}}_{is}^{(u)}$ and the RASP objective value $\tilde{\mathcal{R}}_{is}^{(u)}$, optimality cuts are added to the RASP. Since transportation decisions t_{ij}^s and $t_{ii'j}^s$ are fixed for the VDSPs, the VDSPs provide an upper bound on the recourse objective of the RASP by solving in maximum $|T|$ independent vehicle routing problem. Contrary to the feasibility cuts that are added when subproblem is infeasible, optimality cuts are generated when subproblem is feasible, and it is added to the upper echelon problem when it is validated (Rahmaniani et al., 2017), i.e., $\hat{\mathcal{R}}_{is}^{(u)} > \tilde{\mathcal{R}}_{is}^{(u)}$.

Recall the notations used for feasibility cuts, in an iteration u and a scenario s, let $\tilde{D}_i^{s(u)}$ be customers that obtain relief items directly from facility i, $\tilde{T}_i^{s(u)}$ be facilities that receive relief items from facility i. The gap between the VDSP recourse cost $\hat{\mathcal{R}}_{is}^{(u)}$ and the RASP recourse cost $\tilde{\mathcal{R}}_{is}$ implies that the RASP underestimated the routing cost, since the RASP is a relaxation for the routing problem. In case of $\hat{\mathcal{R}}_{is}^{(u)} > \tilde{R}_{is}$, Benders optimality cut is added to the RASP. Otherwise, if $\hat{\mathcal{R}}_{is}^{(u)} = \tilde{R}_{is}$ for all the i and s, the RASP is optimized with the given PMP location, inventory, assignment, and vehicle decisions. The Benders optimality cuts are presented as

$$\mathcal{R}'_{is} \geq \hat{\mathcal{R}}'^{(u)}_{is} - \sum_{j \in \tilde{D}_i^{s(u)}} R_{ij}^{s(u)} - \sum_{i' \in \tilde{T}_i^{s(u)}} R_{ii'}^{s(u)}, \qquad \forall i \in T, s \in S \tag{11.1}$$

$$R_{ij}^{s(u)} = \begin{cases} MAX_{\tilde{C}_{is}^{(u)},i,j}, & \text{if } t_{ij}^{s} \leq \tilde{t}_{ij}^{s(u)} - 1 \\ 0, & \text{otherwise} \end{cases}, \qquad \forall i \in T, j \in \tilde{D}_i^{s(u)}, s \in S \tag{11.2}$$

$$R_{ii'}^{s(u)} = \begin{cases} MAX_{\tilde{C}_{is}^{(u)},i,i'}, & \text{if } \sum_{j \in D} t_{ii'j}^{s} \leq \sum_{j \in D} \tilde{t}_{ii'j}^{s(u)} - 1 \\ 0, & \text{otherwise} \end{cases}, \qquad \forall i \in T, i' \in \tilde{T}_i^{s(u)}, s \in S \tag{11.3}$$

where $MAX_{\tilde{C}_{is}^{(u)},i,j} = \max\left\{2 \cdot \max_{h \in H} c_{ij}^{h}, 2 \cdot \max_{p,q \in \tilde{C}_{is}^{(u)}, h \in H} \left\{c_{pq}^{h}\right\}\right\}$. The first term $2 \cdot \max_{h \in H} c_{ij}^{h}$ denotes the cost of dispatching another vehicle and the second term $2 \cdot \max_{p,q \in \tilde{C}_{is}^{(u)}, h \in H} \left\{c_{pq}^{h}\right\}$ denotes the cost of inserting j to an existing route. The left-hand side of (11.1) is the estimator of the routing cost of the RASP model. The first term of the right-hand side of (11.1) is the routing cost value yielded by the VDSP in iteration u. The second term of the right-hand side of (11.1) (also (11.2)) defines the maximum reduction of routing cost once the transportation decision $\tilde{t}_{ij}^{s(u)}$ changes. The last term of (11.1) (also (11.3)) is the counterpart of (11.2) for facilities that received relief items from other facilities. The idea of (11.1) is either dispatching additional vehicles or reducing the number of relief items transported out of facility i would lead to a reduction in the routing costs. For (11.1), in case of (i) the transportation decision $t_{ii'j}^{s}$ is equal to or less than $(\tilde{t}_{ij}^{s(u)} - 1)$, the reduction of routing cost is twice the maximum distance between pair of points in the set $\tilde{C}_{is}^{(u)} = \tilde{D}_i^{s(u)} \cup \tilde{T}_i^{s(u)} \cup \{i\}$, which are facilities that i served, customers that i served, and i itself; (ii) the transportation decision $t_{ii'}^{s}$ is same as $\tilde{t}_{ij}^{s(u)}$, the reduction of routing cost is zero; and (iii) the transportation decision $t_{ii'}^{s}$ is more than $\tilde{t}_{ij}^{s(u)}$, since the right-hand side of (11.1) is a lower bound on routing cost, we simply set the reduction value to zero.

4.3. Implementation enhancements

Benders decomposition is known to be slow in convergence. To accelerate it, there are three directions: (i) accelerating the master problem; (ii) accelerating subproblems; and (iii) generating strong cuts. (i) has been addressed by implementing NLBBD-B&C, which searches for the optimal PMP solution on a single branch-and-bound tree, as discussed in Section 4.1. We address (ii) by accelerating the VDSPs and using upper bounds of the VDSPs, and address (iii) by initializing warm-start cuts.

Accelerating VDSPs: Our initial experiments using the MIP solver, specifically Gurobi, for VDSPs show that reaching proven optimums is difficult: due to the poor lower bounds of MIP relaxations. The large proportion of computational time occupied by VDSPs suggests accelerating VDSPs can speed up the overall algorithm. Therefore, the `VRPSolverEasy` (Errami et al., 2024), which is a Python interface of `VRPSolver` (Pessoa et al., 2020), is used for solving VDSPs. `VRPSolver` is a state-of-the-art sophisticated exact solver based on branch-cut-and-price for VRP variants. It focuses on proving optimality for a known solution by improving lower bounds, and is able to solve VRP instances with less than 100 customers in minutes (Errami et al., 2024). `VRPSolverEasy` is easier to use than `VRPSolver`. Unlike `VRPSolver`, which requires knowledge of MIP modeling and branch-cut-and-price algorithms, `VRPSolverEasy` defines VRPs using customers, depots, links, and vehicle types. For simplicity, `VRPSolverEasy` compromises to a subset of VRP variants that can be handled by the `VRPSolver`. Our VDSPs fall within this subset.

Using upper bounds of VDSPs: Our VDSPs can always be solved exactly by `VRPSolverEasy` in minutes. For more difficult instances that cannot reach proven optimum easily, we may choose to terminate the solution process of VDSPs earlier by computational time limit or optimality gap tolerance. However, the validity of (10) and (11) with a suboptimal VDSP (upper bound) need to be analyzed. (10) remains valid since it is generated only when VDSP is infeasible. The validity of (11) is not this straightforward. Let $\tilde{\mathcal{R}}'_{is}$ be the value of the estimator of RASP, $\hat{\mathcal{R}}'^{*}_{is}$ be the optimal VDSP objective, and $\hat{\mathcal{R}}'^{U}_{is}$ be an upper bound of $\hat{\mathcal{R}}'^{*}_{is}$, i.e., $\hat{\mathcal{R}}'^{U}_{is} \geq \hat{\mathcal{R}}'^{*}_{is}$. Assume that during iteration we only know the information of $\hat{\mathcal{R}}'^{U}_{is}$ instead of $\hat{\mathcal{R}}'^{*}_{is}$, and in an iteration it is detected $\hat{\mathcal{R}}'^{U}_{is} > \tilde{\mathcal{R}}'_{is}$. In this case, optimality cut (11) is added to the RASP, resulting current RASP solution being excluded for future iterations. However, the gap between $\hat{\mathcal{R}}'^{U}_{is}$ and $\hat{\mathcal{R}}'^{*}_{is}$ suggests the possibility of $\hat{\mathcal{R}}'^{*}_{is} \leq \tilde{\mathcal{R}}'_{is}$, indicating that (11) should not be generated. A potentially optimal RASP solution is cut off, resulting in a reduced feasibility region compared to the true one. As a result, sub-optimal VDSPs lead to an upper bound for the original problem. In this study, when solving VDSPs using Gurobi, we set a stopping criterion of 500 seconds computational time or 5% optimality gap whichever is first hit; when solving VDSPs using `VRPSolverEasy`, this strategy is not used, since they can reach proven optimums within 350 seconds (with the majority taking less than 100 seconds).

Initializing warm-start cuts: We observed that the lower bounds in initial iterations of NLBBD-C and NLBBD-B&C can be low, leading to high gaps and consequentially slow convergence. This is due to the values of $\mathcal{R}_{ij}$ (estimator for the second-stage costs) in the PMP being guided by Benders cuts. They can be close to zero until a sufficient number of Benders cuts are added after a sufficient number of iterations. In stochastic programming, it is well known that the objective values of perfect information problems (wait-and-see problems) yield lower bounds to the objective value of the stochastic problem.

This can be used to generate cuts at the root node, namely *wait-and-see warm-start cuts* (Guo and Zhu, 2023, Kůdela and Popela, 2017). To derive these cuts, we need to solve $|S|$ (number of scenarios) independent problem each with the scenario being $S = \{s\}$, i.e., perfect information problem. Then the objective of the scenario $\{s\}$ in PMP cannot be smaller than the objective of the perfect information problem, as formulated below for each $s \in S$.

$$\sum_{i \in T} \sum_{k \in K} c^k y_{ik} + \sum_{i \in T} c\ l_i + \sum_{i \in T} \sum_{h \in H} c_1^h w_i^h + \sum_{i \in T} \mathcal{R}_{is} \geq \min_{\substack{(1.2)-(3.7) \\ S=\{s\}}} \left(\sum_{i \in T} \sum_{k \in K} c^k y_{ik} + \sum_{i \in T} c\ l_i + \sum_{i \in T} \sum_{h \in H} c_1^h w_i^h + \sum_{i \in T} \mathcal{R}_{is} \right) \quad (12)$$

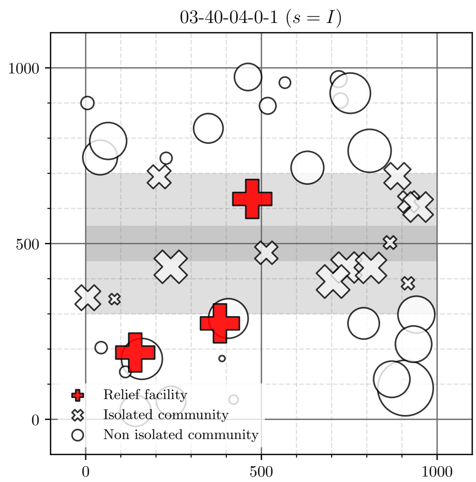
In (12), the left-hand side is the objective of scenario s, and the right-hand side is the objective value of perfect information problem with $S = \{s\}$. Since the right-hand side is a lower bound of the objective of scenario s, we do not have to solve the perfect information problem to optimum but can use a valid lower bound of the perfect information problem, to alleviate the computational burden. We allocate 120 seconds in maximum to each scenario-specific warm-start problem. Note that the lower bounds of perfect information problems can be easily retrieved from the MIP solver.

5. Computational performance

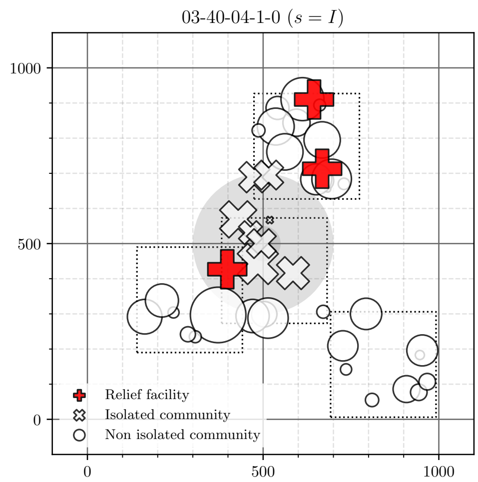
We implemented all the algorithms with Gurobi 9.1's Python application programming interface (API), the version of Python is 3.7.4. The machine that our code ran on was a CentOS 7.9 Linux server with four cores (eight threads) of Intel Xeon Platinum (Cooper Lake) 8369 (3.3 GHz, Turbo Boost speed of 3.8 GHz) CPU and 16GB of RAM. Some of the experiments that are not time-sensitive were run on the NEOS Server (Dolan, 2001). Extensive numerical experiments are conducted in this section, which demonstrates the efficiency of the NLBBD method. Section 5.1 presents the test instances that are used for the following numerical experiments. Section 5.2 compares the performance of the extensive form of SPRP with NLBBD. Section 5.3 compares the convergence behaviors of NLBBD implementations.

5.1. Instances design

Since there is no benchmark instance for our SPRP, we designed instances for numerical experiments. Let a particular class of instances be denoted as $|T|$-$|D|$-$|S|$-$\mathcal{C}$-$\mathcal{T}$, where $|T| \in \{3, 5, 7\}$ is the number of relief facility candidates, $|D| \in \{10, 20, 40\}$ is the number of affected community, $|S| \in \{2, 4\}$ is the number of scenarios, $\mathcal{C} \in \{0 \text{ (uniform)}, 1 \text{ (clustered)}\}$ is the distribution type of affected communities, and $\mathcal{T} \in \{0 \text{ (earthquake)}, 1 \text{ (flood/storm)}\}$ represents the disaster type. The distribution type and disaster type that affect the distribution of isolated affected communities are discussed in detail later. For example, 7-40-4-1-1 represents a flood/storm instance with seven relief facility candidates, 40 clustered affected communities, and four scenarios. These parameter settings are aligned with the literature (see the supplementary material for detail), and thus are expected to accurately reflect real disaster situations. By using instances of various sizes, we can test the scalability of our proposed solution method and gain valuable managerial insights from optimized instances. The characteristics of a particular instance can be divided into scenario-independent ones (location of facilities and affected communities) and scenario-dependent ones (demand of affected communities and whether an affected community is isolated). The scenario-independent ones are introduced followed by the scenario-dependent ones.



(a) Uniformly distributed instance for flood/storm



(b) Clustered instance for earthquake

Figure 3: Instance illustration

Generating relief facility and affected community location: The locations of relief facility candidates are generated from a two-dimensional integer uniform distribution $IU([0, 1000]^2)$. For small- and medium-size instances ($|D| \in \{10, 20\}$), only uniformly distributed ($\mathcal{C} = 0$) affected communities are generated, and the generating method is the same as that for relief facilities. While for large-size instances ($|D| = 40$), additional clustered ($\mathcal{C} = 1$) instances are generated. To generate clustered instances, all the affected communities are firstly assigned to one cluster among four with an equal probability, i.e., 0.25. The four cluster centers are generated from a two-dimensional integer distribution $IU([150, 850]^2)$. Let (o_x, o_y) be the center of a cluster, then affected communities assigned to it are generated from $IU([o_x - 150, o_x + 150] \times [o_y - 150, o_y + 150])$. The distance between pairs of locations is calculated by the Euclidean distance. Figure 3 are two instances with three facility candidates and 40 affected communities, which are uniformly distributed in Figure 3(a) and clustered in Figure 3(b) (four clusters in dotted-line squares).

Generating affected community demand: In practice, the Federal Emergency Management Agency (FEMA) of the United States classifies emergencies and disasters into minor emergencies, limited and potential emergencies, and major disasters by size, type, and number of issues that need to be addressed (FEMA, 2013). Accordingly, there are level 3, level 2, and level 1 (from less severe to more severe) emergency response activation levels to be activated (FEMA, 2019). In China, emergencies are classified into level IV, level III, level II, and level I (from less severe to more severe) by the General Office of the State Council, PRC (2016). Thus it is reasonable to generate instances with $S = \{IV, III, II, I\}$. We assume that the demand of each affected community is determined by population ($p_j \sim IU(50, 300)$) and *severity score* ($ss_s : ss_{IV} \sim U(0.0, 0.4), ss_{III} \sim U(0.2, 0.6), ss_{II} \sim U(0.4, 0.8), ss_I \sim U(0.6, 1.0)$), which denotes the average relief item consumption rate (in units of relief items per person) and is inherently determined by disaster scenarios. Note that $IU(\cdot)$ denotes integer uniform distribution and $U(\cdot)$ denotes uniform distribution. The demand of affected communities j is calculated by $\lceil 1 \times p_j \times ss_s \rceil$ where $\lceil \cdot \rceil$ is the ceiling function and 1 is one unit of emergency preparedness kit that contains the first-72-hours supplies to an individual. Note that for instances with two scenarios, only the level IV scenario and the level I scenario are considered. The scenarios mentioned above are created by aggregating disaster impacts, rather than using simulation results or historical records of disasters directly. This method typically requires fewer scenarios compared to the latter (see supplementary material for scenario numbers in literature). However, we consider it crucial to analyze the scalability of our proposed approach with respect to the number of scenarios $|S|$, which is presented in Section 5.4.

Generating isolated affected communities: The idea for generating isolated communities is based on the simple fact that an affected community has more potential to be isolated if it is closer to the center of a disaster. The "center" of a disaster is determined by its type. The top three most frequently occurring natural disasters from 2000 to 2019 are flood (44% in all recorded events), storm (28% in all recorded events), and earthquake (8% in all recorded events). These top three are responsible for 80% of the world's natural disasters (UNDRR, 2020). We assume that the center of a flood/storm is a line while the center of an earthquake is a point. Then the distance between affected communities and the disaster center can be calculated. The chance of an affected community being isolated can be determined by its distance to the disaster center and the *vulnerability score* (vs_s: $vs_{IV} = 0.10, vs_{III} = 0.25, vs_{II} = 0.50, vs_I = 1.00$), which is inherently determined by disaster scenarios. More specifically, let d_{oj} denotes the distance between affected community j and disaster center o, it is assumed that the center of flood/storm is the line $y = 500$ and the center of earthquake is the point $(500, 500)$ as shown in Figure 3; $F(d_{oj})$ be a step function $F(d_{oj}) = 0.80\chi_{[0,50)}(d_{oj}) + 0.50\chi_{[50,200)}(d_{oj}) + 0.20\chi_{[200,+\infty)}(d_{oj})$ where $\chi_{\mathsf{R}}(d_{oj}) = 1$ if $d_{oj} \in \mathsf{R}$, $\chi_{\mathsf{R}}(d_{oj}) = 0$ if $d_{oj} \notin \mathsf{R}$; and $x \sim B(\mathsf{P})$ be a Bernoulli distribution such that $Prob(x = 1) = \mathsf{P}$ and $Prob(x = 0) = 1 - \mathsf{P}$. Recall that e_j^s is a binary parameter indicating whether an affected community is isolated, e_j^s is generated from $B(F(d_{oj}) \times vs_s)$ for all the affected communities j and scenarios s.

To make sure all the instances are feasible, the capacity of vehicles at least meet the maximum possible affected community demand, i.e., 300 ($= \max(p_j) \times \max(ss_s)$), the large-size facility capacity is set to 4000 ($= \max(|D|) \times \max(p_j) \times \max(ss_s) / \min(|T|)$), and the small-size facility capacity is set to 1/4 of the large-size one, i.e., 1000. In summary, 48 distinct instances are generated, among which there are 12 small-size instances, 12 medium-size instances, 12 uniformly distributed large-size instances, and 12 clustered large-size instances. All the aforementioned parameters are summarized in the supplementary material. Note that we make sure the only difference among instances with the same $|T|$-$|D|$-$*$-$\mathcal{C}$-$*$ is demand quantity and isolated points distribution. In other words, instances with the same facility quantity, affected community quantity, and distribution type have the same facility and community locations and community population.

5.2. Computational efficiency

This subsection evaluates the efficiency of five different solution methods: (i) solving extensive forms of SPRP using Gurobi; (ii) solving extensive forms of SPRP with valid inequalities using Gurobi; (iii) the classical implementation of NLBBD with valid inequalities; (iv) the branch-and-check implementation of NLBBD with valid inequalities; (v) the enhanced (Section 4.3) branch-and-check implementation of NLBBD with valid inequalities. Results are reported in Table 2, Table 3, and Table 4. For each instance, the column SPRP reports the objective value obtained by solving (1.1)–(2.21) and (2.1)–(2.24) using Gurobi. The column SPRP-VI further adds valid inequalities (3.1)–(3.7). The column NLBBD-C reports solutions obtained by the classical NLBBD while NLBBD-B&C presents solutions of branch-and-check implementation. NLBBD-B&C^{++} uses the enhancements described in Section 4.3. For each solution method (SPRP, SPRP-VI, NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++}), Obj. denotes the objective value of the best feasible solution. Gap(%) denotes the optimality gap that is defined as (UB - LB) /

UB. For all the methods, UB is the best integer-feasible solution. For SPRP, SPRP-VI, NLBBD-B&C, and NLBBD-B&C^{++} the LB is the lower bound obtained from the linear relaxation, while for NLBBD-C the LB is the objective value of PMP with an incomplete set of Benders cuts in the last iteration. The time limit for all the solution methods is set to 3,600 seconds. The time limit of 3,600 seconds is set such that there can be easy instances that can be solved to proven optimum, medium instances that may not reach proven optimum but can obtain good quality solutions with lower optimality gap, and hard instances where our approach is the only one capable of yielding feasible solutions (as opposed to Gurobi). For instances that can be optimized within this limit, we report the solution time in the column Time(s).

Table 2: Computational performance (small size)

Instance	SPRP			SPRP-VI			NLBBD-C			NLBBD-B&C			NLBBD-B&C^{++}		
	Obj.	Gap (%)	Time (s)	Obj.	Gap (%)	Time (s)	Obj.	Gap' (%)	Time (s)	Obj.	Gap (%)	Time (s)	Obj.	Gap (%)	Time (s)
03-10-02-0-0	***329424**	0.00	16	***329424**	0.00	25	***329424**	0.00	200	***329424**	0.00	36	***329424**	0.00	45
03-10-02-0-1	***579322**	0.00	14	***579322**	0.00	14	***579322**	0.00	57	***579322**	0.00	20	***579322**	0.00	28
03-10-04-0-0	***295132**	0.00	310	***295132**	0.00	253	***295132**	0.00	472	***295132**	0.00	82	***295132**	0.00	94
03-10-04-0-1	***420081**	0.00	152	***420081**	0.00	156	***420081**	0.00	94	***420081**	0.00	61	***420081**	0.00	73
05-10-02-0-0	***316940**	0.00	77	***316940**	0.00	62	***316940**	0.00	220	***316940**	0.00	43	***316940**	0.00	40
05-10-02-0-1	***443174**	0.00	40	***443174**	0.00	47	***443174**	0.00	53	***443174**	0.00	35	***443174**	0.00	39
05-10-04-0-0	**285998**	1.75	Lmt.	**285998**	1.62	Lmt.	***285998**	0.00	1948	***285998**	0.00	113	***285998**	0.00	133
05-10-04-0-1	***347928**	0.00	1221	***347928**	0.00	744	***347928**	0.00	168	***347928**	0.00	99	***347928**	0.00	94
07-10-02-0-0	***267079**	0.00	358	***267079**	0.00	531	***267079**	0.00	547	***267079**	0.00	98	***267079**	0.00	86
07-10-02-0-1	***377938**	0.00	140	***377938**	0.00	182	***377938**	0.00	168	***377938**	0.00	65	***377938**	0.00	59
07-10-04-0-0	**265369**	2.24	Lmt.	**265369**	2.35	Lmt.	***265369**	0.00	933	***265369**	0.00	277	***265369**	0.00	302
07-10-04-0-1	***320798**	0.00	2716	***320798**	0.00	2943	***320798**	0.00	785	***320798**	0.00	160	***320798**	0.00	159

[1] **Boldface** denotes the best objective value among solution methods in this table; "*" denotes the optimized objective value; "-" denotes no feasible solution is found; "Lmt." means 3,600 seconds.

Performance differences among SPRP, SPRP-VI, NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++} vary with instances scale. In small-size instances (Table 2), NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++} outperform SPRP and SPRP-VI significantly. Most of the small-size instances can be optimized by NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++} in minutes. While within the given time limit, two instances cannot be optimized by SPRP and SPRP-VI, i.e., instances 05-10-04-0-0 and 07-10-04-0-0. It can be observed that the computational effort required by SPRP and SPRP-VI is more sensitive to the number of scenarios compared with NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++}. For instances with four scenarios, NLBBD-B&C and NLBBD-B&C^{++} can save computation time in order of magnitude compared with SPRP and SPRP-VI.

In medium-size instances (Table 3), there are six in 12 NLBBD-B&C^{++} solutions that are better or not worse than SPRP and SPRP-VI. However, the optimality gap of NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++} can be higher than that of SPRP and SPRP-VI, which can be explained from two aspects: upper bound and lower bound obtaining. Firstly, to obtain a feasible upper bound, the NLBBD methods (NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++}) have to solve vehicle routing problems, which can be time-consuming to reach the proven optimum, repetitively, Secondly, lower bounds of NLBBD-B&C and NLBBD-B&C^{++} are obtained from an incomplete PMP linear relaxation. The lower bound will not improve till sufficient Benders cuts are added to PMP. Nevertheless, it is worth noting that when the instances become more complex, e.g., 07-20-04-0-0 and 07-20-04-0-1, the NLBBD methods obtain much better upper bounds and optimality gaps than SPRP and SPRP-VI. This implies the potential of NLBBD in large-size instances. We also observe that the proposed enhancements significantly reduce the optimality gaps. This is evident when comparing the Gap(%) of NLBBD-B&C^{++} and NLBBD-B&C, as shown in Table 3. The *Wilcoxon signed-rank test* (a non-parametric statistical test that is used to determine whether two samples are derived from the same distribution) is conducted on the gap sequences of NLBBD-B&C^{++} and NLBBD-B&C, resulting in a p value of 0.003 (< 0.05), which demonstrates the effectiveness of accelerating VDSPs and using warm-start cuts.

When the community number reaches 40 (Table 4), NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++} can find much better feasible solutions than SPRP and SPRP-VI. In 18 out of 24 instances, NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++} obtain solutions that are better or not worse than SPRP and SPRP-VI. Our NLBBD methods can at least find feasible solutions within the given time limit, while SPRP and SPRP-VI fail to find feasible solutions in seven instances. The optimality gap differences of each NLBBD method are lower than that of SPRP and SPRP-VI, implying the scalability of NLBBD methods. The computational performances of NLBBD-C and NLBBD-B&C vary depending on the size of the instance. For small-size instances, NLBBD-B&C can solve problems to proven optimums with less computational time than NLBBD-C, but for large-size instances, it is difficult to say which one is better. This can be explained by the relative difficulty between the master problem and the subproblem in the LBBD. In fact, 99.68% of the time within the given limit is used for solving VDSPs in large-size instances, while at most 56.11% of the time is used for solving VDSPs in small-size instances. When the master problem is harder relative to the subproblem, the NLBBD-B&C performs better, since it avoids building the branch-and-bound tree from scratch. On the contrary, when the subproblem is harder, the NLBBD-C may outperform NLBBD-B&C. This observation is consistent with what has been reported by Beck (2010), Hooker (2019). The enhancements in NLBBD-B&C^{++} are significant (p value: 0.000) compared to NLBBD-B&C, as shown by the Gap(%) values in Table 4.

Table 3: Computational performance (medium size)

Instance	SPRP Obj.	SPRP Gap (%)	SPRP Time (s)	SPRP-VI Obj.	SPRP-VI Gap (%)	SPRP-VI Time (s)	NLBBD-C Obj.	NLBBD-C Gap' (%)	NLBBD-C Time (s)	NLBBD-B&C Obj.	NLBBD-B&C Gap (%)	NLBBD-B&C Time (s)	NLBBD-B&C^{++} Obj.	NLBBD-B&C^{++} Gap (%)	NLBBD-B&C^{++} Time (s)
03-20-02-0-0	***562114**	0.00	677	***562114**	0.00	837	619215	56.28	Lmt.	602509	40.69	Lmt.	**562114**	1.71	Lmt.
03-20-02-0-1	***1069579**	0.00	1399	***1069579**	0.00	1201	1227096	76.37	Lmt.	**1069579**	20.93	Lmt.	1092847	3.93	Lmt.
03-20-04-0-0	**536067**	2.24	Lmt.	**536067**	1.56	Lmt.	580062	53.02	Lmt.	580062	53.02	Lmt.	580062	19.99	Lmt.
03-20-04-0-1	**815002**	6.43	Lmt.	**815002**	5.93	Lmt.	918504	67.24	Lmt.	918504	66.57	Lmt.	815441	8.28	Lmt.
05-20-02-0-0	**550416**	0.27	Lmt.	***550416**	0.00	3066	638564	57.61	Lmt.	**550416**	7.91	Lmt.	**550416**	2.46	Lmt.
05-20-02-0-1	***752529**	0.00	2067	***752529**	0.00	3580	950069	71.51	Lmt.	***752529**	0.00	2160	***752529**	0.00	3250
05-20-04-0-0	527900	4.56	Lmt.	527900	4.74	Lmt.	618611	55.95	Lmt.	568236	33.67	Lmt.	**527866**	12.98	Lmt.
05-20-04-0-1	628637	6.62	Lmt.	**627272**	6.50	Lmt.	794493	65.70	Lmt.	680712	33.98	Lmt.	628169	9.66	Lmt.
07-20-02-0-0	**511001**	2.54	Lmt.	**511001**	4.61	Lmt.	570364	52.54	Lmt.	511267	22.81	Lmt.	518177	6.98	Lmt.
07-20-02-0-1	**749827**	2.78	Lmt.	750710	9.64	Lmt.	940164	71.21	Lmt.	750710	6.43	Lmt.	**749827**	4.70	Lmt.
07-20-04-0-0	7965224	94.08	Lmt.	6923233	93.29	Lmt.	574220	52.54	Lmt.	**552285**	35.45	Lmt.	560466	22.13	Lmt.
07-20-04-0-1	6346946	90.96	Lmt.	4711994	88.05	Lmt.	775914	64.88	Lmt.	644309	17.91	Lmt.	**627309**	10.74	Lmt.

[1] Same as the notes in Table 2.

Table 4: Computational performance (large size)

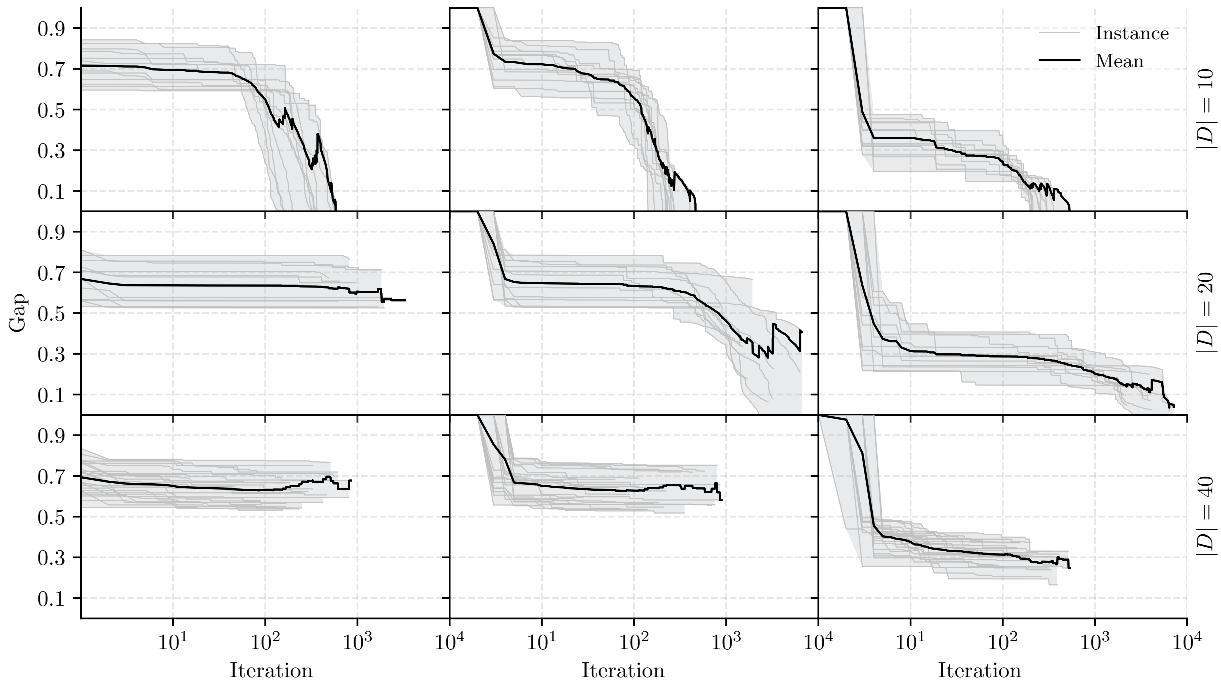
Instance	SPRP Obj.	SPRP Gap (%)	SPRP Time (s)	SPRP-VI Obj.	SPRP-VI Gap (%)	SPRP-VI Time (s)	NLBBD-C Obj.	NLBBD-C Gap' (%)	NLBBD-C Time (s)	NLBBD-B&C Obj.	NLBBD-B&C Gap (%)	NLBBD-B&C Time (s)	NLBBD-B&C^{++} Obj.	NLBBD-B&C^{++} Gap (%)	NLBBD-B&C^{++} Time (s)
03-40-02-0-0	-	-	Lmt.	**1063002**	4.39	Lmt.	1276309	57.00	Lmt.	1232760	55.48	Lmt.	1216365	23.88	Lmt.
03-40-02-0-1	-	-	Lmt.	**1957359**	14.24	Lmt.	2350890	76.65	Lmt.	2220382	75.28	Lmt.	2180927	16.50	Lmt.
03-40-02-1-0	-	-	Lmt.	**1163281**	6.64	Lmt.	1351051	59.38	Lmt.	1312682	58.19	Lmt.	1348985	20.52	Lmt.
03-40-02-1-1	1379661	8.75	Lmt.	**1352669**	7.96	Lmt.	1701777	67.75	Lmt.	1598536	65.67	Lmt.	1684479	24.76	Lmt.
03-40-04-0-0	-	-	Lmt.	-	-	Lmt.	**1200890**	54.00	Lmt.	1219640	54.71	Lmt.	1231935	30.44	Lmt.
03-40-04-0-1	-	-	Lmt.	-	-	Lmt.	**1726230**	68.00	Lmt.	1781101	68.99	Lmt.	1780360	25.03	Lmt.
03-40-04-1-0	-	-	Lmt.	-	-	Lmt.	**1207538**	54.25	Lmt.	1284932	57.01	Lmt.	1249246	31.06	Lmt.
03-40-04-1-1	-	-	Lmt.	-	-	Lmt.	1439285	61.62	Lmt.	1477616	62.62	Lmt.	**1406364**	32.29	Lmt.
05-40-02-0-0	1702939	44.10	Lmt.	1470387	35.32	Lmt.	1259115	56.41	Lmt.	1242644	55.84	Lmt.	**1240835**	30.18	Lmt.
05-40-02-0-1	**1507008**	9.54	Lmt.	-	-	Lmt.	2199013	75.04	Lmt.	1995469	72.40	Lmt.	2056046	32.84	Lmt.
05-40-02-1-0	-	-	Lmt.	-	-	Lmt.	1395550	60.67	Lmt.	**1351805**	59.60	Lmt.	1404687	26.27	Lmt.
05-40-02-1-1	**1294937**	10.24	Lmt.	2572477	55.11	Lmt.	1837247	70.13	Lmt.	1992031	72.45	Lmt.	1948341	37.49	Lmt.
05-40-04-0-0	8462942	89.79	Lmt.	16341162	100.00	Lmt.	1225888	54.94	Lmt.	**1180438**	53.36	Lmt.	1251736	37.16	Lmt.
05-40-04-0-1	-	-	Lmt.	14770808	100.00	Lmt.	1655494	66.63	Lmt.	1725239	68.13	Lmt.	**1586201**	34.80	Lmt.
05-40-04-1-0	8354445	90.22	Lmt.	13451044	94.28	Lmt.	1226059	54.95	Lmt.	**1180927**	53.35	Lmt.	1277156	37.48	Lmt.
05-40-04-1-1	7928030	89.15	Lmt.	-	-	Lmt.	1462007	62.22	Lmt.	1414978	61.07	Lmt.	**1311857**	31.93	Lmt.
07-40-02-0-0	-	-	Lmt.	10783463	92.21	Lmt.	1183105	53.61	Lmt.	**1138574**	51.80	Lmt.	1180025	31.64	Lmt.
07-40-02-0-1	-	-	Lmt.	-	-	Lmt.	1906128	71.21	Lmt.	1951166	71.95	Lmt.	**1892811**	33.05	Lmt.
07-40-02-1-0	-	-	Lmt.	10333030	90.14	Lmt.	1509063	63.63	Lmt.	1535921	64.39	Lmt.	**1459329**	24.80	Lmt.
07-40-02-1-1	-	-	Lmt.	11406871	89.90	Lmt.	1955377	71.93	Lmt.	1968959	71.90	Lmt.	**1887312**	30.45	Lmt.
07-40-04-0-0	-	-	Lmt.	13652625	100.00	Lmt.	1179702	53.17	Lmt.	1188070	53.63	Lmt.	**1172062**	35.75	Lmt.
07-40-04-0-1	8652442	100.00	Lmt.	14809253	100.00	Lmt.	1654175	66.61	Lmt.	**1545285**	64.35	Lmt.	1614093	38.83	Lmt.
07-40-04-1-0	-	-	Lmt.	-	-	Lmt.	1309111	57.80	Lmt.	1251302	56.01	Lmt.	**1241047**	31.90	Lmt.
07-40-04-1-1	-	-	Lmt.	12921254	100.00	Lmt.	**1433350**	61.46	Lmt.	1545427	64.38	Lmt.	1480137	34.45	Lmt.

[1] Same as the notes in Table 2.

In summary, our NLBBD methods (with enhancements) obtain optimal solutions for easy instances and good feasible solutions for hard instances. Additionally, compared with commercial solver, our method uses less RAM, since the original huge (e.g., 265452 rows, 274355 columns and, 1190581 nonzeros for instance 07-40-04-0-0) model is decomposed into smaller models that can be solved independently.

5.3. Convergence of NLBBD implementations

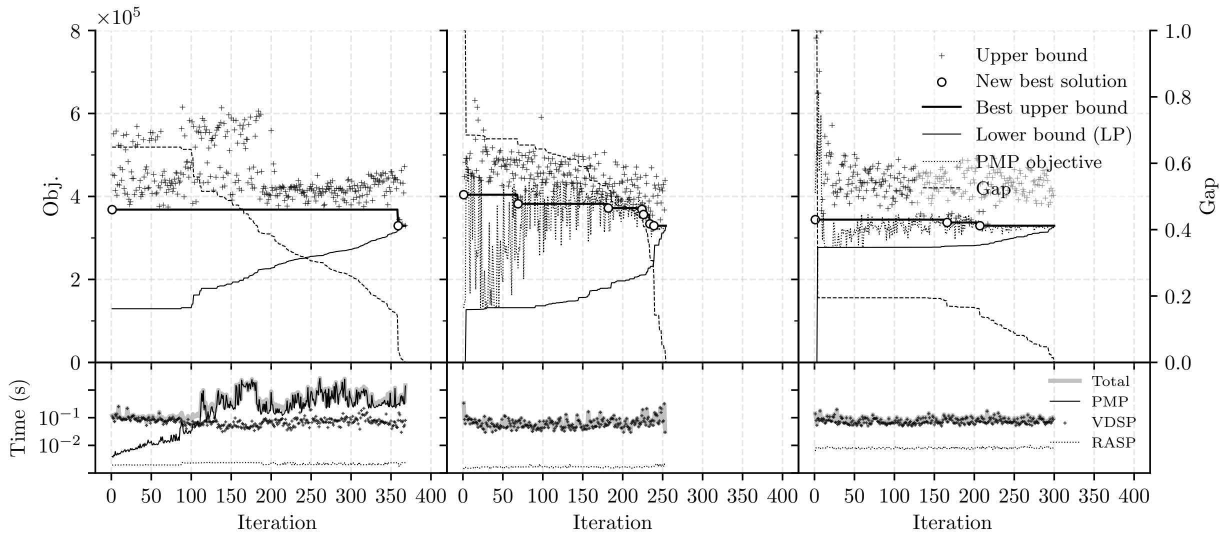
This subsection illustrates the convergence behavior of NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++}. Figure 4 shows how the optimality gap decreases with iterations for instances of varying sizes. As the affected community number is the most crucial factor that affects the algorithm performance, instances are classified into three groups by the number of affected communities. For small-size instances ($|D| = 10$), NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++} can reach optimum within 10^2 to 10^3 iterations. Though the number of iterations is similar, the NLBBD-C can require one order of magnitude longer runtime than NLBBD-B&C and NLBBD-B&C^{++}, as shown in Table 2. This is because classical NLBBD-C typically requires more time to finish one step of iteration than NLBBD-B&C and NLBBD-B&C^{++}. During iterations, NLBBD-C builds a branch-and-bound tree each time a new set of cuts is generated, while NLBBD-B&C and NLBBD-B&C^{++} only require a single branch-and-bound tree. This behavior will be detailed later with Figure 5. Since NLBBD-B&C requires less time to finish one step of the iteration, on medium-size instances ($|D| = 20$), with a given 3,600 seconds time limit, NLBBD-B&C and NLBBD-B&C^{++} can finish more iterations than NLBBD-C. Thus, lower optimality gaps are obtained by NLBBD-B&C and NLBBD-B&C^{++}. For large-size instances ($|D| = 40$), NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++} all finish less than 10^3 iterations. This is because solving VDSPs for large-size instances is more difficult than for small- and medium-size instances, and it takes up the majority



(a) NLBBD-C (b) NLBBD-B&C (c) NLBBD-B&C^{++}

Figure 4: Gap curve

of computation time. Having similar iteration numbers does not necessarily mean that the convergence performance is the same. Through the enhancements discussed in Section 4.3, NLBBD-B&C^{++} achieves significantly lower optimality gaps than NLBBD-C and NLBBD-B&C for medium- and large-size instances. Notably, NLBBD-B&C^{++} can attain low optimality gaps (around 30% to 50%) with fewer iterations (less than ten). In contrast, NLBBD-B&C requires over 10^3 iterations to reach the same gap on medium-size instances, while it never reaches this gap on large-size instances.



(a) NLBBD-C (b) NLBBD-B&C (c) NLBBD-B&C^{++}

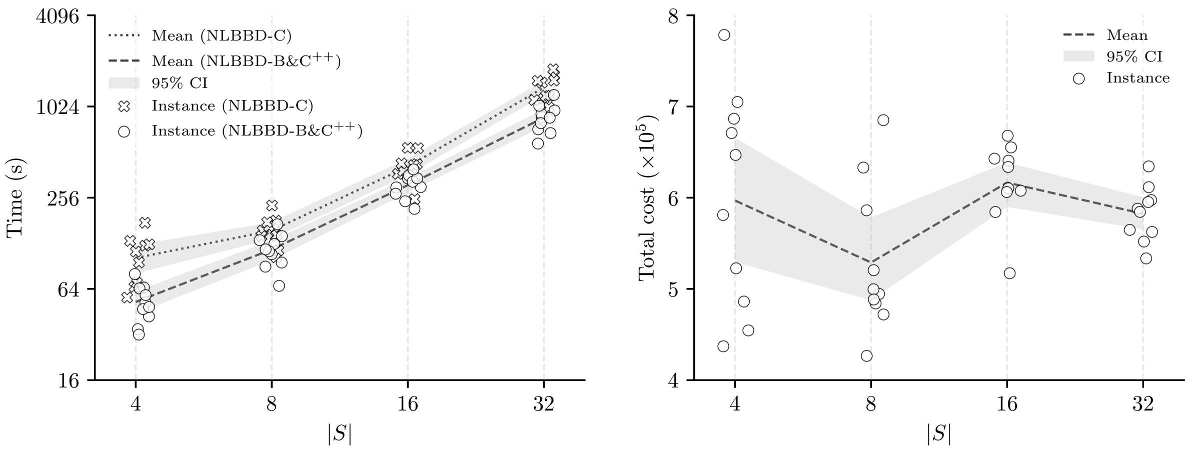
Figure 5: Convergence curve

Figure 5 depicts bounds during iterations for the instance 03-10-02-0-0, which is typical. In Figure 5, the PMP objective is the objective of incomplete PMP; the upper bound is the integer feasible objective value of the current iteration, it is calculated by (master objective - master recourse + sub recourse); the best upper bound is the objective value associated with the currently observed best integer feasible solution, new best solutions are highlighted on the lines with circles; the lower bound (LP) is the objective value of linear programming relaxation for the incomplete PMP model, for NLBBD-C the lower bound (LP) is the PMP objective obtained every time before a Benders cut is added; the gap is the optimality gap calculated by (UB - LB) / UB. Figure 5 shows that NLBBD-B&C and NLBBD-B&C^{++} needs less iteration to converge than NLBBD-C. It is important to note that this result may not be generalizable to all instances since the number of iterations for these methods depends on the search path on the branch-and-bound tree, which is difficult to predict. Nevertheless, both NLBBD-B&C and NLBBD-B&C^{++} exhibit superior computational time compared to NLBBD-C. At the bottom of Figure 5, the times for solving PMP, RASP, and VDSP are presented. NLBBD-C, NLBBD-B&C, and NLBBD-B&C^{++} require similar time to solve RASP and VDSP, but NLBBD-C needs more time to solve PMP after 100 iterations when a large number of Benders cuts have been added. This is because classical NLBBD-C builds a new branch-and-bound tree each time a new Benders cut is added and solves PMP from scratch. In contrast, NLBBD-B&C and NLBBD-B&C^{++} search on a single branch-and-bound tree without performing this operation. It should be noted that in instances with harder VDSP problems, where computational time for solving PMP becomes less significant, NLBBD-C may outperform both NLBBDC-B&C and NLBLD-B&C^{++}. Another observation from Figure 5 is that both NLBBD-C and NLBBD-B&C have a low and flat lower bound at the beginning of the iteration, which actually represents the first-stage prepositioning cost. Initially, both methods need to try numerous combinations of assignments (x_{ij}) before the recourse estimators become nonzero. In contrast, by incorporating warm-start cuts, NLBBD-B&C^{++} bounds recourse estimators and improves overall lower bounds during initial iterations. This highlights the significance of enhancements discussed in Section 4.3.

5.4. Scalability on scenario size

In the previous subsections, we conduct experiments with $|S|$ (scenario number) being two or four, under the assumption that the scenarios are aggregated from historical events or simulations for potential future disasters. In practice, one may expect to solve instances with larger $|S|$, with the scenarios being generated directly from simulation. Therefore, it is worthwhile to analyze how larger numbers of scenarios impact the performance of the proposed method: the efficiency of the solution method and the quality of the estimated cost. It is reasonable to expect that the efficiency of the solution method will decrease as $|S|$ increases, but it is not clear how the efficiency decreases with respect to $|S|$. In terms of cost estimating, which we consider one of the major goals of this study, we expect a relatively small $|S|$ can yield an accurate estimation. Solving an instance with a certain number of scenarios provides a point estimation of the overall cost. When reduplicating experiments on an instance with different scenarios, we can obtain an interval estimation of the overall cost. It is expected that the objective value of each single run should not be far away from other runs, in other words, the estimated confidence interval (CI) of the objective should be narrow. This is known as *in-sample stability* test for stochastic programming. We experiment with $|S| \in \{4, 8, 16, 32\}$. For each $|S|$, we reduplicate ten runs with different scenarios on instance 03-10-04-0-1.

The objectives of each set of experiments with a certain number of $|S|$ can be used to build a *statistical lower bound* for the true stochastic problem. It is expected that the CI is narrower with a larger $|S|$, i.e., better in-sample stability performance. In contrast with the in-sample stability analysis, there is another method namely *out-of-sample analysis*, i.e., given a first-stage solution, how good is it when it is used on instances that are not used for optimizing? After the aforementioned reduplication, we can find a "good" solution and solve multiple out-of-sample instances with the first-stage decisions being fixed to the good solution, to build a *statistical upper bound* (Mak et al., 1999). The *statistical gap*, which is measured by the distance between the stochastic lower bound and the stochastic upper bound, reflects the quality of the solution (the good solution) obtained. A larger number of scenarios usually means a lower stochastic gap. Thus, stochastic gap can be used to determine an appropriate number of scenarios for algorithms with external sampling, by trading off scenario number and runtime length. However, this approach cannot be applied to our problem, since it requires the *relatively complete recourse* assumption (the second-stage problem is always feasible whatever the first-stage decisions and revelation of stochastic parameters are). Therefore, in the following, we only analyze how $|S|$ impacts the computational time and the in-sample stability.



(a) Computational time

(b) Objective value

Figure 6: Impacts of number of scenarios

In Figure 6(a), the relationship between $|S|$ and the computational time is presented in a log-log plot. Here we only present NLBBD-C and NLBBD-B&C^{++}, since NLBBD-B&C will show a similar trend to NLBBD-B&C^{++}. The lines that represent the mean computational time of each $|S|$ show linear trends, suggesting the proposed algorithm has a polynomial time complexity with respect to $|S|$. This is because solving VDSP is the most time-consuming part of the algorithm (as Figure 5 shows), and the number of VDSPs in each iteration is $|T||S|$. A certain proportion of increase on $|S|$ can result in a proportional increase in the number of VDSPs, leading to a proportional increase in the overall computational time. From Figure 6(b), the main finding is that as the size of $|S|$ increases, the CIs become narrower, which is consistent with our expectations. With $|S|$ being 32, a single run can yield a quite good point estimation for the total cost. Figure 6(b) suggests that the estimated cost is sensitive to the number of scenarios. However, interestingly, the difference in mean values of each $|S|$ is small, especially compared with the differences among individual runs with $|S| \in \{4, 8\}$. This suggests that for difficult problems, we can obtain a good estimate by running instances with a small number of scenarios multiple times instead of using a larger number of scenarios. In summary, the proposed algorithm scales well with the size of $|S|$, and even a small $|S|$ is able to provide an accurate estimate of the total cost.

6. Results analysis

To validate the value of horizontal coordination and integrating routing decisions, numerical experiments are conducted in Section 6.1 and Section 6.2, respectively. The trade-off between service level and the total cost is analyzed in Section 6.3.

6.1. Effect of horizontal coordination

We test the sensitivity of benefit from horizontal coordination on some key parameters of the mathematical model, including relief item procurement cost, facility opening cost, unit routing cost, and facility capacity. The results suggest that the facility capacity is the most crucial one. Further experiments show benefits are obtained in both the planning and response phases. In the planning phase, by applying horizontal coordination, feasible plans can be generated for instances that are infeasible without horizontal coordination. In the response phase, horizontal coordination results in lower routing costs. For computational tractability, we only conducted experiments on the small-size instances with ten affected communities, in this subsection and the following subsections.

Firstly, the benefit of horizontal coordination is analyzed in Table 5. Experiments that allow horizontal coordination and disallow horizontal coordination, namely direct transportation, are conducted respectively. We solve direct transshipment models that are modified from the original horizontal coordination ones by fixing the horizontal coordination decision variables $t^s_{ii'j}$ $\forall i \in T, i' \in T, j \in D, s \in S$ to zero. We observe that in our baseline instances, these two families of models yield the same objective value for each instance and they both do not apply horizontal coordination. Our baseline instance is based on the assumption that all the resources, i.e., facility capacity, relief item, and transportation ability, are sufficient. Under this setting, deploying adequate relief item at a relief facility such that it can meet all the demand of affected communities that are assigned to it without accepting relief items from other facilities seem to be a wise decision. But in practice, it is usually not the case that all the relief resources are adequate. For example, for relief items that require specific equipment to store, there can be not enough space to place them. Thus, facility capacity is limited for this type of item. Or in urban areas where the warehouse space is limited, there can be not enough space for relief item storage.

To verify how the available facility capacity influences the overall preparedness planning performance, alternative instances with reduced facility capacity are adapted from the baseline instances with ten affected communities. The capacity of the relief facility is modified according to the maximum total demand of affected communities among all the scenarios. Recall that $m^s_j, \forall j \in D, \forall s \in S$ be the demand of affected community j under scenario s, then for an instance, the maximum total demand among all the scenario is $\max_{s\in S}(\sum_{j\in D} m^s_j)$. Given the number of relief facility candidates $|T|$, we compute the size of the larger facility with $Q^{1'} = \lceil \max_{s\in S}(\sum_{j\in D} m^s_j)/|T| \rceil$ and compute the size of the smaller facility with $Q^{2'} = \lceil 0.25 \times Q^{1'} \rceil$, such that capacity proportion of larger facilities and smaller ones consists with prior experiments. In Table 5, columns SPRP report horizontal coordination model objective value, columns SPRP$_d$ report direct transportation objective value, and DIF(%) is the relative difference calculated by $(SPRP_d - SPRP)/SPRP$. The DIF(%) is always nonnegative since the horizontal coordination model has a large decision space, which contains the direct transportation model decision space. Experiments with $(Q^{1'}, Q^{2'}) := (\lceil 1.2 \times Q^{1'} \rceil, \lceil 1.2 \times Q^{2'} \rceil)$ and $(Q^{1'}, Q^{2'}) := (\lceil 2 \times Q^{1'} \rceil, \lceil 2 \times Q^{2'} \rceil)$are also conducted.

From Table 5, we have the following observations. Firstly, when the capacity of the available facility is restricted, i.e., the total capacity of the relief facility nearly exactly meets the total demand, there is no feasible plan for the model without horizontal coordination. This result also implies that in the context of relief planning with stochastic demand, only considering the worst scenario is insufficient, and, worse still, may result in an infeasible solution. In other words, though the total facility capacity is enough to meet all the demand even in the worst scenario, the stochasticity makes it impossible to allocate relief items to relief facilities properly when horizontal coordination is not allowed. Secondly, when the facility capacity increases slightly (20%) more than the total demand of the worst scenario, the horizontal coordination model can get a solution with a lower cost (7.96% on average) compared with the direct transportation model. Finally, when the facility capacity is about twice the total demand, there is no significant difference between the horizontal coordination model and the direct transportation model. In summary, all the above-mentioned observations validate the flexibility of our horizontal coordination method in the context of humanitarian relief. The horizontal coordination can be particularly meaningful when the warehousing resources are limited.

Table 5: Impacts of horizontal coordination

Instance	$(Q^{1'}, Q^{2'})$			$(\lceil 1.2 \times Q^{1'} \rceil, \lceil 1.2 \times Q^{2'} \rceil)$			$(\lceil 2 \times Q^{1'} \rceil, \lceil 2 \times Q^{2'} \rceil)$		
	SPRP	SPRP$_d$	DIF (%)	SPRP	SPRP$_d$	DIF (%)	SPRP	SPRP$_d$	DIF (%)
03-10-02-0-0	378921.98	-	-	361617.48	386897.36	6.99	342487.10	342487.10	0.00
03-10-02-0-1	628819.36	-	-	612088.04	643884.87	5.19	592957.66	592957.66	0.00
03-10-04-0-0	331927.43	-	-	318454.46	343891.26	7.99	306639.27	306639.27	0.00
03-10-04-0-1	456876.12	-	-	443689.74	477420.46	7.60	431874.55	431874.55	0.00
05-10-02-0-0	400343.58	-	-	372639.77	401118.88	7.64	347758.87	347758.87	0.00
05-10-02-0-1	535672.79	-	-	506873.90	540090.46	6.55	473606.94	473606.94	0.00
05-10-04-0-0	345230.00	-	-	330094.39	370202.60	12.15	305507.85	307753.26	0.73
05-10-04-0-1	417442.28	-	-	397566.96	435565.45	9.56	367889.71	373524.74	1.53
07-10-02-0-0	349541.00	-	-	335939.36	-	-	300353.21	306996.22	2.21
07-10-02-0-1	462787.06	-	-	447647.61	-	-	414219.23	417958.78	0.90
07-10-04-0-0	339946.35	-	-	326989.50	-	-	296254.95	306495.51	3.46
07-10-04-0-1	396569.38	-	-	382418.94	-	-	351684.39	362640.52	3.12
Mean	420339.78	-	-	403001.68	449883.92	7.96	377602.81	380891.12	1.00

[1] "-" represents SPRP$_d$ is infeasible.

There is a trend that the horizontal coordination model outperforms the direct transportation model when the candidate facility number increases. With restricted facility capacity ($(Q^{1'}, Q^{2'})$), none of the direct transportation models yield feasible solutions. When the facility capacity is increased to $(\lceil 1.2 \times Q^{1'} \rceil, \lceil 1.2 \times Q^{2'} \rceil)$, though 3/4 direct transportation instances are feasible, the instances with seven facilities remain infeasible. When the total facility capacity is increased to two times the total demand of the worst scenario, see columns $(\lceil 2 \times Q^{1'} \rceil, \lceil 2 \times Q^{2'} \rceil)$, both types of models are feasible. However, the direct transportation model needs a higher cost for all the instances with seven facilities and two instances with five facilities (1% increase on average). Since all the instances in Table 5 have the same affected community locations, this observation implies that compared with building more facilities, it is more flexible to maintain a relatively smaller group of facilities and use horizontal coordination to rebalance stock in the response phase.

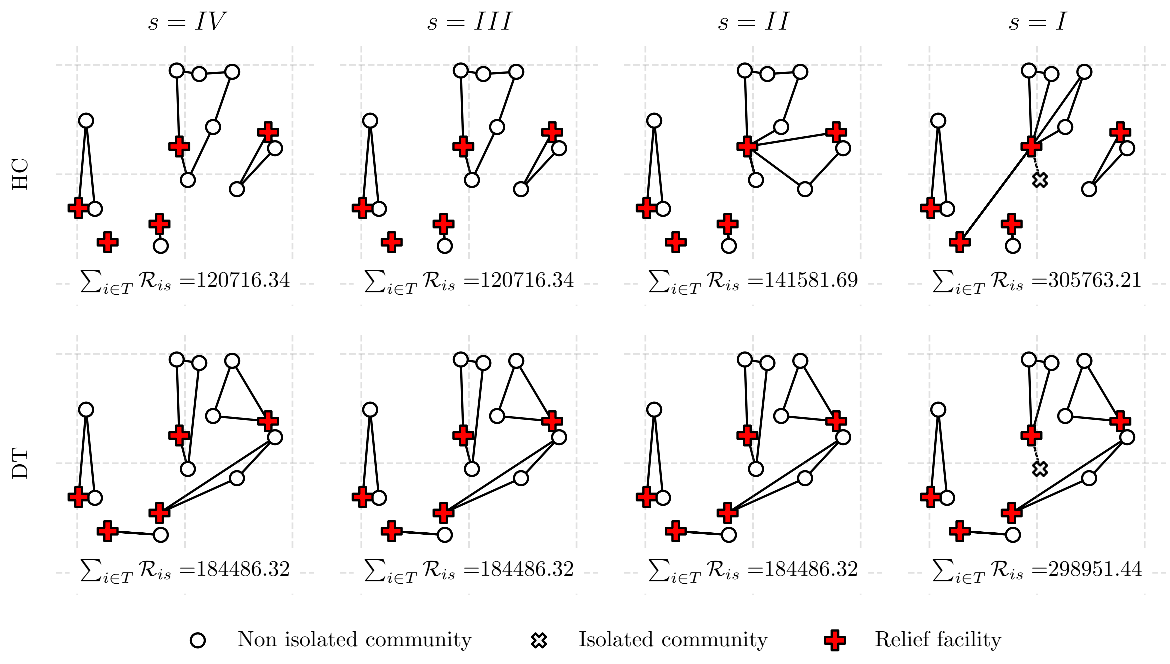


Figure 7: Illustration for horizontal coordination

It can be observed from Table 5, the instance 05-10-04-0-0 ($\lceil 1.2 \times Q^{1'} \rceil = 294, \lceil 1.2 \times Q^{2'} \rceil = 75$) has the most significant difference (12.15% cost saving) between horizontal coordination model and direct transportation model, so the solution for this instance is illustrated in Figure 7 to identify the reason for the cost-saving. In Figure 7, columns s represent scenarios, among which $s = I$ has the highest demand; and the rows "HC" and "DT" represent the horizontal coordination model and direct transportation model respectively. The pre-disaster planning decisions (l_i) for the horizontal coordination model and the direct transportation model are very similar except for the assignment decisions x^*_{ij}. Both the horizontal coordination model and direct transportation model open five relief facilities with the initial inventory being $l_{i(HC)} = \{294, 227, 244, 268, 196\}$ (totally 1229) and $l_{i(DT)} = \{266, 260, 227, 272, 196\}$ (totally 1221). Consequently, it is safe to say, that the only reason that can explain cost-saving by horizontal coordination is the assignment decisions x^*_{ij} and the post-disaster routing decisions. From Figure 7, we have the following observations. Firstly, in the horizontal coordination model, every affected community is served by a relief facility that is close enough, while in the direct transportation model, there are two affected communities that are located in the southeast and one affected community that are located in the south serviced by relief facilities that are far from them in all the scenarios. Secondly, the horizontal coordination model yields results with lower routing cost $\sum_{i\in T}\mathcal{R}_{is}$ in $s=\{IV,III,II\}$ compared with the direct transportation model. Thirdly, in row HC column $s=\{II,I\}$, horizontal coordination is applied to replenish inventory from one relief facility to another. Lastly, in the horizontal coordination model, there is one relief facility, which is located in the southwest, that has never dispatched a vehicle to a community under all the scenarios, but dispatches a vehicle to another facility in $s=I$. These observations validate the flexibility of horizontal coordination. If horizontal coordination is not involved and facility capacity is limited, the model tends to allocate facility capacity to affected communities firstly, and determine vehicle routes secondly, even if the resulting routing cost is high, as shown in row DT columns $s=\{IV,III,II\}$. Otherwise, if each affected community is assigned to its closest relief facility, it is even infeasible, because this facility may not have enough capacity. On the contrary, in the horizontal coordination model, it is pretty sure that an affected community can accept relief items from a relief facility that is close enough. When the inventory turns out to be in shortage for a facility, as shown in row HC columns $s=\{II,I\}$, replenishment is accepted from other facilities. Since inventory shortage is only incurred in extreme scenarios, e.g., $s=\{II,I\}$, consequential horizontal coordination cost is only incurred in these cases, but assignment decisions by the direct transportation model will influence routing cost over all four scenarios, so horizontal coordination is superior to direct transportation.

In summary, the value of horizontal coordination lies in two aspects. Firstly, with limited facility capacity, using only direct transportation may result in infeasibility, while horizontal coordination is always feasible as well as the total facility capacity is sufficient. Secondly, horizontal coordination can obtain benefits from not dispatching vehicles to affected communities that are far from a relief facility. As a matter of fact, the direct transportation method is a special case of the horizontal coordination method, which lets all the horizontal coordination variables be set to zero. Thus, the horizontal coordination method always yields solutions not worse than the direct transportation method.

6.2. Impact of integrating vehicle routing problems

As has been mentioned, existing literature about disaster preparedness always models the second-stage response decision as a transportation problem rather than a vehicle routing problem, though this simplification requires an estimation of the unit transportation cost. In the case of full-truck-load long-haul logistics, the unit transportation cost can be estimated by the back-and-forth routing cost, but this is not the case for the last-mile distribution of relief items. To compare the performance of our prepositioning model with the second-stage vehicle routing decision and second-stage transportation decision, we first estimate the expected unit transportation cost by the solution of selected instances. Consider a single scenario deterministic instance with one relief facility (i_0) and two affected communities (j_1 and j_2), in the optimal solution, the relief facility i_0 dispatches one vehicle that travels to point j_1 then point j_2 and finally returns to i_0. We calculate the expected unit transportation cost with total vehicle routing cost divided by the sum of the distance between i_0 and j_1 plus the distance between i_0 and j_2. More generally, the average unit transportation cost (per unit relief item per unit distance), which is denoted as aut, is estimated as

$$aut=\left\{\sum_{ins\in\mathbb{I}}\frac{\sum_{s\in S_{ins}}\sum_{i\in T}\left[\frac{\sum_{h\in H}\sum_{(p,q)\in A}c^h_{pq}\cdot z^{*(s,i,h)}_{pq}(ins)}{\sum_{j\in D}x^*_{ij}(ins)\cdot\lceil r\cdot m^s_j\rceil\cdot d_{ij}}\right]}{|S_{ins}|\cdot\sum_{i\in T}y^*_i(ins)}\right\}\Big/|\mathbb{I}|\approx 6.5874 \tag{13}$$

where $\mathbb{I}$ is the set of small-size instances, S_{ins} is the scenario set of a particular instance $ins\in\mathbb{I}$, d_{ij} is the distance from facility i to affected community j, $x^*_{ij}(ins)$, $y^*_i(ins)$, and $z^{*(s,i,h)}_{pq}(ins)$ denote the optimal assignment, location, and routing decisions for $ins\in\mathbb{I}$. The term in the square brackets computes the average transportation cost of a facility i in a certain instance ins and scenario s. Then the estimated unit transportation cost aut is obtained by average terms in the square brackets for all the opened facilities in all the $ins\in I$ and $s\in S_{ins}$. For our small-size instances that all reach their optimum, $aut\approx 6.5874$. Then a transportation counterpart for the SPRP can be developed as

$$[\text{SPRP}_t]\min_{(1.2)\text{–}(2.21)}\sum_{i\in T}\sum_{k\in K}c^k y_{ik}+\sum_{i\in T}c\ l_i+\sum_{i\in T}\sum_{h\in H}c^h_1 w^h_i+\sum_{s\in S}\mathsf{P}_s\sum_{i\in T}\sum_{i'\in T}\sum_{j\in D}aut\left[t^s_{ij}\cdot d_{ij}+t^s_{ii'j}\cdot(d_{ii'}+d_{i'j})\right]. \tag{14}$$

In (14), d_{ij} denotes the distance between pair of points $(i,j)\in A$. The first three terms remain the same as that of (1.1), and the last term replaces the routing cost of (1.1) with transportation cost. Finally, when SPRP_t is solved to optimum, we solve SPRP with the fixed first-stage decisions yielded by SPRP_t. The SPRP presented in Section 3 takes location (y_{ik}), inventory (l_i), assignment (x_{ij}), and vehicle (w^h_i) decisions as first-stage decisions. When we fix the first-stage solutions yielded by SPRP_ts to SPRPs and solve the SPRPs, the SPRPs turn out to be infeasible, due to SPRP_ts preparing no vehicle at facilities. This is because among constraints (1.2)–(2.21) for (14), only (1.6) is related to the vehicle decisions, and it is an upper bound of the vehicle number. It is not surprising that (14) yields solutions with $w^h_i=0,\forall i\in T,h\in H$. We also tried to add valid inequalities (3.5) and (3.6), which are lower bounds of the vehicle number, to (14). However, the first-stage solutions yielded by (14) remain infeasible for SPRPs, demonstrating the model with transportation subproblems is not able to capture the required number and location of vehicles. In the following experiments, we use only location (y_{ik}), inventory (l_i), and assignment (x_{ij}) decisions yielded by SPRP$_t$s as first-stage decisions for SPRPs. Other decisions are left to SPRPs. The instances used are ones with $|D| \in \{10, 20\}$, which are optimized or close to the optimum by SPRP.

Table 6: Impacts of integrating vehicle routing problems

Instance	Solution by SPRP				Solution by (14)				Differences	
	Opened facilities	Total capacity	Total inventory	Inventory rate (%)	Opened facilities	Total capacity	Total inventory	Inventory rate (%)	SPRP$_t$ (%)	SPRP$_r$ (%)
$\lvert D\rvert = 10$										
03-10-02-0-0	3 (0, 3)	3000	1244	41.47	3 (0, 3)	3000	1244	41.47	397.19	0.00
03-10-02-0-1	3 (0, 3)	3000	1244	41.47	3 (0, 3)	3000	1244	41.47	182.72	0.00
03-10-04-0-0	3 (0, 3)	3000	1221	40.70	3 (0, 3)	3000	1221	40.70	406.11	0.00
03-10-04-0-1	3 (0, 3)	3000	1221	40.70	3 (0, 3)	3000	1221	40.70	255.57	0.00
05-10-02-0-0	4 (0, 4)	4000	1244	31.10	4 (0, 4)	4000	1244	31.10	261.26	4.36
05-10-02-0-1	4 (0, 4)	4000	1244	31.10	4 (0, 4)	4000	1244	31.10	158.36	6.93
05-10-04-0-0	4 (0, 4)	4000	1249	31.23	4 (0, 4)	4000	1221	30.53	279.23	4.65
05-10-04-0-1	4 (0, 4)	4000	1249	31.23	4 (0, 4)	4000	1221	30.53	211.73	6.59
07-10-02-0-0	5 (0, 5)	5000	1244	24.88	6 (0, 6)	6000	1244	20.73	413.46	7.78
07-10-02-0-1	5 (0, 5)	5000	1244	24.88	6 (0, 6)	6000	1244	20.73	262.85	5.87
07-10-04-0-0	5 (0, 5)	5000	1221	24.42	6 (0, 6)	6000	1221	20.35	370.24	6.43
07-10-04-0-1	5 (0, 5)	5000	1221	24.42	6 (0, 6)	6000	1221	20.35	288.99	5.54
$\lvert D\rvert = 20$										
03-20-02-0-0	2 (1, 1)	5000	2632	52.64	3 (1, 2)	6000	2632	43.87	529.41	0.01
03-20-02-0-1	2 (1, 1)	5000	2632	52.64	3 (1, 2)	6000	2632	43.87	230.78	0.00
03-20-04-0-0	3 (1, 2)	6000	2650	44.17	3 (1, 2)	6000	2650	44.17	550.77	0.00
03-20-04-0-1	3 (1, 2)	6000	2650	44.17	3 (1, 2)	6000	2650	44.17	328.04	0.00
05-20-02-0-0	5 (1, 4)	8000	2632	32.90	5 (0, 5)	5000	2632	52.64	386.77	9.92
05-20-02-0-1	5 (1, 4)	8000	2632	32.90	5 (0, 5)	5000	2632	52.64	256.03	13.15
05-20-04-0-0	5 (0, 5)	5000	2675	53.50	5 (0, 5)	5000	2650	53.00	428.80	7.70
05-20-04-0-1	5 (0, 5)	5000	2675	53.50	5 (0, 5)	5000	2650	53.00	345.00	10.54
07-20-02-0-0	6 (1, 5)	9000	2632	29.24	7 (1, 6)	10000	2632	26.32	513.12	8.01
07-20-02-0-1	7 (0, 7)	7000	2632	37.60	7 (1, 6)	10000	2632	26.32	317.84	4.33
07-20-04-0-0	4 (1, 3)	7000	2650	37.86	7 (0, 7)	7000	2650	37.86	442.55	-5.65
07-20-04-0-1	7 (1, 6)	10000	2650	26.50	7 (0, 7)	7000	2650	37.86	377.66	1.04
Mean	4.25	5375	1941	36.88	4.67	5417	1937	36.90	341.44	4.05

[1] In the columns of "opened facilities", the first number in brackets is the number of large facilities (capacity is 4000), while the second number is the number of small facilities (capacity is 1000). Inventory rate (%) = total inventory / total capacity ×100.

As mentioned earlier, the proposed SPRP has two goals: estimating required costs and providing efficient plans. In the following, we conduct experiments to analyze the estimates and solutions quality of (14), compared with the original SPRP. For each instance, let $Obj.$ denote the objective value of the best-known solution (among the five solution methods) in Table 2 and Table 3 by the original SPRP, Obj_t denotes the objective value of the modified SPRP$_t$ with second-stage transportation problem instead of routing problem. Then we solve the original SPRP with the first-stage decisions fixed to that yielded by the modified SPRP$_t$ and obtain the objective value Obj_r. Two measures "SPRP$_t$(%)" and "SPRP$_r$(%)" are calculated by $(Obj_t - Obj.)/Obj.$ and $(Obj_r - Obj.)/Obj.$, respectively. "SPRP$_t$(%)" reflects the estimation accuracy of the objectives by SPRP$_t$; "SPRP$_r$(%)" reflects the actual performance of the solutions by SPRP$_t$. We compare the solutions obtained by SPRP and SPRP$_t$ in Table 6 followed by a sensitivity analysis on aut in Table 7.

Table 7: Sensitivity to the average unit transportation cost

Instance	$aut \times 0$		$aut \times 25\%$		$aut \times 50\%$		$aut = 6.5874$ (Baseline)		$aut \times 200\%$	
	SPRP$_t$ (%)	SPRP$_r$ (%)	SPRP$_t$ (%)	SPRP$_r$ (%)	SPRP$_t$ (%)	SPRP$_r$ (%)	SPRP$_t$ (%)	SPRP$_r$ (%)	SPRP$_t$ (%)	SPRP$_r$ (%)
03-10-02-0-0	-60.72	-	54.33	0.00	168.61	0.00	397.19	0.00	854.34	0.00
03-10-02-0-1	-77.66	-	-12.24	0.00	52.74	0.00	182.72	0.00	442.67	0.00
03-10-04-0-0	-56.93	22.04	59.46	0.00	175.01	0.00	406.11	0.00	868.30	0.00
03-10-04-0-1	-69.74	30.96	12.03	0.00	93.21	0.00	255.57	0.00	580.29	0.00
05-10-02-0-0	-59.17	17.30	22.12	4.36	101.83	4.36	261.26	4.36	580.11	4.36
05-10-02-0-1	-70.80	58.61	-12.67	6.93	44.34	6.93	158.36	6.93	386.39	6.93
05-10-04-0-0	-55.56	54.84	29.45	4.65	112.71	4.65	279.23	4.65	612.27	4.65
05-10-04-0-1	-63.47	77.76	6.41	6.59	74.85	6.59	211.73	6.59	485.49	6.59
07-10-02-0-0	-51.55	40.43	67.51	7.78	182.83	7.78	413.46	7.78	874.73	7.78
07-10-02-0-1	-65.76	74.43	18.38	5.87	99.87	5.87	262.85	5.87	588.82	5.87
07-10-04-0-0	-52.10	8.53	56.31	6.43	160.95	6.43	370.24	6.43	788.82	6.43
07-10-04-0-1	-60.38	37.09	29.30	5.54	115.86	5.54	288.99	5.54	635.24	5.54
Mean	-61.99	42.20	27.53	4.01	115.24	4.01	290.64	4.01	641.46	4.01

[1] "-" represents SPRP is infeasible with fixed first-stage decisions.

Table 6 presents the first-stage solutions (number of opened facilities, total capacity of opened facilities, initial inventory, and inventory rate) yielded by the original SPRPs (with routing subproblems) and the modified SPRP$_t$s (14), with transportation subproblems). The two measures "SPRP$_t$(%)" and "SPRP$_r$(%)" are also presented. On average, the modified SPRP$_t$s estimate a cost three times higher than the true cost, that estimated by the original SPRPs. When one implements the first-stage solutions by SPRP$_t$s, a 4.05% higher cost can be incurred, compared to the original SPRPs. The SPRP$_t$ not only leads to higher overall cost, but also leads to a waste on the first-stage (before-disaster) investment: larger facility number and higher total capacity. This is critical since in the humanitarian context, which relies on donations, the donors are usually more sensitive to the before-disaster investment than the post-disaster investment. Among instances with $|D| = 10$, the first-stage solutions by SPRP$_t$ and the original SPRP are similar in terms of total inventory. However, SPRP$_t$ can lead to more opened facilities and lower inventory rates (instances 07-10-*-*-*), i.e., waste on before-disaster investment. On instances 05-10-02-*-*, though the initial investments are the same by SPRP$_t$ and the original SPRP, the former leads to higher cost, demonstrating the necessity of incorporating routing decisions on cost saving. Unlike instances with $|D| = 10$, which can reach the optimum by the original SPRP, instances with $|D| = 20$ can be suboptimal, as the instances in Table 3 that do not have zero optimality gap among the five solution methods show. The positive values of "SPRP$_r$(%)" (all but except for 07-20-04-0-0) in Table 6 demonstrate that even these suboptimal solutions outperform the first-stage solutions by SPRP$_t$. SPRP$_t$ can result in an overall cost increase of up to 13.15% (05-20-02-0-1). Therefore, it is worthwhile to solve the model with a routing subproblem.

In Table 7, columns "$aut \times 0$", "$aut \times 25\%$", and "$aut \times 50\%$", reduce the estimation for unit transportation cost by 100%, 75%, and 50%, while column "$aut \times 200\%$" increases the estimated unit transportation cost by 100%. In the column "Baseline", we conduct experiments with unit transportation costs exactly equal to 6.5874. From Table 7, we have three observations. Firstly, by the transportation model, the cost is overestimated in the baseline experiments ($aut = 6.5874$). When the unit transportation cost is modified, this error in cost ranges from -61.99% to 641.46% on average. Secondly, using the planning solution by the transportation model can result in a 4.01% cost increase (experiments on $aut := aut \times 0$ are not considered as it is not likely that transportation cost is completely neglected in practice). Finally, this type of cost increase is not sensitive to the unit transportation cost estimation. In other words, one cannot expect to obtain a solution as good as SPRP by more precise transportation cost estimation. From the planning point of view, our SPRP is a more precise model than the model with a second-stage transportation problem, one can get a cost estimation in the planning phase by applying our model. From the response point of view, our SPRP solution needs a lower overall cost than the transportation model in 3/4 of the instances with $aut = 6.5874$. Since our experiment is based on the optimal solution of SPRP to estimate the unit transportation cost. In other words, it is sufficiently precise for the transportation problem, but the results are not as good as the SPRP solution. In practice, it is difficult to estimate unit transportation costs precisely, especially in the context of humanitarian operations, which implies that using the simplified transportation model may result in worse solutions.

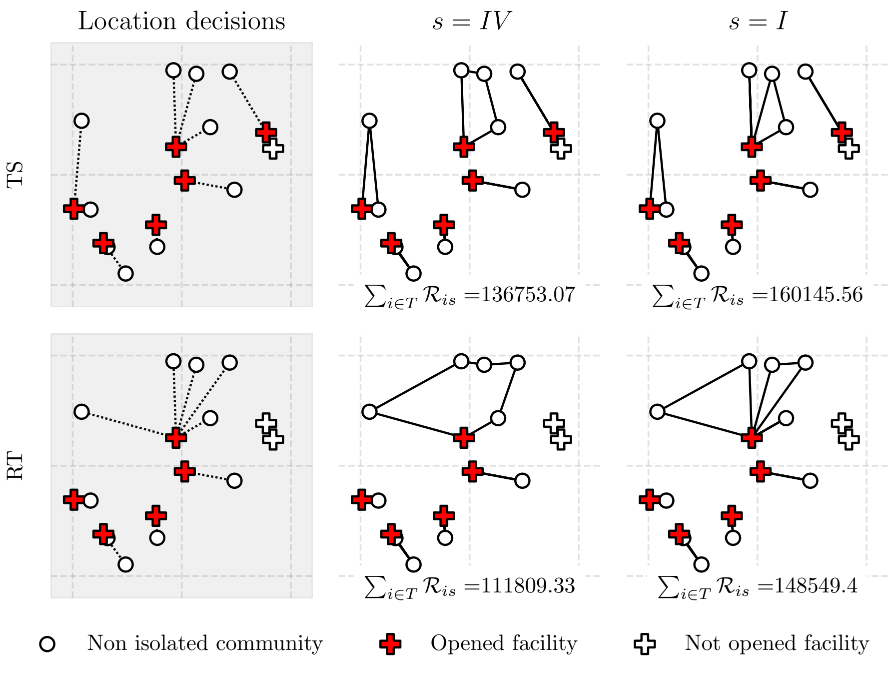


Figure 8: Illustration for integrating routing decisions

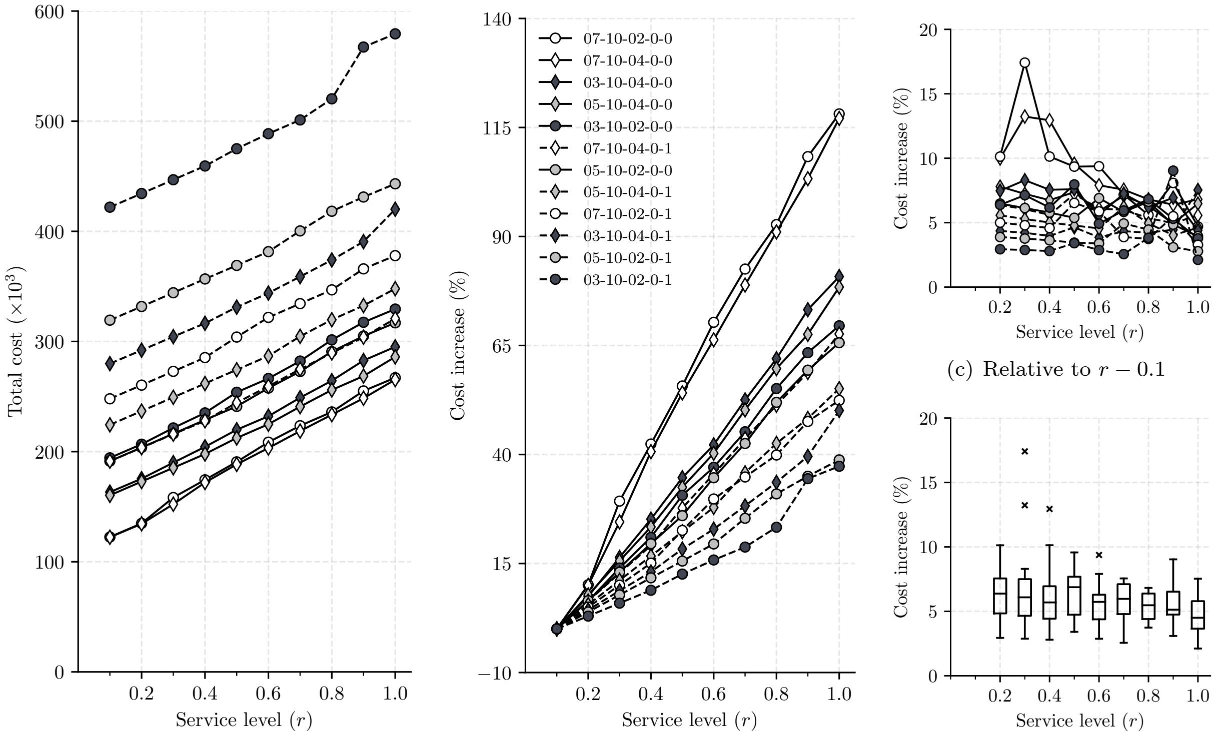
An illustrative example is presented in Figure 8, which shows the value of incorporating vehicle routing problems. The reason we chose instance 07-10-02-0-0 with $aut = 6.5874$ is, among all the baseline experiments in Table 7, it is the one that results in the highest objective difference (7.78%) between models with the second-stage transportation problem and the second-stage routing problem. In Figure 8, the column "Location decisions" represents the pre-disaster facility location decisions, columns $s = \{IV, I\}$ represent the post-disaster routing decisions; and rows "TS" and "RT" represent transportation problem and routing problem respectively. In the planning stage, six out of seven relief facilities are opened by the transportation problem, while only five relief facilities are opened by the routing problem, and the total initial inventory is 1244 units for both TS and RT. The most significant location decision difference is that TS tends to assign a community to its closest relief facility while

RT does not always assign a community to its closest relief facility. Since we know that RT has a better solution than TS, we can conclude it is not worthwhile for TS to open one more relief facility. There is no doubt that additional facility costs have been resulted, worse still, unadvisable location decisions result in suboptimal routing decisions. For all the scenarios, TS results in poorer routing decisions than RT, i.e., 13,6753.07 versus 111,809.33 for scenario $s = IV$ and 160,145.56 versus 148,549.4 for scenario $s = I$.

In summary, the model with routing subproblems (the proposed SPRP) provides a better preparedness plan and more accurate estimates for the overall cost, compared with the model with transportation subproblems. Using the transportation problem to approximate the response-stage problem usually underestimates or overestimates the real cost, even when the average unit transportation cost is fairly precisely estimated. Furthermore, there is an unavoidable bias between the optimal solution and the solution obtained by fixing planning-stage decisions to that yield by transportation problem, no matter how precise the estimated average unit transportation cost is. The value of integrating vehicle routing decisions is thus validated.

6.3. Relationship between service level and objective

The trade-off between service level r and the minimum cost can be valuable for decision-makers. Though our RASP is originally a single objective problem, in reality, the disaster planning cost (1.1) and the service level requirement (2.8) are a pair of conflict objectives, and the improvement of the one may cause degrading on the other. We approximate the Pareto frontier and discuss how the service level influences the overall cost. Pareto frontier is a concept that is usually used for multi-objective optimization. The Pareto frontier is the set of Pareto optimal solutions. Considering a bi-objective problem, a Pareto optimal solution is a solution that cannot be improved in one objective value without degrading the other one. We solve SPRP by varying the service level r from 0.1 to 1, in increments of 0.1. This procedure is also known as the *ϵ-constraint method*. After all ten Pareto optimal solutions are obtained, we plot them on Figure 9.



(a) Absolut value

(b) Relative to $r = 0.1$

(c) Relative to $r - 0.1$

(d) Relative to $r - 0.1$

Figure 9: Cost increasing over service level

In Figure 9, Figure 9(a) presents the relationship between total objective value and service level, which ranges from 0.1 to 1.0; Figure 9(b) presents the cost increase in percent compared with service level $r = 0$; Figure 9(c) and Figure 9(d) present the cost increase in percent compared with service level $r - 0.1$ for $0.2 \leq r \leq 1.0$. Figure 9(a) shows for all the instances, the cost increase is nearly linear, and all the lines have a similar slope. Let us put instance 05-10-02-01 as an example. We fit the curve to a linear function $cost = 140829.94 \times r + 302115.17$, this function can be used to predict the cost of different service levels for decision-makers. In practice, a decision-maker who is responsible for disaster response planning for a particular region, can predict the cost under different service levels by computing the objective value of our model with some particular r but does not have to solve the model every time the cost for a new r is required. Figure 9(a) also shows that with our instances, the number of relief facility candidates (coded with colors) does not influence the overall cost significantly. Figure 9(b) shows that the cost increase for $r = 1.0$ compared with $r = 0.1$ range from around 40% to 120% for all the selected instances. In Figure 9(c) and Figure 9(d), the relative differences for all the instances with given r compared with $r - 1$ are at most 10% except four outliers. Figure 9 demonstrates the value of our model in predicting humanitarian relief costs and identifying the marginal cost for increasing service level.

7. Conclusions

This study focuses on improving the effectiveness and efficiency of disaster preparedness planning. Inspired by commercial logistics operations and humanitarian organization practice, horizontal coordination is integrated into the preparedness planning problem, which allows facilities that are in shortage can be replenished from that in surplus. We formulated the disaster preparedness planning problem with two-stage stochastic programming, in which the first-stage decisions are facility location and relief item preparedness, and the second-stage decisions are vehicle routing problems with the chance of horizontal coordination. The vehicle routing problems in the second stage avoid parameters that are difficult to obtain (e.g., cost per person and cost per delivery), compared with the allocation problems and the transportation problems that have been widely addressed in the literature. To the best of our knowledge, our second-stage routing problem itself is a new variant of vehicle routing with inter-depot routes. The objective of our model is the sum of planning costs (location, procurement) and expected response costs (routing). Also, we introduced the service level, which makes our model a tool for helping decision makers to understand the trade-off between relief supply sufficiency and monetary costs. The service level is defined as the lowest ratio of satisfied demand for all the affected communities in all the cases. Apart from modeling, the solution methodology is another main aspect of this study. Our problem is formulated as a stochastic programming model, which is computationally challenging. To find solutions with optimality certificates, we tailored the original LBBD algorithm for our model, i.e., the nested decomposition and the problem-specific logic-based Benders cuts were developed. A variant of the classical implementation of LBBD, namely branch-and-check is also implemented to prohibit the algorithm from building a branch-and-bound tree from scratch each time a new iteration begins. We also introduced several valid inequalities to reinforce our model.

Extensive numerical results were reported based on the experiments on the designed instances to validate the efficiency of the algorithm and provide managerial insights. Both the classical implementation of NLBBD and the branch-and-check implementation of NLBBD outperform the state-of-the-art commercial solver. For small-size instances where both NLBBD and the commercial solver can find the optimum, the runtime of NLBBD can be less than that of the commercial solver in magnitudes. For large-size instances where the commercial solver cannot even find a feasible solution, the NLBBD algorithms can yield feasible solutions. The value of the valid inequalities is also validated since they help the commercial solver find feasible solutions. The NLBBD-B&C requires less time to finish one iteration than the classical NLBBD-C, thus more iterations can be completed within the same time limit. As a result, in the medium-size instances, the NLBBD-B&C yields solutions with narrower optimality gaps than the classical NLBBD-C. After the computational performance analysis, we analyzed the effect of horizontal coordination and the impact of integrating vehicle routing problems, followed by an example that illustrates the relationship between service level and objective value. The horizontal coordination model can yield feasible solutions on instances that are infeasible with the direct transportation model, and the objective value of the horizontal coordination model is always lower than the direct transportation model on a certain instance. When the second-stage problem of our preparedness planning model is replaced with a transportation problem (in contrast to the routing problem), the overall costs can be underestimated or overestimated even when the unit transportation cost is fairly precisely estimated. In summary, the NLBBD algorithm with optimality cuts and feasibility cuts for the stochastic programming problem contributes to the LBBD literature. The proposed preparedness planning problem with horizontal coordination and routing decisions contributes to the disaster operations management literature.

This study can be further extended in the following directions. First, as described earlier, this work focuses on the pre-disaster planning phase and the very beginning phase after disasters, with two-stage stochastic programming. A more general multi-stage stochastic programming problem that focuses not only on the beginning phase after disasters but also a long period after disasters till full recovery still needs to be investigated. Also, it is natural to extend the NLBBD to multi-stage stochastic settings. There are nested Benders style algorithms for multi-stage stochastic programming, e.g., the stochastic dual dynamic programming (SDDP) (Pereira and Pinto, 1991) and its extension the stochastic dual dynamic integer programming (SDDiP) (Zou et al., 2019). However, both SDDP and SDDiP are limited to a specific class of problems. As LBBD is capable of any class of optimization problems, it would be promising to develop new and more general nested Benders decomposition frameworks based on SDDP and LBBD. Finally, our current implementation of NLBBD includes three simple enhancements that significantly improve the proposed solution method. However, there is still room for further improvement. Exploring additional acceleration techniques to enhance the proposed NLBBD would be interesting.

Funding

This work was supported by the United States National Science Foundation (NSF) [grant number CCF-1948159], the National Natural Science Foundation of China (NSFC) [grant number 72071106], and the Excellent Young Scientists Fund Program (Overseas) of National Natural Science Foundation of China (NSFC).

Notes on contributors

Dr. Penghui Guo earned his Ph.D. in Management Science and Engineering from Nanjing University of Aeronautics and Astronautics (NUAA), China. He earned his master's degree in Management Science and Engineering (Industrial Engineering) from NUAA, and his bachelor's degree in Industrial Engineering from Henan Polytechnic University, China. His primary research focus is on humanitarian logistics, specifically the modeling methods and algorithms under uncertainty for humanitarian relief. He has published several papers in peer-reviewed journals, including *European Journal of Operational Research*.

Dr. Zhijie Sasha Dong is a Junior Chair Professor and Ph.D. advisor in the School of Transportation at Southeast University, Nanjing, China. She received her Ph.D. in Civil Engineering from Cornell University, after completing her master's at Columbia University and bachelor's at Nanjing University. Before joining Southeast University in 2024, she was a dual-appointed associate professor (Industrial Engineering & Construction Management) at the University of Houston and a senior operations research analyst at FedEx. Her research focuses on logistics and supply chain management, emergency management, and system resilience, with a strong emphasis on applying big data, AI, and optimization techniques. She has published extensively in top-tier journals and serves on the editorial boards of several international journals, including *Sustainable Futures* and *Communications Engineering* (Nature journal).

Dr. Jianjun Zhu is a professor in the College of Economics and Management at Nanjing University of Aeronautics and Astronautics, China. He earned his Ph.D. in Systems Engineering from Northeastern University, China. His research focuses on logistics and supply chain management, decision analysis, and system optimization. He has published extensively in journals, including *European Journal of Operational Research* and *Computers & Industrial Engineering*.

References

Adsanver, B., Balcik, B., Bélanger, V., and Rancourt, M.-È. (2024). Operations research approaches for improving coordination, cooperation, and collaboration in humanitarian relief chains: A framework and literature review. *European Journal of Operational Research*, 319(2):384–398.

Albareda-Sambola, M., Fernández, E., and Laporte, G. (2007). Heuristic and lower bound for a stochastic location-routing problem. *European Journal of Operational Research*, 179(3):940–955.

Albareda-Sambola, M., Fernández, E., and Laporte, G. (2009). The capacity and distance constrained plant location problem. *Computers & Operations Research*, 36(2):597–611.

Alem, D., Clark, A., and Moreno, A. (2016). Stochastic network models for logistics planning in disaster relief. *European Journal of Operational Research*, 255(1):187–206.

An, S., Cui, N., Li, X., and Ouyang, Y. (2013). Location planning for transit-based evacuation under the risk of service disruptions. *Transportation Research Part B: Methodological*, 54:1–16.

Arslan, O., Kumcu, G. Ç., Kara, B. Y., and Laporte, G. (2021). The location and location-routing problem for the refugee camp network design. *Transportation Research Part B: Methodological*, 143:201–220.

Aslan, E. and Çelik, M. (2019). Pre-positioning of relief items under road/facility vulnerability with concurrent restoration and relief transportation. *IISE Transactions*, 51(8):847–868.

Baker, J. and Salway, M. (2016). Development of a proposal for a methodology to cost inter-agency humanitarian response plans. Technical report, Inter-Agency Standing Committee.

Balcik, B., Beamon, B. M., Krejci, C. C., Muramatsu, K. M., and Ramirez, M. (2010). Coordination in humanitarian relief chains: Practices, challenges and opportunities. *International Journal of Production Economics*, 126(1):22–34.

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

Baldacci, R., Battarra, M., and Vigo, D. (2008). Routing a Heterogeneous Fleet of Vehicles. In Golden, B., Raghavan, S., and Wasil, E., editors, *The Vehicle Routing Problem: Latest Advances and New Challenges*, Operations Research/Computer Science Interfaces, pages 3–27. Springer US, Boston, MA.

Baskaya, S., Ertem, M. A., and Duran, S. (2017). Pre-positioning of relief items in humanitarian logistics considering lateral transhipment opportunities. *Socio-Economic Planning Sciences*, 57:50–60.

Beck, J. C. (2010). Checking-Up on Branch-and-Check. In Cohen, D., editor, *Principles and Practice of Constraint Programming – CP 2010*, Lecture Notes in Computer Science, pages 84–98, Berlin, Heidelberg. Springer.

Benders, J. F. (1962). Partitioning procedures for solving mixed-variables programming problems. *Numerische Mathematik*, 4(1):238–252.

Boccia, M., Crainic, T. G., Sforza, A., and Sterle, C. (2011). Location-routing models for designing a two-echelon freight distribution system. Technical Report CIRRELT-2011-06, Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation.

Caunhye, A. M., Zhang, Y., Li, M., and Nie, X. (2016). A location-routing model for prepositioning and distributing emergency supplies. *Transportation Research Part E: Logistics and Transportation Review*, 90:161–176.

Chu, Y. and Xia, Q. (2004). Generating Benders cuts for a general class of integer programming problems. In Régin, J.-C. and Rueher, M., editors, *Integration of AI and OR Techniques in Constraint Programming for Combinatorial Optimization Problems*, Lecture Notes in Computer Science, pages 127–141, Berlin, Heidelberg. Springer.

Davis, L. B., Samanlioglu, F., Qu, X., and Root, S. (2013). Inventory planning and coordination in disaster relief efforts. *International Journal of Production Economics*, 141(2):561–573.

DHS (2008). National Response Framework. Government document, United States Department of Homeland Security (DHS).

DHS (2011). National Preparedness System. Technical report, United States Department of Homeland Security (DHS).

DI (2021). Global Humanitarian Assistance Report 2021. Technical report, Development Initiatives.

Dolan, E. D. (2001). The NEOS Server 4.0 administrative guide. Technical Memorandum ANL/MCS-TM-250, Mathematics and Computer Science Division, Argonne National Laboratory.

Doodman, M., Shokr, I., Bozorgi-Amiri, A., and Jolai, F. (2019). Pre-positioning and dynamic operations planning in pre- and post-disaster phases with lateral transhipment under uncertainty and disruption. *Journal of Industrial Engineering International*, 15(1):53–68.

eCFR (2009). Duplication of benefits. Accessed 2023-04-26.

Elçi, Ö. and Hooker, J. (2022). Stochastic planning and scheduling with logic-based Benders decomposition. *INFORMS Journal on Computing*, 34(5):2383–2865.

Elçi, Ö. and Noyan, N. (2018). A chance-constrained two-stage stochastic programming model for humanitarian relief network design. *Transportation Research Part B: Methodological*, 108:55–83.

Erbeyoğlu, G. and Bilge, Ü. (2020). A robust disaster preparedness model for effective and fair disaster response. *European Journal of Operational Research*, 280(2):479–494.

Errami, N., Queiroga, E., Sadykov, R., and Uchoa, E. (2024). VRPSolverEasy: A Python library for the exact solution of a rich vehicle routing problem. *INFORMS Journal on Computing*, 36(4):939–1146.

Fachini, R. F. and Armentano, V. A. (2020). Logic-based Benders decomposition for the heterogeneous fixed fleet vehicle routing problem with time windows. *Computers & Industrial Engineering*, 148:106641.

Fazel-Zarandi, M. M. and Beck, J. C. (2012). Using logic-based Benders decomposition to solve the capacity- and distance-constrained plant location problem. *INFORMS Journal on Computing*, 24(3):387–398.

Fazel-Zarandi, M. M., Berman, O., and Beck, J. C. (2013). Solving a stochastic facility location/fleet management problem with logic-based Benders' decomposition. *IIE Transactions*, 45(8):896–911.

FEMA (2013). Livestock in Disasters - Emergency Management in the United States. Accessed 2021-08-25.

FEMA (2019). Basic Emergency Operations Center Functions - Activation Levels. Accessed 2021-08-25.

General Office of the State Council, PRC (2016). Notice of the General Office of the State Council on Issuing the National Emergency Plan for Natural Disaster Relief. Accessed 2021-08-25.

Grahovac, J. and Chakravarty, A. (2001). Sharing and lateral transshipment of inventory in a supply chain with expensive low-demand items. *Management Science*, 47(4):579–594.

Guo, C., Bodur, M., Aleman, D. M., and Urbach, D. R. (2021). Logic-based Benders decomposition and binary decision diagram based approaches for stochastic distributed operating room scheduling. *INFORMS Journal on Computing*, 33(4):1551–1569.

Guo, P. and Zhu, J. (2023). Capacity reservation for humanitarian relief: A logic-based Benders decomposition method with subgradient cut. *European Journal of Operational Research*, 311(3):942–970.

Hong, X., Lejeune, M. A., and Noyan, N. (2015). Stochastic network design for disaster preparedness. *IIE Transactions*, 47(4):329–357.

Hooker, J. N. (2007). Planning and scheduling by logic-based Benders decomposition. *Operations Research*, 55(3):588–602.

Hooker, J. N. (2019). Logic-based Benders decomposition for large-scale optimization. In Velásquez-Bermúdez, J. M., Khakifirooz, M., and Fathi, M., editors, *Large Scale Optimization in Supply Chains and Smart Manufacturing: Theory and Applications*, Springer Optimization and Its Applications, pages 1–26. Springer International Publishing, Cham.

Hooker, J. N. and Ottosson, G. (2003). Logic-based Benders decomposition. *Mathematical Programming*, 96(1):33–60.

Hu, S. and Dong, Z. S. (2019). Supplier selection and pre-positioning strategy in humanitarian relief. *Omega*, 83:287–298.

IPC-IG (2022). Digital innovations in delivering social protection in rural areas: Lessons for public provisioning during the post-pandemic recovery and beyond. Research Report, International Policy Centre for Inclusive Growth (IPC-IG).

Kamyabniya, A., Lotfi, M. M., Cai, H., Hosseininasab, H., Yaghoubi, S., and Yih, Y. (2019). A two-phase coordinated logistics planning approach to platelets provision in humanitarian relief operations. *IISE Transactions*, 51(1):1–21.

Kara, B. Y. and Savaşer, S. (2017). Humanitarian Logistics. In Batta, R., Peng, J., Smith, J. C., and Greenberg, H. J., editors, *Leading Developments from INFORMS Communities*, INFORMS TutORials in Operations Research, pages 272–309. INFORMS.

Keshvari Fard, M., Ljubić, I., and Papier, F. (2022). Budgeting in international humanitarian organizations. *Manufacturing & Service Operations Management*, 24(3):1261–1885.

Koç, Ç., Bektaş, T., Jabali, O., and Laporte, G. (2016). Thirty years of heterogeneous vehicle routing. *European Journal of Operational Research*, 249(1):1–21.

Kůdela, J. and Popela, P. (2017). Warm-start cuts for generalized Benders decomposition. *Kybernetika*, 53(6):1012–1025.

Kumar, P. and Singh, R. K. (2022). Application of Industry 4.0 technologies for effective coordination in humanitarian supply chains: A strategic approach. *Annals of Operations Research*, 319:379–411.

Kunz, N., Reiner, G., and Gold, S. (2014). Investing in disaster management capabilities versus pre-positioning inventory: A new approach to disaster preparedness. *International Journal of Production Economics*, 157:261–272.

Laporte, G. and Louveaux, F. V. (1993). The integer L-shaped method for stochastic integer programs with complete recourse. *Operations Research Letters*, 13(3):133–142.

Lee, H. L. (1987). A multi-echelon inventory model for repairable items with emergency lateral transshipments. *Management Science*, 33(10):1302–1316.

Lombardi, M., Milano, M., Ruggiero, M., and Benini, L. (2010). Stochastic allocation and scheduling for conditional task graphs in multi-processor systems-on-chip. *Journal of Scheduling*, 13(4):315–345.

Mak, W.-K., Morton, D. P., and Wood, R. K. (1999). Monte Carlo bounding techniques for determining solution quality in stochastic programs. *Operations Research Letters*, 24(1):47–56.

Manopiniwes, W. and Irohara, T. (2017). Stochastic optimisation model for integrated decisions on relief supply chains: Preparedness for disaster response. *International Journal of Production Research*, 55(4):979–996.

Mete, H. O. and Zabinsky, Z. B. (2010). Stochastic optimization of medical supply location and distribution in disaster management. *International Journal of Production Economics*, 126(1):76–84.

Morrice, D. J., Cronin, P., Tanrisever, F., and Butler, J. C. (2016). Supporting hurricane inventory management decisions with consumer demand estimates. *Journal of Operations Management*, 45:86–100.

Ni, W., Shu, J., and Song, M. (2018). Location and emergency inventory pre-positioning for disaster response operations: Min-max robust model and a case study of yushu earthquake. *Production and Operations Management*, 27(1):160–183.

Noham, R. and Tzur, M. (2018). Designing humanitarian supply chains by incorporating actual post-disaster decisions. *European Journal of Operational Research*, 265(3):1064–1077.

OCHA (2018). HRP Costing Methodology Options. Technical report, United Nations Office for the Coordination of Humanitarian Affairs.

OCHA (2021). Haiti: Earthquake 2021 Humanitatian Response. Accessed 2022-03-02.

Özdamar, L. and Ertem, M. A. (2015). Models, solutions and enabling technologies in humanitarian logistics. *European Journal of Operational Research*, 244(1):55–65.

Pacheco, G. G. and Batta, R. (2016). Forecast-driven model for prepositioning supplies in preparation for a foreseen hurricane. *Journal of the Operational Research Society*, 67(1):98–113.

Paul, J. A. and Wang, X. J. (2019). Robust location-allocation network design for earthquake preparedness. *Transportation Research Part B: Methodological*, 119:139–155.

Paul, J. A. and Zhang, M. (2019). Supply location and transportation planning for hurricanes: A two-stage stochastic programming framework. *European Journal of Operational Research*, 274(1):108–125.

Pereira, M. V. F. and Pinto, L. M. V. G. (1991). Multi-stage stochastic optimization applied to energy planning. *Mathematical Programming*, 52(1):359–375.

Pessoa, A., Sadykov, R., Uchoa, E., and Vanderbeck, F. (2020). A generic exact solver for vehicle routing and related problems. *Mathematical Programming*, 183(1):483–523.

Rahmaniani, R., Crainic, T. G., Gendreau, M., and Rei, W. (2017). The Benders decomposition algorithm: A literature review. *European Journal of Operational Research*, 259(3):801–817.

Rawls, C. G. and Turnquist, M. A. (2010). Pre-positioning of emergency supplies for disaster response. *Transportation Research Part B: Methodological*, 44(4):521–534.

Rezaei-Malek, M., Tavakkoli-Moghaddam, R., Cheikhrouhou, N., and Taheri-Moghaddam, A. (2016). An approximation approach to a trade-off among efficiency, efficacy, and balance for relief pre-positioning in disaster management. *Transportation Research Part E: Logistics and Transportation Review*, 93:485–509.

Riise, A., Mannino, C., and Lamorgese, L. (2016). Recursive logic-based Benders' decomposition for multi-mode outpatient scheduling. *European Journal of Operational Research*, 255(3):719–728.

Rodríguez-Pereira, J., Balcik, B., Rancourt, M.-È., and Laporte, G. (2021). A cost-sharing mechanism for multi-country partnerships in disaster preparedness. *Production and Operations Management*, 30(12):4541–4565.

Rottkemper, B., Fischer, K., and Blecken, A. (2012). A transshipment model for distribution and inventory relocation under uncertainty in humanitarian operations. *Socio-Economic Planning Sciences*, 46(1):98–109.

Sabbaghtorkan, M., Batta, R., and He, Q. (2020). Prepositioning of assets and supplies in disaster operations management: Review and research gap identification. *European Journal of Operational Research*, 284(1):1–19.

Salhi, S., Imran, A., and Wassan, N. A. (2014). The multi-depot vehicle routing problem with heterogeneous vehicle fleet: Formulation and a variable neighborhood search implementation. *Computers & Operations Research*, 52:315–325.

Sanci, E. and Daskin, M. S. (2019). Integrating location and network restoration decisions in relief networks under uncertainty. *European Journal of Operational Research*, 279(2):335–350.

Sanci, E. and Daskin, M. S. (2021). An integer L-shaped algorithm for the integrated location and network restoration problem in disaster relief. *Transportation Research Part B: Methodological*, 145:152–184.

Schiffer, M., Schneider, M., Walther, G., and Laporte, G. (2019). Vehicle routing and location routing with intermediate stops: A review. *Transportation Science*, 53(2):319–343.

Setiawan, E., Liu, J., and French, A. (2019). Resource location for relief distribution and victim evacuation after a sudden-onset disaster. *IISE Transactions*, 51(8):830–846.

Stavropoulou, F., Repoussis, P. P., and Tarantilis, C. D. (2019). The vehicle routing problem with profits and consistency constraints. *European Journal of Operational Research*, 274(1):340–356.

Thorsteinsson, E. S. (2001). Branch-and-check: A hybrid framework integrating mixed integer programming and constraint logic programming. In Walsh, T., editor, *Principles and Practice of Constraint Programming — CP 2001*, Lecture Notes in Computer Science, pages 16–30, Berlin, Heidelberg. Springer.

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

TOI (2011). Ghost anganwadi beneficiaries haunt govt. Accessed 2023-04-26.

Topan, E. and van der Heijden, M. C. (2020). Operational level planning of a multi-item two-echelon spare parts inventory system with reactive and proactive interventions. *European Journal of Operational Research*, 284(1):164–175.

Torabi, S. A., Shokr, I., Tofighi, S., and Heydari, J. (2018). Integrated relief pre-positioning and procurement planning in humanitarian supply chains. *Transportation Research Part E: Logistics and Transportation Review*, 113:123–146.

Toyasaki, F., Arikan, E., Silbermayr, L., and Sigala, I. F. (2017). Disaster relief inventory management: Horizontal cooperation between humanitarian organizations. *Production and Operations Management*, 26(6):1221–1237.

Uichanco, J. (2021). A model for prepositioning emergency relief items before a typhoon with an uncertain trajectory. *Manufacturing & Service Operations Management*, 24(2):766–790.

UN (2022). UNdata - Per capita GNI at current prices (US dollars). Accessed 2022-03-07.

UNDP (2022). New digital solution designed to eliminate duplication of financial assistance. Accessed 2023-04-26.

UNDRR (2020). The human cost of disasters: An overview of the last 20 years (2000-2019). Technical report, United Nations Office for Disaster Risk Reduction (UNDRR).

UNICEF (2023). Warehousing and distribution | UNICEF Supply Division. Accessed 2023-04-21.

UNICEF and WFP (2015). UNICEF/WFP return on investment for emergency preparedness study. Technical report, UN Children's Fund / UN World Food Programme.

Van Slyke, R. M. and Wets, R. (1969). L-shaped linear programs with applications to optimal control and stochastic programming. *SIAM Journal on Applied Mathematics*, 17(4):638–663.

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

van Wijk, A. C. C., Adan, I. J. B. F., and van Houtum, G. J. (2019). Optimal lateral transshipment policies for a two location inventory problem with multiple demand classes. *European Journal of Operational Research*, 272(2):481–495.

Velasquez, G. A., Mayorga, M. E., and Ozaltin, O. Y. (2020). Prepositioning disaster relief supplies using robust optimization. *IISE Transactions*, 52(10):1122–1140.

Wang, C. and Chen, S. (2020). A distributionally robust optimization for blood supply network considering disasters. *Transportation Research Part E: Logistics and Transportation Review*, 134:101840.

Wang, J., Cai, J., Yue, X., and Suresh, N. C. (2021a). Pre-positioning and real-time disaster response operations: Optimization with mobile phone location data. *Transportation Research Part E: Logistics and Transportation Review*, 150:102344.

Wang, W., Yang, K., Yang, L., and Gao, Z. (2021b). Two-stage distributionally robust programming based on worst-case mean-CVaR criterion and application to disaster relief management. *Transportation Research Part E: Logistics and Transportation Review*, 149:102332.

Wang, Y., Dong, Z. S., and Hu, S. (2021c). A stochastic prepositioning model for distribution of disaster supplies considering lateral transshipment. *Socio-Economic Planning Sciences*, 74:100930.

Wankmüller, C. and Reiner, G. (2019). Coordination, cooperation and collaboration in relief supply chain management. *Journal of Business Economics*, 90:239–276.

Webster, M., Ginnetti, J., Walker, P., Coppard, D., and Kent, R. (2008). The Humanitarian Costs Of Climate Change. Technical report, Feinstein International Center.

WFP (2019). UN Humanitarian Response Depot. Accessed 2023-04-21.

Zou, J., Ahmed, S., and Sun, X. A. (2019). Stochastic dual dynamic integer programming. *Mathematical Programming*, 175(1):461–502.

Supplemental Online Materials to "Nested logic-based Benders decomposition for disaster preparedness planning with horizontal coordination" by Penghui Guo, Zhijie Sasha Dong, and Jianjun Zhu

S.1. Literature on humanitarian logistics

We reviewed the literature on humanitarian logistics that focuses on preparedness planning and horizontal coordination. The results are summarized in Table S1, which includes 27 papers from Section 2.1 for deterministic, robust, and stochastic preparedness planning problems, and seven papers from Section 2.2 for horizontal coordination. In the 34 papers reviewed, 50% (17/34) included assignment decisions, either explicitly or implicitly. Specifically, 44% (12/27) of the preparedness planning papers and 71% (5/7) of the horizontal coordination papers included assignment decisions. In studies using stochastic programming, which is closely related to our work, 43% (9/21) included assignment decisions, and 44% (4/9) of the papers that included assignment decisions made these decisions in the first stage. This suggests that assigning customers to facilities in the first stage is common.

Table S1: Literature on humanitarian logistics

Paper	Assign.	1st	$\|T\|$	$\|D\|$	$\|S\|$	Paper	Assign.	1st	$\|T\|$	$\|D\|$	$\|S\|$
Preparedness planning						Noham and Tzur (2018)	$\surd$	N	30	50	10
Deterministic:						Torabi et al. (2018)	$\times$	-	24	24	3
An et al. (2013)	$\surd$	-	-	-	-	Ni et al. (2018)	$\times$	-	13	13	100
Arslan et al. (2021)	$\times$	-	-	-	-	Elçi and Noyan (2018)	$\surd^{c}$	N	10	30	51
Robust:						Sanci and Daskin (2019)	$\times$	-	39	39	10
Paul and Wang (2019)	$\times$	-	100	34	8	Paul and Zhang (2019)	$\times$	-	50	30	10
Velasquez et al. (2020)	$\times$	-	30	30	-	Hu and Dong (2019)	$\times$	-	13	25	40
Wang and Chen (2020)	$\surd$	N	4	4	50	Aslan and Çelik (2019)	$\surd$	N	25	37	180
Wang et al. (2021b)	$\surd^{c}$	Y	90	90	118	Erbeyoğlu and Bilge (2020)	$\surd$	N	428	1721	10
Stochastic:						Wang et al. (2021a)	$\times$	-	6	14	30
Morrice et al. (2016)	$\times$	-	-	2	-	Sanci and Daskin (2021)	$\times$	-	10	39	50
Rawls and Turnquist (2010)	$\times$	-	30	30	51	**Horizontal coordination**					
Mete and Zabinsky (2010)	$\surd^{r}$	Y	5	10	6	Baskaya et al. (2017)	$\surd$	-	37	37	-
Hong et al. (2015)	$\times$	-	30	30	51	Kamyabniya et al. (2019)	$\surd^{c}$	-	17	1	-
Alem et al. (2016)	$\times$	-	4	9	40	Setiawan et al. (2019)	$\surd$	-	47	-	-
Pacheco and Batta (2016)	$\surd^{c}$	Y	154	301	14	Rottkemper et al. (2012)	$\times$	-	5	-	-
Rezaei-Malek et al. (2016)	$\surd^{r}$	Y	6	22	8	Davis et al. (2013)	$\surd$	N	47	12	6
Tofighi et al. (2016)	$\times$	-	22	22	4	Doodman et al. (2019)	$\times$	-	12	10	3
Caunhye et al. (2016)	$\surd$	N	4	4	3	Wang et al. (2021c)	$\surd$	N	15	15	15
Manopiniwes and Irohara (2017)	$\surd^{r}$	Y	6	123	7	This paper	$\surd$	Y	7	40	4

[1] Assign., assignment decisions; $\surd$, assignment decisions are explicit; $\times$, no assignment decisions; $\surd^{c}$, assignments are implied by a coverage radius; $\surd^{r}$, assignment decisions are implied by route-related decisions; 1st, whether assignment decisions are in the first stage; Y, assignment decisions are in the first stage; N, assignment decisions are in the second stage; -, not applicable; $|T|$, number of facilities; $|D|$, number of communities; $|S|$, number of scenarios. If multiple instances are used, we report the maximum $|T|$, $|D|$, and $|S|$.

Out of the 31 papers that include numerical studies with a set of facilities, 32% (10/31) have ten or fewer facilities. Similarly, in the 30 papers that specify a set of communities, 83% (25/30) have no more than 40 communities. Our numerical studies show comparable magnitudes for $|T|$ and $|D|$ as reported in the literature. However, our problem is more challenging than those previously studied, which demonstrates that our choices for $|T|$ and $|D|$ are appropriate.

S.2. Notations

Table S2: Notations in Section 4

	PMP	
$\mathcal{R}_{is}$	Var.	Estimators in the PMP for the RASP objectives.
	RASP	
$\tilde{\mathcal{R}}_{is}$	Val.	Values of the RASP objective function.
$\mathcal{R}'_{is}$	Var.	Estimators in the RASP for the VDSP objectives.
w_i^{sh}	Var.	Number of vehicles of type h at facility i in scenario s.
$\bar{y}_{ik}$	Val.	Value of y_{ik} yielded by the PMP.
$\bar{l}_i$	Val.	Value of l_i yielded by the PMP.
$\bar{x}_{ij}$	Val.	Value of x_{ij} yielded by the PMP.

$\bar{w}_i^h$	Val.	Value of w_i^h yielded by the PMP.
VDSP		
$\hat{\mathcal{R}'}_{is}$	Val.	Values of the VDSP objective function.
$\tilde{t}_{ij}^s$	Val.	Value of t_{ij}^s yielded by the RASP.
$\tilde{t}_{ii'j}^s$	Val.	Value of $t_{ii'j}^s$ yielded by the RASP.
$\tilde{D}_i^s = \{j \in D \| e_j^s = 0, \bar{x}_{ij} = 1\}$	Set	Non-isolated communities in scenario s assigned to facility i.
$\tilde{D'}_i^s = D \setminus \tilde{D}_i^s$	Set	Isolated communities in scenario s assigned to facility i.
$\tilde{T}_i^s = \{i' \in T \| \sum_{j\in D} \tilde{t}_{ii'j}^s \geq 1\}$	Set	Facilities that accept relief items from facility i in scenario s.
$\tilde{\mathcal{N}}^{(s,i)} = \{i\} \cup \tilde{D'}_i^s \cup \tilde{D}_i^s \cup \tilde{T}_i^s$	Set	Nodes for the VDSP.
$\tilde{\mathcal{N}}_i^{(s,i)} = \tilde{\mathcal{N}}^{(s,i)} \setminus \{i\}$	Set	Nodes for the VDSP excluding i.
$\tilde{\mathcal{A}}^{(s,i)} = \{(p,q) \| p, q \in \tilde{\mathcal{N}}^{(s,i)}\}$	Set	Arcs for the VDSP.
$\tilde{\mathcal{G}}^{(s,i)} = (\tilde{\mathcal{N}}^{(s,i)}, \tilde{\mathcal{A}}^{(s,i)})$	Set	Graph for the VDSP.
$\tilde{Q}_j^{(s,i)} = \tilde{t}_{ij}^s, \forall j \in \tilde{D'}_i^s \cup \tilde{D}_i^s$	Val.	Delivery quantity from facility i to community j in scenario s.
$\tilde{Q}_{i'}^{(s,i)} = \sum_{j\in D} \tilde{t}_{ii'j}^s, \forall i' \in T_i^s$	Val.	Delivery quantity from facility i to facility i' in scenario s.
PMP's cuts		
u	Val.	Iteration index.
$\bar{D}_i^{(u)} = \{j \in D \| \bar{x}_{ij}^{(u)} = 1\}$	Set	Communities assigned to facility i.
$\bar{D'}_i^{s(u)} = \{j \in D \| e_j^s = 1, \bar{x}_{ij}^{(u)} = 1\}$	Set	Isolated communities in scenario s assigned to facility i.
$\mathcal{I}^{(u)}$	Set	Facilities that do not have a sufficient number of vehicles.
$\mathbb{B}(L, Q)$	Val.	Required number of bins sized Q to store items of sizes from list L.
$\mathbb{B}_i^{s(u)}\left(\lceil r \cdot m_j^s \rceil \forall j \in \bar{D}_i^{(u)}, Q\right)$	Val.	Required number of vehicles by facility i in scenario s.
$\mathbb{B}_i^{s(u)}\left(\lceil r \cdot m_j^s \rceil \forall j \in \bar{D'}_i^{s(u)}, Q\right)$	Val.	Required number of special vehicles by facility i in scenario s.
RASP's cuts		
$\tilde{D}_i^{s(u)} = \{j \in D \| \tilde{t}_{ij}^{s(u)} \geq 1\}$	Set	Communities that accept relief items directly from facility i in scenario s.
$\tilde{D'}_i^{s(u)} = \{j \in \tilde{D}_i^{s(u)} \| e_j^s = 1\}$	Set	Isolated communities in $\tilde{D}_i^{s(u)}$.
$\tilde{T}_i^{s(u)} = \{i' \in T \| \sum_{j\in D} \tilde{t}_{ii'j}^{s(u)} \geq 1\}$	Set	Facilities that accept relief items from facility i in scenario s.
$\mathcal{I}^{s(u)}$	Set	Facilities for which the VDSPs are infeasible in scenario s.
$\mathbb{B}_i^{s(u)}\left(\tilde{t}_{ij}^{s(u)}, \tilde{t}_{ii'j}^{s(u)}, Q\right)$	Val.	Required number of vehicles by facility i in scenario s.
$\mathbb{B}_i^{s(u)}\left(\tilde{t}_{ij}^{s(u)}, Q\right)$	Val.	Required number of special vehicles by facility i in scenario s.
$\tilde{C}_{is}^{(u)} = \tilde{D}_i^{s(u)} \cup \tilde{T}_i^{s(u)} \cup \{i\}$	Set	Node cluster including $\tilde{D}_i^{s(u)}$, $\tilde{T}_i^{s(u)}$, and $\{i\}$.
$MAX_{\tilde{C}_{is}^{(u)}, i, j(\text{or } i')}$	Val.	Cost saving by reducing transportation between i and j(or i').

Var., decision variable; Val., value (parameter).

S.3. Bin packing problem

With a given assignment decision ($\bar{x}_{ij}$) feed by PMP to RASP or a transportation decision ($\tilde{t}_{ij}^s, \tilde{t}_{ii'j}^s$) feed by RASP to VDSPs, we need to solve a *bin packing* problem to determine the minimal number of each type of vehicle for all the opened facilities, such that a Benders feasibility cut can be generated once RASP or VDSP is infeasible. Let I be the set of items, $\theta_i, \forall i \in I$ be the size of the item in I, J be the set of bins, and Q be the capacity of the bin, then the bin packing problem is formulated as

$$\mathbb{B}(\theta_i \forall i \in I, Q) = \tag{S.1.1}$$

$$\text{minimize} \sum_{j\in J} \gamma_j \tag{S.1.2}$$

$$\text{subject to} \sum_{j\in J} \pi_{ij} = 1, \quad \forall i \in I \tag{S.1.3}$$

$$\sum_{i\in I} \pi_{ij} \cdot \theta_i \leq Q \cdot \gamma_j, \quad \forall j \in J \tag{S.1.4}$$

$$\pi_{ij}, \gamma_j \in \{0, 1\}, \quad \forall i \in I, j \in J \tag{S.1.5}$$

where $\gamma_j = 1$ if bin j is selected and $\pi_{ij} = 1$ if item i is put into bin j. (S.1.2) minimizes the number of bins selected. (S.1.3) ensures that an item is put into exactly one bin. (S.1.4) ensures that the total size of items put into a bin cannot exceed the bin's capacity. (S.1.5) is domain of decision variables.

S.4. Proof for validity of logic-based Benders cuts

Theorem 1. *The logic-based Benders feasibility cut (7.1) and (7.2) is valid.*

Proof. It is obvious that (7.1) is valid if (7.2) is valid, and vice versa. For simplification, only the validity of (7.2) is proved. property (i) holds for (7.2), since a subsequent iteration may either change the set of affected communities assigned to i or change the number of vehicles prepared at i. Assume that $w_i^{h(u)}$ and $\bar{D'}_i^{(u)}$ results in an infeasible RASP which imply that $w_i^{h(u)} < \mathbb{B}_i^{s(u)}\left(\lceil r \cdot m_j^s \rceil \forall j \in \bar{D'}_i^{(u)}, Q\right)$, if in a future iteration, solution $w_i^{h(u)}$ and $\bar{D'}_i^{(u)}$ is obtained again which results in $w_i^{h(u)} \geq \mathbb{B}_i^{s(u)}\left(\lceil r \cdot m_j^s \rceil \forall j \in \bar{D'}_i^{(u)}, Q\right)$ when $w_i^{h(u)}$ and $\bar{D'}_i^{(u)}$ are substituted into (7.2). Thus, property (i) is proved by contradiction.

Then we prove property (ii) for (7.2) by contradiction. In the following, we write $\mathbb{B}_i^{s(u)}\left(\lceil r \cdot m_j^s \rceil \forall j \in \bar{D'}_i^{(u)}, Q\right)$ as $\mathbb{B}_i^{s(u)}$ for simplification. Let $w_i^{h(v)}$ be a feasible solution in iteration $v > u$ that does not satisfy (7.2), then

$$\sum_{h \in H'} w_i^{h(v)} < \mathbb{B}_i^{s(u)} - \sum_{j \in \bar{D'}_i^{s(u)}} (1 - x_{ij}). \tag{S.2}$$

Let $\bar{D'}_i^{s(v)}$ be the set of isolated affected communities that assigned to i in iteration v, $\delta_1 = \bar{D'}_i^{s(u)} \setminus \bar{D'}_i^{s(v)}$, $\delta_2 = \bar{D'}_i^{s(u)} \cap \bar{D'}_i^{s(v)}$, and $\delta_3 = \bar{D'}_i^{s(v)} \setminus \bar{D'}_i^{s(u)}$ be affected communities that only belonging to $\bar{D'}_i^{s(u)}$, affected pints that belonging to $\bar{D'}_i^{s(u)} \cup \bar{D'}_i^{s(v)}$, and affected communities that only belonging to $\bar{D'}_i^{s(v)}$ respectively. (S.2) can be written as

$$\sum_{h \in H'} w_i^{h(v)} < \mathbb{B}_i^{s(u)} - |\delta_1|, \tag{S.3}$$

since the contribution of δ_2 and δ_3 to the second right-hand side term of (S.2) is zero. If we let $\mathbb{B}_i^{s*}$ be the minimum number of vehicle required for δ_2, it is obvious that $\mathbb{B}_i^{s*} \geq \mathbb{B}_i^{s(u)}$. If the number of vehicle that used for $\bar{D'}_i^{s(v)}$ is exactly $\sum_{h \in H'} w_i^{h(v)}$ and each affected communities in δ_1 is assigned an additional single vehicle then $\sum_{h \in H'} w_i^{h(v)} + |\delta_1| \geq \mathbb{B}_i^{s*}$ holds. From the aforementioned inequality, we obtain that

$$\sum_{h \in H'} w_i^{h(v)} + |\delta_1| \geq \mathbb{B}_i^{s(u)} \tag{S.4}$$

which contradicts (S.3). Thus a feasible solution must not satisfy (7.2). In other words, (7.2) must not exclude any feasible solution. Property (ii) holds for (7.2).

In conclusion, property (i) and (ii) holds for (7.1) and (7.2), (7.1) and (7.2) is a valid logic-based Benders feasibility cut. □

Theorem 2. *The logic-based Benders optimality cut (8) is valid.*

Proof. We first prove property (iii) by contradiction. Cut (8) is generated only if $\bar{\mathcal{R}}_{is}^{(u)} < \tilde{\mathcal{R}}_{is}^{(u)}$ (see Line 17 in Algorithm 1). If the same assignment decision $\bar{x}_{ij}^{(u)}$ and corresponding $\bar{\mathcal{R}}_{is}^{(u)}$ are again yield by PMP in a future iteration, substituting ($\bar{x}_{ij}^{(u)}$, $\bar{\mathcal{R}}_{is}^{(u)}$) to (8) results in $\bar{\mathcal{R}}_{is}^{(u)} \geq \tilde{\mathcal{R}}_{is}^{(u)}$, which is contradicted to $\bar{\mathcal{R}}_{is}^{(u)} < \tilde{\mathcal{R}}_{is}^{(u)}$. Thus current not globally optimal solution must be excluded by (8). Property (iii) holds for (8).

Then we prove property (iv). Let $(\bar{y}_{ik}^*, \bar{l}_i^*, \bar{x}_{ij}^*, \bar{w}_i^*)$ and corresponding PMP recourse cost $\bar{\mathcal{R}}_{is}^*$ denote a globally optimal solution for PMP. Let $\tilde{\mathcal{R}}_{is}^*$ denotes the RASP recourse cost, then $\bar{\mathcal{R}}_{is}^* = \tilde{\mathcal{R}}_{is}^*, \forall i \in T, s \in S$. Assume that (8) is generated in an iteration, we need to prove for any pair of (i, s) (8) holds if $(\bar{y}_{ik}^*, \bar{l}_i^*, \bar{x}_{ij}^*, \bar{w}_i^*)$ and $\bar{\mathcal{R}}_{is}^*$ are substituted to it. For simplification, in the following we let $MAX_{j, \bar{D}_i^{(u)}}$ denote the maximum cost reduction that remove j from $\delta_1 + \delta_2$. If relief facility i is not opened in the optimal solution, that is $\bar{y}_{ik}^* = 0$. In this case, no affected community j is assigned to i , that is $\bar{x}_{ij}^* = 0, \forall j \in D$. Substituting $\bar{\mathcal{R}}_{is}^*$ and $\bar{x}_{ij}^* = 0$ to (8) results in

$$\bar{\mathcal{R}}_{is}^* \geq \tilde{\mathcal{R}}_{is}^{(u)} - \left|\bar{D}_i^{(u)}\right| \cdot MAX_{j, \bar{D}_i^{(u)}}.$$

This inequality holds since the left-hand side (LHS) is trivially nonnegative while the right-hand side (RHS) is nonpositive. More precisely, in the RHS, $\tilde{\mathcal{R}}_{is}^{(u)}$ cannot exceed $|\bar{D}_i^{(u)}| \cdot \max_{p \in \bar{D}_i^{(u)} \cup \{i\}} \{2 \cdot \max_{h \in H} c_{pj}^h\}$ since the second term is the maximum possible reduction that assign all the $j \in \bar{D}_i^{(u)}$ to another facility.

If relief facility i is opened in the optimal solution, that is $\bar{y}_{ik}^* \neq 0$, imply that at least one affected community $j \in D$ is assigned to i. There are two cases as follows. Note that we do not consider assign $j \in D \setminus \bar{D}_i^{(u)}$ to i since $j \notin \bar{D}_i^{(u)}$ have nothing to do with (8). In other words, the contribution of this type of affected community to RHS of (8) is 0, taking these affected communities into consideration will not influence the validity of (8).

Case 1: All the affected communities in $\bar{D}_i^{(u)}$ are assigned to facility i in the optimal solution, that is $\bar{x}_{ij}^* = 1, \forall j \in \bar{D}_i^{(u)}$. Substituting $\bar{\mathcal{R}}_{is}^*$ and $\bar{x}_{ij}^* = 1, \forall j \in \bar{D}_i^{(u)}$ to (8) results in $(\tilde{\mathcal{R}}_{is}^* =)\bar{\mathcal{R}}_{is}^* \geq \tilde{\mathcal{R}}_{is}^{(u)}$. This inequality holds since $\tilde{\mathcal{R}}_{is}^*$ is the cost for visiting $j \in \bar{D}_i^*$, $\tilde{\mathcal{R}}_{is}^{(u)}$ is the cost for visiting $j \in \bar{D}_i^{(u)}$, and $\bar{D}_i^*$ is a super set of $\bar{D}_i^{(u)}$. The equality holds when $\bar{D}_i^* = \bar{D}_i^{(u)}$.

Case 2: A subset of affected communities in $\bar{D}_i^{(u)}$ are assigned to facility i in the optimal solution. Recall that $\bar{D}_i^*$ be the affected communities that are assigned to facility i in the optimal solution. Let $\delta_1 = \bar{D}_i^{(u)} \setminus \bar{D}_i^*$ denotes the affected community that belonging to $\bar{D}_i^{(u)}$ but not belonging to $\bar{D}_i^*$, similarly, $\delta_1 = \bar{D}_i^{(u)} \cap \bar{D}_i^*$ denotes the affected communities that belonging to both $\bar{D}_i^{(u)}$ and $\bar{D}_i^*$, and $\delta_1 = \bar{D}_i^* \setminus \bar{D}_i^{(u)}$ denotes the affected community that belonging to $\bar{D}_i^*$ but not belonging to $\bar{D}_i^{(u)}$. Assume that the globally optimal solution is cut off by (8), we have

$$\begin{aligned}\bar{\mathcal{R}}_{is}^*(\delta_2+\delta_3) <& \tilde{\mathcal{R}}_{is}^{(u)}(\delta_1+\delta_2) - \sum_{j\in\delta_1+\delta_2}(1-x_{ij})\cdot MAX_{j,\bar{D}_i^{(u)}} \\ =& \tilde{\mathcal{R}}_{is}^{(u)}(\delta_1+\delta_2) - |\delta_1|\cdot MAX_{j,\bar{D}_i^{(u)}}\end{aligned}. \tag{S.5}$$

On the other hand

$$\bar{\mathcal{R}}_{is}^*(\delta_2) \geq \bar{\mathcal{R}}_{is}^*(\delta_1+\delta_2) - |\delta_1|\cdot MAX_{j,\bar{D}_i^{(u)}} \tag{S.6}$$

holds since $MAX_{j,\bar{D}_i^{(u)}}$ is the maximum cost reduction once an affected community is removed from a cluster. We known that $\bar{\mathcal{R}}_{is}^*(\delta_2+\delta_3) \geq \bar{\mathcal{R}}_{is}^*(\delta_2)$ holds trivially since δ_2 is a subset of $\delta_2+\delta_3$, then (S.6) can be written as

$$\bar{\mathcal{R}}_{is}^*(\delta_2+\delta_3) \geq \bar{\mathcal{R}}_{is}^*(\delta_1+\delta_2) - |\delta_1|\cdot MAX_{j,\bar{D}_i^{(u)}}. \tag{S.7}$$

$\tilde{\mathcal{R}}_{is}^{(u)}(\delta_1+\delta_2) = \bar{\mathcal{R}}_{is}^*(\delta_1+\delta_2)$ holds since both sides denote the optimal cost for visiting δ_1 and δ_2, thus (S.5) contradicts (S.7), a globally optimal solution must not be excluded by (8).

In conclusion, property (iii) and (iv) holds for (8), (8) is a valid Benders optimality cut. □

Theorem 3. *The logic-based Benders feasibility cut* (10.1) *and* (10.2) *is valid.*

Proof. We only prove that property (i) and (ii) holds for (10.1), and the validity of (10.2) can be proved similarly. We first prove property (i). In a particular iteration u, $\mathbb{B}_i^{s(u)}\left[\tilde{D}_i^{s(u)}(\tilde{t}_{ij}^{s(u)}), \tilde{T}_i^{s(u)}(\tilde{t}_{ii'j}^{s(u)}), Q\right]$ is the minimum vehicle requirement with the given transportation decisions $\tilde{t}_{ij}^{s(u)}$, $\tilde{t}_{ii'j}^{s(u)}$ and vehicle capacity Q. Let $\sum_{h\in H} w_i^{sh(u)}$ be the number of vehicles prepared at facility i. If (10.1) is added, then $\sum_{h\in H} w_i^{sh(u)} < \mathbb{B}_i^{s(u)}\left[\tilde{D}_i^{s(u)}(\tilde{t}_{ij}^{s(u)}), \tilde{T}_i^{s(u)}(\tilde{t}_{ii'j}^{s(u)}), Q\right]$. Assume that the vehicle decisions and transportation decisions remain unchanged, we got inequality directly from (10.1) by substituting $w_i^{sh(u)}$, $\tilde{t}_{ij}^{s(u)}$, and $\tilde{t}_{ii'j}^{s(u)}$ to (10.1) as $\sum_{h\in H} w_i^{sh(u)} \geq \mathbb{B}_i^{s(u)}\left[\tilde{D}_i^{s(u)}(\tilde{t}_{ij}^{s(u)}), \tilde{T}_i^{s(u)}(\tilde{t}_{ii'j}^{s(u)}), Q\right] - \sum_{j\in\tilde{D}_i^{s(u)}} W_{ij}^s - \sum_{i'\in\tilde{T}_i^{s(u)}} W_{ii'}^s$ where $\sum_{j\in\tilde{D}_i^{s(u)}} W_{ij}^s = \sum_{i'\in\tilde{T}_i^{s(u)}} W_{ii'}^s = 0$. Thus, property (i) is proved by contradiction.

Then we prove property (ii) by contradiction. Let $w_i^{sh(v)}$ be vehicle decision that in iteration v which is after iteration u, $(v > u)$. Assume that $w_i^{sh(v)}$ is globally feasible for RASP, and was cut off by the cut plane generated in iteration u, we have the following inequality

$$\sum_{h\in H} w_i^{sh(v)} < \mathbb{B}_i^{s(u)}\left[\tilde{D}_i^{s(u)}(\tilde{t}_{ij}^{s(u)}), \tilde{T}_i^{s(u)}(\tilde{t}_{ii'j}^{s(u)}), Q\right] - \sum_{j\in\tilde{D}_i^{s(u)}} W_{ij}^s - \sum_{i'\in\tilde{T}_i^{s(u)}} W_{ii'}^s \tag{S.8}$$

The definition for W_{ij}^s and $W_{ii'}^s$ are same as (10.3) and (10.4). Let $\delta_1\cup\delta_2\cup\delta_3$ be customers and facilities that assigned to facility i in either iteration u or v. In the following, we let

- $\delta_1 = \{j\in\tilde{D}_i^{s(u)} | \tilde{t}_{ij}^{s(v)} \leq \tilde{t}_{ij}^{s(u)} - 1\} \cup \{i'\in\tilde{D}_i^{s(u)} | \sum_{j\in D}\tilde{t}_{ii'j}^{s(v)} \leq \sum_{j\in D}\tilde{t}_{ii'j}^{s(u)} - 1\}$ be customers and facilities that accept a lower quantity of relief items in iteration v than in iteration u. δ_1 also include facilities that are removed from i in iteration v.
- $\delta_2 = \{j\in\tilde{T}_s^{s(u)} | \tilde{t}_{ij}^{s(v)} \geq \tilde{t}_{ij}^{s(u)}\} \cup \{i'\in\tilde{T}_s^{s(u)} | \sum_{j\in D}\tilde{t}_{ii'j}^{s(v)} \geq \sum_{j\in D}\tilde{t}_{ii'j}^{s(u)}\}$ be customers and facilities that accept a higher or equal quantity of relief items in iteration v than in iteration u.
- $\delta_3 = \{j | j\in\tilde{D}_s^{s(v)}\setminus\tilde{D}_i^{s(u)}\} \cup \{i' | i'\in\tilde{T}_s^{s(v)}\setminus\tilde{T}_s^{s(u)}\}$ be customers and facilities that are assigned to facility i in iteration v but not in iteration u.

Since the contributions to (S.8) from δ_2, δ_3 are 0, from (S.8) and definition for $\delta_1, \delta_2, \delta_3$ we have

$$\sum_{h\in H} w_i^{sh(v)} < \mathbb{B}_i^{s(u)}\left[\tilde{D}_i^{s(u)}(\tilde{t}_{ij}^{s(u)}), \tilde{T}_i^{s(u)}(\tilde{t}_{ii'j}^{s(u)}), Q\right] - |\delta_1| \tag{S.9}$$

Let $\mathbb{B}_i^{s*}(\tilde{D}_i^{s(u)}\cup\tilde{T}_i^{s(u)}\cup\tilde{D}_i^{s(v)}\cup\tilde{T}_i^{s(v)}, Q)$ be the minimum number of vehicle for $\tilde{D}_i^{s(u)}\cup\tilde{T}_i^{s(u)}\cup\tilde{D}_i^{s(v)}\cup\tilde{T}_i^{s(v)}$, the relief item acceptance quantity for customer $j\in\tilde{D}_i^{s(u)}\cup\tilde{D}_i^{s(v)}$ be $\max\{\tilde{t}_{ij}^{s(u)}, \tilde{t}_{ij}^{s(v)}\}$, the relief item acceptance quantity for facility $i'\in\tilde{T}_i^{s(u)}\cup\tilde{T}_i^{s(v)}$ be $\max\{\sum_{j\in D}\tilde{t}_{ii'j}^{s(u)}, \sum_{j\in D}\tilde{t}_{ii'j}^{s(v)}\}$, then

$$\begin{aligned}&\mathbb{B}_i^{s*}\left[\tilde{D}_i^{s(u)}\cup\tilde{T}_i^{s(u)}\cup\tilde{D}_i^{s(v)}\cup\tilde{T}_i^{s(v)}, Q\right] = \mathbb{B}_i^{s*}\left[\delta_1(\tilde{t}_{ij}^{s(u)}, \tilde{t}_{ii'j}^{s(u)}), \delta_2(\tilde{t}_{ij}^{s(v)}, \tilde{t}_{ii'j}^{s(v)}), \delta_3(\tilde{t}_{ij}^{s(v)}, \tilde{t}_{ii'j}^{s(v)}), Q\right] \\ \geq& \mathbb{B}_i^{s*}\left[\delta_1(\tilde{t}_{ij}^{s(u)}, \tilde{t}_{ii'j}^{s(u)}), \delta_2(\tilde{t}_{ij}^{s(u)}, \tilde{t}_{ii'j}^{s(u)}), Q\right] = \mathbb{B}_i^{s(u)}\left[\tilde{D}_i^{s(u)}(\tilde{t}_{ij}^{s(u)}), \tilde{T}_i^{s(u)}(\tilde{t}_{ii'j}^{s(u)}), Q\right]\end{aligned} \tag{S.10}$$

due to additional customers and facilities in δ_3, and additional transportation quantity to customers and facilities is δ_2. If we assign $\tilde{D}_i^{s(v)}$ and $\tilde{T}_i^{s(v)}$ with transportation decisions $\tilde{t}_{ij}^{s(v)}$ and $\tilde{t}_{ii'j}^{s(v)}$ to $\sum_{h\in H} w_i^{sh(v)}$ vehicles, and assign each customer j or facility i' in δ_1 associated with acceptance quantity $\tilde{t}_{ij}^{s(u)} - \tilde{t}_{ij}^{s(v)}$ or $\sum_{j\in D}(\tilde{t}_{ii'j}^{s(u)} - \tilde{t}_{ii'j}^{s(v)})$ to one vehicle, $|\delta_1|$ in all, then

$$\sum_{h\in H} w_i^{sh(v)} + |\delta_1| \geq \mathbb{B}_i^{s*}\left[\tilde{D}_i^{s(u)} \cup \tilde{T}_i^{s(u)} \cup \tilde{D}_i^{s(v)} \cup \tilde{T}_i^{s(v)}, Q\right] \tag{S.11}$$

Since (S.10) and (S.11) result in

$$\sum_{h\in H} w_i^{sh(v)} + |\delta_1| \geq \mathbb{B}_i^{s(u)}\left[\tilde{D}_i^{s(u)}(\tilde{t}_{ij}^{s(u)}), \tilde{T}_i^{s(u)}(\tilde{t}_{ii'j}^{s(u)}), Q\right] \tag{S.12}$$

which contradicts (S.9), the assumption that globally feasible $w_i^{sh(v)}$ for RASP can be cut off by cut plane generated in iteration u is not true. In other words, any globally feasible solution will not be excluded by (10.1).

In conclusion, property (i) and (ii) holds for (10.1) and (10.2) , (10.1) and (10.2) is a valid logic-based Benders feasibility cut. □

Theorem 4. *The logic-based Benders optimality cut (11.1)–(11.3) is valid.*

Proof. We first prove property (iii) by contradiction. Let $(\tilde{\mathcal{R}'}_{is}^{(u)}, \tilde{t}_{ij}^{s(u)}, \tilde{t}_{ii'j}^{s(u)})$ be an RASP solution that generates cut (11.1)–(11.3) (imply that $\tilde{\mathcal{R}'}_{is}^{(u)} < \hat{\mathcal{R}'}_{is}^{(u)}$, since this is the condition for optimality cut generating, see Line 25 in Algorithm 3). Substituting $(\tilde{\mathcal{R}'}_{is}^{(u)}, \tilde{t}_{ij}^{s(u)}, \tilde{t}_{ii'j}^{s(u)})$ to (11.1) results in $\tilde{\mathcal{R}'}_{is}^{(u)} \geq \hat{\mathcal{R}'}_{is}^{(u)}$ (both (11.2) and (11.3) equals to 0), which is contradicted to $\tilde{\mathcal{R}'}_{is}^{(u)} < \hat{\mathcal{R}'}_{is}^{(u)}$. Therefore, (11.1)–(11.3) must exclude current solution $(\tilde{\mathcal{R}'}_{is}^{(u)}, \tilde{t}_{ij}^{s(u)}, \tilde{t}_{ii'j}^{s(u)})$. So property (iii) holds for (11.1)–(11.3), current RASP solution will be excluded by (11.1)–(11.3) if it is not globally optimal with a fixed PMP solution.

Then we prove property (iv) by contradiction. Let transportation decision $\tilde{t}_{ij}^{s*}$, $\tilde{t}_{ii'j}^{s*}$, and corresponding recourse cost $\tilde{\mathcal{R}'}_{is}^{*}$ be the optimal resource allocation solution that obtained with a given PMP solution. Assume that $(\tilde{t}_{ij}^{s*}, \tilde{t}_{ii'j}^{s*}, \tilde{\mathcal{R}'}_{is}^{*})$ violates (11.1), then for each pair of (i, s) we have

$$\tilde{\mathcal{R}'}_{is}^{*} < \hat{\mathcal{R}'}_{is}^{(u)} - \sum_{j\in \tilde{D}_i^s} R_{ij}^s - \sum_{i'\in \tilde{T}_i^s} R_{ii'}^s \tag{S.13.1}$$

$$R_{ij}^s = \begin{cases} MAX_{\tilde{C}_{is}^{(u)},i,j}, & \text{if } \tilde{t}_{ij}^{s*} \leq \tilde{t}_{ij}^{s(u)} - 1 \\ 0, & \text{otherwise} \end{cases}, \qquad \forall j \in \tilde{D}_i^{s(u)} \tag{S.13.2}$$

$$R_{ii'}^s = \begin{cases} MAX_{\tilde{C}_{is}^{(u)},i,i'}, & \text{if } \sum_{j\in D} \tilde{t}_{ii'j}^{s*} \leq \sum_{j\in D} \tilde{t}_{ii'j}^{s(u)} - 1 \\ 0, & \text{otherwise} \end{cases}, \qquad \forall i' \in \tilde{T}_i^{s(u)} \tag{S.13.3}$$

There are four cases as follows The relationship between transportation decisions in iteration u $(\tilde{t}_{ij}^{s(u)}, \tilde{t}_{ii'j}^{s(u)})$ and in optimal solution $(\tilde{t}_{ij}^{s*}, \tilde{t}_{ii'j}^{s*})$ belonging to one of the following cases

Case 1: If $\tilde{t}_{ij}^{s*} \geq \tilde{t}_{ij}^{s(u)}, \forall j \in \tilde{D}_i^{s(u)}$ and $\sum_{j\in D} \tilde{t}_{ii'j}^{s*} \geq \sum_{j\in D} \tilde{t}_{ii'j}^{s(u)}, \forall i' \in \tilde{T}_i^{s(u)}$, we have $\tilde{\mathcal{R}'}_{is}^{*} < \tilde{\mathcal{R}'}_{is}^{(u)}$ from (S.13.1). But this cannot be true since in the optimal solution more relief items are transported than iteration u, which implies that $\tilde{\mathcal{R}'}_{is}^{*} \geq \tilde{\mathcal{R}'}_{is}^{(u)}$, in this case.

Case 2: If $\exists j \in \tilde{D}_i^{s(u)}, \tilde{t}_{ij}^{s*} \leq \tilde{t}_{ij}^{s(u)} - 1$ and $\exists i' \in \tilde{T}_i^{s(u)}, \sum_{j\in D} \tilde{t}_{ii'j}^{s*} \leq \sum_{j\in D} \tilde{t}_{ii'j}^{s(u)} - 1$, the proof is more involved. Consider that we increase the transportation quantity in the optimal solution to that of iteration u, without loss of generality, only horizontal coordination is considered, there are three subcases as follows for every $i' \in \tilde{T}_i^{s(u)}$ as

Subcase (1): $\sum_{j\in D} \tilde{t}_{ii'j}^{s*} = 0$, means i' is not visited in the optimal solution. The additional cost for visit i' is at most $\max\left\{2\cdot \max_{h\in H} c_{ii'}^h, 2\cdot \max_{p,q\in \tilde{C}_{is}^{(u)},h\in H}\left\{c_{pq}^h\right\}\right\}$ in which the first term denotes the cost of dispatch another vehicle while the second term denotes the cost of inserting i' to an existing route.

Subcase (2): $\sum_{j\in D} \tilde{t}_{ii'j}^{s*} > 0$ and the total demand after the quantity increase can be met with route in the optimal solution, the cost increase is 0.

Subcase (3): $\sum_{j\in D} \tilde{t}_{ii'j}^{s*} > 0$ and the total demand after the quantity increase cannot be met with the route in the optimal solution, in another word, an additional vehicle has to be used. The cost increase is $max_{h\in H} c_{ii'}^h$.

Hence from the above cases,

$$\begin{aligned}
&\tilde{\mathcal{R}'}_{is}^{*} + \sum_{j\in \tilde{D}_i^s} R_{ij}^s + \sum_{i'\in \tilde{T}_i^s} R_{ii'}^s \\
=&\tilde{\mathcal{R}'}_{is}^{*} + \sum_{k\in case_1} MAX_{\tilde{C}_{is}^{(u)},i,k} + \sum_{k\in case_2} 0 + \sum_{k\in case_3} MAX_{\tilde{C}_{is}^{(u)},i,k} \\
\geq&\tilde{\mathcal{R}'}_{is}^{*} + \sum_{k\in case_1} MAX_{\tilde{C}_{is}^{(u)},i,k} + \sum_{k\in case_2} 0 + \sum_{k\in case_3} max_{h\in H} c_{ik}^h \\
\geq&\tilde{\mathcal{R}'}_{is}^{(u)}
\end{aligned} \tag{S.14}$$

in which $k \in \tilde{D}_i^{s(u)} \cup \tilde{T}_i^{s(u)}$. (S.14) can be rearranged to $\tilde{\mathcal{R}'}_{is}^{*} \geq \tilde{\mathcal{R}'}_{is}^{(u)} - \sum_{j\in \tilde{D}_i^s} R_{ij}^s - \sum_{i'\in \tilde{T}_i^s} R_{ii'}^s$, thus (S.14) contradicts (S.13.1), the optimal solution cannot be exclude by (S.13.1).

Case 3: If $\tilde{t}_{ij}^{s*} \geq \tilde{t}_{ij}^{s(u)}, \forall j \in \tilde{D}_i^{s(u)}$ and $\exists i' \in \tilde{T}_i^{s(u)}, \sum_{j \in D} \tilde{t}_{ii'j}^{s*} \leq \sum_{j \in D} \tilde{t}_{ii'j}^{s(u)} - 1$, we have results similar to case 2 as the contribution of direct transportation to (S.13.1) is zero.

Case 4: If $\exists j \in \tilde{D}_i^{s(u)}, \tilde{t}_{ij}^{s*} \leq \tilde{t}_{ij}^{s(u)} - 1$ and $\sum_{j \in D} \tilde{t}_{ii'j}^{s*} \geq \sum_{j \in D} \tilde{t}_{ii'j}^{s(u)}, \forall i' \in \tilde{T}_i^{s(u)}$, results similar to case 3 can be obtained due to symmetry.

Therefore, the optimal solution must not be cut off by (11.1)–(11.3).

In conclusion, property (iii) and (iv) holds for (11.1)–(11.3), (11.1)–(11.3) is a valid logic-based Benders optimality cut. □

S.4.1. Relationships among the cuts

The logic-based Benders cuts in our study are inspired by the cuts in the literature (Fazel-Zarandi and Beck, 2012, Hooker, 2007, Fachini and Armentano, 2020), but are tailored for our problem with some improvement. The cuts in our study can be generally formulated as $\mu \geq \hat{\mu} + \Delta\,(X - \bar{X})$, where μ is a variable in the upper-echelon problem (master problem), $\hat{\mu}$ is a value yielded by the lower-echelon problem (subproblem), X is set of upper-echelon variables, $\bar{X}$ is the values of the upper-echelon variables fed to the lower-echelon problem, and Δ is the cutting coefficient (gradient) provided by the lower-echelon problem for the upper-echelon problem. In other words, among the parameters in $\mu \geq \hat{\mu} + \Delta\,(X - \bar{X})$, μ and X are variables that need to be optimized, $\bar{X}$ is the value flow from the master problem to the subproblem, and $\hat{\mu}$ and Δ are pair of values flow from the subproblem to the master problem. Our cuts differ from ones by Fazel-Zarandi and Beck (2012), Hooker (2007), Fachini and Armentano (2020) in terms of the problems handled (μ, X, and $\bar{X}$) and the information from the subproblem to the master problem ($\hat{\mu}$ and Δ), as Table S3 shows.

Table S3: Relationships among the logic-based Benders cuts

Type	RASP's cuts	PMP's cuts	Reference	Reference
Feasibility cuts	(10.1)–(10.4)	← (7.1) and (7.2)	← Fazel-Zarandi and Beck (2012)	← Hooker (2007)
Problem	Location-routing	Allocation	Facility location	Scheduling
$\hat{\mu}$	MIP (VDSP)	MIP (bin packing)	Algorithmic	Constraint Programming
Δ	1 × number of communities	1 (number of vehicle)	1 (number of vehicle)	Processing time
Optimality cuts	(11.1)–(11.3)	← (8) and (9)	← Fachini and Armentano (2020)	
Problem	Location-routing	Allocation	vehicle routing	
$\hat{\mu}$	MIP (VDSP)	MIP (RASP)	MIP (TSP with time windows)	
Δ	See (11.1)–(11.3)	See (8) and (9)	See (24) in their paper	

Two streams are presented in Table S3: the origin of the feasibility cuts and the origin of the optimality cuts. The feasibility cuts for the PMP is inspired by Fazel-Zarandi and Beck (2012) for facility location problem, which is in turn inspired by Hooker (2007) for scheduling problem. The latter two are both deterministic problems, whereas our problem is stochastic. To calculate $\hat{\mu}$, Hooker (2007) uses constraint programming, Fazel-Zarandi and Beck (2012) runs a sequence of first-fit decreasing heuristic for bin packing problem, while we solve a MIP to directly obtain the solution of the bin packing problem. (10.1)–(10.4) shows similar structure with (7.1) and (7.2). They both require bin packing problems to be solved. However, the "items" in the bin packing problems are distinct: the size of items in (7.1) and (7.2) are the demand of affected communities, while the size of items in (10.1)–(10.4) is determined by the allocation decisions. Another stream is the optimality cuts. Though the optimality cut for the PMP is inspired by Fachini and Armentano (2020), and (11.1)–(11.3) shares similarity with (8) and (9), the problems handled by us and Fachini and Armentano (2020) are distinct. They solve a deterministic single-depot heterogeneous fleet vehicle routing problem with time windows.

S.5. Implementing extensive form and NLBBD

The SPRP(-VI), which contains nonlinear constraints (2.19) and (2.20), is solved with Gurobi. (2.20) defines auxiliary decision variables $Q_p^{(s,i)}$ that are derived from t_{ij}^s and $t_{ii'j}^s$. Since t_{ij}^s and $t_{ii'j}^s$ are integer, $Q_p^{(s,i)}$ are also integer. Thus, (2.19) contains multiplication of integer and binary ($z_{pq}^{(s,i,h)}$) variables, which can be linearized manually or automatically by setting Gurobi's `"PreQLinearize"` parameter to `"1"`. The original SPRP(-VI) is an MIQCP, for which Gurobi solves each node on the branch-and-bound tree with *barrier method*. Since there is no nonlinear term in models of the NLBBD method, Gurobi solves nodes on the branch-and-bound tree for the models in the NLBBD with *simplex method*. For fairness of comparison, we let Gurobi linearize the SPRP(-VI), such that the internal behavior of the solver can be consistent between SPRP(-VI) and

NLBBD.

Algorithm 1: Solving PMP with NLBBD-C

Data: PMP_0 generated from (4.1) to (4.7), (4.10); time limit in seconds (*Lmt.*); relative tolerance (*Tol.*).
Result: Optimal solution $PMP*, RASP*$, and $VDSP*$ (or infeasibility proof).

def `NLBBD-C`(PMP_0, *Lmt.*, *Tol.*)**:**
  $Gap \leftarrow +\infty$, $BestMasterUB \leftarrow +\infty$, $u \leftarrow 0$;
  while $Gap > Tol.$ **and** $Runtime \leq Lmt.$ **:**
    $RASP* \leftarrow$ None, $VDSP* \leftarrow$ None;
    Solve PMP_u with an MIP solver;
    if *PMP_u is optimal* **:**
      Let $\bar{y}_{ik}^{(u)}, \bar{l}_i^{(u)}, \bar{x}_{ij}^{(u)}$, and $\bar{w}_i^{h(u)}$ denote the solution of PMP_u ;
      $RASP, VDSP \leftarrow$ `nestedBenders`$(\bar{y}_{ik}^{(u)}, \bar{l}_i^{(u)}, \bar{x}_{ij}^{(u)}, \bar{w}_i^{h(u)})$; // Call Algorithm 3
      if *RASP is optimal* **:**
        $MasterLB \leftarrow PMP_u\_obj$;
        // $\sum_{i\in T}\sum_{k\in K} c^k \bar{y}_{ik}^{(u)} + \sum_{i\in T} c\, \bar{l}_i^{(u)} + \sum_{i\in T}\sum_{h\in H} c_1^h \bar{w}_i^{h(u)} + \mathbb{E}_{s\in S}\big(\sum_{i\in T}\bar{\mathcal{R}}_{is}^{(u)}\big)$
        $MasterUB \leftarrow PMP_u\_obj - PMP_u\_recourse + RASP\_obj$;
        // $\sum_{i\in T}\sum_{k\in K} c^k \bar{y}_{ik}^{(u)} + \sum_{i\in T} c\, \bar{l}_i^{(u)} + \sum_{i\in T}\sum_{h\in H} c_1^h \bar{w}_i^{h(u)} + \mathbb{E}_{s\in S}\big(\sum_{i\in T}\tilde{\mathcal{R}}_{is}\big)$
        if $MasterUB < BestMasterUB$ **:**
          $BestMasterUB \leftarrow MasterUB$;
        $Gap = |BestMasterUB - MasterLB| / |BestMasterUB|$;
        if $Gap > Tol.$ **:**
          for $i, s \in T \times S$ **:**
            if $\bar{\mathcal{R}}_{is}^{(u)} < \tilde{\mathcal{R}}_{is}$ **:**
              $PMP_u \leftarrow PMP_u$ + (8); // Add Benders optimality cut to PMP
        else:
          $PMP* \leftarrow PMP_u$, $RASP* \leftarrow RASP$, $VDSP* \leftarrow VDSP$; // PMP is optimized
      else:
        $PMP_u \leftarrow PMP_u +$ (7.1) and (7.2); // Add Benders feasibility cut to PMP
    else:
      return *PMP is infeasible*; // PMP is infeasible
    $u \leftarrow u + 1$;
  return *PMP*, RASP*, VDSP*, Gap.*

Two versions of NLBBD, the classical NLBBD-C (Algorithm 1) and the modern NLBBD-B&C (Algorithm 2), are implemented. The difference between NLBBD-C and NLBBD-B&C is that NLBBD-C builds a master problem branch-and-bound tree from scratch each time a Benders cut is generated, while NLBBD-B&C generates Benders cut and adds it to the single branch-and-bound tree once an integer incumbent solution is found. The differences between NLBBD-C and NLBBD-B&C are underlined in Algorithm 2. For simplicity, consider a Benders master problem and a Benders subproblem associated with it (rather than our nested version). The classic Benders method solves the master problem to optimum and feeds the optimal solution to the subproblem, then generates Benders cuts. After a cut is generated, the classic Benders method rebuilds a master problem model with the cut from scratch. The modern branch-and-check method solves subproblems when an integer incumbent solution is found. Then a Benders cut is added to the branch-and-bound tree of the master problem as a `"LazyCut"`. The branch-and-check relies on the `"callback"` function provided by state-of-the-art solvers. `"callback"` provides the function of adding user cuts or lazy cuts to the master model when an incumbent solution is found. Unfortunately, both Gurobi and CPLEX do not support adding cuts with auxiliary decision variables, i.e., (10.1), (10.2) and (11.1). Therefore, it is impossible to implement branch-and-check for the nested RASP problem. We rebuild the RASP model with Benders cuts generated by VDSPs. This works well since it usually takes less than one second to build and optimize a RASP. The NLBBD-C can be time-consuming for building up master models and potentially results in a repetitive search on the branch-and-bound tree. The NLBBD-C can be inefficient when the master problem is harder relative to the subproblem. On the other hand, since the NLBBD-B&C has to solve subproblems each time an integer feasible master problem solution is found, it solves subproblems more frequently than the NLBBD-C. The choice of NLBBD-C and NLBBD-B&C depends on the relative difficulty between the master problem and the subproblem.

Algorithm 2: Solving PMP with NLBBD-B&C

Data: PMP generated from (4.1) to (4.7), (4.10); time limit in seconds ($Lmt.$); relative tolerance ($Tol.$).
Result: Optimal solution $PMP*, RASP*$, and $VDSP*$ (or infeasibility proof).

def `NLBBD-B&C`(PMP, $Lmt.$, $Tol.$):
  $Gap \leftarrow +\infty$, $BestMasterUB \leftarrow +\infty$, $PMP* \leftarrow$ None, $RASP* \leftarrow$ None, $VDSP* \leftarrow$ None;
  Solve PMP using an MIP solver with $Lmt.$ and $Tol.$ as parameters, and `cb` as callback function;
  return $PMP*$, $RASP*$, $VDSP*$, Gap.

def `cb()`: // callback function
  if *an incumbent integer feasible solution is found for PMP* **:**
    Let $\bar{y}_{ik}, \bar{l}_i, \bar{x}_{ij}$, and $\bar{w}_i^h$ denote the incumbent solution of PMP;
    $RASP, VDSP \leftarrow$ `nestedBenders`$(\bar{y}_{ik}, \bar{l}_i, \bar{x}_{ij}, \bar{w}_i^h)$; // Call Algorithm 3
    if *RASP is optimal* **:**
      $MasterLB \leftarrow PMP(LP)\_obj$; ; // Objective value of PMP LP relaxation
      $MasterUB \leftarrow PMP\_obj - PMP\_recourse + RASP\_obj$;
      // $\sum_{i\in T}\sum_{k\in K} c^k \bar{y}_{ik} + \sum_{i\in T} c\,\bar{l}_i + \sum_{i\in T}\sum_{h\in H} c_1^h \bar{w}_i^h + \mathbb{E}_{s\in S}\big(\sum_{i\in T}\tilde{\mathcal{R}}_{is}\big)$
      if $MasterUB < BestMasterUB$ **:**
        $BestMasterUB \leftarrow MasterUB$;
      $Gap = |BestMasterUB - MasterLB|/|BestMasterUB|$;
      if $Gap > Tol.$ **:**
        for $i, s \in T \times S$ **:**
          if $\bar{\mathcal{R}}_{is} < \tilde{\mathcal{R}}_{is}$ **:**
            $PMP \leftarrow PMP + (8)$; // Add Benders optimality cut to PMP as LazyCut
      else:
        $PMP* \leftarrow PMP$, $RASP* \leftarrow RASP$, $VDSP* \leftarrow VDSP$; // PMP is optimized
    else:
      $PMP \leftarrow PMP+$ (7.1) and (7.2); // Add Benders feasibility cut to PMP as LazyCut

Algorithm 3: Solving nested RASP with LBBD

Data: PMP_u fed by Algorithm 1; time limit ($Lmt'.$); relative tolerance ($Tol'.$).
Result: Optimal $RASP*$ and $VDSP*$ with the given PMP_u decisions (or infeasibility proof).

def `nestedBenders`$(y_{ik}, l_i, x_{ij}, w_i^h)$:
  Construct $RASP_0$ from (5.1) to (5.5), (5.8), (5.9) with PMP_u decisions $\bar{y}_{ik}^{(u)}, \bar{l}_i^{(u)}, \bar{x}_{ij}^{(u)}$, and $\bar{w}_i^{h(u)}$;
  $SubGap \leftarrow +\infty$, $BestSubUB \leftarrow +\infty$, $v \leftarrow 0$;
  while $SubGap > Tol'.$ **and** $Runtime \leq Lmt'.$ **:**
    $RASP* =$ None, $SolvedVDSP \leftarrow$ `dict()`;
    Solve $RASP_v$ with an MIP solver;
    if *$RASP_v$ is optimal* **:**
      $VDSPsAllFeasible =$ True;
      for $i, s \in T \times S$ **:**
        Construct $VDSP_{is}$ form (6.1)–(6.10) with $RASP_v$ decisions $\tilde{t}_{ij}^{s(v)}, \tilde{t}_{ii'j}^{s(v)}$, and $\tilde{w}_i^{sh(v)}$;
        Solve $VDSP_{is}$ with an MIP solver;
        if *$VDSP_{is}$ is infeasible* **:**
          $VDSPsAllFeasible =$ False;
          $RASP_v \leftarrow RASP_v +$ (10.1) to (10.4); // Add Benders feasibility cut to RASP
        else:
          $SolvedVDSP[i, s] \leftarrow \mathrm{VDSP}_{is}$;
      if *VDSPsAllFeasible* **:**
        $SubLB \leftarrow \mathbb{E}_{s\in S}\big(\sum_{i\in T}\tilde{\mathcal{R}}_{is}^{(v)}\big)$;
        $SubUB \leftarrow \mathbb{E}_{s\in S}\big(\sum_{i\in T}\hat{\mathcal{R}}_{is}\big)$;
        if $BestSubUB > SubUB$ **:**
          $BestSubUB \leftarrow SubUB$;
        $SubGap = |BestSubUB - SubLB|/|BestSubUB|$;
        if $SubGap > Tol'.$ **:**
          for $i, s \in T \times S$ **:**
            if $\tilde{\mathcal{R}}_{is}^{(v)} < \hat{\mathcal{R}}_{is}$ **:**
              $RASP_v \leftarrow RASP_v +$ (11.1) to (11.3); // Add Benders optimality cut to RASP
        else:
          $RASP* \leftarrow RASP_v$, $VDSP* \leftarrow SolvedVDSP$; // RASP is optimized
    else:
      return *RASP is infeasible*; // RASP is infeasible
    $v \leftarrow v + 1$;
  return $RASP*$, $VDSP*$.

S.6. Instance parameters

We determine the vehicle- and facility-related fixed and variable costs as follows. Ni et al. (2018) estimated the proportion of facility fixed cost, relief item cost, transportation cost of road links, and transportation cost of helicopter links at $c^1 : c : c_u^1 : c_u^2 = 200 : 2.8 : 1 : 10$. The price for a basic 72-hour emergency preparedness kit is obtained from the website of American Red Cross website , which is roughly \$80. We add 25% operational and holding cost to the \$80, i.e., the resulting unit relief item cost is \$100. It is assumed that $c^1 : c^2 = 7 : 2.5$. All the costs can be calculated proportionally, and they are then rounded to the greatest place value. That is $c^1 = 7000, c^2 = 2500, c = 100, c_u^1 = 40$, and $c_u^2 = 400$.

Table S4: Instance parameters

Parameters	Description	Value
Variable parameters		
$\lvert T \rvert$	Number of relief facility candidate	{3, 5, 7}
$\lvert D \rvert$	Number of affected community	{10, 20, 40}
$\lvert S \rvert$	Number of scenario	{2, 4}
$\mathcal{C}$	Uniform or clustered	{0, 1}
$\mathcal{T}$	Flood/storm or earthquake	{0, 1}
Deterministic parameters		
p_j	Population of $j \in D$	$IU(50, 300)$
ss_s	Severity score of scenario $s \in S$	$ss_{IV} \sim U(0.0, 0.4), ss_{III} \sim U(0.2, 0.6),$ $ss_{II} \sim U(0.4, 0.8), ss_I \sim U(0.6, 1.0)$
m_j^s	Demand of $j \in D$ in scenario $s \in S$	$\lceil 1 \times p_j \times ss_s \rceil$
vs_s	Vulnerability score of scenario $s \in S$	$vs_{IV} = 0.10, vs_{III} = 0.25,\ vs_{II} = 0.50, vs_I = 1.00$
e_j^s	Whether $j \in D$ is isolated in scenario $s \in S$	$B(F(d_{oj}) \times vs_s)$
K	Types of relief facility	$Q^1 = 4000, c^1 = 7000;\ Q^2 = 1000, c^2 = 2500$
c	Unit relief item cost	100
$H \setminus H'$	Types of normal vehicle	$n^1 = 40, Q = 300, c_1^1 = 0, c_u^1 = 40$
H'	Types of special vehicle	$n^2 = 40, Q = 300, c_1^2 = 0, c_u^2 = 400$
c_{pq}^h	Traveling cost of $(p, q) \in A$ by $h \in H$	$c_u^h \times d_{pq}$

[1] $IU(\cdot)$ denotes integer uniform distribution; $U(\cdot)$ denotes uniform distribution; d_{pq} denotes the Euclidean distance between $p \in N$ and $q \in N$; $F(d_{oj})$ be a step function $F(d_{oj}) = 0.80\chi_{[0,50)}(d_{oj}) + 0.50\chi_{[50,200)}(d_{oj}) + 0.20\chi_{[200,+\infty)}(d_{oj})$ where $\chi_{\mathsf{R}}(d_{oj}) = 1$ if $d_{oj} \in \mathsf{R}$, $\chi_{\mathsf{R}}(d_{oj}) = 0$ if $d_{oj} \notin \mathsf{R}$; $B(\cdot)$ denotes Bernoulli distribution.