Dynamic emergency medical services network design: A novel probabilistic envelope constrained stochastic program and decomposition scheme

C. Peng, E. Delage,
J. Li

G–2018–57

July 2018

GERAD HEC Montréal
3000, chemin de la Côte-Sainte-Catherine
Montréal (Québec) Canada H3T 2A7

Tél. : 514 340-6053
Téléc. : 514 340-5665
info@gerad.ca
www.gerad.ca

Dynamic emergency medical services network design: A novel probabilistic envelope constrained stochastic program and decomposition scheme

Chun Peng [a,b]

Erick Delage [a]

Jinlin Li [b]

[a] *GERAD & HEC Montréal, Montréal (Québec), Canada, H3T 2A7*

[b] *Beijing Institute of Technology, Beijing 100081, China*

chun.peng@hec.ca
erick.delage@hec.ca
jinlinli@bit.edu.cn

July 2018
Les Cahiers du GERAD
G–2018–57

Abstract: This paper considers a dynamic Emergency Medical Services (EMS) network design problem and introduces two novel two-stage stochastic programming formulations that account for uncertainty about emergency demand. Similarly to some recent work on emergency demand coverage, we consider both a constraint on the probability of covering the realized emergency demand and minimize the expected cost of doing so, yet unlike other formulations, our optimization models account for the dynamics throughout a full day of operations and allow the EMS system managers to control the degradation of coverage under the more severe scenarios. In order to do so, we present both a two-stage chance-constrained stochastic programming formulation and a variant of this model, which employs probabilistic envelope constraints. These give rise to large mixed-integer programs, which can be tackled directly or using a conservative approximation scheme. To improve the numerical efficiency of the exact approach, we implement the Branch-and-Benders-Cut method, which improves significantly the solution time when compared to a state-of-the art Branch-and-Bound algorithm proposed in the recent literature for a simplified version of these problems. Finally, a practical study is conducted using historical data from Northern Ireland Ambulance Service Health and Social Care Trust and sheds some light on optimal EMS network configuration for this region and necessary trade-offs that must be made between emergency demand coverage and expected cost. These insights are confirmed through an out-of-sample analysis.

Keywords: Two-stage chance-constrained stochastic programming, probabilistic envelope constraint, Emergency Medical Services, Branch-and-Benders-Cut, dynamic ambulance location, time-dependent uncertainty

Acknowledgments: This research was partially supported by the Natural Sciences and Engineering Research Council of Canada [Grant RGPIN-2016-05208], the Canada Research Chair program [Grant 950-230057], the Groupe d'études et de recherche en analyse des décisions, and the National Natural Science Foundation of China [71432002, 91746210]. The authors also thank Northern Ireland Ambulance Service Health & Social Care Trust for providing the data used in our case study, and thank Carlos A. Zetina for his support during the implementation of the Branch-and-Benders-Cut method.

1 Introduction

Emergency Medical Services (EMS) network design plays an important role in providing adequate and high-quality medical services for the public to answer as many emergency requests as possible under dynamic and complex conditions. The locations of ambulance bases and the emergency vehicle themselves on this network are especially important considering their effect on response time and level of coverage. Recently, EMS network design has attracted substantial amount of attention in the literature through the form of ambulance location and relocation problems, humanitarian relief and disaster management problems, etc. (as reported in Brotcorne et al. (2003), Aringhieri et al. (2017), Ahmadi-Javid et al. (2017)). In this paper, we study a dynamic EMS network design problem that focuses on both strategic decisions, including ambulance base locations and fleet size, and operational decisions that includes the assignment of vehicle to emergency requests and their relocation on the network throughout the day.

Most of the literature on EMS network design currently formulates the problem as a deterministic coverage location problem (see recent reviews by Brotcorne et al. 2003, Aringhieri et al. 2017, Ahmadi-Javid et al. 2017). Specifically, these models enforce that when an emergency request occurs, it can be covered by an ambulance within a certain response time or distance. However, such models disregard the fact that even within a 24-hour cycle the rate of emergency requests, the travel time, the speed of vehicles, and their availability can vary drastically. Dynamic relocation of emergency vehicle on the network can therefore improve the performance of EMS by preventing areas from becoming unprotected, as well as improving the utilization of available resources. The models presented in this paper incorporate the dynamic evolution of these parameters in order to identify time-dependent location strategies that make the best trade-off between the flexibility of EMS network and the financial commitment.

Another key feature of the environment in which EMS network operate is the pervasive presence of uncertainty. This is especially the case for factors like the amount of emergency requests at any given time, traffic conditions, operational cost, etc. Disregarding this uncertainty in a EMS network design model is likely to lead to surprises regarding operational costs and might even mean that targets for demand coverage won't be met. In this regard, our proposed models will follow in the steps of Beraldi et al. (2004) and more recently Beraldi and Bruni (2009) who consider imposing a constraint (a.k.a. chance constraint as introduced in Charnes and Cooper (1959)) on the probability of covering the demand given a joint probability distribution on the amount of requests in each location of the network. This allows the EMS system managers to evaluate the additional cost related to improving the reliability level of the network. This model however has the deficiency that its optimal operating strategy recommends covering none of the emergency requests for the most severe scenarios. We extend the work of Beraldi and Bruni (2009) firstly and foremost by considering probabilistic envelope constraints (PEC) (see Xu et al. (2012)), instead of single chance constraints, which allow the EMS system managers to control how much coverage he is willing to offer as a function of how unlikely the scenario is. Other contributions are more of a numerical and empirical nature and will be discussed next.

To summarize, this paper addresses the modeling and resolution of an EMS network design problem that trades-off between expected cost and level of reliability while taking into account both strategic (ambulance base of location, fleet size) and operational (assignment and relocation of vehicle) decisions. More specifically, the contribution can be described as follows:

- We extend the static model presented in Beraldi and Bruni (2009) by considering a multi-period environment in which vehicles can be relocated in order to better account for time-dependencies of emergency requests, operational costs, and availability of emergency vehicles, and by imposing a probabilistic envelope constraint on the system-wide coverage level that is achieved under all realizable scenarios, instead of only the more likely ones that are addressed by chance constraint paradigm. This allows us to better characterize the trade-offs one needs to make between the cost of responding to each potential request and its likelihood of occurring.
- We develop an efficient exact algorithm for the chance constrained version of our model that significantly outperforms the Branch-and-Bound algorithm used in Beraldi and Bruni (2009). The algorithm is based on a Branch-and-Benders-Cut (B&BC) scheme and is accelerated using a set of valid inequalities and optimality cuts. This allows us to solve instances of the chance constrained stochastic program with

up to 100 ambulance base locations, 150 customer locations, 6 time periods, and 200 scenarios in less than 5 minutes.

- We further develop an efficient approximation scheme for our PEC stochastic program that returns a solution guaranteed to be feasible. We observe empirically that the conservative solution of this approximation are nearly optimal for the problem instances considered in this study. This leads us to believe that the PEC stochastic program can be used to derive useful insights in problems of realistic size.
- We perform a case study based on a historical dataset obtained from the Northern Ireland Ambulance Service Health & Social Care Trust (NIASHSCT) which allows us to demonstrate how our proposed stochastic programming models can be applied in a real data-driven environment. This allows us to provide new insights on how probabilistic envelopes can be designed and on the trade-offs that exist between coverage, reliability, and expected total cost. Finally, to the best of our knowledge, we are the first to perform an out-of-sample evaluation of optimal strategies obtained for EMS network design problems. This motivates us to propose a novel online procedure that identifies ambulance assignments for scenarios that are not observed in the in-sample dataset and confirm empirically its statistical consistency.

The rest of this paper is structured as follows. Section 2 presents a brief overview of related literature. Section 3 describes the deterministic, chance-constrained (CC), and probabilistic envelope constrained (PEC) version of our dynamic EMS network design model. Section 4 introduces the accelerated B&BC scheme for the chance-constrained stochastic program. Section 5 presents how to adapt this B&BC scheme to the PEC stochastic program and also describes our proposed conservative approximation scheme. The numerical performance of the different algorithms and practical study of NIASHSCT are presented in Section 6 and 7 respectively. Finally, we give concluding remarks and future work in Section 8.

2 Literature review

The earliest versions of ambulance location and EMS network design models can be considered as extensions of classical covering location models such as the set covering location problem (SCLP) (Toregas et al. 1971), the maximal covering location problem (MCLP) (Church and Velle 1974), the double standard model (DSM) (Gendreau et al. 1997) and their variants. These early models are static and deterministic ones with decision variables that model both the location of bases and the fleet sizes while the objective typically aim at maximizing the coverage or minimizing the number of facilities. In particular, one of the first set covering location model introduced by Toregas et al. (1971) minimize the number of ambulances needed to cover all demand points, while ignoring other relevant aspects of these problems. Another classical ambulance location model is called MCLP and introduced by Church and Velle (1974). This model maximizes the proportion of the population that can be covered with a limited number of ambulances. In the DSM, two coverage standards need to be verified. First, a fixed proportion of population needs to be served in less than r_1 distance, while it must still be possible to serve any member of the population within r_2 distance, where $r_2 \geq r_1$. Extensions of the SCLP and MCLP models have also appeared more recently in (Başar et al. 2011, Schneeberger et al. 2016, Paul et al. 2017). A more detailed review of covering location problems can be found in Li et al. (2011).

In contrast to static location models, dynamic location model consider long term effects of ambulance deployment decisions. To the best of our knowledge, the first dynamic ambulance location model is proposed by Gendreau et al. (2001), where the authors account for double coverage standard (as in DSM) and penalize frequent relocation of vehicles. More recently, Moeini et al. (2015) propose a modification to the earlier model that appears to provide better cover for emergency request by using an adjustment parameter that accounts for its fluctuation. Similarly, Schmid and Doerner (2010) extend the work of Gendreau et al. (2001) and Gendreau et al. (1997) by accounting for capacity over a certain horizon of time. Degel et al. (2015) also formulate a multi-objective and multi-period covering location model by incorporating time-dependent parameters and decisions, such as fleet size, location and relocation. It is worth mentioning that a number of studies have made use of exact or approximate dynamic programming (see Maxwell et al. (2010), Schmid (2012)) yet their application is limited due to the well-known curse of dimensionality.

The earliest version of stochastic programming model can be traced back to Daskin (1983) whose seminal work takes the form of the Maximal EXpected Covering Location Problem (MEXCLP). In MEXCLP, the author assumes that each emergency vehicle is busy with a certain probability and that its availability is independent from other vehicles. He then formulates a model in which the expected coverage is maximized. van den Berg and Aardal (2015) later extend the MEXCLP by formulating a time-dependent model with start-up and relocation costs, which was later extended to the multi-period setting in van den Berg and Aardal (2015). At the same period, Ansari et al. (2015) account for uncertainty travel times and preferences in the assignment of ambulances from different stations to demand location while Maleki et al. (2014) adapt the maximal expected coverage relocation problem (MECRP) in order to optimize the redeployment of ambulances. In order to improve fairness in semi-rural/semi-urban communities, Chanta et al. (2014) propose a bi-objective covering location model EMS covering location model that trades-off between expected number of emergency calls and amount of disparity between different geographic regions. Finally, van den Berg et al. (2016) extend the MEXCLP by considering fractional coverage and Chan et al. (2016) account for coverage probabilities that decayed with respect to distance in the context of the deployment of public automated external defibrillator (AED).

Another important family of stochastic programming formulation appears to have been introduced for the first time by Ball and Lin (1993) and employs so-called "chance constraints" to control the probability that a vehicle is unable to respond to a demand call within a certain amount of time. In Beraldi et al. (2004), the authors develop a stochastic programming model with joint probabilistic constraints that aims at ensuring a reliable service level of service for emergency demand under a known distribution of random emergency demand while minimizing the total cost. They employ a so-called p-efficient points of the joint probability distribution to reformulate the chance constraints using mixed-integer representation. Similarly, Zhang and Li (2015) also mitigate emergency demand uncertainty by proposing a chance constrained formulation that can be reformulated as a second-order cone program (SOCP) by assuming that only the mean and covariance information is known for the random emergency demand.

A number of recent contributions have also formulated the EMS network design problem in the form of a two-stage model where location and fleet size decisions are considered in the first stage while the assignment decisions of ambulances to demand requests are made in a second stage. Such a model is proposed for instance by Naoum-Sawaya and Elhedhli (2013) and used an objective function that trade-offs between relocation costs and expected coverage, in the form of an expected penalty cost for unserved demand. The model in Boujemaa et al. (2017) is similar in spirit yet introduces the notion of a two-tiered system with two types of vehicles that can be employed on the network. Nickel et al. (2016) handle the trade-off between cost and coverage by imposing a lower bound on the expected coverage. An important issue associated to all of these approaches (and similarly all MEXCLP approaches) is that by using expected coverage as a measure of the risk of unmet demand, the decision model does not provide any control over the likelihood that the coverage goes below certain critical levels. This issue is naturally addressed in the chance constraints formulation typically at the price of computationally inefficiency. It is worth mentioning that Noyan (2010) does propose the use of "integrated chance constraint" or second-order stochastic dominance constraints, which also circumvent this issue. One needs to know however that these risk measures are much less interpretable than chance constraints and still give rise to significant computational difficulties when considering two-stage formulations.

Our work is heavily inspired by the work of Beraldi and Bruni (2009) who consider a two-stage stochastic programming formulation that minimizes the expected total location and operation cost of responding to emergency requests while imposing a chance constraint on coverage of the total emergency request demand. These authors design a Branch-and-Bound algorithm to solve the resulting mixed integer linear program (MILP). We extend this work in four ways. Firstly, we extend this approach to a multi-period setting in which ambulances can be relocated to other stations (as proposed in Schmid and Doerner (2010) and van den Berg and Aardal (2015) for expected covering models) between the subsequent time blocks. Secondly, we consider a more general version of the chance constraint that allows us to also control coverage in scenarios that are considered more extreme while these are ignored in Beraldi and Bruni (2009)'s work thus causing an optimistic bias in terms of optimal expected cost. Thirdly, we propose an exact solution scheme that significantly improves the numerical performance of the algorithm proposed by these authors, allowing us to solve problems of similar sizes than what is currently achieved for two-stage expected coverage problems.

Finally, while Beraldi and Bruni (2009) conduct numerical experiments using test problems for the Two-Stage Capacitated Facility Location problem, we evaluate our network design model in a more realistic case study that uses a dataset covering historical ambulance service requests from a region in northern Ireland and draw a number of practical insights.

3 Model formulation

In this section, we firstly present in Section 3.1 the deterministic version of dynamic EMS network design model and related notation. Then, we incorporate emergency demand uncertainty and propose in Section 3.2 two probabilistic constrained stochastic programs, i.e. chance-constrained stochastic program and probabilistic envelope constrained stochastic program.

3.1 Deterministic model and notation

We consider a dynamic EMS network design problem that addresses ambulance location, fleet size, allocation and relocation decisions, and where emergency requests occur randomly. Specifically, we partition the whole region into a set $\mathcal{I}$ of subregions or zones, based on postal code or geographic division. These zones may refer to a community, a village, a road, etc. Thus, we assume that emergency requests originate randomly from any of these zones. Moreover, we consider a one-day horizon divided in T periods (i.e. 4 hours per time period). Given a set $\mathcal{J}$ of potential ambulance base locations, the strategic decision involves the identification of which subset of these locations will be operated at different times of the day and the number of ambulances assigned to each of these location throughout the day. This implicit involves decisions about the size of the ambulance fleet and how these ambulances are relocated from one time period to the other in order to serve the regions with higher expected demand. We capture these decisions using binary variables x_j^t, which controls whether the ambulance base $j \in \mathcal{J}$ will be open in time period t, while a set of integer variables y_j^t represent the number of emergency vehicles allocated to ambulance base j in time period t. A second set of integer decision variables r_{mj}^t denote the number of emergency vehicles planned to be relocated from base $m \in \mathcal{J}$ to base $j \in \mathcal{J}$ between time period t and $t+1$. At the operational level, each time an emergency request occurs, it is vital to determine from which base an ambulance will be dispatched to the scene in order to provide the necessary services. These decisions are modeled using integer variables z_{ij}^t representing the number of emergency requests in zone i answered by ambulances at base j in time period t.

Unlike most related work discussed in Section 2 under a static context, we incorporate time-dependent parameters and decisions, such as emergency demand, available fleet size, related cost parameters, some of which are also in Schmid and Doerner (2010) and van den Berg and Aardal (2015). The latter consider deterministic models where the total coverage is maximized. Similarly to a number of deterministic models proposed for this problem (namely Beraldi et al. (2004), Beraldi and Bruni (2009), Noyan (2010), Naoum-Sawaya and Elhedhli (2013), Nickel et al. (2016), Boujemaa et al. (2017)), we assume that the objective of the system manager is to minimize the total cost for operating the EMS network, which decomposes as: a fixed-charge cost f_j^t for operating a base at location j at time t, an operating and maintenance cost g_j per emergency vehicle at ambulance base location $j \in \mathcal{J}$, an emergency service cost $c^t l_{ij}$ per emergencies served in zone i from base j at time t, which accounts for marginal travel cost c^t and distance l_{ij}, finally α^t is the unit cost for ambulance relocation in time period $t \in \{1, \ldots, T\}$. This gives rise to the following deterministic EMS network design model:

$$\text{[DM]} \quad \underset{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{r}}{\text{minimize}} \quad \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} f_j^t x_j^t + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} g_j y_j^t + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} c^t l_{ij} z_{ij}^t + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{J}} \alpha^t r_{mj}^t \tag{1a}$$

$$\text{subject to} \quad y_j^t \le P_j^t x_j^t \qquad \forall j \in \mathcal{J},\ t \in \{1, \ldots, T\} \tag{1b}$$

$$x_j^t \ge x_j^{t-1} \qquad \forall j \in \mathcal{J}, t \in \{1, \ldots, T\} \tag{1c}$$

$$y_j^t + \sum_{m\in\mathcal{J}} r_{mj}^t - \sum_{m\in\mathcal{J}} r_{jm}^t = y_j^{t+1} \qquad \forall j \in \mathcal{J}, t \in \{1, \ldots, T-1\} \tag{1d}$$

$$y_j^T + \sum_{m\in\mathcal{J}} r_{mj}^T - \sum_{m\in\mathcal{J}} r_{jm}^T = y_j^1 \qquad \forall j \in \mathcal{J} \tag{1e}$$

$$\sum_{i\in\mathcal{I}} z_{ij}^t \leq \lambda y_j^t \qquad \forall j \in \mathcal{J},\ t \in \{1,\dots,T\} \tag{1f}$$

$$\sum_{j\in\mathcal{J}} z_{ij}^t \leq d_i^t \qquad \forall i \in \mathcal{I},\ t \in \{1,\dots,T\} \tag{1g}$$

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z_{ij}^t \geq \beta \sum_{i\in\mathcal{I}} d_i^t \qquad \forall t \in \{1,\dots,T\} \tag{1h}$$

$$x_j^t \in \{0,1\}, y_j^t, r_{mj}^t, z_{ij}^t \in \mathbb{N} \qquad \forall i \in \mathcal{I}, j, m \in \mathcal{J}, t \in \{1,\dots,T\}. \tag{1i}$$

Constraint (1b) imposes the maximum capacity of vehicles on shift at each ambulance base j and time period t depending on whether the base is open or closed. Constraint (1c) imposes that an open base remains open for the rest of the day. Constraints (1d) and (1e) are dynamic balance equations, which ensure that the relocation plan of emergency vehicles is feasible over the planning horizon T. It also imposes that initial and final distribution of ambulances should be the same for consistency from one day to the other. Constraint (1f) states that the maximum number of emergency request that need to be served from each base depending on the number of ambulances present and the per period service rate $\lambda \in \mathbb{N}$ of an ambulance. Constraint (1g) imposes that, in each zone i, one cannot serve more emergencies than the number of requests submitted d_i^t. Finally, constraint (1h) imposes an overall minimum service level of β for the EMS system, in other words the network should be designed and managed so that at each period of the day a certain proportion $\beta \in [0, 1]$ of the total emergency demand be served.

It is worth noting that problem (1) combines ideas from classical mathematical formulation such as the capacitated fixed-charge location problem, the p-median problem, and transportation problem and integrates them in a dynamic setting. In doing so, some assumptions will be useful for later analysis. For instance, we make the assumption that every zone can be served from every base location. Although this is not realistic, one can correct for this deficiency by considering that distances to unreachable zones are arbitrarily large, thus making the travel financially infeasible. We also assume that all emergency requests are served in the same amount of time yet we believe that this should not have an important effect, due to the large number of total requests served, on the conclusion that will be drawn when comparing the quality of different strategic decisions. Finally, one might also note that it is possible to model the fact that some ambulances become off-duty during some periods of the day in order to reduce total cost by assigning them to an artificial node with $g_j = 0$ and are considered arbitrarily far from all zones.

3.2 Two probabilistic constrained stochastic programming models

An important limitation of the DM model consists in the assumption that all parameters are exactly known at the moment of designing the network. Indeed, in practice it is difficult to predict exactly where and when the emergency requests will occur. For this reason, we now propose a model that assumes that these requests occur randomly according to some distribution $\boldsymbol{d} \sim \mathbb{Q}$. In this context, it is reasonable to assume that the assignment of ambulances will be done once the demand of each period is known. In other words, each decision $z_{ij}^t : \mathbb{R}^{|\mathcal{J}|} \to \mathbb{R}$ can be adjusted to the observed emergency demand $\boldsymbol{d}^t$. Following the ideas proposed in Beraldi and Bruni (2009), this gives rise to the following two-stage chance-constrained stochastic program (CCSP):

[CCSP]

$$\underset{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}(\cdot),\boldsymbol{r}}{\text{minimize}}\ \mathbb{E}_{\mathbb{Q}}\left[\sum_{t=1}^{T}\sum_{j\in\mathcal{J}} f_j^t x_j^t + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} g_j y_j^t + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} c^t l_{ij} z_{ij}^t(\boldsymbol{d}^t) + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{J}} \alpha^t r_{mj}^t\right] \tag{2a}$$

subject to (1b) – (1e)

$$\sum_{i\in\mathcal{I}} z_{ij}^t(\boldsymbol{d}^t) \leq \lambda y_j^t \qquad \forall j \in \mathcal{J},\ t \in \{1,\dots,T\}, \text{ almost surely} \tag{2b}$$

$$\sum_{j\in\mathcal{J}} z_{ij}^t(\boldsymbol{d}^t) \le d_i^t \qquad \forall i\in\mathcal{I},\ t\in\{1,\dots,T\}, \text{ almost surely} \tag{2c}$$

$$\mathbb{P}_{\mathbb{Q}}\left(\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z_{ij}^t(\boldsymbol{d}^t) \ge \beta\sum_{i\in\mathcal{I}} d_i^t\right) \ge 1-\eta \qquad \forall t\in\{1,\dots,T\} \tag{2d}$$

$$z_{ij}^t(\boldsymbol{d}^t)\in\mathbb{N} \qquad \forall i\in\mathcal{I}, j\in\mathcal{J}, t\in\{1,\dots,T\}, \text{ almost surely} \tag{2e}$$

$$x_j^t\in\{0,1\}; y_j^t, r_{mj}^t\in\mathbb{N} \qquad \forall i\in\mathcal{I}, j,m\in\mathcal{J}, t\in\{1,\dots,T\}, \tag{2f}$$

for some $\eta\in[0,1]$, where the expected total cost of operation is minimized and where constraint (2d) is a so-called chance constraint that controls the reliability of the coverage, i.e. the likelihood of serving the minimum proportion β of request that is required according to $\mathbb{Q}$ at time t. For example, when $\eta=0$ then the coverage must be satisfied with probability one, and in particular it must be satisfied under any scenario that has a strictly positive likelihood of occurrence. As argued in Beraldi and Bruni (2009), this might however lead to an over-dimensioned system hence the need for the CCSP model, which can serve as a tool to evaluate different alternatives in terms of cost-reliability trade-off. Unfortunately, the above model is necessarily overly optimistic in assessing the expected total cost of managing the system given that it completely disregards how the system should respond to requests in the more extreme scenarios that have less than η probability of occurring. This is illustrated by the following example.

Example 1 *Consider a simple network with a single ambulance base, one zone that generates emergency requests, and a single period horizon. Let d_1 be drawn according to two possible scenarios: in scenario #1 $d_1=50$ and has 90% chance of occurring while, in scenario #2, $d_1=200$ with 10% chance. If one imposes a 90% chance of covering all emergency requests $\beta=1$, then, assuming that the service rate is $\lambda=50$, it is clear that it becomes optimal to plan for a single ambulance to be present all day. In particular the solution would take the shape $x_1=1$, $y_1=1$, and $z_{11}(50)=50$ while $z_{11}(200)=0$. The optimal expected cost would be $f_1+g_1+45cl_{11}$ reflecting the fact that the available ambulance is left idle at the base under the more extreme scenarios.*

To correct for this deficiency of the CCSP model, we replace constraint (2d) with a probabilistic envelope constraint as popularized in Xu et al. (2012) in order to provide some control on the level of coverage achieved under all potential scenarios of emergency requests. Similarly, this gives rise to the following two-stage probabilistic envelope constrained stochastic program (PECSP):

[PECSP]

$$\underset{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z}(\cdot),\boldsymbol{r}}{\text{minimize}}\ \mathbb{E}_{\mathbb{Q}}\left[\sum_{t=1}^{T}\sum_{j\in\mathcal{J}} f_j^t x_j^t + \sum_{t=1}^{T} g_j y_j^t \sum_{j\in\mathcal{J}} + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} c^t l_{ij} z_{ij}^t(\boldsymbol{d}^t) + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{J}} \alpha^t r_{mj}^t\right] \tag{3a}$$

subject to (1b) − (1e), (2b), (2c), (2e), (2f)

$$\mathbb{P}_{\mathbb{Q}}\left(\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z_{ij}^t(\boldsymbol{d}^t) \ge \beta(\eta)\sum_{i\in\mathcal{I}} d_i^t\right) \ge 1-\eta \qquad \forall t\in\{1,\dots,T\}, \eta\in[0,1], \tag{3b}$$

where constraint (3b) now covers all reliability levels $\eta\in[0,1]$ using a controlled coverage envelope function $\beta(\eta)$. Actually, the case that $\eta=1$ does not appear in the constraint since a chance constraint that impose 0% reliability is always redundant. Note that without loss of generality, we will assume that $\beta(\eta)$ is non-decreasing (i.e. the less extreme the scenario is, the more strict the targeted coverage), otherwise one can replace this function with $\beta'(\eta):=\sup\{\beta(\eta'):\eta'\le\eta\}$ without affecting the set of feasible solutions. It is also straightforward to see that PECSP generalizes CCSP since the latter can be obtained by using the following coverage envelope function

$$\beta(\eta)=\begin{cases}\bar{\beta} & \text{if } \eta\ge\bar{\eta}\\ 0 & \text{otherwise.}\end{cases}$$

Looking back at Example 1, one can see how the identified deficiency can be resolved using our PECSP model. In particular, this new model allows one to describe what level of coverage is expected for the more

extreme scenario. In the case that the coverage needed for all scenarios is always above 50%, e.g. $\beta(\eta) = 50\%$ for $\eta < 10\%$, then the optimal solution would recommend a larger fleet of ambulance since $200 \times 50\% > 50$. If the reliability level is more relaxed, e.g. $\beta(\eta) = 5\%$ for $\eta < 10\%$, then the same number of vehicles would be proposed yet the expected total cost would reflect the fact that a minimum proportion of requests need to be satisfied in all scenarios.

The downside of employing PECSP is the challenge that it raises from a computational perspective. Indeed, chance constrained stochastic programs are typically considered to be computationally intractable, except in rare cases with special structure. In general, chance constraints define a non-convex feasible set and can either be conservatively approximated as in Nemirovski and Shapiro (2006) or reformulated as mixed-integer programs when the distribution is discrete or when a sample average approximation is used (Pagnoncelli et al. 2009). The challenge is even more significant when employing our probabilistic envelope constraint, which conceptually speaking imposes an infinite continuum of chance constraints. To the best of our knowledge, to this date the most efficient solution scheme employs distributionally robust versions of these constraints (see Xu et al. (2012)). Yet, while these versions might not be very interesting from a practical point of view (as it might lead to overly conservative solutions), it is also not clear how the method proposed in Xu et al. (2012) could be used for constraint (3b) given that in this constraint the effect between the envelope function and uncertainty is not additive but rather multiplicative. We leave this issue for future research.

In Sections 4 and 5, we explain how to obtain a mixed-integer linear programming reformulation of both CCSP and PECSP and describe decomposition schemes that can be employed to improve numerical efficiency. A conservative approximation scheme is also proposed for PECSP in order to allow the resolution of EMS network design problems of realistic sizes.

4 Solution scheme for scenario-based CCSP

In this section, we present a scenario-based version of CCSP, derive a mixed-integer linear programming reformulation of the problem, and propose different variants of a Branch-and-Benders-Cut (B&BC) method to solve the problem efficiently.

4.1 Scenario-based CCSP

In a scenario-based approach, one assumes that the random emergency requests $\boldsymbol{d}$ are drawn from a finite set of N scenarios $\{d_\omega\}_{\omega\in\Omega}$, where for simplicity $\Omega = \{1,\dots,N\}$, hence the distribution $\mathbb{Q}$ can be characterized using a probability vector $[p_1\ p_2\ \dots\ p_N]$ such that $\boldsymbol{p} \geq 0$ and $\sum_{\omega\in\Omega} p_\omega = 1$. For each scenario $\omega \in \Omega$, we let $d^t_{i\omega}$ denote the number of emergency requests in zone i at time period t, while $z^t_{ij\omega}$ will denote the number of emergency requests in zone i at time period t served from ambulance base j. Under these conditions, one can rewrite the CCSP as:

$$\underset{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{r}}{\text{minimize}} \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} f^t_j x^t_j + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} g_j y^t_j + \sum_{\omega\in\Omega}\sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} p_\omega c^t l_{ij} z^t_{ij\omega}(\boldsymbol{d}^t) + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{J}} \alpha^t r^t_{mj} \tag{4a}$$

subject to (1b) – (1e), (2f)

$$\sum_{i\in\mathcal{I}} z^t_{ij\omega} \leq \lambda y^t_j \qquad \forall j\in\mathcal{J}, t\in\{1,\dots,T\}, \omega\in\Omega \tag{4b}$$

$$\sum_{j\in\mathcal{J}} z^t_{ij\omega} \leq d^t_{i\omega} \qquad \forall i\in\mathcal{I}, t\in\{1,\dots,T\}, \omega\in\Omega \tag{4c}$$

$$\sum_{\omega\in\Omega} p_\omega \mathbf{1}\left\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta \sum_{i\in\mathcal{I}} d^t_{i\omega}\right\} \geq 1-\eta \qquad \forall t\in\{1,\dots,T\} \tag{4d}$$

$$z^t_{ij\omega} \in \mathbb{N} \qquad \forall i\in\mathcal{I}, j\in\mathcal{J}, t\in\{1,\dots,T\}, \omega\in\Omega\,, \tag{4e}$$

where $\mathbf{1}\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta \sum_{i\in\mathcal{I}} d^t_{i\omega}\}$ is an indicator function that returns one if the required coverage is achieved at time period t under scenario ω, and otherwise returns zero.

We next introduce a set of binary variables $\boldsymbol{\rho} \in \{0,\,1\}^{N\times T}$ that will be used to assess and count whether the coverage frequencies imposed in constraint (4d) are satisfied. This allows us to replace the constraint with:

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta(1-\rho^t_\omega)\sum_{i\in\mathcal{I}} d^t_{i\omega} \qquad \forall t\in\{1,\ldots,T\}, \omega\in\Omega \tag{5}$$

$$\sum_{\omega\in\Omega} p_\omega \rho^t_\omega \leq \eta \qquad \forall t\in\{1,\ldots,T\}\,. \tag{6}$$

In particular, $\rho^t_\omega = 1$ indicates that the system manager plans to violate the coverage constraint under scenario ω at time t.

Proposition 1 *The CCSP problem is equivalent to the following mixed integer linear program in terms of optimal value and set of optimal solutions for $\boldsymbol{x}$, $\boldsymbol{y}$, and $\boldsymbol{r}$:*

[CCSP2]

$$\underset{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{r},\boldsymbol{\rho}}{\text{minimize}} \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} f^t_j x^t_j + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} g_j y^t_j + \sum_{\omega\in\Omega}\sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} p_\omega c^t l_{ij} z^t_{ij\omega} + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{J}} \alpha^t r^t_{mj} \tag{7a}$$

subject to (1b) – (1e), (4b), (4c), (6)

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq (1-\rho^t_\omega)\left\lceil \beta\sum_{i\in\mathcal{I}} d^t_{i\omega}\right\rceil \qquad \forall t\in\{1,\ldots,T\}, \omega\in\Omega \tag{7b}$$

$$z^t_{ij\omega} \geq 0 \qquad \forall i\in\mathcal{I}, j\in\mathcal{J}, t\in\{1,\ldots,T\},\ \omega\in\Omega \tag{7c}$$

$$r^t_{mj} \geq 0 \qquad \forall j,m\in\mathcal{J}, t\in\{1,\ldots,T\} \tag{7d}$$

$$x^t_j\in\{0,1\}, y^t_j\in\mathbb{N}, \rho^t_\omega\in\{0,\,1\} \qquad \forall j\in\mathcal{J}, t\in\{1,\ldots,T\}, \omega\in\Omega\,. \tag{7e}$$

Moreover, a full optimal solution for CCSP can be retrieved by solving a series of small MILPs.

Note that in CCSP2, we were able to relax the integrality constraint on the assignment variables $\boldsymbol{z}$ and the relocation variables $\boldsymbol{r}$ without affecting the quality of the solutions obtained for $\boldsymbol{x}$ and $\boldsymbol{y}$ and optimal value of the problem. This is due to the fact that when first stage decisions $\boldsymbol{x}$, $\boldsymbol{y}$ and $\boldsymbol{\rho}$ are fixed, the convex hull of the joint feasible set for $\boldsymbol{z}$ and $\boldsymbol{r}$ can be obtained simply by employing a linear relaxation (modulo the tightening of constraint (5) using a simple rounding scheme). The detailed proof can be seen in Appendix A.1. In what follows, we describe a Branch-and-Benders-Cut (B&BC) method that can be used to solve this MILP together with ways of improving its convergence time.

4.2 Enhanced Branch-and-Benders-Cut method

While stochastic programming models are generally known to be computationally challenging, a method known as the L-shaped method or Benders decomposition (BD), introduced by Benders (1962) has achieved good numerical performance in a number of applications including facility location and network design (see for example in Adulyasak et al. (2015), Martins de Sá et al. (2015), Gendron et al. (2016), Noyan et al. (2016), Dalal and Üster (2018), Rahmaniani et al. (2018) and Rahmaniani et al. (2017) for a recent review of the topic). The main idea of BD is that the MILP that emerges in a stochastic program might decompose into a pure integer master problem (MP) and a number of smaller linear sub-problems (SPs). If so, the two problems can be solved iteratively, introducing each time in the MP a group of additional constraints, known as Benders cuts. This is repeated until the lower bound obtained from the master problem reaches the upper bound that can be computed based on the solutions of the SPs.

In this section, we describe an implementation of this approach that employs the Branch-and-Benders-cut method in order to improve numerical efficiency by integrating Benders cuts directly in the Branch-and-Bound

algorithm that solves the MP. We start in Section 4.2.1 by describing the Bender's decomposition scheme for CCSP2 after introducing redundant constraints that ensures that only optimality cuts will be returned by the SPs. We then summarize the B&BC scheme in Section 4.2.2. We finally discuss in sections 4.2.3 and 4.2.4 variants of the algorithm that involve a set of valid inequalities that can be used to tighten the MP and a way of tightening the optimality cuts returned from the SPs.

4.2.1 Benders decomposition

A traditional application of BD aims at separating the complicating integer variables $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\rho})$ from the non-complicating variables $\boldsymbol{z}$ and $\boldsymbol{r}$ in order to accelerate the resolution of CCSP2. In particular, we consider the following reformulation of CCSP2:

$$\text{[CCSP2']} \quad \underset{\boldsymbol{x},\boldsymbol{y},\boldsymbol{\rho},\theta^r,\boldsymbol{\theta}^z}{\text{minimize}} \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} f_j^t x_j^t + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} g_j y_j^t + \sum_{\omega\in\Omega}\sum_{t=1}^{T} \theta_{t\omega}^z + \theta^r \tag{8a}$$

$$\text{subject to } (1b), (1c), (6), (7e)$$

$$\theta^r \geq h^r(y) \tag{8b}$$

$$\theta_{t\omega}^z \geq h_{\omega,t}^z(y,\rho) \qquad \forall t \in \{1,\dots,T\}, \omega \in \Omega \tag{8c}$$

$$\sum_{j\in\mathcal{J}} y_j^t = \sum_{j\in\mathcal{J}} y_j^{t+1} \qquad \forall j \in \mathcal{J}, t \in \{1,\dots,T-1\} \tag{8d}$$

$$\sum_{j\in\mathcal{J}} \lambda y_j^t \geq \lceil \beta \sum_{i\in\mathcal{I}} d_{i\omega}^t \rceil (1-\rho_\omega^t) \qquad \forall t \in \{1,\dots,T\}, \omega \in \Omega \tag{8e}$$

where

$$\text{[CCSP-SP}^r\text{]} \quad h^r(\boldsymbol{y}) := \min_{\boldsymbol{r}\geq 0} \sum_{t=1}^{T}\sum_{m\in\mathcal{J}}\sum_{j\in\mathcal{J}} \alpha^t r_{mj}^t \tag{9a}$$

$$\text{subject to } y_j^t + \sum_{m\in\mathcal{J}} r_{mj}^t - \sum_{m\in\mathcal{J}} r_{jm}^t = y_j^{t+1} \quad \forall j \in \mathcal{J}, t \in \{1,\dots,T-1\} \quad (\mu_{jt}^r) \tag{9b}$$

$$y_j^T + \sum_{m\in\mathcal{J}} r_{mj}^T - \sum_{m\in\mathcal{J}} r_{jm}^T = y_j^1 \quad \forall j \in \mathcal{J} \quad (\mu_{jT}^r) \tag{9c}$$

and where

$$\text{[CCSP-SP}_{\omega,t}^z\text{]} \quad h_{\omega,t}^z(\boldsymbol{y},\boldsymbol{\rho}) := \min_{\boldsymbol{z}_\omega^t \geq 0} \sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} p_\omega c^t l_{ij} z_{ij\omega}^t \tag{10a}$$

$$\text{subject to } \sum_{i\in\mathcal{I}} z_{ij\omega}^t \leq \lambda y_j^t \qquad \forall j \in \mathcal{J} \qquad (\mu_{jt\omega}^1) \tag{10b}$$

$$\sum_{j\in\mathcal{J}} z_{ij\omega}^t \leq d_{i\omega}^t \qquad \forall i \in \mathcal{I} \qquad (\mu_{it\omega}^2) \tag{10c}$$

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z_{ij\omega}^t \geq (1-\rho_\omega^t)\lceil \beta \sum_{i\in\mathcal{I}} d_{i\omega}^t \rceil \qquad (\mu_{t\omega}^3)\,. \tag{10d}$$

In Problem CCSP2', we also include the redundant constraints (8d) and (8e) in order to ensure that $h^r(\boldsymbol{y})$ and $h_{\omega,t}^z(\boldsymbol{y},\boldsymbol{\rho})$ can be considered finite valued.

Proposition 2 *CCSP2' is equivalent to CCSP2. Moreover, given any solutions triplet $(\boldsymbol{x},\boldsymbol{y},\boldsymbol{\rho})$ that satisfy (8d) and (8e), problems CCSP-SP^r and CCSP-SP$_{\omega,t}^z$ are always feasible and bounded.*

A detailed proof of Proposition 2 can be found in Appendix A.2.

In order to obtain a decomposition between the optimization over $\boldsymbol{x}$, $\boldsymbol{y}$, $\boldsymbol{\rho}$, and the optimization over $\boldsymbol{z}$ and $\boldsymbol{r}$, we consider approximating the piecewise linear convex functions $h^r(\boldsymbol{y})$ and $h_{\omega,t}^z(\boldsymbol{y},\boldsymbol{\rho})$ using a subset

of their supporting hyperplanes. In particular, considering $h^r(\boldsymbol{y})$ it is possible for any fixed $\bar{\boldsymbol{y}}$ that satisfies constraints (8d) to identify a supporting hyperplane at $\bar{\boldsymbol{y}}$ by solving the dual problem associated to CCSP-SP^r.

$$[\text{CCSP-DSP}^r] \quad h^r(\bar{\boldsymbol{y}}) := \max_{\boldsymbol{\mu}^r \geq 0} \quad \sum_{t=1}^{T-1} \sum_{j \in \mathcal{J}} \mu^r_{jt}(\bar{y}^t_j - \bar{y}^{t+1}_j) + \sum_{j \in \mathcal{J}} \mu^r_{jT}(\bar{y}^T_j - \bar{y}^1_j) \tag{11a}$$

$$\text{subject to} \quad \alpha^t + \mu^r_{jt} - \mu^r_{mt} \geq 0 \qquad \forall m, j \in \mathcal{J}, t \in \{1, \ldots, T\}\,, \tag{11b}$$

where $\{\mu^r_{jt}\}_{j \in \mathcal{J}, t \in \{1,\ldots,T\}}$ are the dual variables associated with constraints (9b) and (9c). Strong duality necessarily applies for CCSP-SP^r since Proposition 2 guarantees that this problem is feasible. Furthermore, since it is also bounded, the existence of an optimal solution for CCSP-DSP^r is guaranteed. A supporting hyperplane at $\bar{\boldsymbol{y}}$ therefore necessarily takes the form:

$$\sum_{t=1}^{T-1} \sum_{j \in \mathcal{J}} \bar{\mu}^{r*}_{jt}(y^t_j - y^{t+1}_j) + \sum_{j \in \mathcal{J}} \bar{\mu}^{r*}_{jT}(y^T_j - y^1_j)\,,$$

where $\bar{\boldsymbol{\mu}}^{r*}$ is an optimal solution to CCSP-DSP^r.

Similarly, in the case of CCSP-SP$^z_{\omega,t}$, for a given solution $(\bar{\boldsymbol{y}}, \bar{\boldsymbol{\rho}})$ that satisfies (8e), the dual problem takes the form:

$$[\text{CCSP-DSP}^z_{\omega,t}] \quad h^z_{\omega,t}(\bar{\boldsymbol{y}}, \bar{\boldsymbol{\rho}}) = \max_{\boldsymbol{\mu}^1_{t\omega}, \boldsymbol{\mu}^2_{t\omega}, \mu^3_{t\omega}} \quad -\sum_{j \in \mathcal{J}} \lambda \bar{y}^t_j \mu^1_{jt\omega} - \sum_{i \in \mathcal{I}} d^t_{i\omega} \mu^2_{it\omega} + \mu^3_{t\omega}(1 - \bar{\rho}^t_\omega)\lceil \beta \sum_{i \in \mathcal{I}} d^t_{i\omega} \rceil \tag{12a}$$

$$\text{subject to} \quad -\mu^1_{jt\omega} - \mu^2_{it\omega} + \mu^3_{t\omega} \leq p_\omega c^t l_{ij} \quad \forall i \in \mathcal{I}, j \in \mathcal{J} \tag{12b}$$

$$\boldsymbol{\mu}^1_{t\omega} \geq 0, \boldsymbol{\mu}^2_{t\omega} \geq 0, \mu^3_{t\omega} \geq 0\,, \tag{12c}$$

where $\boldsymbol{\mu}^1_{t\omega} \in \mathbb{R}^{|\mathcal{J}|}$, $\boldsymbol{\mu}^2_{t\omega} \in \mathbb{R}^{|\mathcal{I}|}$, $\mu^3_{t\omega} \in \mathbb{R}$ are the dual vectors for constraints (10b)–(10d) respectively. Hence, a supporting hyperplane can be obtained in the form:

$$-\sum_{j \in \mathcal{J}} \lambda y^t_j \bar{\mu}^{1*}_{jt\omega} - \sum_{i \in \mathcal{I}} d^t_{i\omega} \bar{\mu}^{2*}_{it\omega} + \bar{\mu}^{3*}_{t\omega}(1 - \rho^t_\omega)\lceil \beta \sum_{i \in \mathcal{I}} d^t_{i\omega} \rceil\,,$$

where $\bar{\boldsymbol{\mu}}^{1*}$, $\bar{\boldsymbol{\mu}}^{2*}$, and $\bar{\boldsymbol{\mu}}^{3*}$ are optimal solutions of CCSP-DSP$^z_{t,\omega}$.

Given a subset of supporting hyperplanes for CCSP-SP^r and CCSP-SP$^z_{\omega,t}$, which we will represent using $\{\bar{\boldsymbol{\mu}}^{r\tau}\}_{\tau \in G^r}$ and $\{(\bar{\boldsymbol{\mu}}^{1\tau}_{t\omega}, \bar{\boldsymbol{\mu}}^{2\tau}_{t\omega}, \bar{\mu}^{3\tau}_{t\omega})\}_{\tau \in G^z_{t\omega}}$ for some index sets G^r and $G^z_{t\omega}$, one can formulate a Master Problem that returns a triplet $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\rho})$, which is feasible in CCSP2' and which optimal value provides a lower bound for CCSP2':

[CC-MP]

$$\underset{\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\rho}, \theta^r, \boldsymbol{\theta}^z}{\text{minimize}} \quad \sum_{t=1}^{T} \sum_{j \in \mathcal{J}} f^t_j x^t_j + \sum_{t=1}^{T} \sum_{j \in \mathcal{J}} g_j y^t_j + \sum_{t=1}^{T} \sum_{\omega \in \Omega} \theta^z_{t\omega} + \theta^r \tag{13a}$$

$$\text{subject to (1b), (1c), (6), (7e), (8d), (8e)}$$

$$\theta^r \geq \sum_{j \in \mathcal{J}} \sum_{t=1,\ldots,T-1} \bar{\mu}^{r\tau}_{jt}(y^t_j - y^{t+1}_j) + \sum_{j \in \mathcal{J}} \bar{\mu}^{r\tau}_{jT}(y^T_j - y^1_j),\ \forall \tau \in G^r \tag{13b}$$

$$\theta^z_{t\omega} \geq -\sum_{j \in \mathcal{J}} \lambda y^t_j \bar{\mu}^{1\tau}_{t\omega j} - \sum_{i \in \mathcal{I}} d^t_{i\omega} \bar{\mu}^{2\tau}_{t\omega i} + \bar{\mu}^{3\tau}_{t\omega} \beta \sum_{i \in \mathcal{I}} d^t_{i\omega}(1 - \rho^t_\omega),\ \forall \tau \in G^z_{t\omega}, t \in \{1, \ldots, T\}, \omega \in \Omega\,. \tag{13c}$$

Furthermore, for any feasible solution $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \bar{\boldsymbol{\rho}}, \bar{\theta}^r, \bar{\boldsymbol{\theta}}^z)$ of CC-MP, one can obtain an upper bound for CCSP2' through computing:

$$uobj = \sum_{t=1}^{T} \sum_{j \in \mathcal{J}} f^t_j \bar{x}^t_j + \sum_{t=1}^{T} \sum_{j \in \mathcal{J}} g_j \bar{y}^t_j + \sum_{t=1}^{T} \sum_{\omega \in \Omega} h^z_{\omega,t}(\bar{\boldsymbol{y}}, \bar{\boldsymbol{\rho}}) + h^r(\bar{\boldsymbol{y}}). \tag{14}$$

Algorithm 1 Traditional implementation of BD

Input A tolerance $\epsilon \geq 0$ and maximum run time *stoptime*
Initialize $v = 0, LB = -\infty, UB = +\infty, \theta^r = \theta^z_{t\omega} = 0, G^r = G^z_{t\omega} = \emptyset$ for all $t \in \{1, \ldots, T\}$, and $\omega \in \Omega$.
while (*runtime* $\leq$ *stoptime* and $UB - LB > \epsilon$) **do**
  Set $v = v + 1$, solve CC-MP.
  Record optimal solution $(\boldsymbol{x}^v, \boldsymbol{y}^v, \boldsymbol{\rho}^v, \theta^{rv}, \boldsymbol{\theta}^{zv})$ and optimal objective $lobj^v$.
  Update $LB = \max\{LB, lobj^v\}$.
  Fix $\bar{\boldsymbol{x}} := \boldsymbol{x}^v$, $\bar{\boldsymbol{y}} := \boldsymbol{y}^v$, and $\bar{\boldsymbol{\rho}} := \boldsymbol{\rho}^v$ and solve CCSP-DSP^r and all CCSP-DSP$^z_{\omega,t}$ to obtain $h^z_{t\omega}(\bar{\boldsymbol{y}}, \bar{\boldsymbol{\rho}})$ and $h^r(\bar{\boldsymbol{y}})$ together with $(\bar{\boldsymbol{u}}^{r*}, \bar{\boldsymbol{u}}^{1*}, \bar{\boldsymbol{u}}^{2*}, \bar{u}^{3*})$.
  Add $\bar{\boldsymbol{\mu}}^{r*}$ to G^r, and, for all $t \in \{1, \ldots, T\}$ and $\omega \in \Omega$, add $(\bar{\boldsymbol{\mu}}^1_{t\omega}, \bar{\boldsymbol{\mu}}^{2*}_{t\omega}, \bar{\mu}^{3*}_{t\omega})$ to $G^z_{t\omega}$
  Update $UB = \min\{UB, uobj^v\}$ where $uobj^v$ is as described in (14).
end while
return UB and corresponding optimal solution $(\boldsymbol{x}^{v^*}, \boldsymbol{y}^{v^*}, \boldsymbol{\rho}^{v^*})$ for which $uobj^{v^*} = UB$.

We present the detailed procedure of the traditional BD approach in Algorithm 1. Note that the iteration count is denoted by v. It is well known that Algorithm 1 always terminates in a finite number of iterations (even when $stoptime = \infty$) as long as one makes sure that each $\bar{\boldsymbol{u}}^{r*}$ and $(\bar{\boldsymbol{u}}^{1*}_{t\omega}, \bar{\boldsymbol{u}}^{2*}_{t\omega}, \bar{u}^{3*}_{t\omega})$ are vertices of the feasible set described in each CCSP-DSP problem. This is due to the fact that the number of vertices of each of these feasible sets is finite so that in the worst-case, one will eventually add all of them to the CC-MP and finally obtain that $lobj^v = uobj^v$. Unfortunately, in practice this implementation of BD can be excessively slow to converge due to the fact that each iteration involves solving an CC-MP that takes the form of a MILP and of a size that grows with the number of iterations. For this reason, an implementation of BD has become popular, which integrates the addition of supporting hyperplanes (a.k.a. Benders cuts) inside the Branch-and-Bound algorithm that is used to solve CC-MP.

4.2.2 Branch-and-Benders-Cut method

Instead of solving a Master Problem using the Branch-and-Bound algorithm at every iteration, the Branch-and-Benders-Cut method only requires traversing once the Branch-and-Bound tree structure. This is done by introducing the new supporting hyperplanes of $h^r(\boldsymbol{y})$ and $h^z_{\omega,t}(\boldsymbol{y}, \boldsymbol{\rho})$ (i.e. Benders cuts) only at nodes of the Branch-and-Bound tree for which the linear relaxation of the MILP identifies an integer solution (in the CPLEX solver this can be done through the *lazyconstraint callback routine*). The Branch-and-Bound tree search then resumes until the optimality is attained or a stopping criterion is met. Our implementation B&BC allowed us to reduce solution time significantly for CCSP2'. The procedure of the proposed B&BC method is presented in Algorithm 2.

4.2.3 Strengthened valid inequalities

Although constraint (8e) is useful for restricting the feasible set of CC-MP, the constraint becomes less informative once the integrality constraints of CC-MP are relaxed. This can therefore lead to excessive steps of branching in the Branch-and-Bound algorithm. It therefore becomes interesting to identify valid inequalities that are redundant for the MILP version of CC-MP but end up tightening its linear relaxation. Proposition 3 introduces a set of valid inequalities that are inspired from the work of Luedtke et al. (2010).

Proposition 3 *For any fixed t, let the ordered scenario set $\Omega'_t := \{\omega'_k\}_{k=1}^N$ be such that for all $k < k'$ we have that $\sum_{i\in\mathcal{I}} d^t_{i\omega'_k} \leq \sum_{i\in\mathcal{I}} d^t_{i\omega'_{k'}}$. Given some $\eta \in [0, 1)$, let $\bar{k} \in \{1, 2, \ldots, N\}$ be such that $\sum_{k=\bar{k}}^N p_{\omega'_k} > \eta$ and $\sum_{k=\bar{k}+1}^N p_{\omega'_k} \leq \eta$, where we consider that $\sum_{k=N+1}^N p_{\omega'_k} = 0$. Then for all $t \in \{1, \ldots, T\}$, constraints (6) and (8e) imply that :*

$$\sum_{j\in\mathcal{J}} \lambda y^t_j \geq \lceil \beta \sum_{i\in\mathcal{I}} d^t_{i\omega'_{\bar{k}}} \rceil . \tag{15}$$

This proposition provides a set of T valid inequalities for CC-MP. Note that for each of these constraint, i.e. for each t, the set Ω needs to be reordered appropriately and that index $\bar{k}$ might depend on t. The proof of Proposition 3 is provided in Appendix A.3.

Algorithm 2 Branch-and-Benders-Cut (B&BC) implementation

Input A tolerance $\epsilon \geq 0$ and maximum run time *stoptime*
Initialize $UB = +\infty$, $LB = -\infty$, $\mathcal{N} = \{\text{root}\}$ where root is the linear relaxation of MP.
Initialize $G^r = G^z_{t\omega} = \emptyset$ for all $t \in \{1, \ldots, T\}$, and $\omega \in \Omega$.
while ($\mathcal{N}$ is nonempty and $UB - LB > \epsilon$ and $runtime \leq stoptime$) **do**
  Select a node $o' \in \mathcal{N}$ and remove it from $\mathcal{N} \leftarrow \mathcal{N}/o'$.
  Solve linear relaxation of o' with G^r and $G^z_{t\omega}$ sets to obtain optimal solution $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \bar{\boldsymbol{\rho}}, \bar{\theta}^r, \bar{\boldsymbol{\theta}}^z)$ and consider optimal value as $lobj$.
  if $lobj < UB$ **then**
    if $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{\rho})$ is not integer valued **then**
      Update $LB := \max(LB, lobj)$.
      Branch on a non-integer variable, resulting in nodes o^* and o^{**}.
      $\mathcal{N} \leftarrow \mathcal{N} \cup \{o^*, o^{**}\}$.
    else
      Solve CCSP-DSP^r and CCSP-DSP$^z_{\omega,t}$ to obtain $h^z_{\omega,t}(\bar{\boldsymbol{y}}, \bar{\boldsymbol{\rho}})$ and $h^r(\bar{\boldsymbol{y}})$ together with $(\bar{\boldsymbol{u}}^{r*}, \bar{\boldsymbol{u}}^{1*}, \bar{\boldsymbol{u}}^{2*}, \bar{u}^{3*})$.
      Compute $uobj$ as described in (14).
      if $uobj - lobj > \epsilon/2$ **then**
        Add $\bar{\boldsymbol{\mu}}^{r*}$ to G^r, and, for all $t \in \{1, \ldots, T\}$ and $\omega \in \Omega$, add $(\bar{\boldsymbol{\mu}}^1_{t\omega}, \bar{\boldsymbol{\mu}}^{2*}_{t\omega}, \bar{\mu}^{3*}_{t\omega})$ to $G^z_{t\omega}$
        $\mathcal{N} \leftarrow \mathcal{N} \cup \{o'\}$.
      end if
      if $uobj - lobj \leq \epsilon/2$ **then**
        $UB = uobj$, $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\rho}^*, \theta^{r*}, \boldsymbol{\theta}^{z*}) = (\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}}, \bar{\boldsymbol{\rho}}, \bar{\theta}^r, \bar{\boldsymbol{\theta}}^z)$.
      end if
    end if
  end if
end while
return UB and corresponding optimal solution $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\rho}^*, \theta^{r*}, \boldsymbol{\theta}^{z*})$.

4.2.4 Strengthened optimality cuts

Looking back at the definition of $h^z_{\omega,t}(\boldsymbol{y}, \boldsymbol{\rho})$, one might notice that as long as $\rho^t_\omega \in \{0, 1\}$ the optimal value of this problem remains the same when replacing constraint (10c) with

$$\sum_{j \in \mathcal{J}} z^t_{ij\omega} \leq d^t_{i\omega}(1 - \rho^t_\omega) \qquad \forall i \in \mathcal{I},\ t \in \{1, \ldots, T\}, \omega \in \Omega\,.$$

This is due to the fact that when $\rho^t_\omega = 1$ then the two constraints are exactly the same, while if $\rho^t_\omega = 0$, then $z^t_\omega = 0$ becomes an optimal solution of CCSP-SP$^z_{\omega,t}$ due to the fact that the objective function is non-decreasing in $\boldsymbol{z}$. It is therefore possible to consider a modified version of CCSP-SP$^z_{\omega,t}$ for which the dual sub-problem becomes:

$$\begin{aligned}
[\text{CCSP-DSP}^z_{\omega,t}{}'] \quad h^z_{\omega,t}(\bar{\boldsymbol{y}}, \bar{\boldsymbol{\rho}}) = \max_{\boldsymbol{\mu}^1_{t\omega}, \boldsymbol{\mu}^2_{t\omega}, \mu^3_{t\omega}} \quad & -\sum_{j \in \mathcal{J}} \lambda \bar{y}^t_j \mu^1_{jt\omega} - \sum_{i \in \mathcal{I}} d^t_{i\omega}(1 - \bar{\rho}^t_\omega)\mu^2_{it\omega} + \mu^3_{t\omega}(1 - \bar{\rho}^t_\omega)\lceil \beta \sum_{i \in \mathcal{I}} d^t_{i\omega} \rceil \\
\text{subject to} \quad & -\mu^1_{jt\omega} - \mu^2_{it\omega} + \mu^3_{t\omega} \leq p_\omega c^t l_{ij} \quad \forall i \in \mathcal{I}, j \in \mathcal{J} \\
& \boldsymbol{\mu}^1_{t\omega} \geq 0, \boldsymbol{\mu}^2_{t\omega} \geq 0, \mu^3_{t\omega} \geq 0\,,
\end{aligned}$$

Hence, this reformulated sub-problem gives rise to a supporting hyperplane that will provide a tighter constraint when CC-MP will be relaxed.

4.2.5 Benders strategy solution mode in CPLEX 12.7

At the time of writing this article, the latest version of CPLEX (namely version 12.7) did offer a solution mode in which a Benders decomposition scheme is autonomously implemented by the software during resolution of a MILP. Based on documentation, the decomposition assigns all integer variables to a master problem and all continuous variables to one or more sub-problems. This decomposition can be further directed using the programming interface. While we commend the efforts made by IBM to streamline the use of this efficient decomposition scheme, the preliminary experiments we made using this feature of the software provided us evidence that a custom implementation of B&BC was significantly more efficient for this problem.

5 Solution scheme for scenario-based PECSP

In this section, we study the numerical resolution of PECSP in a context where $\mathbb{Q}$ is modeled as a discrete distribution. As done in Section 4 with CCSP, we firstly formulate the scenario-based PECSP in the form of a mixed-integer linear program and adapt the Branch-and-Benders-Cut algorithm. Moreover, in order to overcome the additional computational challenges that arise with the PECSP, we propose a conservative approximation that we expect can more easily handle large-scale problems.

5.1 Scenario-based PECSP

Similarly as was done in Section 4, we start by assuming that $\mathbb{Q}$ is discrete and characterized using a probability vector $\boldsymbol{p} \in \mathbb{R}^N$ such that $\boldsymbol{p} \geq 0$ and $\sum_{\omega\in\Omega} p_\omega = 1$. While it is still possible in this context to employ $z^t_{ij\omega}$ to model the conditional assignment plan of emergency requests in zone i answered by base j at time t under scenario ω and reuse much of the formulation of CCSP2, a difficulty arises in reformulating the probabilistic envelope constraint (3b). Namely, under a discrete distribution $\mathbb{Q}$, it first can be reduced to :

$$\sum_{\omega\in\Omega} p_\omega \mathbf{1}\left\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta(\eta)\sum_{i\in\mathcal{I}} d^t_{i\omega}\right\} \geq 1-\eta \qquad \forall \eta\in[0,1],\, t\in\{1,\dots,T\}\,. \tag{16}$$

Yet, it is not possible to introduce the binary variables that count whether the number of scenarios where the minimum level of coverage is not achieved since this accounting needs to be done for each reliability level η in the continuous range [0, 1]. For this reason, we make a simplifying assumption about the discrete distribution $\mathbb{Q}$, which allows to identify a finite set of constraints that capture the feasible set defined through constraint (16).

Assumption 1 *The discrete distribution $\mathbb{Q}$ is a uniform distribution over all $\omega \in \Omega$. Namely, $p_\omega = 1/N$ for all $\omega \in \Omega$.*

Proposition 4 *Under Assumption 1, constraint (16) is equivalent to the constraint that there exists a $\boldsymbol{\rho} \in \{0,1\}^{(N+1)\times N\times T}$ such that*

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \sum_{k=0}^{N-1} \beta_k(\rho^t_{k+1,\omega}-\rho^t_{k\omega})\sum_{i\in\mathcal{I}} d^t_{i\omega} \qquad \forall t\in\{1,\dots,T\},\omega\in\Omega \tag{17}$$

$$\sum_{\omega\in\Omega}\rho^t_{k\omega} = k \qquad \forall t\in\{1,\dots,T\},k\in\{0,\dots,N-1\} \tag{18}$$

$$\rho^t_{k\omega} \leq \rho^t_{k+1,\omega} \qquad \forall t\in\{1,\dots,T\},\omega\in\Omega,k\in\{0,\dots,N-2\} \tag{19}$$

$$\rho^t_{N\omega} = 1 \qquad \forall t\in\{1,\dots,T\},\omega\in\Omega\,, \tag{20}$$

where $\beta_k := \sup_{\eta<(k+1)/N}\beta(\eta)$.

Using Proposition 4, which proof is deferred to Appendix A.4, we are able to introduce a finite dimensional MILP reformulation for PECSP:

[PEC-SP2]

$$\underset{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{r},\boldsymbol{\rho}}{\text{minimize}} \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} f^t_j x^t_j + \sum_{j\in\mathcal{J}}\sum_{t=1}^{T} g_j y^t_j + (1/N)\sum_{\omega\in\Omega}\sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} c^t l_{ij} z^t_{ij\omega} + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{J}} \alpha^t r^t_{mj} \tag{21a}$$

subject to (1b) − (1e), (4b), (4c), (7c), (7d), (18) − (20)

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \sum_{k=1}^{N-1}(\rho^t_{k+1,\omega}-\rho^t_{k\omega})\lceil\beta_k\sum_{i\in\mathcal{I}} d^t_{i\omega}\rceil \qquad \forall t\in\{1,\dots,T\},\omega\in\Omega \tag{21b}$$

$$x^t_j\in\{0,1\}\,, y^t_j\in\mathbb{N}, \rho^t_{k\omega}\in\{0,1\} \qquad \forall j\in\mathcal{J},t\in\{1,\dots,T\},\omega\in\Omega,k\in\{0,\dots,N\}\,. \tag{21c}$$

The arguments that one needs to use for relaxing the integrality constraints on $\boldsymbol{z}$ and $\boldsymbol{r}$ and rounding up the values $\beta_k \sum_{i\in\mathcal{I}} d^t_{i\omega}$ are exactly similar as in Section 4.

Although we derive a MILP based reformulation, it still poses a great computational challenge given that the number of integer variables and number of constraints now grow at the rate of $O(N^2T)$ in PEC-SP2. Therefore, we propose an exact solution method in Section 5.2 and conservative approximation method in Section 5.3.

5.2 Exact solution method

We briefly summarize how to set-up the problem in order for applying the B&BC method presented in Section 4.2. In particular, we can decompose the following reformulation using a MP and a set of SPs:

$$\text{[PEC-SP2']} \underset{\boldsymbol{x},\boldsymbol{y},\boldsymbol{\rho},\theta^r,\boldsymbol{\theta}^z}{\text{minimize}} \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} f^t_j x^t_j + \sum_{t=1}^{T}\sum_{j\in\mathcal{J}} g_j y^t_j + \sum_{\omega\in\Omega}\sum_{t=1}^{T}\theta^z_{t\omega} + \theta^r \tag{22a}$$

$$\text{subject to } (1b), (1c), (8d), (18)-(20), (21c)$$

$$\theta^r \geq h^r(\boldsymbol{y}) \tag{22b}$$

$$\theta^z_{t\omega} \geq h^z_{\omega,t}(\boldsymbol{y},\boldsymbol{\rho}) \qquad \forall t\in\{1,\dots,T\}, \omega\in\Omega \tag{22c}$$

$$\sum_{j\in\mathcal{J}} \lambda y^t_j \geq \sum_{k=1}^{N-1} (\rho^t_{k+1,\omega} - \rho^t_{k\omega}) \lceil \beta_k \sum_{i\in\mathcal{I}} d^t_{i\omega} \rceil \qquad \forall t\in\{1,\dots,T\}, \omega\in\Omega \tag{22d}$$

$$\sum_{j\in\mathcal{J}} \lambda y^t_j \geq \kappa_t \qquad \forall t\in\{1,\dots,T\}, \tag{22e}$$

where, for each t, we consider the ordered list $\Omega'_t := \{\omega'_k\}_{k=1}^N$ defined in Proposition 3 and let $\kappa_t := \max_{k=0,\dots,N-1} \lceil \beta_k \sum_{i\in\mathcal{I}} d^t_{i,\omega'_{N-k}} \rceil$. While $h^r(\boldsymbol{y})$ is exactly as defined in (9), $h^z_{\omega,t}(\boldsymbol{y},\boldsymbol{\rho})$ needs to be redefined as:

$$\text{[PECSP-SP}^z_{\omega,t}\text{]} \quad h^z_{\omega,t}(\boldsymbol{y},\boldsymbol{\rho}) := \min_{\boldsymbol{z}^t_w \geq 0} \sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} (1/N) c^t l_{ij} z^t_{ij\omega} \tag{23a}$$

$$\text{subject to } \sum_{i\in\mathcal{I}} z^t_{ij\omega} \leq \lambda y^t_j \qquad \forall j\in\mathcal{J} \qquad (\mu^1_{jt\omega}) \tag{23b}$$

$$\sum_{j\in\mathcal{J}} z^t_{ij\omega} \leq d^t_{i\omega} \qquad \forall i\in\mathcal{I} \qquad (\mu^2_{it\omega}) \tag{23c}$$

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \sum_{k=1}^{N-1} (\rho^t_{k+1,\omega} - \rho^t_{k\omega}) \lceil \beta_k \sum_{i\in\mathcal{I}} d^t_{i\omega} \rceil \qquad (\mu^3_{t\omega})\,. \tag{23d}$$

This gives rise to the following dual sub-problem:

$$\text{[PECSP-DSP}^z_{\omega,t}\text{]}$$

$$h^z_{\omega,t}(\bar{\boldsymbol{y}},\bar{\boldsymbol{\rho}}) := \max_{\boldsymbol{\mu}^1_{i\omega},\boldsymbol{\mu}^2_{i\omega},\mu^3_{t\omega}} -\sum_{j\in\mathcal{J}} \lambda \bar{y}^t_j \mu^1_{jt\omega} - \sum_{i\in\mathcal{I}} d^t_{i\omega}\mu^2_{it\omega} + \mu^3_{t\omega} \sum_{k=1}^{N-1} (\rho^t_{k+1,\omega} - \rho^t_{k\omega}) \lceil \beta_k \sum_{i\in\mathcal{I}} d^t_{i\omega} \rceil \tag{24a}$$

$$\text{subject to } -\mu^1_{jt\omega} - \mu^2_{it\omega} + \mu^3_{t\omega} \leq (1/N) c^t l_{ij} \quad \forall i\in\mathcal{I}, j\in\mathcal{J} \tag{24b}$$

$$\boldsymbol{\mu}^1_{t\omega} \geq 0,\ \boldsymbol{\mu}^2_{t\omega} \geq 0,\ \mu^3_{t\omega} \geq 0. \tag{24c}$$

The implementation details of traditional Benders decomposition and the B&BC method are exactly similar to what is presented in Algorithm 1 and Algorithm 2 respectively. Note again that imposing constraint (22d) in the Master Problem ensures that the PECSP-SPs are always feasible and bounded when new Benders cuts are needed by the algorithms while constraint (22e) implements the valid inequality proposed in Section 4.2.3. Finally, it does not appear that tighter optimality cuts as presented in Section 4.2.4 can be identified for PEC-SP2'.

5.3 Conservative approximation method

Although the exact B&BC solution method discussed above can be employed to solve PEC stochastic reformulation, it still struggles with limited scenarios, and is of great challenges to deal with large-scale MILP problems. In view of such computational challenge, we also exploit the special structure of our PECSP reformulation and propose an efficient conservative approximation of PEC-SP2, which significantly reduces the computational burden.

Proposition 5 *Under Assumption 1, the probabilistic envelope constraint (16) is conservatively approximated by*

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega'_{N-k}} \geq \lceil \beta_k \sum_{i\in\mathcal{I}} d^t_{i\omega'_{N-k}} \rceil \qquad \forall t\in\{1,\dots,T\}, k\in\{0,\dots,N-1\}, \tag{25}$$

where $\beta_k := \sup_{\eta<(k+1)/N}\beta(\eta)$, *and the ordered scenario set* $\Omega'_t := \{\omega'_k\}_{k=1}^N$ *is defined such that for all* $k<k'$ *we have* $\sum_{i\in\mathcal{I}} d^t_{i\omega'_k} \leq \sum_{i\in\mathcal{I}} d^t_{i\omega'_{k'}}$.

The proof can be found in Appendix A.5.

This gives rise to the following conservative approximation formulation for PECSP:

[CAPECSP]

$$\underset{\boldsymbol{x},\boldsymbol{y},\boldsymbol{z},\boldsymbol{r}}{\text{minimize}} \sum_{t=1}^T\sum_{j\in\mathcal{J}} f^t_j x^t_j + \sum_{t=1}^T\sum_{j\in\mathcal{J}} g_j y^t_j + (1/N)\sum_{\omega\in\Omega}\sum_{t=1}^T\sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} c^t l_{ij} z^t_{ij\omega} + \sum_{t=1}^T\sum_{j\in\mathcal{J}}\sum_{m\in\mathcal{J}} \alpha^t r^t_{mj} \tag{26a}$$

$$\text{subject to } (1b)-(1e),(4b),(4c),(7c),(7d),(25)$$

$$x^t_j\in\{0,1\}, y^t_j\in\mathbb{N} \qquad \forall j\in\mathcal{J}, t\in\{1,\dots,T\}. \tag{26b}$$

Problem (26) avoids the need to introduce extra binary variables $\rho^t_{\omega k}$ in the model. Furthermore, it can still be treated using the B&BC method by exploiting the following decomposition:

[CAPECSP']

$$\underset{\boldsymbol{x},\boldsymbol{y},\boldsymbol{\theta}^z,\theta^r}{\text{minimize}} \sum_{t=1}^T\sum_{j\in\mathcal{J}} f^t_j x^t_j + \sum_{t=1}^T\sum_{j\in\mathcal{J}} g_j y^t_j + (1/N)\sum_{\omega\in\Omega}\sum_{t=1}^T \theta^z_{t\omega} + \theta^r \tag{27a}$$

$$\text{subject to } (1b),(1c),(8d),(26b)$$

$$\theta^r \geq h^r(y) \tag{27b}$$

$$\theta^z_{t\omega} \geq h^z_{\omega,t}(y) \qquad \forall t\in\{1,\dots,T\},\omega\in\Omega \tag{27c}$$

$$\sum_{j\in\mathcal{J}}\lambda y^t_j \geq \max_{k\in\{1,\dots,N-1\}} \lceil \beta_k \sum_{i\in\mathcal{I}} d^t_{i\omega'_{N-k}} \rceil \qquad \forall t\in\{1,\dots,T\}, \tag{27d}$$

where $h^r(\boldsymbol{y})$ is exactly as defined in (9) but $h^z_{\omega,t}(\boldsymbol{y})$ needs to be redefined as:

[CAPECSP-SP$^z_{\omega,t}$]

$$h^z_{\omega,t}(\boldsymbol{y}) := \min_{\boldsymbol{z}^t_\omega\geq 0} \sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} (1/N) c^t l_{ij} z^t_{ij\omega} \tag{28a}$$

$$\text{subject to } \sum_{i\in\mathcal{I}} z^t_{ij\omega} \leq \lambda y^t_j \qquad \forall j\in\mathcal{J} \qquad (\mu^1_{jt\omega}) \tag{28b}$$

$$\sum_{j\in\mathcal{J}} z^t_{ij\omega} \leq d^t_{i\omega} \qquad \forall i\in\mathcal{I} \qquad (\mu^2_{it\omega}) \tag{28c}$$

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \sum_{k=0}^{N-1} \mathbf{1}\{\omega=\omega'_{N-k}\}\lceil \beta_k \sum_{i\in\mathcal{I}} d^t_{i\omega}\rceil \qquad (\mu^3_{t\omega}). \tag{28d}$$

Note that once again the inequalities (8d) and (27d) are needed to ensure that the problems associated to the evaluation of $h^r(\boldsymbol{y})$ and $h^z_{\omega,t}(\boldsymbol{y})$ are both feasible and bounded. We omit the details of the dual sub-problem associated with $h^z_{\omega,t}(\boldsymbol{y})$ for conciseness.

6 Numerical experiments

To evaluate the numerical performance of our implementation of the B&BC method, we employ a set of randomly generated CCSP and PECSP instances of different sizes. The parameters in each instance are generated as follows. First, each fixed-charge cost f_j^t is generated independently and uniformly from the interval [1000, 1200], each operating maintenance cost g_j from [100, 120], each unit transportation cost c^t from [0.5, 1], and each ambulance redeployment penalty α^t from [3.5, 5]. We then generate each possible emergency request location and each ambulance base location independently and uniformly over the $[0, 10]^2$ square. We also assume that the distribution is uniform over the set of scenario, which size depends on the problem instance. At each request location and for each scenario ω, the number of requests $d_{i\omega}^t$ that can be served is drawn independently and uniformly from the set $\{1, \ldots, 5\}$. The maximum number of on-shift emergency vehicles P_j^t that can be housed at each period t is uniformly drawn $\{5, 6, 7\}$ while the maximum number of requests that can be served by an ambulance in a period is fixed to $\lambda = 4$. In the CCSP model we predefined the coverage level β to be 0.9, while the reliability level is set to $\eta = 0.05$ unless stated otherwise as in Table 1. Finally, in the case of the PECSP model we fix the coverage envelope to $\beta(\eta) = \eta$.

All of our algorithms are programmed in C and use CPLEX 12.71 through the associated API. All runs were conducted on a Desktop machine with Intel(R) Xeon(R) 3.30 GHz processor and 32 GB RAM in a Windows 64-bit system. Only one thread is used in all implementations. For all instances, the algorithms that employ the B&BC method are run until an optimality gap below 1% is reached. In the case of experiments where the CCSP model is solved, we set a maximum runtime of 3600 seconds (1 hour) for all algorithms, while we let the maximum runtime be 7200 seconds for the PECSP model. In both case, CPU time is reported in seconds.

In what follows we start in Section 6.1 by comparing the performances of our proposed implementation of the B&BC method (which we refer to as the B&BC algorithm) to the Branch-and-Bound algorithm proposed in Beraldi and Bruni (2009) on a version of the CCSP model that considers only a single period $T = 1$. We then in Section 6.2 present an exhaustive computational study of the performance of the different variants of the B&BC method for both the CCSP model and the PECSP model where $T = 6$ and evaluate the quality of solutions obtained from the conservative approximation approach presented in Section 5.3.

6.1 Algorithmic comparison with Branch-and-Bound in Beraldi and Bruni (2009)

In Beraldi and Bruni (2009), the authors propose a Branch-and-Bound algorithm to solve a problem of the same form as CCSP yet where a single period is considered. It is therefore possible to assess whether the resolution time decreases significantly with the use of B&BC. To perform this analysis we ran experiments with instances of sizes that are comparable to the sizes discussed in Beraldi and Bruni (2009). Specifically, we considered $|\mathcal{I}| \in \{40, 80, 150\}$, $|\mathcal{J}| \in \{20, 50, 100\}$, and $N \in \{10, 20, 30, 40, 100, 200\}$ while the reliability level was considered with $\eta \in \{0.05, 0.10, 0.15\}$ and $\beta = 0.9$. Our implementation of the B&BC method includes the valid inequalities and tighter optimality cut presented in sections 4.2.3 and 4.2.4. Each experiment involves running both algorithms on the same five random instances of each class of problems.

Table 1 presents the minimum, maximum, and average CPU time required to solve the 5 instances of each class of problem. We also report the proportion of problems that could not be solved exactly in the allotted time (3600s). When some instances cannot be solved to $\epsilon = 1\%$ during this time, the reported average is cumulated over only the instances that were successfully solved. A quick read of this table confirms that the B&BC method achieves significantly better performance than the Branch-and-Bound (BB) algorithm. In fact, the BB algorithm is unable to solve to optimality most of the problem instances that are considered within 3600 seconds. In contrast, our implementation of the B&BC method allows us to solve nearly all problems (267 out 270 problems) within this amount of time. In fact, we notice that the BB algorithm's performance quickly degrades as the number of scenarios considered is increased and similarly, although less drastically, as η increases.

Studying the two algorithms more carefully led us to identify the following argument to explain the lower efficiency of the BB algorithm from Beraldi and Bruni (2009). One needs to know that the BB algorithm starts with the search of all feasible and "non-dominated" assignments for $\boldsymbol{\rho}$ by traversing, using a customized

rule, a tree where each node defines a specific assignment of binary values for $\boldsymbol{\rho}$. Once all of these candidates are identified, the algorithm iteratively solves the MILP associated to each of these candidates. While this can be efficient when η is small and the number of feasible candidates is reasonable, it quickly becomes impossible to assemble such a list for larger η's and more importantly larger N (since the number of branches at each node grows linearly in N). Even once this first step is completed, the algorithm still needs to solve a number of mixed-integer linear programs. In contrast, our B&BC algorithm only solves a single MILP. Moreover, we believe that by treating all integer variables simultaneously, we strongly benefit from an array of perfected heuristic available in CPLEX solver to select which node to focus on and which integer variable to branch on. Overall, it appears clear to us that the B&BC method is much better suited to solve problems of realistic sizes as we discuss next.

6.2 Numerical performance of our B&BC algorithm

This section reports on a set of experiments performed using classes of problems of different sizes, defined by the respective sizes of $|\mathcal{I}|$, $|\mathcal{J}|$, T, and N, for both the CCSP and PECSP model. The definition of each class of problems is described in Table 2 together with details about the total number of integer and continuous decision variables for both the first and second stage, and total number of constraints. Each class of problem contains 10 randomly generated instances of the given size. Overall, the set of test instances includes 240 problems with number of ambulance base locations ranging from 40 to 150, number of emergency request locations ranging from 20 to 100, with 6 time periods, and with discrete distribution supported on a number of scenarios ranging from 50 to 200. We believe that this set of problems covers well the size of problems that should emerge in practice for small to medium sized territories.

The objective of this set of experiments is to compare the numerical performance of different type of implementation of the B&BC method. In particular, we evaluated the CPU time needed to obtain a 1% sub-optimal solution using the following variations of the method:

- Raw-B&BC: this refers to applying Algorithm 2 to solve either CCSP2' (in CCSP study) or PECSP2' (in PECSP study).
- VI-B&BC: this refers to applying Algorithm 2 with the valid inequalities (15) (in CCSP study) and (22e) (in PECSP study).
- VI+OC-B&BC: this refers to applying Algorithm 2 to solve CCSP2' with both the valid inequalities and the strengthened optimality cut from sections 4.2.3 and 4.2.4.
- Raw-CA-B&BC: this refers to solving Algorithm 2 to solve CAPECSP' without constraint (27d).
- VI-CA-B&BC: this refers to solving Algorithm 2 to solve CAPECSP' with constraint (27d).

Note that the Raw-B&BC, VI-B&BC, and VI+OC-B&BC implementations are designed for the CCSP model while Raw-B&BC, VI-B&BC, Raw-CA-B&BC, and VI-CA-B&BC are designed for the PECSP model. In experiments that are not reported here, we also evaluated the performance of other implementations of the B&BC method yet these failed to exhibit performance levels that could be compared to the ones presented in what follows. In particular, we studied the use of Pareto optimality cuts (Magnanti and Wong 1981) and the performance of the integrated Bender strategy in CPLEX (as proposed in Section 4.2.5). The performance of the former is worse than our VI-B&BC, for which we do not report here. In the case of the latter (Bender strategy in CPLEX), the optimality gap typically stayed above 10% after two hours of computing time for small problems of class C13, and 7 out of 10 instances were unsolved optimally after an hour computing time for small problems of class C1.

Similarly as in Section 6.1, Table 3 presents the maximum, and average CPU time required to solve the 10 instances of each class of the CCSP model (C1-C12). We also report the proportion of problems that could not be solved in the allotted time (3600s). When some instances cannot be solved during this time, the reported average is calculated over the subset of instances that were solved. In the case of Raw-B&BC, we also report in Table 4 additional statistics about the optimality gap obtained after 3600 seconds for the instances where a 1% tolerance was not reached. The bottom row of Table 3 finally reports the average of each of these statistics over the set of twelve problem classes. The first thing that one might notice from this table is how the Raw-B&BC algorithm fails to solve most (i.e. 103 out of 120 instances) of the instances. In

Table 1: Numerical performance comparison with Branch-and-Bound method in Beraldi and Bruni (2009).We report the average (avg), maximum (max), minimum (min) CPU time and the proportion of unsolved instances (as is shown in [·]) within 3600 seconds over each five instances.

$\|\mathcal{I}\|$	$\|\mathcal{J}\|$	η	N	Branch-and-Bound			B&BC		
				avg	max	min	avg	max	min
40	20	.05	10	70	134	2	.21	.24	.17
			20	60	177	41	.23	.34	.16
			30	1,483	352	189	.25	.42	.19
			40	3,005*	>3,600[.4]	1,672	.31	.39	.19
			100	-	-	-	.70	1.15	.35
			200	-	-	-	1.36	1.69	.96
		.10	10	144	379	5	.22	.28	.13
			20	489	1374	115	.31	.54	.13
			30	-	-	-	.14	.18	.11
			40	-	-	-	.29	.37	.19
			100	-	-	-	.56	.94	.30
			200	-	-	-	1.28	2.65	.80
		.15	10	141	332	3	.17	.22	.12
			20	1,761*	>3,600[.4]	317	.21	.25	.20
			30	-	-	-	.19	.25	.14
			40	-	-	-	.46	.60	.31
			100	-	-	-	1.14	2.85	.51
			200	-	-	-	1.31	1.86	.79
80	50	.05	10	106	253	4	1.18	1.97	.55
			20	301	1,276	29	1.22	1.61	.79
			30	2,791*	>3,600[.6]	1,271	2.66	4.11	1.23
			40	-	-	-	5.48	8.80	1.99
			100	-	-	-	10.7	15.5	4.60
			200	-	-	-	59.6	200	13.1
		.10	10	107	323	11	.58	.69	.50
			20	1,577*	>3,600[.2]	412	1.17	1.40	1.02
			30	-	-	-	1.78	2.18	1.35
			40	-	-	-	2.70	4.14	1.93
			100	-	-	-	8.37	11.9	6.38
			200	-	-	-	737*	>3,600[.2]	17.8
		.15	10	144	593	11	.65	.92	.41
			20	2,236*	>3,600[.2]	1,328	1.12	1.31	.87
			30	-	-	-	2.35	3.47	1.97
			40	-	-	-	3.23	4.42	2.49
			100	-	-	-	7.27	14.5	2.98
			200	-	-	-	731*	>3,600[.2]	9.00
150	100	.05	10	779	1,900	157	16.8	32.3	7.94
			20	1,133	3,054	307	27	43.7	17.8
			30	3,061*	>3,600[.8]	906	48.5	120	22.1
			40	-	-	-	67.1	127	31.1
			100	-	-	-	108	211	65.6
			200	-	-	-	277	543	156
		.10	10	927	2,830	25	8.38	12.8	3.51
			20	2,472	2,980	1,277	14.3	16.1	11.9
			30	-	-	-	25.4	34.7	19.8
			40	-	-	-	34	42.9	20.0
			100	-	-	-	128	224	49.5
			200	-	-	-	842*	>3,600[.2]	65.6
		.15	10	1,587	2,672	109	15.6	32.2	5.35
			20	3,384*	>3,600[.6]	2,901	22.5	36.2	16.2
			30	-	-	-	47.9	77.4	32.7
			40	-	-	-	46.3	60.5	36.3
			100	-	-	-	145	203	73.9
			200	-	-	-	243	314	186

"-" means that none of the five instances were solved within 3600 seconds;
"*" means the average was computed only on instances for which solution was obtained in less than 3600 seconds.

Table 2: Test data classes and their size for scenario-based two-stage stochastic program with respect to the number of emergency demand zones ($|\mathcal{I}|$), base locations ($|\mathcal{J}|$), time periods (T) and scenarios (N), in which first stage variables are x, y, ρ and r and the second stage variables are z, respectively.

Class	$\|\mathcal{I}\|$	$\|\mathcal{J}\|$	T	N	First Stage		Second Stage	
					# of variables	# of constraints	# of variables	# of constraints
C1	40	20	6	50	2,940	366	240,000	18,300
C2	40	20	6	100	3,240	366	480,000	36,600
C3	40	20	6	150	3,540	366	720,000	54,900
C4	40	20	6	200	3,840	366	960,000	73,200
C5	80	50	6	50	15,900	906	1,200,000	39,300
C6	80	50	6	100	16,200	906	2,400,000	78,600
C7	80	50	6	150	16,500	906	3,600,000	117,900
C8	80	50	6	200	16,800	906	4,800,000	157,200
C9	150	100	6	50	61,500	1,806	4,500,000	75,300
C10	150	100	6	100	61,800	1,806	9,000,000	150,600
C11	150	100	6	150	62,100	1,806	13,500,000	225,900
C12	150	100	6	200	62,400	1,806	18,000,000	301,200
C13	20	10	6	20	3,120	300	24,000	6,000
C14	20	10	6	40	10,320	420	48,000	16,800
C15	20	10	6	60	22,320	540	72,000	32,400
C16	20	10	6	100	60,720	780	120,000	78,000
C17	40	20	6	20	5,040	480	96,000	9,600
C18	40	20	6	40	12,240	600	192,000	24,000
C19	40	20	6	60	24,240	720	288,000	43,200
C20	40	20	6	100	62,640	960	480,000	96,000
C21	80	50	6	20	18,000	1,020	480,000	18,000
C22	80	50	6	40	25,200	1,140	960,000	40,800
C23	80	50	6	60	37,200	1,260	1,440,000	68,400
C24	80	50	6	100	75,600	1,500	2,400,000	138,000

Table 3: Numerical performance for CCSP with respect to three implementations summarized at the beginning of Section 6.2, in which we report the average (avg), maximum (max) CPU time (in seconds) and the unsolved proportion (prop) within 3600 seconds over ten instances for each data class (C1-C12), $\beta = 0.9, \eta = 0.05$.

Class	Raw-B&BC			VI-B&BC			VIOC-B&BC		
	avg	max	prop	avg	max	prop	avg	max	prop
C1	-	-	1	2	4	.00	2	5	.00
C2	27*	>3,600	.9	3*	>3,600	.1[2.13]	3*	>3,600	.1[2.18]
C3	76*	>3,600	.9	4	6	.00	4	6	.00
C4	89*	>3,600	.6	13	60	.00	6	11	.00
C5	166*	>3,600	.8	13	16	.00	12	15	.00
C6	151*	>3,600	.9	23	38	.00	23	45	.00
C7	-	-	1	79	528	.00	74	465	.00
C8	77*	>3,600	.9	53	95	.00	48	91	.00
C9	554*	>3,600	.6	342*	>3,600	.1[1.26]	80	163	.00
C10	1426*	>3,600	.8	175	241	.00	144	203	.00
C11	-	-	1	218	378	.00	213	283	.00
C12	1137*	>3,600	.9	269	279	.00	260	208	.00
average			.86	100*		.02	73*		.008

"-" indicates that no instance were solved in less 3600 seconds;
"*" means the average was computed only on instances for which solution was obtained in less than 3600 seconds.
[·] in column of *prop* means the average closing optimality gap (in %) for instances beyond 3600 seconds.

the case of VI-B&BC and VI+OC-B&BC algorithms, the performance is surprisingly better given that the failure rate is below 2%. One might also notice that even with the easier problems, the Raw-B&BC algorithm seems to struggle much more than the other two approaches in terms of average and maximum CPU time. This seems to confirm that the use of the valid inequalities proposed in Section 4.2.3 plays a critical role in improving the numerical efficiency. The table also reveals that the improved optimality cuts proposed in Section 4.2.4 lead to an average reduction of about 25% of CPU time. Overall, one can certainly draw

the conclusion that both the VI-B&BC and VI+OC-B&BC algorithms are valuable approaches for solving larger-scale CCSP models.

Table 4: Optimality gap of Raw-B&BC for CCSP in terms of the average (avg), maximum (max) and minimum (min) optimality gap within 3600 seconds, and the unsolved proportion (prop) over ten instances for each data class (C1-C12).

Class	avg %	max %	min %	prop
C1	3.93	6.40	2.37	1
C2	3.36	5.91	1.59	.9
C3	4.12	6.99	3.05	.9
C4	4.38	5.81	2.32	.6
C5	2.02	3.22	1.06	.8
C6	2.07	2.76	1.01	.9
C7	2.44	3.31	1.50	1
C8	3.15	6.44	1.13	.9
C9	1.18	1.42	1.04	.6
C10	1.69	2.23	1.06	.8
C11	2.42	3.83	1.44	1
C12	2.70	4.97	1.47	.9
average	1.66	4.41	1.58	.71

Table 5 presents the numerical performance of four algorithms for PECSP models of class C13-C24 using a format that is exactly similar to Table 3 except for the fact that maximum runtime is now 7200 seconds. The observations here are similar as before concerning the superior performance of VI-B&BC over Raw-B&BC. We see for instance that the success rate passes from 2% to 46% using VI-B&BC. Yet, these results raise a serious concern that exact methods can reliably be used to solve PECSP model of realistic sizes. The algorithms that are based on the conservative approximation CAPECSP appear much more promising in this regard. All problem instances except one were solved within the allotted 7200 seconds. Furthermore, over this set of 119 instances, the maximum resolution time was about 7 minutes. One also remarks that the use of valid inequalities (22e) in the B&BC method is especially useful to accelerate completion time.

The fact that the conservative model CAPECSP can be solved so efficiently raises the question of what is the loss in terms of quality. Table 6 sheds some light on this issue by presenting the average of the lower bound identified by the exact method VI-B&BC in 7200s and its remaining optimality gap for problem instances of each class (C13-C24). It also presents for both the average optimal value of the solutions produced by the CAPECSP model, the average optimality gap (based on the obtained lower bounds). Finally, we express in the sixth and ninth columns the difference between the average gap for the conservative model and the average gap reached by the exact method. Overall, the sizes of average optimality gaps for Raw-CA-B&BC and VI-CA-B&BC seem to confirm that the solution from the conservative model are of good quality. In fact, based on the fact that for 7 out of the 12 problem classes the difference in optimality gap for VI-CA-B&BC is negative and the average value is only 1.6%, one might even argue that the conservative solutions are of marginally better quality than the feasible solutions reached for the exact model after 2 hours of computations.

7 A Northern Ireland Ambulance service Health and Social Care Trust case study

In this case study, we consider the design of an EMS network in the region of Northern Ireland, located in the north-east of island. We start by describing the context in which we are assuming that this EMS network is being operated together with the historical dataset that was used to perform our analysis. We then present and discuss the structure of the optimal networks obtained from solving the CCSP and PECSP models. Section 7.3 illustrates our findings regarding the trade-offs that can be made in this region between expected total installation and operation cost and coverage reliability. We conclude this case study by performing an out-of-sample analysis that confirms that our conclusions should remain relatively valid when looking forward in time.

Table 5: Numerical performance for PECSP with respect to four implementations summarized at the beginning of Section 6.2, in which we report the average (avg), maximum (max) CPU time (in seconds) and the unsolved proportion (prop) over ten instances for each data class (C13-C24).

Class	Raw-B&BC			VI-B&BC			Raw-CA-B&BC			VI-CA-B&BC		
	avg	max	prop	avg	max	prop	avg	max	prop	avg	max	prop
C13	-	[20.3]	1[10.8]	112	498	.00	2	6	.00	1	1	.00
C14	-	[46.7]	1[33.2]	3,764	6,553	.1[1.78]	4	11	.00	1	5	.00
C15	-	[48.2]	1[44.7]	6,547	6,928	.4[3.11]	14	31	.00	9	32	.00
C16	-	[56.2]	1[47.9]	-	[18.84]	1[8.90]	28	134	.00	12	35	.00
C17	1,822*	[18.2]	.7[7.08]	552	1,694	.3[1.11]	6	27	.00	6	35	.00
C18	-	[43.8]	1[36.9]	4,876*	[3.98]	.7[2.28]	96	619	.00	5	20	.00
C19	-	[49.0]	1[42.4]	6,930*	[8.67]	.7[2.83]	25	64	.00	6	11	.00
C20	-	[57.6]	1[49.5]	5,224*	[3.98]	1[2.28]	66	182	.00	75	378	.00
C21	-	[19.1]	1[7.67]	752	2,407	.4[1.43]	52	158	.1[1.60]	10	15	.1[1.18]
C22	-	[44.8]	1[40.4]	2,236	4,903	.4[1.20]	65	340	.00	45	145	.00
C23	-	[49.6]	1[45.7]	6,400*	[5.97]	.7[3.54]	278	743	.1[2.20]	61	142	.00
C24	-	[65.3]	1[51.9]	-	[12.54]	1[6.52]	497	1,542	.1[1.11]	155	242	.00
average	-	[43.2]	.98[34.8]	-	-	.54[3.18]	94	321	.03[.04]	32	88	.01[.09]

"-" indicates that no instance were solved in less 7200 seconds;
"*" means the average was computed only on instances for which solution was obtained in less than 7200 seconds.
[·] in column of *max* denotes the maximum closing optimality gap if more than half of instances are unsolved optimally;
[·] in column of *prop* means the average closing optimality gap for instances beyond 7200 seconds.

Table 6: The performance of conservative approximation for PECSP when comparing with exact method of VI-B&BC, in which avg gap for VI-B&BC is averaged closing optimality gap over ten instances for each data class, avg CA opt is the average objectives for Raw-CA-B&BC and VI-CA-B&BC, avg gap for Raw-CA-B&BC and VI-CA-B&BC can be computed by $\mathbb{E}$[(CA-opt − LB)/LB × 100%], and Δ measures the average gap difference (in percentile points) between our conservative approximation and exact method.

Class	VI-B&BC		Raw-CA-B&BC			VI-CA-B&BC		
	avg LB	avg gap%	avg CA opt	avg gap%	Δ%	avg CA opt	avg gap%	Δ%
C13	33,459	0.96	33,834	1.12	0.16	33,869	1.23	0.27
C14	32,016	1.02	32,445	1.34	0.32	32,447	1.35	0.33
C15	33,871	1.77	34,526	1.93	0.16	30,901	8.77	7.00
C16	32,661	8.90	34,834	6.65	-2,25	34,815	6.60	-2.30
C17	63,900	0.90	58,133	9.02	8.12	64,590	1.08	0.18
C18	64,405	1.85	65,713	2.03	0.18	65,737	2.07	0.22
C19	63,411	2.31	64,243	1.31	-1.00	64,282	1.37	-0.94
C20	64,406	1.84	63,910	0.77	-1.07	63,916	0.76	-1.08
C21	128,061	1.03	128,883	0.64	-0.39	128,880	0.64	-0.39
C22	127,039	0.84	127,605	0.45	-0.39	127,633	0.47	-0.37
C23	133,069	3.05	135,032	1.48	-1.57	135,110	1.53	-1.52
C24	127,879	6.32	129,146	0.99	-5.33	129,189	1.02	-5.30

7.1 Context

The total population of Northern Ireland is about 1,870,451. We consider the territory to be divided into 80 emergency demand regions, each of which consists of the area where the same 4-digit ZIP code is used. We also consider 63 potential locations for ambulance based on existing EMS service stations. Figure 1 illustrates these geographical details.

We use historical emergency demand request data gathered by the Northern Ireland Ambulance Service Health & Social Care Trust over the period ranging from April 2015 until December 2016 (21 months).

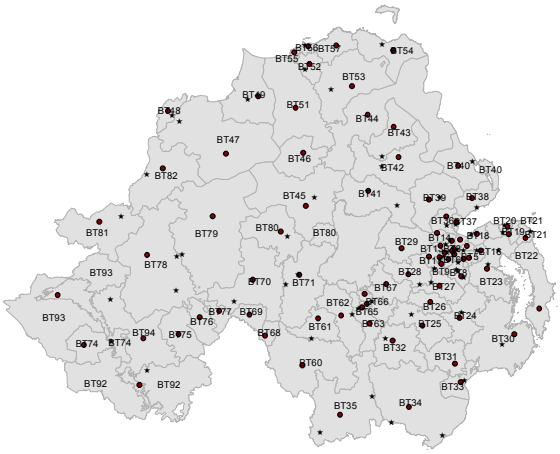


Figure 1: ZIP code based geographical partition for Northern Ireland. Solid dots represent 80 emergency demand zones and stars represent 63 potential base locations respectively ($|\mathcal{I}| = 80, |\mathcal{J}| = 63$).

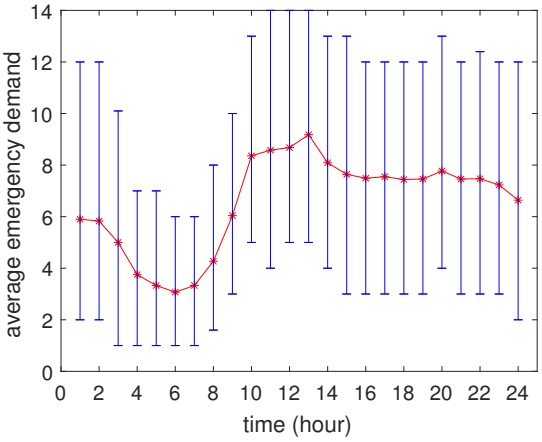


Figure 2: Average emergency demand in 24-hour cycle, based on emergency demand data during 03/2015-12/2016. The bars represent the 5% and 95% quantiles on the hourly emergency demand respectively.

More than 113,000 emergency requests are reported in Northern Ireland over this period. Figure 2 shows statistics about the number of requests on this territory at different hourly periods of a day during these 21 months. We can observe that the statistics of emergency requests are highly time-dependent. For example, emergency demand stays below 5 requests per hour 4:00am and 8:00am, while from 10:00am until 12:00am, it stays above 8 requests per hour. Moreover, we observed that this demand is unevenly distributed over the Northern Ireland region. In our CCSP and PECSP models, we consider that a day is divided into 6 periods of 4 hours each (i.e. 12:00am - 4:00am, 4:00am - 8:00am, ..., 8:00pm - 12:00am) and let each day of our historical dataset identify a possible scenario for the emergency demand. Distances are computed using geographical distances obtained using the ArcGIS software, between the center of the ZIP code zone and the location of the ambulance base. This appears reasonable given that most of the ZIP code regions are relatively small (i.e. an area of 48 square miles on average). We let the time-dependent fixed-charge cost be $f := [200\ \ 300\ \ 350\ \ 450\ \ 500\ \ 500]^\intercal$ to model decreasing marginal cost as the ambulance base is run for a longer time period, given that bases that are open need to stay open until the end of the day. We further let the marginal travel cost take the form $c := [2\ \ 2\ \ 1\ \ 1\ \ 1.5\ \ 1.5]^\intercal$ and let ambulance relocation costs as $\alpha := [5\ \ 5\ \ 4\ \ 4\ \ 4.5\ \ 4.5]^\intercal$ to model higher salaries for late night or early morning shifts. The operating and maintenance cost g_j is set uniformly to 40. We assume that each ambulance can serve at most $\lambda := 2$ emergency requests in each time period while the maximum capacity of on-shift vehicles at each ambulance base is assumed to be $P_j := [2\ \ 1\ \ 2\ \ 3\ \ 2\ \ 1]^\intercal$ irrespective of the location j in order to simulate periods of the day where hospitals are more congested and cannot host as many ambulances.

In all experiments, the distribution of demand requests consists of the empirical distribution of a set of observed daily demand realizations from the historical dataset. In the experiments conducted in Sections 7.2 and 7.3, the CCSP model uses a subset of 366 contiguous days (January 2016 - December 2016) while the PECSP model uses only 100 observations (January 2016 - mid April 2016). In Section 7.4, we will investigate the quality of solutions obtained from such empirical distributions by varying both sample size. In all three sections, PECSP is approximated using CAPECSP in order to accelerate computations and the two models are solved using VIOC-B&BC and VI-CA-B&BC respectively.

For the CCSP model, we will usually consider that the EMS manager wants to ensure that 95% of the emergency requests can be satisfies with more than 95% probability (i.e., $\beta = 0.95$ and $\eta = 0.05$). The exception will be in Section 7.3 where the sensitivity of total expected cost to the choice of both β and η will be studied.

In the case of the CAPECSP model, we constructed probabilistic coverage envelope functions based on the following reasoning. Given a reference first stage solution $(\bar{\boldsymbol{x}}, \bar{\boldsymbol{y}})$, e.g. an optimal solution to the DM formulation, we evaluate the coverage profile $\bar{\beta}(\eta)$ that is achieved by this candidate solution. Considering

that this coverage profile is considered as a benchmark for the decision maker, we assume that she/he is interested in finding a solution that reduces by a factor of two the proportion of uncovered population for each reliability level η. This gives rise to the probabilistic envelope $\beta^{cvx}_{0.5}(\eta) := 0.5(1 + \bar{\beta}(\eta))$ (presented in Figure 3), since for each η we want $1 - \beta^{cvx}_{0.5}(\eta) = 0.5(1 - \bar{\beta}(\eta))$. We refer the reader to Appendix B.1 for more details on how to obtain $\bar{\beta}(\eta)$. Similarly as for the CCSP model, in Section 7.3 we study the sensitivity of the total expected cost to the choice of the $\beta(\eta)$ function.

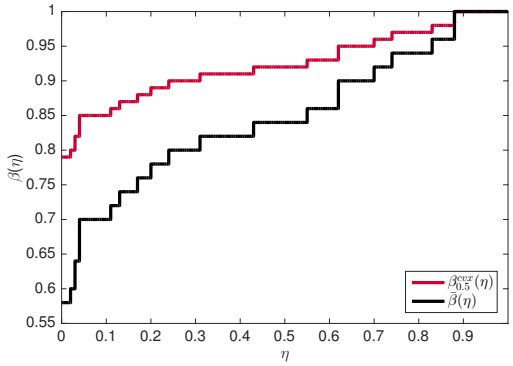


Figure 3: Probabilistic envelope function $\beta^{cvx}_{0.5}(\eta)$ used in CAPECSP model for Sections 7.2 and 7.4. The figure also presents our reference coverage profile $\bar{\beta}(\eta)$ constructed from the solution of the DM problem.

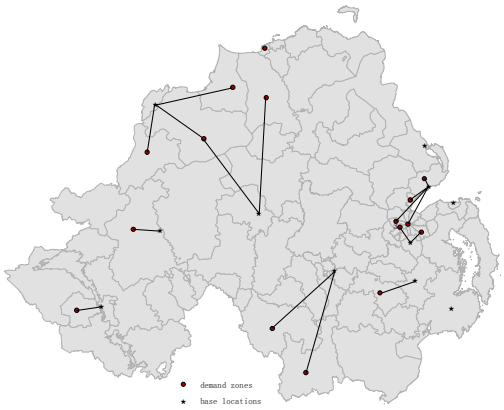
Remark 1 *We note that although we argue that the context described above is a realistic one for Northern Ireland, the parameters f, c, α, λ are in fact artificially chosen in this case study. While we believe the practical insights that will be drawn in our analysis should hold under more accurate choices of parameters, additional efforts would be needed to calibrate the CCSP and CAPECSP models in a way that can provide precise guidelines on how to improve the EMS system in this region.*

7.2 Time-dependent optimal EMS network configuration

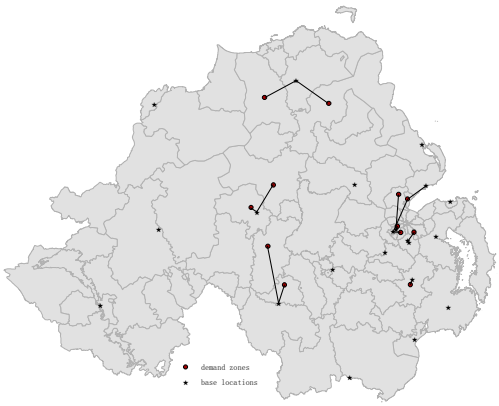
We start by presenting in Table 7 some information about the structure of the two optimal strategies obtained from solving the CCSP and CAPECSP models. In particular, one might recognize that a large number of ambulance bases are needed at time $t = 2$ and that these bases remain open (based on our modeled requirement) although they are not all occupied in future periods. The table also presents information about the optimal activity, which is illustrated in Figures 4 and 5, under some scenarios ω_1 and ω_2 for the CCSP and CAPECSP models respectively. In the case of CCSP, we clearly see that scenario ω_1 was identified as an extreme scenario for time $t = 3$ where the number of emergency requests reaches 34. Consequently, the plan for this scenario and time period simply recommends keeping all ambulances idle at there base. Scenario ω_2 is also considered a hard scenario to cover by the CAPECSP model yet we see that the use of a probabilistic envelope leads to an optimal strategy that still recommends covering a reasonable proportion of the requests, albeit achieves an 80.95% coverage for the hard scenario of $t = 1$. This is achieved at the price of incurring 2 500$ overhead in terms of expected total cost and some reduced coverage for the more optimistic scenarios. Finally, we note that the possible trade-off between expected cost and coverage at any reliability level η can easily be explored using the PECSP or CAPECSP models. This is in fact the subject of our next section.

7.3 Trade-off between cost management and coverage

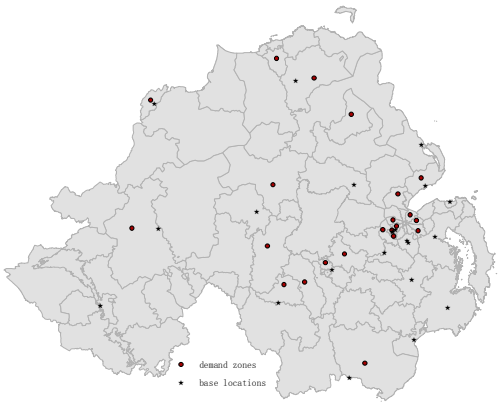
We next performed a sensitivity analysis to investigate what are the trade-offs between expected total cost and emergency request coverage performance in the region of Northern Ireland. In particular, we first focus on the CCSP model and consider the effect of changing the reliability level and the coverage target, parametrized by η and β respectively. The result of this analysis is presented in Figure 6 where one can already remark that a lower permitted violation probability η and a higher coverage target β necessarily leads to an increase in expected total cost. Looking more closely at Figure 6(a), one might realize that the



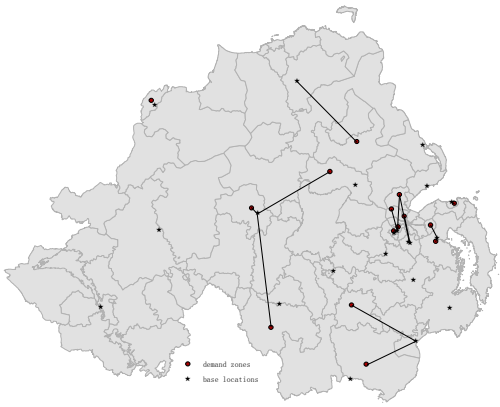
(a) $t = 1$, $\tilde{\beta}^1 = 0.95$



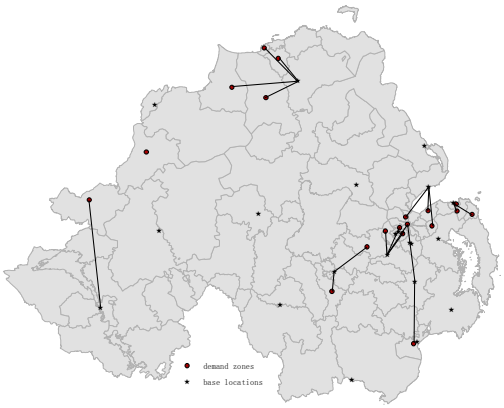
(b) $t = 2$, $\tilde{\beta}^2 = 1$



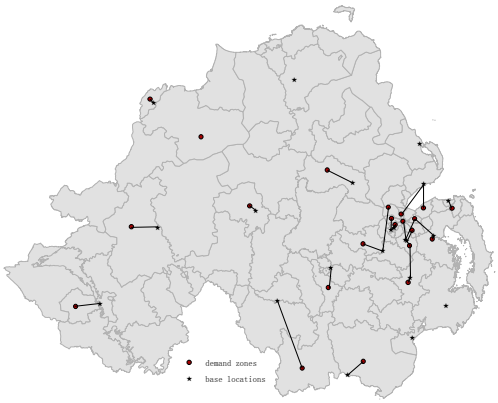
(c) $t = 3$, $\tilde{\beta}^3 = 0$



(d) $t = 4$, $\tilde{\beta}^4 = 0.9545$



(e) $t = 5$, $\tilde{\beta}^5 = 0.9545$



(f) $t = 6$, $\tilde{\beta}^6 = 0.9688$

Figure 4: Time-dependent optimal EMS network configuration for CCSP under scenario ω_1 over 6 time periods ($\beta = 0.95, \eta = 0.05$). The total emergency demand is 20, 14, 34, 22, 22 and 32 respectively. We also report the empirical coverage for each time period by $\tilde{\boldsymbol{\beta}}^{\boldsymbol{t}}$ under scenario ω_1. Solid dots represent the emergency demand zones, stars represents the opened bases and the black line represents the assignment respectively.

(a) $t = 1$, $\tilde{\beta}^1 = 0.8095$

(b) $t = 2$, $\tilde{\beta}^2 = 0.8636$

(c) $t = 3$, $\tilde{\beta}^3 = 0.9143$

(d) $t = 4$, $\tilde{\beta}^4 = 0.9143$

(e) $t = 5$, $\tilde{\beta}^5 = 1$

(f) $t = 6$, $\tilde{\beta}^6 = 0.8684$

Figure 5: Time-dependent optimal EMS network configuration for CAPECSP under scenario ω_2 over 6 time periods with $\boldsymbol{\beta}^{\boldsymbol{cvx}}_{\mathbf{0.5}}(\boldsymbol{\eta})$. The total emergency demand are 42, 22, 35, 35, 27 and 38 respectively. We also report the empirical coverage for each time period by $\tilde{\boldsymbol{\beta}}^{\boldsymbol{t}}$ under scenario ω_2. Solid dots represent the emergency demand zones, stars represents the opened bases and the black line represents the assignment respectively.

Table 7: Characteristics of optimal strategies proposed by the CCSP and CAPECSP models. The scenarios $\boldsymbol{\omega_1}$ and $\boldsymbol{\omega_2}$ are representative scenarios respectively for CCSP and CAPECSP where lower coverage is achieved.

CCSP: Expected total cost = 52 794$						
Time periods	1	2	3	4	5	6
Total number of opened ambulance bases	11	21	21	21	21	21
Total number of occupied bases	11	21	11	8	11	21
Total number of bases providing service in ω_1	8	8	0	7	7	17
Total emergency requests in ω_1	20	14	34	22	22	32
Effective coverage level for ω_1	0.95	1	**0**	0.95	0.95	0.97
CAPECSP: Expected total cost = 55 294$						
Time periods	1	2	3	4	5	6
Total number of opened ambulance bases	11	22	22	22	22	22
Total number of occupied bases	11	22	14	11	14	22
Total number of bases providing service in ω_2	10	11	13	11	11	18
Total emergency requests in ω_2	42	22	35	35	27	38
Effective coverage level for ω_2	**0.81**	0.86	0.91	0.91	1	0.87

marginal cost per percentile point (p.p.) of additional reliability becomes more expensive in the region where $\eta \leq 0.05$. Specifically, for $\beta = 0.95$ the rate is around -3 030 $ per p.p. in this region compared to a rate of around -267 $ per p.p. for $\eta \in [0.05,\ 0.9]$. This is due to the fact that the EMS system must now be prepared for the more extreme scenario, which typically occur in Fall and Winter, requiring a larger fleet size to achieve a higher reliability, in turn which incurs larger fixed cost for resources that are unused under most scenarios. The total expected cost also drastically falls to zero when η gets to 1 due to the fact that the EMS system managers can then simply completely shut down its operations.

In Figure 6(b) one can observe that the effect of increasing the targeted coverage β on expected total cost is nearly linear and grows at a rate between 503 $ per p.p. and 597 $ per p.p. depending on the magnitude of η. It also appears that the marginal cost is higher when the permitted violated probability is smaller possibly due to the fact that the emergency requests that need to be covered in the more extreme scenario are either more numerous or more spread out on the territory. The fact that the marginal cost appears nearly constant when η is fixed is particularly interesting. Intuitively, this is due to the fact that as β is increased the EMS system manager can continue to focus his efforts on the same scenarios of emergency demand, gradually investing in more resources in order to increase his cover of these scenarios. This is unlike the case of a marginal change of η, which can force the EMS system managers to deal with new and more extreme scenarios thus creating a non-linear effect on the amount of resources needed.

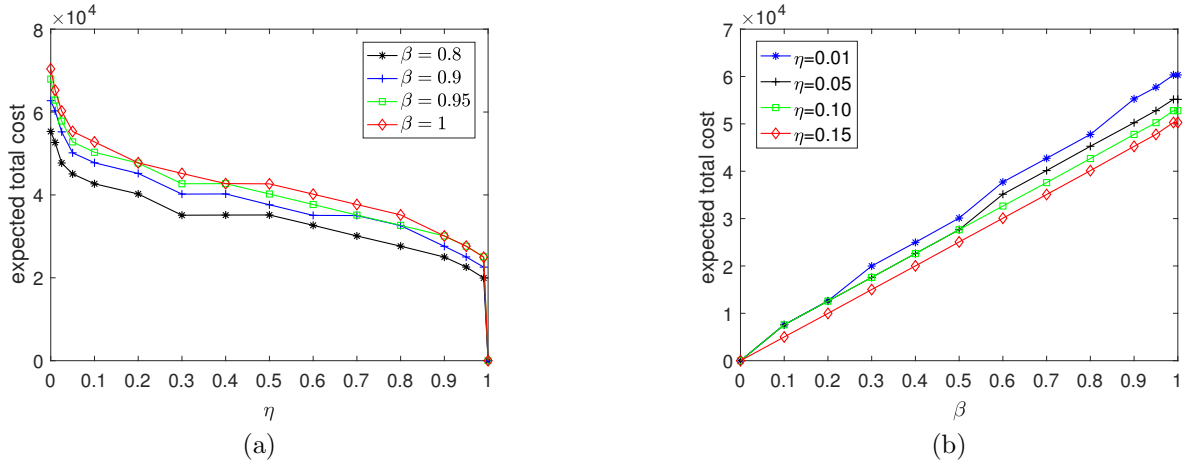
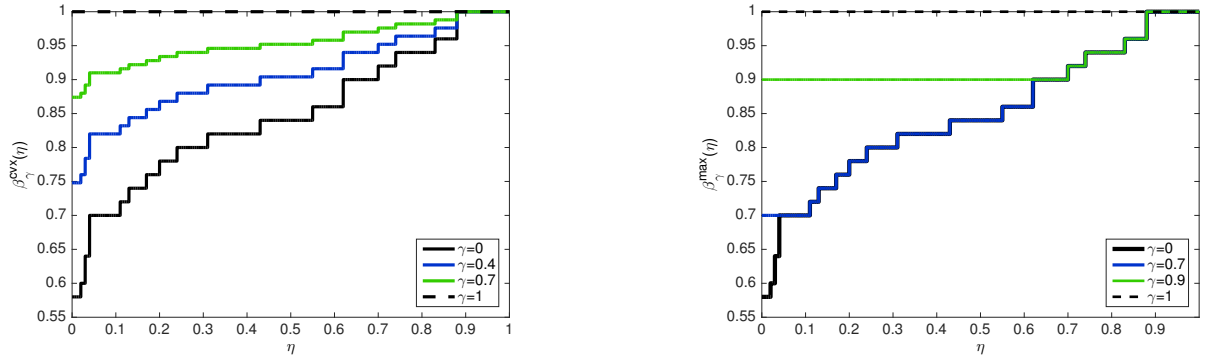


Figure 6: Sensitivity of optimal expected total cost to targeted coverage and reliability in the CCSP model. (a) presents the optimal expected total cost as a function of the permitted violation probability $\boldsymbol{\eta}$ for different coverage level $\boldsymbol{\beta}$. (b) presents the expected total cost as a function of the coverage level $\boldsymbol{\beta}$ for different values of permitted violation probabilities $\boldsymbol{\eta}$.

We also analysed the sensitivity of the optimal expected total cost of the CAPECSP model for different choices of probabilistic coverage envelope $\beta(\eta)$. To do so, we considered two parametric families of envelope. The first family extends $\beta^{cvx}_{0.5}(\eta)$ by employing different convex combination of $\bar{\beta}(\eta)$ and 1, which denotes a full coverage of emergency requests with probability one. Specifically, we consider $\beta^{cvx}_{\gamma}(\eta) := \gamma + (1-\gamma)\bar{\beta}(\eta)$. Note that $\beta^{cvx}_{\gamma}(\eta)$ allows to generate a continuum of increasingly restrictive envelopes between $\bar{\beta}$ and 1 as γ goes from 0 to 1, i.e. if $\gamma_1 \leq \gamma_2$ then $\beta^{cvx}_{\gamma_1}(\eta) \leq \beta^{cvx}_{\gamma_2}(\eta)$ for all η. Alternatively, we consider a second monotone parametric family of probabilistic coverage envelopes that takes the form $\beta^{max}_{\gamma}(\eta) := \max\big(\gamma,\ \bar{\beta}(\eta)\big)$, which simply ensures that the demand coverage is at least γ over all scenarios. Examples of both of these parametric families of probabilistic coverage envelopes are presented in Figure 7. Our sensitivity analysis consists in evaluating the optimal expected total cost achieved in the CAPECSP model with both $\beta^{cvx}_{\gamma}(\eta)$ and $\beta^{cvx}_{\gamma}(\eta)$ and for values of $\gamma \in \{0,\ 0.1,\ 0.2,\ 0.3,\ 0.4,\ 0.5,\ 0.6,\ 0.7,\ 0.8,\ 0.9,\ 1\}$.



(a) (b)

Figure 7: Examples of probabilistic coverage envelopes obtained using the parametric families $\beta^{cvx}_{\gamma}(\eta)$ (in (a)) and $\beta^{max}_{\gamma}(\eta)$ (in (b)).

Figure 8 presents the trade-off between expected coverage required by the probabilistic envelope, i.e. $\int_0^1 \beta(\eta)d\eta = \frac{1}{N}\sum_{k=0}^{N-1}\beta_k$, and the expected total cost. In words, expected coverage reflects the overall probability that an emergency request, drawn randomly from any scenario and uniformly from the requests made in that scenario, ends up being served by an ambulance in the solution that satisfies the probabilistic envelope constraint. For example, the expected coverage in the CCSP model is $0.95 \cdot 0.95 + 0.05 \cdot 0 = 0.9025$ when $\beta = 0.95$ and $\eta = 0.05$. First, one can notice from Figure 8 that given a targeted expected coverage, the choice of probabilistic envelope has a significant impact on the investment that needs to be made. In particular, the CAPECSP model that employs $\beta^{max}_{\gamma}(\eta)$ identifies more expensive solutions. This is because all of the investment goes in improving coverage in the most extreme scenarios, which is more expensive (in \$ per p.p.) than in the less extreme ones since extreme scenarios have larger total demand. In particular, looking at the curve associated to $\beta^{max}_{\gamma}(\eta)$, we note that it costs about 9 946\$ to raise the worst-case coverage from 57% to 70%, which effectively accounts for less than 1 p.p. of expected coverage. Otherwise, we see that when using $\beta^{cvx}_{\gamma}(\eta)$ the marginal cost for improving expected coverage is nearly constant and close to 2 000 \$ per p.p.

7.4 Out-of-Sample analysis

This section focuses on a out-of-sample analysis of the quality of solutions obtained from the CCSP and CAPECSP models. Generally speaking this part of our study serves two purposes. The first one is to highlight the peculiarities associated to the implementation of a solution of the CCSP/PECSP/CAPECSP models given that these are approximated using scenarios. In particular, one should expect in practice that the realized emergency demand scenario is not a member of the assumed scenario list. The question therefore arises regarding how much coverage should be provided since the optimal policy, namely $\boldsymbol{\rho}^*$, is not defined for this scenario. The second objective of this study aims at verifying whether the performance, in terms of expected total cost and probabilistic coverage, that is estimated for an optimal EMS network design using historical data can be generalized into the future.

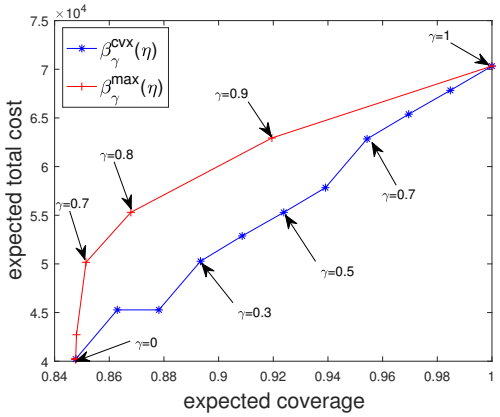


Figure 8: Trade-off between expected total cost and expected coverage for CAPECSP model using either $\beta_\gamma^{cvx}(\eta)$ or $\beta_\gamma^{max}(\eta)$ with respect to different values of γ.

We first shed some light on how the solution of our CCSP, PECSP, or CAPECSP models can be implemented in practice. Specifically, we focus on the CCSP model but a similar approach can be used for the PECSP and CAPECSP. Given an optimal solution $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{r}^*, \boldsymbol{z}^*)$ to the CCSP that employs an empirical distribution based on the realizations $\{\boldsymbol{d}_k\}_{k=1}^N$, one should commit to implementing the plan prescribed by $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{r}^*)$. The peculiarity comes at any time period $\bar{t}$ where a realization $\bar{\boldsymbol{d}}^{\bar{t}}$ occurs and is most likely not part of $\{\boldsymbol{d}_k^{\bar{t}}\}_{k=1}^N$. In order to identify the right assignment to proceed with, one can solve once again the CCSP model but with $(\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{r}) = (\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{r}^*)$ being fixed, and with a new set of empirical realizations $\{\boldsymbol{d}'_k\}_{k=1}^N$ where each $\boldsymbol{d}_k^{t\,\prime} = \boldsymbol{d}_k^t$ except for some randomly chosen $\bar{k}$ for which $\boldsymbol{d}_{\bar{k}}^{\bar{t}} = \bar{\boldsymbol{d}}^{\bar{t}}$. One can then implement the new optimal assignment described as $\hat{\boldsymbol{z}}^{\bar{t}}(\bar{\boldsymbol{d}}^{\bar{t}}) := \boldsymbol{z}_{\bar{k}}^{\bar{t}\,**}$. The argument that supports this procedure is that one should consider the new scenario as a scenario that should have been considered equally likely as any other anticipated scenario and to verify how would this new scenario be treated by the CCSP decision model. Depending on the characteristics of the observed scenario, the modified CCSP model will decide whether it should be considered as a scenario that is too extreme to cover or not. In what follows, our out-of-sample experiments will confirm that this procedure produces a policy that has similar performance to the performance anticipated by the in-sample version of CCSP.

An out-of-sample experiment consists of the following. We first take a random subset of N days from the "training" period of January 2016 to December 2016. The CCSP or CAPECSP model is then solved using the empirical distribution over the N observations. The optimal policy is then implemented (as described above) on every day of the "testing" period spanning from April 2015 to December 2015 (275 days) and statistics are obtained regarding the average total cost and empirical coverage. In the case of the CCSP, we let $N \in \{50,\ 100,\ 200\}$ while for the CAPECSP model we focus on $N \in \{20,\ 60,\ 100,\ 200\}$. Each figure presents statistics cumulated over 10 different experiments.

Figures 9 and 10 present statistics of the in-sample and out-of-sample cost incurred by both CCSP and CAPECSP respectively. These statistics are presented in the form of box plots indicating the minimum, maximum, and first, second, and third quartiles of the values obtained over the 10 simulations. In both cases we can observe that, when the training set is small (i.e. less than 50 observations), the expected cost varies significantly both in-sample and out-of-sample depending on the particular subset of training data that is used. This is due to to the fact that small sets of observations are inaccurate representations of the distribution captured by the training (and even testing set). Namely, depending on the specific subset of scenarios that is used, which might contain mostly easy or mostly hard scenarios to cover, the CCSP and CAPECSP might be overly optimistic or overly pessimistic regarding how much needs to be invested in the EMS network to meet the coverage requirements. It also appears that the out-of-sample expected cost is generally lower than its in-sample estimate (see both Figure 9(b) and Figure 10(b), and the further Table 8). While this is could be interpreted as good news, in the case of CCSP the reduction in cost comes with a slight degradation of coverage out-of-sample. This does not seem to be the case for CAPECSP for which

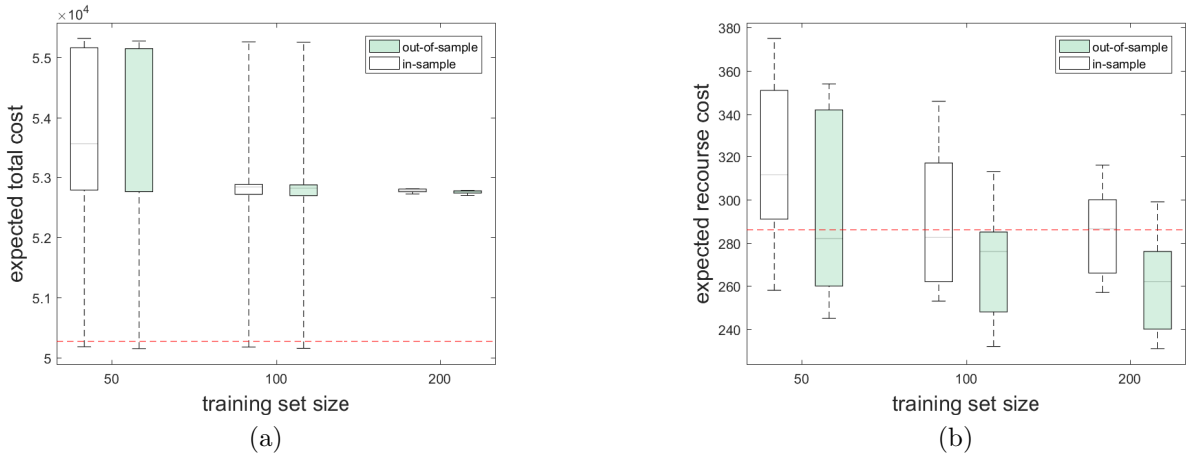


Figure 9: Out-of-sample analysis of expected cost in CCSP with $\beta = 0.95$ and $\eta = 0.05$. (a) presents box plots for the expected total cost achieved in-sample and out-of-sample by solutions of the CCSP model over 10 experiments. The red dashed line is the best expected total cost that can be achieved out-of-sample. (b) presents similar box plots for the average recourse cost.The red dashed line is the out-of-sample expected recourse cost achieved by the best out-of-sample policy.

the reduction is better explained by an out-of-sample distribution that is slightly easier to cover. It is also notable that, even when 200 observations are used, both the CCSP and CAPECSP models identify policies that achieve an out-of-sample expected total cost that is respectively 5% and 22% sub-optimal with respect to the best policy that could have been used out-of-sample. This could perhaps again be due to the fact that the in-sample and out-of-sample distributions of emergency request are structurally different given that they cover different periods, i.e. 2015 vs. 2016. The larger sub-optimality for CAPECSP could be explained by the fact that the probabilistic envelope constraint is more sensitive to the distribution than the chance constraint thus making the expected total cost similarly more sensitive.

Table 8 provides additional information about the mean and standard deviation of in-sample and out-of-sample expected recourse cost for both CCSP and CAPECSP. In this table, we also report the average relative generalization error that can be computed as

$$\text{avg-error} := \frac{1}{10}\sum_{l=1}^{10} \frac{\text{in-sample expected recourse cost}_l - \text{out-of-sample expected recourse cost}_l}{\text{in-sample expected recourse cost}_l},$$

where $l \in \{1, \ldots, 10\}$ is the index of the 10 different experiments. It is particularly noteworthy that for all values of N, for both the CCSP and the CAPECSP models, the average relative generalization error appears to stay below 10% hence roughly below 0.1% of the total expected cost. This indicates that the expected cost assessment is fairly accurate in this case study. The second point of observation has to do with the fact that this error does not seem to decrease with the number of observations N, which might confirm that performance improvements won't be achieved by using more data but rather by understanding better the dynamics behind how emergency requests distribution evolves through time.

Table 8: Expected recourse cost statistics for CCSP and CAPECSP with respect to in-sample and out-of-sample problem over ten simulations, where we report mean, standard deviation (stdev) and average relative generalization error (avg-error).

	CCSP						CAPECSP				
sample	in-sample		out-sample		avg-error	sample	in-sample		out-sample		avg-error
size	mean	stdev	mean	stdev		size	mean	stdev	mean	stdev	
50	316	42	295	40	6.59%	20	287	41	260	37	8.80%
100	288	33	270	25	6.05%	60	274	43	246	41	10.2%
200	285	23	262	28	8.07%	100	242	25	221	23	8.66%
						200	243	16	220	13	9.64%

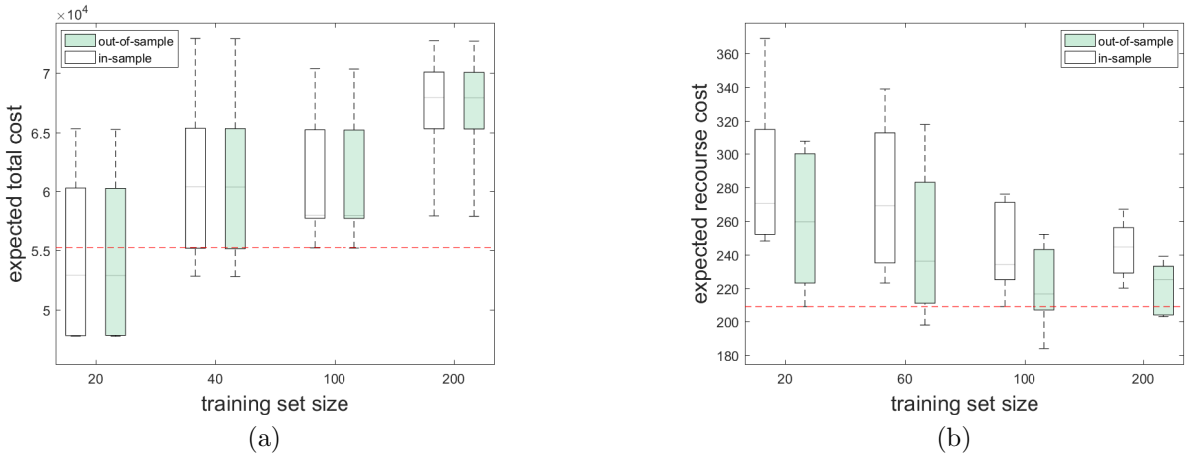


Figure 10: Out-of-sample analysis of expected cost in CAPECSP with $\beta^{cvx}_{0.5}(\eta)$. (a) presents box plots for the expected total cost achieved in-sample and out-of-sample by solutions of the CAPECSP model over 10 experiments. The red dashed line is the best expected total cost that can be achieved out-of-sample. (b) presents similar box plots for the average recourse cost.The red dashed line is the out-of-sample expected recourse cost achieved by the best out-of-sample policy.

We next turn to evaluating the out-of-sample performance of the chance constraints. In this regard, Figure 11 presents statistics about the reliability of the coverage for the six different time periods and three different training set sizes. To be precise, in each experiments, we tested the implementation of the optimal in-sample CCSP policy on 275 test scenarios and computed the probability of covering $\beta \sum_i d^t_{i\omega}$ for each period on this training set. The box plots that appear in the figure indicate the minimum, maximum, mean, and first and third quartiles of these values over the 10 experiments. First, we can observe from this figure that the in-sample policy has a reasonably good out-of-sample performance in terms of reliability. In most experiments and for most time periods, the policy is able to achieve the 95% reliability that it was aiming for, and often reaches 100% reliability, which might explain the sub-optimal empirical expected total cost discussed previously. In the case where this target is not reached we see that most of the time, the violation is below 5%. The figure also exhibits a daily pattern. The policy is typically less reliable in the morning and the evening then during the middle of the day. This might be due to the fact that the model is more flexible in the middle of the day with large ambulance station capacities, and lower relocation and travel costs, which allows the model to more easily satisfy the requests. Finally, we believe it is possible to qualitatively state that as more training samples are used in CCSP, the empirical reliability starts concentrating more around the 95% target.

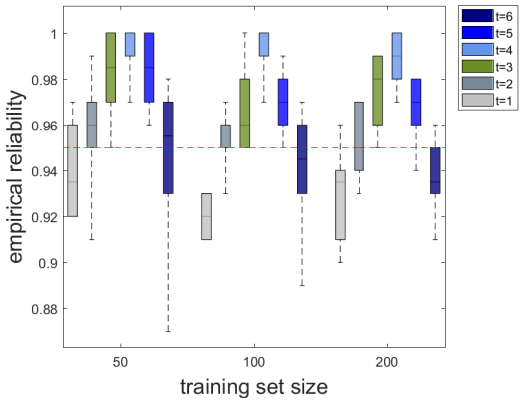


Figure 11: Out-of-sample analysis of reliability. The figure presents box plots for the reliability achieved in 10 out-of-sample experiments by the solution of the CCSP model with $\boldsymbol{\beta = 0.95}$ and $\boldsymbol{\eta = 0.05}$ for different training set sizes and in different time periods. The red dashed line identifies the target reliability level $\mathbf{1 - \eta = 95\%}$.

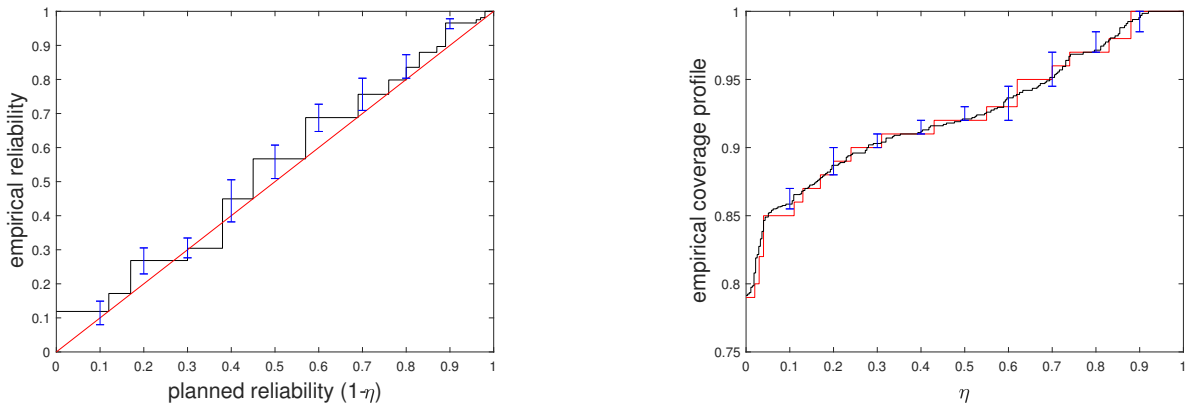
Finally, we briefly investigate the out-of-sample performance of the PEC policy. To do so, we focus on the performance achieved for the first time period $t = 1$ and with a training set of size $N = 100$, although our observations are similar for other values of t. We investigate two out-of-sample measurements. First, in Figure 12(a) we present the difference between the planned reliability and the actual reliability achieved out-of-sample. Specifically, the figure presents statistics about the empirical probability of reaching every level of targeted coverage and compares it to the intended reliability: i.e.

$$f_{12(a)}(\eta) := \frac{1}{275}\sum_{\omega'=1}^{275} \mathbf{1}\left\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} \hat{z}_{ij}^t(\boldsymbol{d}_{\omega'}^t) \geq \beta(\eta)\sum_{i\in\mathcal{I}} d_{i\omega'}^t\right\},$$

where $\{\boldsymbol{d}_{\omega'}\}_{\omega'=1}^{275}$ is the out-of-sample test set and $\hat{\boldsymbol{z}}^t(\boldsymbol{d}_{\omega'}^t)$ is the out-of-sample implementation of the CAPECSP policy as described earlier in this section. Ideally, one would want $f_{12(a)}(\eta) \approx 1-\eta$ in order to conclude that the policy is perfectly calibrated, i.e. that it achieves the reliability level without exceeding it in order to reduce unnecessary expenses. Overall, Figure 12(a) seems to indicate that the PEC policy has a slightly higher out-of-sample reliability than needed, which could explain the sub-optimal expected total costs reached by this policy. We also present in Figure 12(b) statistics about the out-of-sample empirical coverage profile. This is calculated in each experiment using

$$f_{12(b)}(\eta) := \sup\left\{\beta \;\middle|\; \frac{1}{275}\sum_{\omega'=1}^{275} \mathbf{1}\left\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} \hat{z}_{ij}^t(\boldsymbol{d}_{\omega'}^t) \geq \beta\sum_{i\in\mathcal{I}} d_{i\omega'}^t\right\} \geq 1-\eta\right\},$$

and can be compared to the targeted probabilistic envelope $\beta_{0.5}^{cvx}(\eta)$ (red curve in the figure) used in CAPECSP. This figure also confirms that the out-of-sample performance in terms of empirical coverage is of surprisingly good quality. This provides empirical evidence that the online procedure that is proposed to decide of ambulance assignments in an out-of-sample scenario is consistent with respect to the in-sample CAPECSP formulation.



(a) (b)

Figure 12: Out-of-sample analysis of the empirical reliability in (a) and empirical coverage profile in (b). The figures present the average, minimum and maximum values for the empirical reliability and coverage profiles achieved by the solutions to the CAPECSP model in 10 out-of-sample experiments at time $\boldsymbol{t = 1}$ and using a training set of $\boldsymbol{N = 100}$ observations. The red solid line in (a) indicates the targeted reliability, while the red one in (b) indicates the probabilistic coverage envelope $\boldsymbol{\beta_{0.5}^{cvx}(\eta)}$ used by the CAPECSP model.

8 Concluding remarks

In this paper, we extend the chance constrained EMS network design model presented in Beraldi and Bruni (2009) to a dynamic setting and propose a novel probabilistic envelope constrained formulation that allows the EMS system manager to control the relative level of EMS coverage achieved under every possible scenarios of emergency request demand. This gives rise to two multi-period stochastic programming models, CCSP and PECSP. In order to solve instances of the proposed models that have realistic sizes, we develop a solution

scheme that is based on the Branch-and-Benders-Cut method and a set of enhancements that is based on valid inequalities and strengthened optimality cuts. In the case of PECSP, we also propose a conservative approximation model CAPECSP that can be solved using B&BC significantly faster. We compare the numerical performance of our solution schemes to the Branch-and-Bound scheme presented in Beraldi and Bruni (2009) and Benders strategy in CPLEX. Our results demonstrate that the gain in performance is significant especially for larger-scale instances.

We also presented a case study that exploited data from NIASHSCT, where we are able to illustrate the differences between the solutions obtained from CCSP and CAPECSP models, the use of a reference coverage profile in the design of probabilistic envelopes, and the recommended procedure for implementation of the delayed assignment decisions. Our out-of-sample analysis confirms that the policies obtained using 200 historical observations perform reasonably well on test data where characteristics of the distribution might have changed due to the passage of time.

We suspect that the models and empirical analysis presented in this paper should benefit a number of other types of applications in the public sector, e.g. fire station location problems, humanitarian relief network design, etc. In particular, we believe that probabilistic envelope constraints are the most natural way of encoding the expectations that stakeholders have in terms of reliability when faced against uncertain operating conditions. While it is clear that this is the first application of PECs in the EMS network design in literature, to the best of our knowledge it is also the first application of such constraints in general two-stage stochastic programming models.

Finally, we highlight two potential directions for future work. Firstly, while our models currently assume that emergency requests are made from a discrete set of locations, it would be interesting to consider continuous locations instead. One could perhaps employ methods from machine learning (i.e. kernel density estimation) or distributionally robust optimization (using Wasserstein ambiguity sets) to do so. It would also be interesting to consider uncertainty regarding travel times and/or processing time of the emergency request although this would significantly increase the complexity of our models.

Appendix A Proof of propositions

A.1 Proof of Proposition 1

The first step of this proof consists in demonstrating that constraint (5) is equivalent to constraint (7b) in CCSP. This can be seen from the fact that $\boldsymbol{z}$ is integer in CCSP, hence any assignment that satisfies (5) must have $\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z_{ij} \geq \lceil \beta(1-\rho^t_\omega)\sum_{i\in\mathcal{I}} d^t_{i\omega}\rceil = (1-\rho^t_\omega)\lceil \beta\sum_{i\in\mathcal{I}} d^t_{i\omega}\rceil$ since ρ^t_ω is binary. The reverse is more straightforward, namely since constraint (5) is a relaxation of (7b).

The next step is to demonstrate that by relaxing the integrality constraint on $\boldsymbol{r}$ and $\boldsymbol{z}$, we still produce optimal integer solutions. Let $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\rho}^*, \boldsymbol{r}^*, \boldsymbol{z}^*)$ be any optimal solution of CCSP2, we will show that there exists integer valued assignments $\bar{\boldsymbol{r}}^*$ and $\bar{\boldsymbol{z}}^*$ for which $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\rho}^*, \bar{\boldsymbol{r}}^*, \bar{\boldsymbol{z}}^*)$ is also necessarily optimal for CCSP2. Since CCSP2 is a relaxation of CCSP, we will therefore conclude that both problems are equivalent.

We focus on the case of $\bar{\boldsymbol{z}}^*$ since the case of $\bar{\boldsymbol{r}}^*$ is similar. Specifically, for any fixed time period t and scenario ω, we know that any optimal solution to the following problem produces equivalent optimal solutions for CCSP2:

$$
\begin{aligned}
\underset{\boldsymbol{z}^t_\omega \geq 0}{\text{minimize}} \quad & \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} p_\omega\, c\, l_{ij} z^t_{ij\omega} && \\
\text{subject to} \quad & \sum_{i\in\mathcal{I}} z^t_{ij\omega} \leq \lambda y_j^{t^*} && \forall j \in \mathcal{J} \\
& \sum_{j\in\mathcal{J}} z^t_{ij\omega} \leq d^t_{i\omega} && \forall i \in \mathcal{I} \\
& \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \lceil \beta \sum_{i\in\mathcal{I}} d^t_{i\omega}\rceil (1-\rho^{t\,*}_\omega)\,. &&
\end{aligned}
$$

One can actually show that the above problem can be reformulated as the following minimum cost network flow problem (MCNFP).

$$
\begin{aligned}
\underset{\boldsymbol{z}^t_\omega \geq 0, \boldsymbol{w} \geq 0, \boldsymbol{s} \geq 0, q \geq 0}{\text{minimize}} \quad & \sum_{i \in \mathcal{I}} \sum_{j \in \mathcal{J}} p_\omega \; c \; l_{ij} z^t_{ij\omega} & \\
\text{subject to} \quad & \sum_{i \in \mathcal{I}} z^t_{ij\omega} + s_j = \lambda y_j^{t*} & \forall j \in \mathcal{J} \\
& w_i - \sum_{j \in \mathcal{J}} z^t_{ij\omega} = 0 & \forall i \in \mathcal{I} \\
& q - \sum_i w_i = -\lceil \beta \sum_{i \in \mathcal{I}} d^t_{i\omega} \rceil (1 - \rho_\omega^{t\,*}) & \\
& -q - \sum_{i \in \mathcal{I}} s_i = \lceil \beta \sum_{i \in \mathcal{I}} d^t_{i\omega} \rceil (1 - \rho_\omega^{t\,*}) - \sum_{j \in \mathcal{J}} \lambda y_j^{t*} & \\
& w_i \leq d_i & \forall i \in \mathcal{I}\,,
\end{aligned}
$$

where $\boldsymbol{s} \in \mathbb{R}^{|\mathcal{J}|}$, $\boldsymbol{w} \in \mathbb{R}^{|\mathcal{I}|}$, and $q \in \mathbb{R}$. Based on the integrality theorem for this family of problem, it is known (see Bertsimas and Tsitsiklis (1997)) that there necessarily exists an integer valued optimal solution in contexts where all parameters in the set of constraints of the MCNFP are integers. This is necessarily the case for the instances that arise using CCSP2 since we assume that λ is an integer, and since any feasible $\boldsymbol{y}$ and $\boldsymbol{\rho}$ also are. We can therefore conclude that there is an integer assignment for $\bar{\boldsymbol{z}}^*$ that can replace the original optimal assignment while preserving the optimality of $(\boldsymbol{x}^*, \boldsymbol{y}^*, \boldsymbol{\rho}^*, \boldsymbol{r}^*, \bar{\boldsymbol{z}}^*)$ with respect to CCSP.

A similar argument can be used for the case of $\boldsymbol{r}$ where the MCNFLP more naturally takes the form of:

$$
\begin{aligned}
\underset{\boldsymbol{r} \geq 0}{\text{minimize}} \quad & \sum_{t=1}^{T} \sum_{m \in \mathcal{J}} \sum_{j \in \mathcal{J}} \alpha^t r^t_{mj} & \\
\text{subject to} \quad & \sum_{m \in \mathcal{J}} r^t_{mj} - \sum_{m \in \mathcal{J}} r^t_{jm} = y_j^{t+1} - y_j^t & \forall j \in \mathcal{J}, t \in \{1, \ldots, T-1\} \\
& \sum_{m \in \mathcal{J}} r^T_{mj} - \sum_{m \in \mathcal{J}} r^T_{jm} = y_j^1 - y_j^T & \forall j \in \mathcal{J}\,.
\end{aligned}
$$

This completes our proof. □

A.2 Proof of Proposition 2

This proof is divided into two parts. We first show that CCSP2 is equivalent to CCSP2'. We then confirm that solutions that satisfy constraints (8d) and (8e) give rise to versions of the CCSP-SP^r and CCSP-SP$^z_{\omega,t}$ models that necessarily feasible and bounded.

In demonstrating the equivalence between CCSP2 and CCSP2', we start by remarking that CCSP2' follows from exploiting an epigraph reformulation of the operations $\sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}} p_\omega c^t l_{ij} z^t_{ij\omega}$ and $\sum_{t=1}^{T} \sum_{m \in \mathcal{J}} \sum_{j \in \mathcal{J}} \alpha^t r^t_{mj}$ in the objective function, and adding constraints (8d) and (8e) to the model. We therefore only need to show that the latter two constraints are redundant in CCSP2. First, we can obtain constraint (8e) directly from the constraints (4b) and (7b):

$$
\sum_{j \in \mathcal{J}} \lambda y_j^t \geq \sum_{j \in \mathcal{J}} \sum_{i \in \mathcal{I}} z^t_{ij\omega} \geq (1 - \rho^t_\omega) \lceil \beta \sum_{i \in \mathcal{I}} d^t_{i\omega} \rceil \,, \; \forall t \in \{1, \ldots, T\}, \forall \omega \in \Omega\,.
$$

Moreover, constraints (1d) and (1e) imply that

$$
\sum_{j \in \mathcal{J}} y_j^{t+1} = \sum_{j \in \mathcal{J}} \left(y_j^t + \sum_{m \in \mathcal{J}} r^t_{mj} - \sum_{m \in \mathcal{J}} r^t_{jm} \right) = \sum_{j \in \mathcal{J}} y_j^t \,, \; \forall t \in \{1, \ldots, T\}\,.
$$

We are left with showing that both CCSP-SP^r and CCSP-SP$^z_{\omega,t}$ are always feasible and bounded when constraints (8d) and (8e) are satisfied.

In the case of CCSP-SP^r, boundedness follows from the fact that $\boldsymbol{r} \geq 0$ and that $\boldsymbol{\alpha} \geq \mathbf{0}$. As for the feasibility of this slave problem, this follows from the fact that $\sum_{j\in\mathcal{J}} y_j^t = \sum_{j\in\mathcal{J}} y_j^{t+1}$ for all t, and that we assumed vehicles can be transferred from any location to any other location in a single period.

In the case of CCSP-SP$^z_{\omega,t}$, boundedness again follows from the non-negativity of $\boldsymbol{z}^t_\omega$ and monotonicity of the objective function. As for feasibility, since we know that $\sum_{j\in\mathcal{J}} \lambda y_j^t \geq (1-\rho^t_\omega)\lceil \beta \sum_{i\in\mathcal{I}} d^t_{i\omega}\rceil$, we know that in this transportation problem there is enough offer to cover the demand. Since every source $i \in \mathcal{I}$ can serve every demand node $j \in \mathcal{J}$, we conclude that a feasible assignment can necessarily be found.

This completes our proof. □

A.3 Proof of Proposition 3

For fixed t, let the ordered scenario set $\Omega'_t := \{\omega'_k\}_{k=1}^N$ be such that for all $k < k'$ we have that $\sum_{i\in\mathcal{I}} d^t_{i\omega'_k} \leq \sum_{i\in\mathcal{I}} d^t_{i\omega'_{k'}}$. One can simply rewrite constraints (6) and (8e) as

$$\sum_{j\in\mathcal{J}} \lambda y_j^t \geq \lceil \beta \sum_{i\in\mathcal{I}} d^t_{i\omega'}\rceil (1-\rho^t_{\omega'}) \qquad \forall t \in \{1,\dots,T\}, \omega' \in \Omega'_t \tag{29}$$

$$\sum_{\omega'\in\Omega'_t} p_{\omega'}\rho^t_{\omega'} \leq \eta \qquad \forall t \in \{1,\dots,T\}. \tag{30}$$

We will prove our claim by contradiction. In order to do so, let $(\boldsymbol{y}, \boldsymbol{\rho})$ satisfy constraints (29) and (30) yet violate constraint (15), namely

$$\sum_{j\in\mathcal{J}} \lambda y_j^t < \beta \sum_{i\in\mathcal{I}} d^t_{i\omega'_{\bar{k}}} \qquad \forall k \geq \bar{k}\,.$$

Since constraint (29) holds, then it must be that $\rho^t_{\omega_k} = 1$ for all $k \geq \bar{k}$. This in turn can be used to show that

$$\eta \geq \sum_{\omega'\in\Omega'_t} p_{\omega'}\rho^t_{\omega'} \geq \sum_{k:k\geq\bar{k}} p_{\omega'_k}\rho^t_{\omega'_k} = \sum_{k=\bar{k}}^{N} p_{\omega'_k} > \eta\,.$$

Since this necessarily leads to a contradiction $\eta < \eta$, we must conclude that any solution that satisfy (29) and (30) must also satisfy

$$\sum_{j\in\mathcal{J}} \lambda y_j^t \geq \beta \sum_{i\in\mathcal{I}} d^t_{i\omega'_{\bar{k}}}.$$

□

A.4 Proof of Proposition 4

This result is obtained by remarking that for any $t \in \{1,\dots,T\}$ and any $k \in \{0,\dots,N-1\}$ the constraint

$$\sum_{\omega\in\Omega} (1/N)\mathbf{1}\left\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta(\eta)\sum_{i\in\mathcal{I}} d^t_{i\omega}\right\} \geq 1-\eta \quad \forall \eta \in [k/N,\,(k+1)/N)$$

is equivalent to

$$\sum_{\omega\in\Omega} (1/N)\mathbf{1}\left\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta(\eta)\sum_{i\in\mathcal{I}} d^t_{i\omega}\right\} \geq 1-k/N \quad \forall \eta \in [k/N,\,(k+1)/N)\,.$$

Since the *sum* operator can only reach values in $\{j/N\}_{j\in\{0,\dots,N\}}$. Hence, the constraint can be reformulated equivalently as

$$\inf_{\eta\in[k/N,\,(k+1)/N)} \sum_{\omega\in\Omega}(1/N)\mathbf{1}\left\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta(\eta)\sum_{i\in\mathcal{I}} d^t_{i\omega}\right\} \geq 1-k/N\,,$$

yet since $\mathbf{1}\left\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta\sum_{i\in\mathcal{I}} d^t_{i\omega}\right\}$ is left continuous and decreasing with respect to β, it is necessary and sufficient to verify that

$$\sum_{\omega\in\Omega}(1/N)\mathbf{1}\left\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \sup_{\eta\in[k/N,\,(k+1)/N)} \beta(\eta)\sum_{i\in\mathcal{I}} d^t_{i\omega}\right\} \geq 1-k/N\,.$$

This constraint is exactly equivalent to

$$\sum_{\omega\in\Omega}(1/N)\mathbf{1}\left\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta_k\sum_{i\in\mathcal{I}} d^t_{i\omega}\right\} \geq 1-k/N\,,$$

which can then be reformulated using binary variables $\boldsymbol{\rho}$. We thus conclude that constraint (16) is equivalent to the condition that there exists $\boldsymbol{\rho}\in\{0,\,1\}^{N\times N\times T}$ that satisfies

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta_k(1-\rho^t_{k\omega})\sum_{i\in\mathcal{I}} d^t_{i\omega} \qquad \forall t\in\{1,\dots,T\},\omega\in\Omega,k\in\{0,\dots,N-1\} \tag{31}$$

$$(1/N)\sum_{\omega\in\Omega}\rho^t_{k\omega} \leq k/N \qquad \forall t\in\{1,\dots,T\},k\in\{0,\dots,N-1\}\,. \tag{32}$$

The next step of this proof consists of showing that the inequality in constraint (32) can be replaced with an equality without affecting the feasibility of $\boldsymbol{z}$. In particular, for any feasible $\boldsymbol{z}$, given some $\bar{\boldsymbol{\rho}}$ that confirms the feasibility of $\boldsymbol{z}$ using (31) and (32), if $(1/N)\sum_{\omega\in\Omega}\bar{\rho}^t_{k\omega} < k/N$ for some k and some t then it must be that $(1/N)\sum_{\omega\in\Omega}\bar{\rho}^t_{k\omega} \leq (k-1)/N$ and that there is a $\bar{\omega}$ such that $\bar{\rho}^t_{k\bar{\omega}} = 0$, otherwise $(1/N)\sum_{\omega\in\Omega}\bar{\rho}^t_{k\omega} = 1 > (N-1)/N \geq k/N$. One can actually show that $\hat{\boldsymbol{\rho}}$ constructed by copying $\bar{\boldsymbol{\rho}}$ at every index except for $\hat{\rho}^t_{k\bar{\omega}} := 1$ will also satisfy both constraints. Namely,

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\bar{\omega}} \geq \beta_k(1-\bar{\rho}^t_{k\bar{\omega}})\sum_{i\in\mathcal{I}} d^t_{i\bar{\omega}} \geq \beta_k(1-\hat{\rho}^t_{k\bar{\omega}}+1)\sum_{i\in\mathcal{I}} d^t_{i\bar{\omega}} \geq \beta_k(1-\hat{\rho}^t_{k\bar{\omega}})\sum_{i\in\mathcal{I}} d^t_{i\bar{\omega}}$$

$$(1/N)\sum_{\omega\in\Omega}\hat{\rho}^t_{k\omega} = (1/N)\sum_{\omega\in\Omega}\bar{\rho}^t_{k\omega} + (1/N) \leq (k-1)/N + 1/N \leq k/N\,.$$

Based on this argument, we can conclude that there always exists a $\bar{\boldsymbol{\rho}}$ that confirms feasibility of $\boldsymbol{z}$ based on constraints (31) and (32) and also satisfies constraint (18).

The third step of this proof is to show that constraint (19) can also be imposed without affecting the feasibility of $\boldsymbol{z}$. In particular, for any feasible $\boldsymbol{z}$, given some $\bar{\rho}$ that confirms the feasibility of $\boldsymbol{z}$ using (31) and (18), if $\bar{\rho}^t_{k\omega} > \bar{\rho}^t_{k+1,\omega}$ for some $\bar{t}$, $\bar{\omega}$, and $\bar{k}$, then one can simply construct a $\hat{\boldsymbol{\rho}}$ that mimics $\bar{\boldsymbol{\rho}}$ except for $\hat{\rho}^{\bar{t}}_{\bar{k}\bar{\omega}} := \bar{\rho}^{\bar{t}}_{\bar{k}+1,\bar{\omega}} = 0$ and for $\hat{\rho}^{\bar{t}}_{\bar{k}\bar{\omega}'} := 1$, where $\bar{\omega}'$ is any scenario for which $\bar{\rho}^{\bar{t}}_{\bar{k}\bar{\omega}'} = 0 < \bar{\rho}^{\bar{t}}_{\bar{k}+1,\bar{\omega}'}$. One can verify that such an index $\bar{\omega}'$ always exists because of constraint (18). One can as a last step confirm that all constraints are satisfied for $\hat{\boldsymbol{\rho}}$:

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^{\bar{t}}_{ij\bar{\omega}} \geq \bar{\beta}_{\bar{k}+1}(1-\bar{\rho}^{\bar{t}}_{\bar{k}+1,\bar{\omega}})\sum_{i\in\mathcal{I}} d^{\bar{t}}_{i\bar{\omega}} \geq \bar{\beta}_{\bar{k}}(1-\bar{\rho}^{\bar{t}}_{\bar{k}+1,\bar{\omega}})\sum_{i\in\mathcal{I}} d^{\bar{t}}_{i\bar{\omega}} = \bar{\beta}_{\bar{k}}(1-\hat{\rho}^{\bar{t}}_{\bar{k},\bar{\omega}})\sum_{i\in\mathcal{I}} d^{\bar{t}}_{i\bar{\omega}}$$

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^{\bar{t}}_{ij\bar{\omega}'} \geq \bar{\beta}_{\bar{k}}(1-\bar{\rho}^{\bar{t}}_{\bar{k},\bar{\omega}'})\sum_{i\in\mathcal{I}} d^{\bar{t}}_{i\bar{\omega}'} \geq \bar{\beta}_{\bar{k}}(1-\hat{\rho}^{\bar{t}}_{\bar{k},\bar{\omega}'})\sum_{i\in\mathcal{I}} d^{\bar{t}}_{i\bar{\omega}'}$$

$$\sum_{\omega\in\Omega}\hat{\rho}^{\bar{t}}_{\bar{k}\omega} = \sum_{\omega\in\Omega}\bar{\rho}^{\bar{t}}_{\bar{k}\omega} - \bar{\rho}^{\bar{t}}_{\bar{k}\bar{\omega}} + \hat{\rho}^{\bar{t}}_{\bar{k}\bar{\omega}} - \bar{\rho}^{\bar{t}}_{\bar{k}\bar{\omega}'} + \hat{\rho}^{\bar{t}}_{\bar{k}\bar{\omega}'} = \sum_{\omega\in\Omega}\bar{\rho}^{\bar{t}}_{\bar{k}\omega} - 1 + 0 - 0 + 1 = k$$
$$\hat{\rho}^{\bar{t}}_{\bar{k}\bar{\omega}} = 0 \leq \hat{\rho}^{\bar{t}}_{\bar{k}+1,\bar{\omega}}$$
$$\hat{\rho}^{\bar{t}}_{\bar{k}\bar{\omega}'} = 1 \leq 1 = \hat{\rho}^{\bar{t}}_{\bar{k}+1,\bar{\omega}'}\,.$$

This allows us to conclude that there always exists a $\bar{\boldsymbol{\rho}}$ that confirms feasibility of $\boldsymbol{z}$ based on constraints (31) and (18) and also satisfies constraint (19).

The final step consists in demonstrating that for all $\boldsymbol{\rho}$ that satisfy constraint (19), constraint (31) is equivalent to constraint (17). This can be shown by arguing that for all t and all ω:

$$\begin{aligned}
&\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta_k(1-\rho^t_{k\omega})\sum_{i\in\mathcal{I}} d^t_{i\omega}\;\forall\, k\in\{0,\dots,N-1\}\\
&\qquad\Leftrightarrow \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \left(\max_{k\in\{0,\dots,N-1\}}\beta_k(1-\rho^t_{k\omega})\right)\sum_{i\in\mathcal{I}} d^t_{i\omega}\\
&\qquad\Leftrightarrow \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \left(\max_{k:\rho^t_{k\omega}=0}\beta_k\right)\sum_{i\in\mathcal{I}} d^t_{i\omega}\\
&\qquad\Leftrightarrow \sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \left(\bar{\beta}_{N-1}(1-\rho^t_{N-1,\omega}) + \sum_{k\in\{0,\dots,N-2\}}\beta_k(\rho^t_{k+1,\omega}-\rho^t_{k\omega})\right)\sum_{i\in\mathcal{I}} d^t_{i\omega}\,,
\end{aligned}$$

where the second equivalence follows from the fact that β_k is a non-decreasing sequence, and the third equivalence follows from the fact that $\boldsymbol{\rho}$ satisfies constraint (19).

This completes the proof. □

A.5 Proof of Proposition 5

First, constraint (25) can be relaxed by dropping the rounding operation in the right-hand side thus obtaining

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega'_{N-k}} \geq \beta_k\sum_{i\in\mathcal{I}} d^t_{i\omega'_{N-k}} \qquad \forall t\in\{1,\dots,T\}, k\in\{0,\dots,N-1\}\,. \tag{33}$$

Then, based on the definition of Ω'_t, we have that for all $k<k'$, $\sum_{i\in\mathcal{I}} d^t_{i\omega'_k} \leq \sum_{i\in\mathcal{I}} d^t_{i\omega'_{k'}}$. This allows us to state that any assignment for $\boldsymbol{z}$ that satisfies constraint (33), must also satisfy

$$\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega'_{N-k}} \geq \beta_k\sum_{i\in\mathcal{I}} d^t_{i\omega'_{k'}} \qquad \forall t\in\{1,\dots,T\}, 1\leq k'\leq N-k, k\in\{0,\dots,N-1\}\,. \tag{34}$$

Hence, we must also have that

$$\frac{1}{N}\sum_{\omega\in\Omega}\mathbf{1}\left\{\sum_{i\in\mathcal{I}}\sum_{j\in\mathcal{J}} z^t_{ij\omega} \geq \beta_k\sum_{i\in\mathcal{I}} d^t_{i\omega}\right\} \geq \sum_{k'=1}^{N-k} 1/N = 1-\frac{k}{N}. \tag{35}$$

Yet, we already showed in the proof of Proposition 4 (see Appendix A.4) that under Assumption 1, constraint (35) and constraint (16) are equivalent.

This completes our proof. □

Appendix B Additional supplemental material

B.1 Constructing the coverage profile associated to a solution

Given an solution $(\boldsymbol{x}^*, \boldsymbol{y}^*)$, one can computed the coverage profile that this solution achieved for a specific set of equiprobable scenarios $\{\boldsymbol{d}_\omega\}_{\omega\in\Omega}$ with $|\Omega| = N$ as follows. First, for each $\omega \in \Omega$, one should solve the following linear program

$$
\begin{aligned}
\bar{\beta}^*_\omega := \quad & \underset{\boldsymbol{z},\beta}{\text{maximize}} && \beta && \\
& \text{subject to} && \sum_{i\in\mathcal{I}} z^t_{ij} \le \lambda(y^t_j)^* && \forall j \in \mathcal{J}, t \in \{1,\dots,T\} \\
& && \sum_{j\in\mathcal{J}} z^t_{ij} \le d^t_i && \forall i \in \mathcal{I}, t \in \{1,\dots,T\} \\
& && \sum_{j\in\mathcal{J}}\sum_{i\in\mathcal{I}} z^t_{ij} \ge \beta \sum_{i\in\mathcal{I}} d^t_i && \forall t \in \{1,\dots,T\} \\
& && z^t_{ij} \in \mathbb{N} && \forall i \in \mathcal{I}, j \in \mathcal{J}, t \in \{1,\dots,T\}.
\end{aligned}
$$

Letting $\bar{\beta}^*_{[k]}$ represent the k-th smallest element of the list $\{\bar{\beta}^*_\omega\}_{\omega\in\Omega}$, it is then possible to construct $\bar{\beta}(\eta) := \bar{\beta}^*_{[\lfloor \eta N+1 \rfloor]}$, where $\lfloor\cdot\rfloor$ is the round down to nearest integer operation. Note that, in our case study, we used the $N = 100$ observations from January 2016 to middle of April 2016.

References

Adulyasak, Y., Cordeau, J.-F., Jans, R., 2015. Benders decomposition for production routing under demand uncertainty. Operations Research 63 (4), 851–867.

Ahmadi-Javid, A., Seyedi, P., Syam, S. S., 2017. A survey of healthcare facility location. Computers & Operations Research 79, 223–263.

Ansari, S., McLay, L. A., Mayorga, M. E., 2015. A maximum expected covering problem for district design. Transportation Science 51 (1), 376–390.

Aringhieri, R., Bruni, M. E., Khodaparasti, S., van Essen, J., 2017. Emergency medical services and beyond: Addressing new challenges through a wide literature review. Computers & Operations Research 78, 349–368.

Ball, M. O., Lin, F. L., 1993. A reliability model applied to emergency service vehicle location. Operations Research 41 (1), 18–36.

Başar, A., Çatay, B., Ünlüyurt, T., 2011. A multi-period double coverage approach for locating the emergency medical service stations in istanbul. Journal of the Operational Research Society 62 (4), 627–637.

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

Beraldi, P., Bruni, M. E., 2009. A probabilistic model applied to emergency service vehicle location. European Journal of Operational Research 196 (1), 323–331.

Beraldi, P., Bruni, M. E., Conforti, D., 2004. Designing robust emergency medical service via stochastic programming. European Journal of Operational Research 158 (1), 183–193.

Bertsimas, D., Tsitsiklis, J., 1997. Introduction to linear optimization. Athena Scientific.

Boujemaa, R., Jebali, A., Hammami, S., Ruiz, A., Bouchriha, H., 2017. A stochastic approach for designing two-tiered emergency medical service systems. Flexible Services and Manufacturing Journal, 1–30.

Brotcorne, L., Laporte, G., Semet, F., 2003. Ambulance location and relocation models. European Journal of Operational Research 147 (3), 451–463.

Chan, T. C., Demirtas, D., Kwon, R. H., 2016. Optimizing the deployment of public access defibrillators. Management Science 62 (12), 3617–3635.

Chanta, S., Mayorga, M. E., McLay, L. A., 2014. Improving emergency service in rural areas: a bi-objective covering location model for ems systems. Annals of Operations Research 221 (1), 133–159.

Charnes, A., Cooper, W. W., 1959. Chance-constrained programming. Management Science 6 (1), 73–79.

Church, R., Velle, C. R., 1974. The maximal covering location problem. Papers in Regional Science 32 (1), 101–118.

Dalal, J., Üster, H., 2018. Combining worst case and average case considerations in an integrated emergency response network design problem. Transportation Science 152 (1), 171–188.

Daskin, M. S., 1983. A maximum expected covering location model: formulation, properties and heuristic solution. Transportation Science 17 (1), 48–70.

Degel, D., Wiesche, L., Rachuba, S., Werners, B., 2015. Time-dependent ambulance allocation considering data-driven empirically required coverage. Health Care Management Science 18 (4), 444–458.

Gendreau, M., Laporte, G., Semet, F., 1997. Solving an ambulance location model by tabu search. Location Science 5 (2), 75–88.

Gendreau, M., Laporte, G., Semet, F., 2001. A dynamic model and parallel tabu search heuristic for real-time ambulance relocation. Parallel Computing 27 (12), 1641–1653.

Gendron, B., Scutellà, M. G., Garroppo, R. G., Nencioni, G., Tavanti, L., 2016. A branch-and-benders-cut method for nonlinear power design in green wireless local area networks. European Journal of Operational Research 255 (1), 151–162.

Li, X., Zhao, Z., Zhu, X., Wyatt, T., 2011. Covering models and optimization techniques for emergency response facility location and planning: a review. Mathematical Methods of Operations Research 74 (3), 281–310.

Luedtke, J., Ahmed, S., Nemhauser, G. L., 2010. An integer programming approach for linear programs with probabilistic constraints. Mathematical Programming 122 (2), 247–272.

Magnanti, T. L., Wong, R. T., 1981. Accelerating benders decomposition: Algorithmic enhancement and model selection criteria. Operations Research 29 (3), 464–484.

Maleki, M., Majlesinasab, N., Sepehri, M. M., 2014. Two new models for redeployment of ambulances. Computers & Industrial Engineering 78, 271–284.

Martins de Sá, E., Contreras, I., Cordeau, J.-F., Saraiva de Camargo, R., de Miranda, G., 2015. The hub line location problem. Transportation Science 49 (3), 500–518.

Maxwell, M. S., Restrepo, M., Henderson, S. G., Topaloglu, H., 2010. Approximate dynamic programming for ambulance redeployment. INFORMS Journal on Computing 22 (2), 266–281.

Moeini, M., Jemai, Z., Sahin, E., 2015. Location and relocation problems in the context of the emergency medical service systems: a case study. Central European Journal of Operations Research 23 (3), 641–658.

Naoum-Sawaya, J., Elhedhli, S., 2013. A stochastic optimization model for real-time ambulance redeployment. Computers & Operations Research 40 (8), 1972–1978.

Nemirovski, A., Shapiro, A., 2006. Convex approximations of chance constrained programs. SIAM Journal on Optimization 17 (4), 969–996.

Nickel, S., Reuter-Oppermann, M., Saldanha-da Gama, F., 2016. Ambulance location under stochastic demand: A sampling approach. Operations Research for Health Care 8, 24–32.

Noyan, N., 2010. Alternate risk measures for emergency medical service system design. Annals of Operations Research 181 (1), 559–589.

Noyan, N., Balcik, B., Atakan, S., 2016. A stochastic optimization model for designing last mile relief networks. Transportation Science 50 (3), 1092–1113.

Pagnoncelli, B., Ahmed, S., Shapiro, A., 2009. Sample average approximation method for chance constrained programming: theory and applications. Journal of Optimization Theory and Applications 142 (2), 399–416.

Paul, N. R., Lunday, B. J., Nurre, S. G., 2017. A multiobjective, maximal conditional covering location problem applied to the relocation of hierarchical emergency response facilities. Omega 66, 147–158.

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

Rahmaniani, R., Crainic, T. G., Gendreau, M., Rei, W., 2018. Accelerating the benders decomposition method: Application to stochastic network design problems. SIAM Journal on Optimization 28 (1), 875–903.

Schmid, V., 2012. Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. European Journal of Operational Research 219 (3), 611–621.

Schmid, V., Doerner, K. F., 2010. Ambulance location and relocation problems with time-dependent travel times. European Journal of Operational Research 207 (3), 1293–1303.

Schneeberger, K., Doerner, K. F., Kurz, A., Schilde, M., 2016. Ambulance location and relocation models in a crisis. Central European Journal of Operations Research 24 (1), 1–27.

Toregas, C., Swain, R., ReVelle, C., Bergman, L., 1971. The location of emergency service facilities. Operations Research 19 (6), 1363–1373.

van den Berg, P., Kommer, G., Zuzáková, B., 2016. Linear formulation for the maximum expected coverage location model with fractional coverage. Operations Research for Health Care 8, 33–41.

van den Berg, P. L., Aardal, K., 2015. Time-dependent MEXCLP with start-up and relocation cost. European Journal of Operational Research 242 (2), 383–389.

Xu, H., Caramanis, C., Mannor, S., 2012. Optimization under probabilistic envelope constraints. Operations Research 60 (3), 682–699.

Zhang, Z.-H., Li, K., 2015. A novel probabilistic formulation for locating and sizing emergency medical service stations. Annals of Operations Research 229 (1), 813–835.