CIRRELT

Centre interuniversitaire de recherche sur les réseaux d'entreprise, la logistique et le transport

Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation

# Recent Advances in Emergency Medical Services Management

**Valérie Bélanger**
**Angel Ruiz**
**Patrick Soriano**

**July 2015**

**CIRRELT-2015-28**

Document de travail également publié par la Faculté des sciences de l'administration de l'Université Laval, sous le numéro FSA-2015-006.

# Recent Advances in Emergency Medical Services Management

**Valérie Bélanger[1,2,*], Angel Ruiz[1,3], Patrick Soriano[1,2]**

[1] Interuniversity Research Centre on Enterprise Networks, Logistics and Transportation (CIRRELT)

[2] Department of Management Sciences, HEC Montréal, 3000 chemin de la Côte-Sainte-Catherine, Montréal, Canada H3T 2A7

[3] Department of Operations and Decision Systems, 2325 de la Terrasse, Université Laval, Québec, Canada G1V 0A6

**Abstract.** This article discusses different modelling approaches considered to address problems related to ambulances' location and relocation as well as dispatching decisions arising in Emergency Medical Services (EMS) management. Over the past 10 years, a considerable amount of research has been devoted to this field and specifically, to the development of strategies that more explicitly take into account the uncertainty and dynamism inherent to EMS. This article thus reviews the most recent advances related to EMS management from an operational level standpoint. It briefly reviews early works on static ambulances' location problems and presents new modelling and solution strategies proposed to address it. However, it intends to concentrate on relocation strategies and dispatching rules, and discuss the interaction between these two types of decisions. Finally, conclusions and perspectives are presented.

**Keywords**. Emergency medical services, location, relocation.

**Acknowledgements.** This research was funded by the Fond de recherche du Québec - Nature et les technologies (FQRNT) through the Team research project program [grant PR-122269] and by the Natural Sciences and Engineering Research Council of Canada (NSERC) through the Discovery grants program [grants OPG 0293307 and OPG 0177174] and the Alexander Graham Bell Canada Graduate Scholarships-Doctoral Program awarded by the first author. We want to thank both organizations for their support.



---

* Corresponding author: Valerie.Belanger@cirrelt.ca

Dépôt légal – Bibliothèque et Archives nationales du Québec
Bibliothèque et Archives Canada, 2015

## 1. Introduction

Emergency Medical Services (EMS) are critical elements of modern health systems. These organizations are responsible for the pre-hospital component of health systems, which consists of medical care and transportation activities performed from the reception of an emergency call to the release of a patient or its transfer to a hospital. EMS thus play an important role in modern health systems and their ability to efficiently respond to emergency calls can have a significant impact on patient's health and recovery.

For most EMS, the process leading to the intervention of a paramedical team consists of the following steps: (1) reception of an emergency call, (2) call screening, (3) vehicle dispatching, (4) vehicle traveling from its current location to the emergency site, (5) on-site treatment and, (6) patient's transportation to a health facility and/or release. Although EMS's vehicles may also provide medical transportation for non-urgent patients, this paper focuses exclusively on emergency response activities. To adequately support this process, EMS have to mobilize a set of resources (i.e. paramedics, ambulances, emergency medical responders, etc.) and manage them efficiently. Several questions arise regarding the strategies to be deployed in order to efficiently provide services. How many units of each resource should be used? How should the ambulance fleet be managed? Where should the ambulances be located? All such decisions can be classified according to the classical three decision-making levels: strategic, tactical and operational.

As presented in Figure 1, strategic decisions address, among others things, the location of ambulance stations and ambulance fleet dimensioning. The tactical level involves decisions such as the location of the potential standby sites, as well as crew pairing and scheduling. Finally, operational decisions, often considered in real-time, are concerned with short term decisions such as relocation strategies and dispatching rules. Dispatching decisions are taken in real-time considering predetermined rules, and there is a clear trend towards addressing relocation decisions dynamically as well. Real-time fleet management strategies thus receive an increasing amount of researchers' attention.

Researchers have studied different aspects of EMS management. Until recently, most works had focused on strategic and tactical issues, decisions of a static nature. For instance, the static location problem determines the set of standby sites where ambulances will be positioned while waiting to be dispatched to respond to emergency calls. Once implemented, the corresponding location plan will remain unchanged, i.e. each ambulance will return to its designated standby site after completing a mission. Nevertheless, it becomes clear that it may be beneficial to change ambulances' locations during a day (i.e. to relocate them) in order to account for the evolution of the situation faced by EMS. In the last years, a lot of effort has been dedicated to the development of approaches that more explicitly consider the uncertainty and dynamism inherent to EMS context, leading to a considerable amount of new models and management strategies. These new models, particularly the ones devoted to vehicles relocation, have been mainly developed to deal with situations where ambulances can be deployed on a large set of locations over the territory, which includes, but is not limited, to hospitals or medical facilities. In this context, the designated standby site of an ambulance can be changed during a day without involving significant costs or modifications to the system itself.

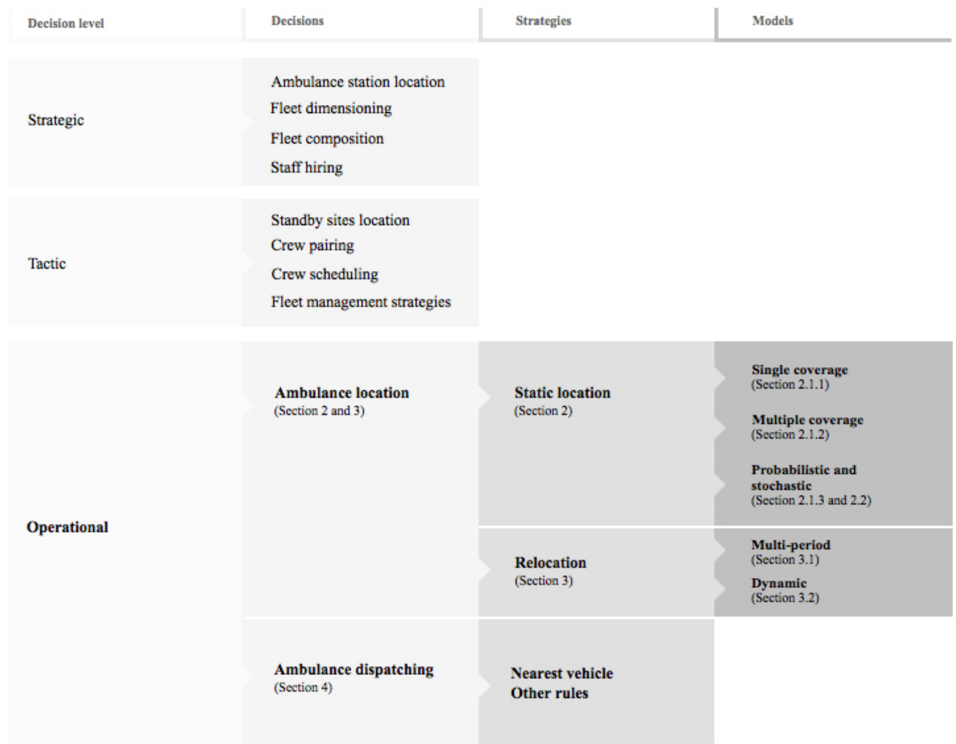


Figure 1: Decisions problems related to EMS management

As highlighted in Figure 1, the aim of this survey is to review and discuss the most recent advances in EMS management from an operational level standpoint. Contrarily to previous surveys (Brotcorne et al., 2003; Goldberg, 2004), it concentrates on relocation strategies and dispatching decisions, and discusses the interaction between these two decisions from a practical perspective. It also completes existing reviews by considering specific location models that weren't covered in previous surveys, providing a precise, exhaustive and up-to-date picture on the research on location and dispatching decisions. Finally, it contributes summarizing tables (inspired by the ones in Brotcorne et al. (2003)) to present in a synthetic form all the models and variants discussed in the paper. The article is organized as follows. Section 2 briefly reviews early key works on static ambulance location as an anchor to understand the most recent modelling and solution methodologies. Section 3 traces and surveys the development of multi-period and dynamic relocation strategies considered to tackle the system's evolution over time. Section 4 presents and discusses dispatching rules. Section 5 concludes the paper by providing conclusions and perspectives.

## 2. Static ambulance location

To ensure an adequate service level to the population, EMS use a given number of ambulances located strategically over the territory they serve. The *static* ambulance location problem aims to select the standby sites to use, and the number of ambulances that should be located at each of them while satisfying a set of constraints. A location plan thus defines a set of standby sites (i.e. parking lot, ambulance station, hospital, and so on.) to locate one or more ambulances and assumes that each ambulance will return to its designated standby site after completing a mission. Three main approaches have been considered so far to address the location problem: mathematical programming, simulation and queueing theory. As presented in the surveys of Goldberg (2004) and Bélanger et al. (2012), each of these approaches has its own *pros and cons*. Simulation and queueing theory are able to adequately capture the dynamics and uncertainty of EMS, but their use have commonly been limited to evaluate the performance of a given solution or strategy over a set of different scenarios. On the other hand, mathematical programming seeks to produce a solution, but doing so generally requires simplifying assumptions to ensure the development of tractable models.

This survey focuses on the latter approach. We refer the reader to the recent review of Aboueljinane et al. (2013) for an interesting survey of simulation studies in the context of EMS and to Larson (1974, 1975) for a description of one of the most used queueing theory based approaches for the evaluation of location plans.

### *2.1. Early works*

As discussed in Brotcorne et al. (2003), many studies have been concerned with the *static* emergency vehicle location problem thus resulting in the development of diverse static ambulance location models. These models can be divided into three main categories following their chronological evolution: (1) single coverage deterministic models, (2) multiple coverage deterministic models and (3) probabilistic and stochastic models. Indeed, location models have evolved over the years to integrate more realistic aspects of the problem such as demand uncertainty, availability of vehicles, traffic congestion, and so on. ReVelle (1989), Marianov and ReVelle (1995) and Brotcorne et al. (2003) presented interesting surveys of mathematical models applied to emergency vehicles location. Recently, Başar et al. (2012) proposed a taxonomy for emergency service location's problems, which includes a systematic analysis of model characteristics, but did not trace their evolution nor described them as we do in this survey.

This section summarizes and presents early works that we deem more relevant to the presentation of the new location and relocation strategies discussed later on. We refer the reader to the surveys listed above for a more detailed description of these models, their formulation, as well as the methodologies proposed to solve them. In the first subsection, we recall two seminal mathematical formulations to illustrate and support the presentation of recent advances and new approaches related to static location.

Before going on with model descriptions, let us introduce the notation that will be used throughout the paper. Ambulance location models can be defined on a graph $G = (V, E)$ where $V = N \cup M$, $N = (v_1, ..., v_n)$ and $M = (v_{n+1}, ..., v_{n+m})$ are two vertex sets representing, respectively, demand zones and potential standby sites, and $E = \{(v_i, v_j) : v_i, v_j \in V, i < j\}$ is the edge set. To each edge $(v_i, v_j) \in E$ is associated a travel time or distance $d_{ij}$. The population associated with demand zone $v_i \in V$ is equal to $a_i$. Since most of the models use the notion of coverage, the sets $M_i$ and $M'_i$ correspond to the sets of standby sites that can cover a demand zone $v_i$ respectively within a defined time limit $S$ and $S'$, $S' > S$. The set $N_j$ will rather correspond to the set of demand zone that can be reached by a vehicle located in $v_j$ within $S$. Finally, when the number of ambulances is given, it is equal to $P$.

*2.1.1. Deterministic single coverage models*

Toregas et al. (1971) were the first to explicitly formulate the emergency vehicle location problem using the notion of coverage: a demand zone is said to be covered if and only if it can be reached by at least one vehicle within a prescribed time or distance frame. The location set covering problem (LSCP), proposed by Toregas et al. (1971) sought to minimize the number of vehicles such that all zones are adequately covered. It uses binary variables $x_j$ which are equal to 1 if and only if an ambulance is located at $v_j \in M$:

**LSCP**

$$\min \sum_{j=1}^{m} x_j \tag{1}$$

subject to

$$\sum_{j \in M_i} x_j \geq 1, \ i = 1, ..., n, \tag{2}$$

$$x_j \in \{0, 1\}, \ j = 1, ..., m. \tag{3}$$

In practice, the number of vehicles needed to achieve such a complete coverage can be significant, even not realistic in practice. Moreover, in many cases, managers are more interested in determining the best possible usage of their given vehicle fleet. Considering these practical limitations, Church and ReVelle (1974) formulated the maximal covering location problem (MCLP) that seeks to maximize the population covered by a vehicle fleet of size $P$. Denoting $y_i$, a binary variable equal to 1 if and only if demand zone $v_i$ is covered by at least one vehicle within $S$, this model is formulated as:

**MCLP**

$$\max \sum_{i=1}^{n} a_i y_i \tag{4}$$

subject to

$$\sum_{j \in M_i} x_j \geq y_i, \ i = 1, ..., n, \tag{5}$$

$$\sum_{j=1}^{m} x_j = P, \tag{6}$$

$$x_j \in \{0, 1\}, \ j = 1, ..., m, \tag{7}$$

$$y_i \in \{0, 1\}, \ i = 1, ..., n. \tag{8}$$

The MCLP was later studied by Eaton et al. (1985) when selecting the location of ambulances in Austin (Texas) and by Galvão and ReVelle (1996) who proposed a lagrangean heuristic to improve its resolution.

It is worth noting that both LSCP and MCLP considered only one type of vehicle. In real-life contexts, it is however common to use more than one type of vehicle. In this regard, Schilling et al. (1979) proposed three models that aim to maximize the population simultaneously covered by two types of vehicles: tandem equipment allocation model (TEAM), multi-objective tandem equipment allocation model (MOTEAM) and facility-location, equipment-emplacement technique (FLEET).

Although relatively simple in their formulation, deterministic single coverage models LSCP and MCLP lead to a significant number of variants and extensions, playing an important role in the development of the models that will be discussed hereafter. They also made a valuable contribution by their application in practice.

*2.1.2. Deterministic multiple coverage models*

Deterministic single coverage models consider the fact that a vehicle is always available upon reception of an emergency call. Clearly, this is not always the case in practice. Indeed, a vehicle may not be available to answer a demand when the time elapsed between the arrival of two consecutive calls from zones covered by the same vehicle is too short, i.e. the vehicle is still serving the first call when the second call is received. Solutions determined using single coverage models may not be robust when considered in real-life situations. To increase the solution's robustness, multiple coverage models have emerged. These models seek to increase the likelihood of having a demand zone covered by one available vehicle by increasing the number of vehicles available to cover the zone. Even if these models are neither probabilistic nor stochastic, they represent an improvement over single coverage models since they indirectly consider the random nature of emergency demands through vehicle availability.

Daskin and Stern (1981) was the first to explicitly consider multiple coverage. The hierarchical objective set covering problem (HOSC) proposed by Daskin and Stern (1981) to locate emergency vehicles in Austin (Texas) considers not only the first coverage, but all subsequent coverages. First, the HOSC minimizes the number of vehicles needed to ensure a complete coverage (i.e. that all demand zones are covered at least once), and secondly, maximizes the number of vehicles that can cover a zone. In this case, each supplementary vehicle has the same impact on the objective function eventually leading to some perverse effects. Indeed, it may not seem really interesting in practice to cover a zone with more than two vehicles if the likelihood of having these two vehicles simultaneously busy is low. Moreover, since the HOSC does not explicitly consider the population of each demand zone, it will tend to regroup vehicles around zones that can be easily covered, leaving harder to reach zones covered only once.

Eaton et al. (1986) tried to overcome the weaknesses of the HOSC by considering the population density. The Dominican ambulance deployment problem (DADP) then seeks to maximize the population that can be covered by more than one vehicle, each additional vehicle still making the same contribution to the objective function. It also minimizes the number of vehicles needed to guarantee a complete coverage. In parallel to the work of Eaton et al. (1986), Hogan and ReVelle (1986) also intended to remedy the drawbacks of the HOSC by considering the population density, but also by giving a hierarchical importance to the different coverage levels. Two models are thus formulated for this purpose: backup coverage model 1 (BACOP1) and backup coverage model 2 (BACOP2). Both models basically seek to maximize the population covered twice, given a number of vehicles to locate.

Later, Gendreau et al. (1997) presented the double standard model (DSM) which considers both the concept of double coverage and different coverage radii. The DSM seeks to determine the location of a fixed number of vehicles in order to maximize the population covered twice within a prescribed time frame $S$. The model should also ensure that at least a proportion of the population is covered within $S$ and that all the population is covered within $S'$, $S' > S$. Doerner et al. (2005) applied the DSM to locate ambulances in Austria. The authors observed that some vehicles are overused in many models, i.e. they covered a large proportion of the population within the time limit. Such an assignment can be unrealistic in practice. They thus propose to integrate a penalty term in the objective function to limit the number of inhabitants that can be assigned to an ambulance. Laporte et al. (2009) reported different applications of the DSM in Canada, Austria and Belgium. These studies show that the DSM can be used to tackle real-life problems and that the algorithms developed in such contexts allow to solve the problem efficiently. In a study conducted to optimize ambulance deployment in Shanghai, China, Su et al. (2015) refined the DSM with a new objective function that minimizes both the costs of delayed services and the operational costs. They also included, in this modified formulation of the problem, a constraint that limits ambulances' workload.

Finally, Storbeck (1982) proposed a flexible formulation based on goal programming to determine ambulance location. The maximal-multiple location covering problem (MMLCP) aims to locate a fixed fleet of vehicles in order to simultaneously minimize the population that will be left uncovered and maximize the number of demand zones covered by more than one vehicle.

### *2.1.3. Probabilistic and stochastic models*

Deterministic multiple coverage models were elaborated considering the fact that the robustness of a location plan should increase when multiple coverage is ensured. Even if they represent a significant improvement over single coverage models, they still have their limits. Indeed, ensuring the double coverage does not, in practice, guarantee a satisfying service level. On the other hand, it may not be necessary to seek double coverage if the system under study is not congested at all. To overcome these limitations, some authors have decided to consider more explicitly the different sources of uncertainty. Probabilistic and stochastic models have been developed to provide a more accurate representation of real-life situations.

The first probabilistic models presented in this section are referred to as expected covering location models. These models seek to establish the set of vehicle locations that maximizes the expected coverage. As will be described later on, the expected coverage generally considers the vehicle availability. This is in fact one of the approaches proposed to address the uncertainty related to service requests. In some cases, the expected coverage may also consider the travel time stochasticity. As shown in Erkut et al. (2009), considering more explicitly the uncertainty in location models presents clear advantages.

Daskin (1982, 1983) was among the first to integrate vehicle availability in a maximal coverage location model. The busy fraction of a vehicle, noted $q$ and defined as the probability that a vehicle is unavailable to respond to an emergency call, is then considered in the expression of expected coverage. The busy fraction is assumed to be known and takes the same value for all vehicles, independently of their location. Hence, if a demand zone $i$ is covered by $k$ vehicles, the expected covered demand is given by $E_k = a_i(1-q^k)$ where $1-q^k$ represents the probability that at least one vehicle is available and $a_i$ is the number of demands arising from demand zone $i$. The maximum expected covering location problem (MEXCLP) presented by Daskin (1982, 1983) thus aimed to locate a given number of vehicles in order to maximize the expected coverage, which depends on the busy fraction. The MEXCLP was later used by Fujiwara et al. (1987) to locate ambulances in Bangkok, Thailand. Several extensions of the MEXCLP have also been proposed. Bianchi and Church (1988) formulated MOFLEET, a MEXCLP variant that independently considers the location of standby sites and the assignment of vehicles to them. Daskin et al. (1988) proposed a more general version of the MEXCLP that can handle different coverage levels. A multi-period version of the MEXCLP that will be discussed hereafter has been developed by Repede and Bernardo (1994). Finally, the MEXCLP is studied by Aytug and Saydam (2002) who proposed a genetic algorithm to solve it.

The MEXCLP and its variants consider three main assumptions: the busy fraction is known and the same for all vehicles, the busy fraction is independent of the vehicle location, and each vehicle operates independently. As mentioned by Batta et al. (1989), these assumptions are generally not met in practice. Depending on the context, this can lead to a significant gap between the predicted and the actual system's performances. To provide a better estimate of the expected coverage, Batta et al. (1989) proposed two variants of the MEXCLP that allow the relaxation of some of its basic assumptions. Their first model, the adjusted MEXCLP (AMEXCLP), is very similar to the MEXCLP but it considers in the objective function a corrective factor from queuing theory (Larson, 1975) that allows to relax the vehicle independency assumption. In the second model, Batta et al. (1989) proposed the use of the hypercube model (Larson, 1974, 1975) to estimate the expected coverage given a pre-determined location plan. In this case, the relaxation of the three basic assumptions is allowed as well as the integration of calls that have been placed in queue in the expected coverage computation. Following the same idea, Galvão et al. (2005) developed a simulated annealing heuristic to find an optimal location plan that uses the hypercube model to compute the expected coverage.

Each model presented so far considers deterministic travel times. In practice, the travel time between two locations may however vary from one intervention to another due, for instance, to traffic congestion. Daskin (1987) suggested to determine the location of emergency vehicles, their assignment to demand zones as well as the route they should follow to reach a demand considering stochastic travel times. In this case, the expected coverage is expressed as a function of the probability that a vehicle can reach a demand zone within the prescribed time frame assuming that vehicles are always available. The model proposed by Daskin (1987) thus maximized the expected coverage while minimizing the average response time in the specific case where two emergency vehicles must be sent to an emergency call.

Goldberg et al. (1990) proposed a model that pursues similar objectives as those of the MEXCLP and its variants. This model still determines the location of a given number of vehicles that maximizes the expected coverage, but now considers stochastic travel times. In addition, dispatching decisions are assumed to be based on a preference list, i.e. a list of vehicles sorted with respect to their priority of dispatch. The expected coverage is thus expressed using the probability that a demand zone can be reached within the time frame, given the preference list. The probability that a vehicle is busy is assumed to be independent of the system state and is computed using queueing theory. The model proposed by Goldberg et al. (1990) allows to compute several performance measures given a set of fixed vehicles locations. It can be used as a descriptive model or within an optimization framework. Goldberg and Paz (1991) later developed an exchange-based heuristic that uses the model proposed by Goldberg et al. (1990). The concept of expected coverage is extend to EMS organizations that use two types of vehicles in Mandell (1988). Finally, Ingolfsson et al. (2008) proposed a model inspired by the one of Goldberg et al. (1990) but in which the chute time variability is also considered. The chute time is defined as the time elapsed between the arrival of the call and the dispatching of a vehicle to the corresponding emergency.

Each model presented in this section aims to maximize the expected coverage by taking into account the probability that a vehicle can serve a demand zone within a prescribed time frame. Chance-constrained programming has been considered as a second approach to address the different sources of uncertainty arising in the emergency vehicle location. Unlike expected coverage models, chance-constrained ones consider the uncertainty in the formulation of a set of probabilistic constraints that will guarantee the system reliability. A lower bound on the number of vehicles needed to ensure an adequate coverage of each demand zone is determined consequently. The methodology used to compute this lower bound varies from one model to another.

Following this idea, ReVelle and Hogan (1988) formulated the probabilistic location set covering problem (PLSCP), a probabilistic version of the LSCP. It seeks to minimize the number of vehicles required to ensure that at least one vehicle is available for each demand with a given level of reliability. In this case, probabilistic constraints are formulated using zone-specific busy fractions. The busy fraction of a given zone is then expressed as the ratio between the service time for demands arising in the zone and the availability of the vehicles that can ensure their coverage. Following the same idea, ReVelle and Hogan (1989) proposed a probabilistic version of the MCLP called the maximal availability location problem (MALP). The MALP determines the location of a given number of vehicles that maximizes the population covered by at least one available vehicle within the prescribed time frame, with a given level of reliability. Two versions of the MALP have been formulated. These two versions basically differ in the way the busy fraction is expressed. In the first version, the busy fraction is assumed to be the same for all demand zones. The number of vehicles required to guarantee the service's reliability is consequently the same for all demand zones. The second version rather uses zone-specific busy fractions as in the PLSCP. ReVelle and Marianov (1991) also proposed a probabilistic version of FLEET referred to as PROFLEET. The model aims to maximize the number of calls that can be simultaneously covered by two types of vehicles with a given level of reliability. Recently, Shariat-Mohaymany et al. (2012) exploited the idea of the PLSCP to formulate a model that limits the demand assigned to each vehicle, in addition to providing a minimum reliability level for each demand zone.

MALP, PLSCP and PROFLEET assume that busy fractions are vehicle independent. However, this assumption can, in some cases, impact the busy fraction estimate and consequently, the predicted system performances. To relax this assumption and thus provide a more accurate estimate of the actual system's performances, Marianov and ReVelle (1994, 1996) proposed two models, Q-PLSCP and Q-MALP, that are, in fact, straightforward extensions of the PLSCP and MALP. In such models, queueing theory is used to compute the number of vehicles needed to ensure the system's reliability. More recently, Harewood (2002) developed a multi-objective variant of the Q-MALP. Galvão et al. (2005) also proposed an extension of the MALP that uses queueing theory to better represent real-life situations. This model integrates a corrective factor within the probabilistic constraints formulation that allows the consideration of vehicle-specific busy fractions rather that zone-specific ones as well as cooperation between vehicles.

Up to now, most of the models proposed considered the availability of vehicles as the main source of uncertainty. This is in fact a first approach to consider indirectly the uncertainty related to the arrival of demands. In their model called REL-P, Ball and Lin (1993) decided to consider the random behaviour of service requests more directly. It is thus assumed that each demand is generated according to a given probability distribution. They formulated a constraint that limits the probability that the number of calls arising in a given region and for a given time period is greater than the number of vehicles available to cover the zone, with a given level of reliability. The objective function of the REL-P minimizes the total cost incurred to guarantee the service reliability for each demand zone. Borràs and Pastor (2002) later proposed a variant of REL-P that uses a vehicle-specific busy fractions rather than a zone-specific ones as it is the case in REL-P.

### *2.2. Recent models and new approaches*

Most of the early models indirectly consider the uncertainty related to both demands and travel times. Many of them also use the coverage to assess the system's performances. More recently, researchers have continued to work on the static ambulance location problem with the aim of addressing more explicitly some of the issues related to the randomness. Alsalloum and Rand (2006) refined the expected coverage definition to consider random travel times. Beraldi et al. (2004), Beraldi and Bruni (2009) and Zhang and Jiang (2014) decided to consider more explicitly the assignment of vehicles to emergency demands within a stochastic framework. This section thus depicts these recent models new approaches.

Similarly to Ball and Lin (1993), Beraldi et al. (2004) studied the vehicle fleet location and dimensioning, as well as the assignment of vehicles to demand zones. They also intended to consider more explicitly service requests randomness. They then assumed that each vehicle is able to serve at most a given number of emergency calls during the planning horizon. The model proposed by Beraldi et al. (2004) is thus closely related to classical location-allocation models. Nevertheless, it still takes into account that a vehicle is able to serve an emergency demand if and only if it can reach it within a prescribed time frame. In both their deterministic and probabilistic models, $y_j$ denotes a binary variable equal to 1 if, and only if, the site $j$ is used, $x_{ij}$, the number of vehicles located in $j$ that will serve the demand zone $i$, $p_j$, the limit on the number of vehicles that can be located in $j$, $c_{ij}$, the cost of allocating a demand zone $i$ to the site $j$, $f_j$, site $j$ opening costs. Other variables and parameters have been defined at the beginning of the current section. The deterministic model corresponding to the problem under study and that serves as the basis for their probabilistic model is:

$$\min \sum_{i=1}^{n} \sum_{j=1}^{m} c_{ij} x_{ij} + \sum_{j=1}^{m} f_j y_j \tag{9}$$

subject to:

$$\sum_{j \in M_i} x_{ij} \geq a_i, i = 1, ..., n, \tag{10}$$

$$\sum_{i \in N_j} x_{ij} \leq p_j y_j, j = 1, ..., m, \tag{11}$$

$$x_{ij} \geq 0,\ integer, i = 1, ..., n, j = 1, ..., m, \tag{12}$$

$$y_j \in \{0, 1\}, j = 1, ..., m. \tag{13}$$

This model intends to determine the ambulance standby sites and the assignment of ambulances to demand zone that minimize the total cost (9), including opening and assignment costs. It should also ensure that each demand zone is covered by at least one vehicle within a given time frame (10) and that the limit for the number of vehicles that can be positioned at each standby site is satisfied (11). Nevertheless, the reliability of EMS systems often relies on their capacity to serve random service requests, which corresponds, from a mathematical perspective, to the satisfaction of demand related constraints (10). To ensure demand satisfaction with a given level of reliability, Beraldi et al. (2004) proposed the formulation of probabilistic constraints. Considering a global reliability level rather than local ones as it is the case with the REL-P, the probabilistic constraints proposed by Beraldi et al. (2004) can be expressed by:

$$P(\sum_{j \in M_i} x_{ij} \geq \zeta_i, i = 1, ..., n) \geq \alpha \tag{14}$$

where $\zeta_i$ represents the random variable related to the demand placed in $i$. Considering that each demand follows an independent random distribution, a deterministic equivalent can be formulated, where the probabilistic constraints are replaced by a set of deterministic constraints. The stochastic model proposed by Beraldi et al. (2004) can thus be formulated in a similar way than its deterministic counterpart where the constraint (10) is replaced by the set of following constraints:

$$\sum_{i=1}^{n} \sum_{k=1}^{k_i} a_{ik} z_{ik} \geq \beta \tag{15}$$

$$\sum_{j \in M_i} x_{ij} = l_i + \sum_{k=1}^{k_i} z_{ik}, \tag{16}$$

$$z_{ik} \in \{0, 1\} \tag{17}$$

where

$$a_{ik} = ln(F_i(l_i + k)) - ln(F_i(l_i + k - 1)), \tag{18}$$

$$\beta = ln(\alpha) - ln(F(l)), \tag{19}$$

$$l_i = F_i^{-1}(\alpha) \tag{20}$$

and where $F_i$ is the probability distribution of $\zeta_i$. The reliability of the system is then ensured by the satisfaction of this set of constraints.

More recently, Alsalloum and Rand (2006) formulated an extension of the MCLP to locate an ambulance fleet in Riyadh (Saudi Arabia). The MCLP simultaneously integrates the concept of expected coverage and the formulation of probabilistic constraints, and is concerned with the standby sites selection and the assignment of vehicles to demand zones. It determines the best standby site selection based on the expected coverage, as well as the number of vehicles to locate at each of the selected standby sites in a way that each demand zone can find at least one available vehicle when a demand is placed. In this case, rather than considering a strictly binary coverage measure, the authors proposed to consider a coverage probability. The coverage probability denoted $P_{ij}$ is defined as the probability that a demand zone $i$ can be reached by a vehicle located at a standby site $j$ within a prescribed time frame. This probability is strongly influenced by the travel time between the demand zone and the standby site considered. According to the authors, this allows for a better assessment of the expected coverage, especially when the aggregation level is high. The availability of vehicles is considered through a probabilistic constraints where the minimum number of vehicles required to achieve a given level of reliability is obtained using queueing theory. To model the MLCP, some variables and parameters are defined: $y_{ij}$ is a binary variable equal to 1 if, and only if, $P_{ij} \geq \rho$ where $\rho$ is a predetermined bound and $j$ is the location for which the distance between $i$ and $j$ is the smallest, $x_{jk}$is a binary variable equal to 1 if, and only if, $k$ vehicles are located in $j$, $\lambda_i$, the arrival rate of demand placed in $i$, $a_i$, the proportion of demand arising in $i$, and $r_k$, a boundary value on the arrival rate that requires the addition of a vehicle (from $k$ to $k+1$) determined by the means of queueing theory. The proposed goal programming model is then:

$$\min P_0 d_0^- + P_1 \sum_{j=1}^{m} d_j^+ \tag{21}$$

subject to:

$$\sum_{i=1}^{n}\sum_{j=1}^{m} a_i P_{ij} y_{ij} + d_0^- = 1, \tag{22}$$

$$\sum_{1<k<p_j} r_k x_{jk} - \sum_{i=1}^{n} \lambda_i y_{ij} - d_j^+ = 0, \ j = 1, ..., m, \tag{23}$$

$$\sum_{j=1}^{m} y_{ij} \leq 1, \ i = 1, ..., n, \tag{24}$$

$$\sum_{k=1}^{p_j} x_{jk} \leq 1, \ j = 1, ..., m, \tag{25}$$

$$\sum_{j=1}^{m}\sum_{k=1}^{p_j} x_{jk} = P, \tag{26}$$

$$y_{ij}, x_{jk} \in \{0,1\}, i = 1,...,n, j = 1,...,m, k = 1,...,p_j. \tag{27}$$

More explicitly, this model simultaneously aims to minimize the population that is not expected to be adequately covered and the number of vehicles to locate. It needs to guarantee that each demand zone can find at least one vehicle when a demand is placed, that the capacity of each standby site is satisfied, and that a given number of standby sites are utilized. $P_0$ and $P_1$ represent the weight related to each objective while $d_0^-$ and $d_j^+$ represent the deviation corresponding to each of these objectives. Deviations are computed within constraint (22) and (23). Finally, constraints (24), (25) and (26) respectively ensure the coverage of each demand zone, the satisfaction of the capacity constraint of each standby site $j$ and the location of exactly $P$ sites. The authors used the model to determine the optimal location of an ambulance fleet in the city of Riyadh, in Saudi Arabia.

In addition, Beraldi and Bruni (2009) proposed an ambulance location model that is neither based on the expected coverage nor on the definition of probabilistic constraints. The model is rather formulated as a two-stage stochastic program. To the best of our knowledge, this is the first attempt to apply two-stage stochastic programming to the static ambulance location. Two decisions stages are thus considered: the first stage selects the location of standby sites, and the second determines the assignment of incoming service requests to standby sites once the uncertainty about service requests is revealed. Denoting $x_j$, the number of vehicles located in $j$, $z_j$, a binary variable equal to 1 if, and only if, the site $j$ is selected, $a(w)$, the random vector corresponding to demand realization, $y_{ij}(w)$, a binary variable equal to 1 if, and only if, demand zone $i$ is assigned to vehicles located in $j$ when $a(w)$ is known, $c_j$, the cost of locating a vehicle in $j$, $f_j$, the opening cost of $j$, $d_{ij}$, the distance or the cost involved when a demand arising in $i$ is served from site $j$, $N_j$, the set of demand zones that can be covered by the site $j$ within the prescribed distance or time frame $S$ and $M_i$, the set of sites that can ensure the coverage of a demand zone $i$ within $S$, Beraldi and Bruni's model is formulated as follows:

$$\min \sum_{j=1}^{m} (c_j x_j + f_j z_j) + E_w[Q(x,z,w)] \tag{28}$$

$$x_j \leq p_j z_j, j = 1,...,m, \tag{29}$$

$$z_j \in \{0,1\}, x_j \ integer, \ j = 1,...,m \tag{30}$$

where

$$Q(x,z,w) = \min_y \sum_{i=1}^{n} \sum_{j=1}^{m} d_{ij} y_{ij}(w), \tag{31}$$

$$\sum_{i \in N_j} a_i(w) y_{ij}(w) \leq x_j, \ j = 1,...,n, \tag{32}$$

$$\sum_{j \in M_i} y_{ij}(w) \geq 1, \ i = 1,...,n, \tag{33}$$

$$y_{ij}(w) \leq x_j, \ i = 1,...,n, \ \ j = 1,...,m, \tag{34}$$

$$y_{ij}(w) \in \{0,1\}, \ i = 1,...,n, \ \ j = 1,...,m. \tag{35}$$

In particular, the objective function includes two terms: the first one aims to minimize the location costs, and the second, the recourse action costs, is defined as the cost of serving the demand. The first stage constraints stipulate that a limited number of vehicles must be located at each standby sites. Meanwhile, the second stage constraints force the number of vehicles at a given standby site to be sufficient to cover demands assigned to it, and ensure that each demand is assigned to at least one vehicle once the uncertainty is revealed. Each demand arising in $i$ must by assigned to vehicles located in $j$ if, and only if, site $j$ is selected (34). The authors also proposed to integrate probabilistic constraints to the model to ensure the system's reliability through demand-related constraints satisfaction. To solve the problem, an exact solution approach as well as three heuristics have been developed. The solution methodologies are validated using a set of different size instances, and considering sets of scenarios varying from 10 to 40. Zhang and Jiang (2014) followed the same idea to formulate an ambulance location model that minimizes operating cost of ambulances and transportation costs, as well as demands not fulfilled in time. However, they resorted to robust programming instead of two-stage stochastic programming to address demand uncertainty. The proposed methodology was validated using historical data from a major city in China, which includes from 30 to 70 candidate stations.

More recently, some researchers have considered the use of simulation tools to better approximate the expected performance of the system within a simulation-optimization framework. Indeed, simulation offers an opportunity to evaluate system's performances without the assumptions needed to adequately elaborate most of the previous models. Mason (2013) therefore reports the use of a simulation model to evaluate a set of neighbouring solutions in a local search heuristic that seeks to determine the best possible set of standby sites for an ambulance fleet. Zhen et al. (2014) also considered the use of simulation to evaluate solutions in the genetic algorithm developed in their study. In this case, the number of vehicles to assign to each standby sites whose location has been already fixed must be determined. This methodology was successfully applied to the context of Shanghai, in China where 80 vehicles and 12 standby sites were considered.

## 3. Multi-period and dynamic relocation problems

*Static* ambulances location problems basically aim to select the set of stations or standby points where vehicles will be waiting between two assignments to emergency calls. Then, a vehicle will be assumed to return to its home base after its mission's completion. Under some circumstances, it may however be more interesting to modify a vehicle's home base during a day to better consider the evolution of the system over time. The relocation problem thus consists in relocating available vehicles among potential stations or standby points to ensure an adequate service to the population. The evolution of the system can be the result of demand's pattern fluctuations throughout a day due, for instance, to population's movements. To take those fluctuations into account, a workday is divided into several time periods. Different location plans are established for each time period, then vehicles are moved between periods to reach the next location plan. This problem is referred to as the *multi-period* relocation problem. The evolution of the system can also be the result of system state's variations. For instance, when some vehicles are responding to emergency demands, the system has to operate with a reduced vehicle fleet. Considering this fact, the system state is assumed to vary as the number of available vehicles will evolve through ambulance dispatches and ends of mission. Vehicles will then be relocated when the system changes and requires it in order to maintain a proper service level. In this case, since relocation decisions depend on the system's state, the problem is referred to as the *dynamic* relocation problem.

Both multi-period and dynamic ambulance relocation problems are in fact closely related to their static counterpart. Nevertheless, relocation problems have some characteristics of their own. First of all, the static ambulance location problem is generally considered at the tactical level. The relocation problem is rather considered at the operational level and, in some cases, even solved in real-time. Indeed, EMS managers often have to make very quick decisions when it comes to dispatching and relocation decisions to ensure a proper service level. In addition to the decision-making level difference, relocation problems usually include a set of practical constraints that aims to ensure the system stability, which is not the case in static location problems. Upper bound on the number of relocated vehicles or on the traveled distance can then be taken into account. In this manner, the trade-off between service level and relocation costs can be considered.

In the past few years, a lot of attention has been devoted to the development of approaches to solve both the multi-period and the dynamic relocation problems, and new studies continue to appear regularly. Indeed, as shown in Bélanger et al. (2014), in many cases, a better service level can be achieved using more flexible management strategies such as relocation. In this section, we trace the evolution of these approaches from the early ones to the most recents, and discuss their respective particularities and characteristics.

### *3.1. Multi-period relocation models*

When addressing the ambulance location problem arising in Louisville (Kentucky), Repede and Bernardo (1994) observed that the location models proposed so far did not consider demand variations over time. However, in the case of Louisville as in many other contexts, demand patterns may vary according to the time period. Repede and Bernardo (1994) therefore formulated the maximal expected coverage location model with time variation (TIMEXCLP), a multi-period variant of the MEXLCP. To the best of our knowledge, the TIMEXCLP is the first multi-period ambulance relocation model. As in the MEXCLP, the TIMEXCLP maximizes the expected coverage, but also considers variations in the demand patterns and the number of available vehicles with respect to the time periods. However, it did not explicitly account for relocation costs between periods, which makes the TIMEXCLP a straightforward extension of the MEXCLP. Both formulations are thus very similar, except for a time index added to adequately address the different time periods. The TIMEXCLP is solved only once a priori (i.e. before operating the system), and selected location plans are applied according to the time period. Cost between periods, which includes start-up and relocation cost, was recently been included in the work of van den Berg and Aardal (2015).

More than a decade after Repede and Bernardo (1994), Rajagopalan et al. (2008) proposed a multi-period variant of the PLSCP called the dynamic available coverage location model (DACL). The DACL seeks the minimum number of vehicles required to guarantee the coverage of each demand zone, with a given level of reliability and considering several time periods. A corrective factor is also included in the formulation of the probabilistic constraints to ensure the system's reliability, which is not the case in the PLSCP. As in the TIMEXCLP, the DACL did not integrate any constraints that take into account the relocation of vehicles between periods. A reactive tabu search metaheuristic is proposed by Rajagopalan et al. (2008) to solve this problem. Later, Saydam et al. (2013) extended the DACL to also consider the minimization of the number of relocated vehicles between periods. The metaheuristic developed by Rajagopalan et al. (2008) was then adapted to address this new objective. In both contexts, location plans are computed once a priori for all periods and applied when needed.

Başar et al. (2011) considered the problem of determining where and when ambulance stations should be open over a multi-period planning horizon. More precisely, the multi-period backup double covering model (MPBDCM) proposed by Başar et al. (2011) is a multi-period combination of BACOP and DSM. It aims to determine which stations to use at each period so that the population covered by two distinct stations is maximized for all periods. The main difference between the MPBDCM and previous multi-period relocation problems is that it considers that, when a station is open at one period, it must remain open until the end of the planning horizon. This can be justified when the station status changes involve significant costs or inconveniences. It also allows to consider longer time periods. To adequately formulate the MPBDCM, the following variables and parameters need to be defined. Denoting $x_{jt}$, a binary variable equal to 1 if, and only if, site $j$ is used at period $t$, $u_{it}$, a binary variable equal to 1 if demand zone $i$ is covered two times respectively within $S$ and $S'$ at period $t$, and considering $p_t$, a parameter denoting the maximum number of sites that can be used at period $t$ and $a_{it}$, the population of demand zone $i$ at period $t$, the MPBDCM is:

**MPBDCM**

$$\max \sum_{t=1}^{T} \sum_{i=1}^{n} a_{it} u_{it} \tag{36}$$

subject to:

$$\sum_{j=1}^{m} x_{jt} \leq p_t, \ t = 1, ..., T, \tag{37}$$

$$\sum_{j \in M_i} x_{jt} - u_{it} \geq 0, \ i = 1, ..., n, \ t = 1, ..., T, \tag{38}$$

$$\sum_{j \in M_i'} x_{jt} - 2u_{it} \geq 0, \ i = 1, ..., n, \ t = 1, ..., T, \tag{39}$$

$$x_{jt} - x_{j,t-1} \geq 0, \ j = 1, ..., m, \ t = 1, ..., T, \tag{40}$$

$$x_{jt}, u_{it} \in \{0, 1\}, \ i = 1, ..., n, \ j = 1, ..., m, \ t = 1, ..., T. \tag{41}$$

The MPBDCM seeks to maximize the population covered twice within $S$ and $S'$ respectively (36). It must also ensure that the maximum number of stations used at each period is respected (37), that each demand zone is adequately covered (38) (39), and that a station remain open until the end of the planning horizon once it has been open (40). To solve this model, Başar et al. (2011) developed a tabu search metaheuristic, and applied it in the case of the city of Istanbul, Turkey.

Finally, Schmid and Doerner (2010) presented the multi-period double standard model (mDSM), a multi-period extension of the DSM. The mDSM differs from its static counterpart by the fact that it now considers the travel time variation between periods due, for instance, to traffic congestion. Indeed, the authors observe that in many cases, travel times may vary significantly throughout a day. However, only a few models address this issue, which can lead to an inadequate estimation of the system's performance. To overcome this drawback, Schmid and Doerner (2010) proposed the mDSM that considers period dependent travel times. A set of locations from which a demand zone can be reached within a time frame is then defined for each period. Nevertheless, the number of available vehicles, the capacity of potential standby sites as well as the demand patterns remain the same over the entire planning horizon. The mDSM is thus relatively similar to the DSM besides the fact that it integrates a penalty term in the objective function to limit the number of relocated vehicles between periods. Moreover, it considers a limit on the number of demands that a vehicle can serve, which was not the case in the first version of the DSM. This constraint is integrated in the DSM only in Doerner et al. (2005). Schmid and Doerner (2010) proposed a variable neighbourhood search metaheuristic to solve the mDSM, then used it to address the ambulance location problem encountered in Vienna, Austria.

The different models presented so far considered that the demand, as well as the number of available vehicles and travel times, may vary throughout a day. Although they may offer a better representation of real-life situations, they do not allow to consider explicitly the system's state changes following vehicle dispatches or ends of missions. To deal with these situations, the dynamic relocation must rather be considered.

### *3.2. Dynamic relocation models*

The first ambulance relocation model that explicitly accounts for EMS dynamic nature has been proposed by Gendreau et al. (2001). The ambulance relocation problem ($RP^t$) is based on the DSM proposed by the same authors. The $RP^t$ still maximizes the population covered by at least two vehicles within the prescribed time frame, but simultaneously intends to minimize relocation costs. For this purpose, the objective function includes a penalty term that takes into account the relocation history of vehicles. This penalty term aims to avoid too long and round-trips, as well as to move the same ambulance repeatedly, and is updated each time a relocation is performed. Denoting $u_i$, a binary variable that is equal to 1 if the demand zone $i$ is covered at least twice, $x_{jk}$ a binary variable equal to 1 if, and only if, the vehicle $k$ is located in $j$, and $M^t_{kl}$, the penalty term related to the relocation of vehicle $k$ from its current location to location $j$ at time $t$, the $RP^t$ function is formulated as follows:

$$\max \sum_{i=1}^{n} a_i u_i - \sum_{j=1}^{m} \sum_{k=1}^{p} M^t_{jk} x_{jk} \tag{42}$$

Moreover, considering $y_i$ a binary equal to 1 if, and only if, demand zone $i$ is covered at least once, the set of constraints is:

$$\sum_{j \in M'_i} \sum_{k=1}^{p} x_{jk} \geq 1, \; i = 1, ..., n, \tag{43}$$

$$\sum_{i=1}^{n} a_i y_i \geq \alpha \sum_{i=1}^{n} d_i, \tag{44}$$

$$\sum_{j\in M_i}\sum_{k=1}^{p} x_{jk} \geq y_i + u_i,\ \ i = 1, ..., n, \tag{45}$$

$$y_i \geq u_i,\ \ i = 1, ..., n, \tag{46}$$

$$\sum_{j=1}^{m} x_{jk} = 1,\ \ k = 1, ..., p, \tag{47}$$

$$\sum_{k=1}^{p} x_{jk} \leq p_j,\ \ j = 1, ..., m, \tag{48}$$

$$y_i, u_i \in \{0, 1\}\ i = 1, ..., n, \tag{49}$$

$$x_{jk} \in \{0, 1\},\ \ j = 1, ..., m, k = 1, ..., p. \tag{50}$$

In practice, the $RP^t$ should be solved each time a vehicle is dispatched to an emergency call. However, in EMS contexts, relocation decisions must generally be made in real-time, and the computational time needed to solve such a relocation problem may be too long to consider each time a vehicle is dispatched to a call. To overcome this difficulty, the authors proposed to take advantage of the time available between two calls to determine the relocation plan associated with each possible dispatching decision. This way, when the identity of the dispatched vehicle becomes known, the corresponding relocation plan can be applied directly. Therefore, since the $RP^t$ must be solved for each possible dispatching decisions, an efficient solution methodology is required. The authors thus proposed a tabu search metaheuristic inspired by the one proposed to solve the DSM (Gendreau et al., 1997). In addition, they proposed the use of parallel computing to solve the different relocation problems in a timely manner. The proposed methodology has been successfully tested on the data of Montreal, Canada. Moeini et al. (2013) recently extended the $RP^t$ to integrate different requirements in terms of coverage, i.e. only some of the demand zones will really benefit from double coverage. As mentioned by the authors, the development of such a model is justified in the context of the study carry out in Val-de-Marne (a county in France) where the intensity of emergency service demands is quite low. In this case, approximately 20 to 30 calls per day required the dispatch of one of the 8 ambulances used to provide the service.

Gendreau et al. (2006) proposed another dynamic relocation model that is particularly relevant to physician's cars location that they named the maximal expected relocation problem (MECRP). The proposed approach intended to determine the appropriate relocation plan for each possible system's states, defined as the number of available physician cars. Such an approach is possible since the number of physician's cars is relatively small. Therefore, the MECRP aims to maximize the expected coverage for all possible system's states. It also includes a limit on the number of vehicles that can possibly be relocated between states. Denoting $x_{jk}$, a binary variable equal to 1 if, and only if, a vehicle located in $j$ at state $k$, $y_{ik}$, a binary variable equal to 1 if, and only if, demand zone $i$ is covered by at least one vehicle at state $k$, $u_{jk}$, a binary variable equal to 1 if, and only if, the location $j$ is no longer used when the system goes from state $k$ to state $k+1$, and $q_k$, the probability of reaching state $k$, $k = 0, ..., P$ where $P$ is the total number of vehicles, the

MECRP is formulated as follows:

$$\max \sum_{k=1}^{P} \sum_{i=1}^{n} a_i q_k y_{ik} \tag{51}$$

subject to:

$$\sum_{j \in M_i} x_{jk} \geq y_{ik}, \ i = 1, ..., n, \ k = 0, ..., P, \tag{52}$$

$$\sum_{j=1}^{m} x_{jk} = k, \ k = 1, ..., P, \tag{53}$$

$$x_{jk} - x_{j,k+1} \leq u_{jk}, \tag{54}$$

$$\sum_{j=1}^{m} u_{jk} \leq \alpha_k, \ k = 1, ..., P-1, \tag{55}$$

$$x_{jk} \in \{0,1\}, u_{jk} \in \{0,1\}, \ j = 1, ..., m, \ k = 1, ..., P, \tag{56}$$

$$y_{ik} \in \{0,1\}, \ i = 1, ..., n, \ k = 1, ..., P. \tag{57}$$

It is worth noting that MECRP constraints are similar to those formulated in the MCLP. Nevertheless, constraints must be added to control the number of vehicles relocated between states. The MECRP is thus solved only once a priori, and the relocation plan corresponding to each state will be applied when needed. This model is solved and validated by the means of CPLEX using the data of Montreal, Canada. Maleki et al. (2014) formulated two models that specify the movements of ambulances from hospitals to stations and from stations to other stations, using the output of the MECRP. They applied this methodology to locate ambulances in four districts of Isfahan, Iran.

Andersson and Värbrand (2007) proposed a dynamic ambulance relocation model that differs from the one of Gendreau et al. (2001) by the way it assesses the system's performance. Indeed, this model considers the preparedness measure, defined as the capacity of the system to answer future demands, rather than a coverage measure as in most of the previous models. The preparedness of a given demand zone $i$ considers $a_i$, a weight that mirrors the demand for ambulance in the zone (here the population), $K_i$, a given number of vehicles that will be used in the computation of the preparedness of the zone (generally the closest vehicles), $t_i^k$, the travel time for each considered vehicle $k$ to the zone $i$, and $\gamma^k$, a contribution factor of each considered vehicle. Considering these parameters, the demand zone $i$ preparedness, $\varrho_i$, is given by the following equation:

$$\varrho_i = \frac{1}{a_i} \sum_{k=1}^{K_i} \frac{\gamma^k}{t_i^k}. \tag{58}$$

In practice, the level of preparedness for each demand zone is verified regularly, and the relocation of vehicles is launched when it drops below a predetermined value. To determine the best relocation plans, the authors proposed to solve a model that they referred to as DYNAROC. This model seeks to minimize the maximal travel time required to perform the relocation, i.e. for any of the relocated ambulances. As it was done in the $RP^t$, DYNAROC includes a set of practical constraints that limit relocation travel times and the number of relocated vehicles. Moreover, a limit is considered on the level of preparedness to achieve for each demand zone after a relocation. Denoting $x_i^k$, a binary variable equal to 1 if, and only if, the vehicle $k$ is relocated to a standby site in zone $i$, $N_k$, the set of zones that can be reach by the vehicle $k$ within a prescribed delay $S$, and $P_{max}$, a parameter denoting the maximal number of relocated vehicles, DYNAROC is:

$$\min z \tag{59}$$

subject to:

$$z \geq \sum_{i \in N_k} t_i^k x_i^k, \ k = 1, ..., P, \tag{60}$$

$$\sum_{i \in N_k} x_i^k \leq 1, \ k = 1, ..., P, \tag{61}$$

$$\sum_{k=1}^{P} \sum_{i \in N_k} x_i^k \leq P_{max}, \tag{62}$$

$$\frac{1}{a_i} \sum_{l=1}^{K_i} \frac{\gamma^k}{t_i^l(x_1^1, ..., x_N^P)} \geq \varrho_{min}, \ i = 1, ..., n, \tag{63}$$

$$x_j^k \in \{0, 1\}, \ i = 1, ..., n, \ k = 1, ..., P. \tag{64}$$

To solve this model, the authors proposed a tree-search heuristic, and then tested it using the data from Stockholm, Sweden.

In a similar way that what was done in Gendreau et al. (2006) for physician's cars location, Nair and Miller-Hooks (2009) presented a location-relocation model that considers the evolution of the system's state over time. In this case, the system's state at a given time is defined by the incoming call probability distributions, the number of available vehicles and the travel time within the road network at that particular time. The multi-objective model proposed by Nair and Miller-Hooks (2009) includes two objectives that maximize the double coverage and minimize location-relocation costs respectively. Denoting $R$, the set of all possible system's states, $|R| = r_{max}$, $u_{ir}$, a binary variable equal to 1 if, and only if, the demand zone $i$ is covered by at least two vehicles within a prescribed delay at state $r$, $p_{ir}$, the probability that a call arising from demand zone $i$ at state $r$, $x_{jr}$, a binary variable equal to 1 if, and only if, a vehicle is located in $j$ a state $r$, $c_{jr}$, the cost associated with such location, $w_{jl}^{rr'}$, a binary variable that is set to 1 if a vehicle is relocated in $j$ to $l$ when the system's state goes from $r$ to $r'$, $RR_{jl}^{rr'}$, the cost of such relocation, and $PP_{rr'}$, the probability of going from state $r$ to $r'$, these two objectives are:

$$Z_1 = \max \sum_{r=1}^{r_{max}} \sum_{i=1}^{n} p_{ir} u_{ir} \tag{65}$$

$$Z_2 = \min \sum_{r=1}^{r_{max}} \sum_{i=1}^{n} c_{ir} y_{ir} + \sum_{r=1, r \neq r'}^{r_{max}} \sum_{r'=1}^{r_{max}} \sum_{j=1}^{m} \sum_{l=1}^{m} PP_{rr'} RR_{jl}^{rr'} w_{jl}^{rr'} \tag{66}$$

The model' constraints are very similar to the ones of the DSM (Gendreau et al., 1997) and similar models, but adapted to consider the studied context. In the same way as it was done for the MECRP, the model is solved once a priori to establish a relocation plan for each possible system's state. Results obtained using the data of the city of Montreal, Canada, show that the use of such strategy can improve system's performances by 1.3 % to 6.4 % according to the number of available vehicles.

As opposed to previous models, in the work of Maxwell et al. (2009), dynamic programming is considered to formulate the dynamic ambulance relocation problem. The problem studied by the authors is however limited to the relocation of vehicles that just completed their mission. From a human resources standpoint, it allows to reduce the inconveniences related to relocation. Moreover, from a mathematical standpoint, it significantly reduces the number of possible decisions. When a vehicle completes a mission, the relocation problem considered consists in determining its next standby site such that the number of calls that can be reached within a given time frame is maximized. More precisely, the optimality equation seeks the policy that minimizes the discounted total expected cost given an initial state, where the cost is defined as the number of calls that cannot be reached in a timely manner. The model uses $s$, the system's state defined according to the current time and event, to a vector $A$ that describes each vehicle's state, and to a vector $C$ that describes each call's state, $X(s)$, the set of all feasible decisions at state $s$, $c(s_k, x_k, s_{k+1})$, the transition cost from state $s_k$ to state $s_{k+1}$ given a decision $x_k$, $f(s, x, w(s, x))$, the transfer function that depends on the system's state, the decision made and random elements $w(s, x)$, $\alpha$, a fixed discounted factor with $\alpha \in [0, 1[$ and $\tau(s)$ the time at which the system visits state $s$. The policy that minimizes the discounted total expected cost given an initial state $s$ can be determined by computing the value function through the optimality equation:

$$J(s) = \min_{x \in X(s)} \left\{ E[c(s, x, f(s, x, w(s, x))) + \alpha^{\tau(f(s,x,w(s,x))) - \tau(s)} J(f(s, x, w(s, x)))] \right\}. \quad (67)$$

It is worth noting that in this case, the cardinality of the $X(s)$ is relatively small since only one vehicle is eligible for relocation. The transition cost allows to compute the number of calls that will be reached beyond the time frame: $c(s_k, x_k, s_{k+1})$ is equal to 1 if the next event $e(s_{k+1})$ is of the form *ambulance $i$ arrives at scene of call $j$*, and if the corresponding call is urgent and the time frame is exceed and 0, otherwise. Nonetheless, the evaluation of the corresponding value function is still a challenging task. Indeed, the high dimensionality of the state variable leads to a large number of possible values of the system's state. Classical dynamic programming algorithms cannot be directly applied. To overcome this situation, the authors proposed the use of approximate dynamic programming. The new challenge thus consists in selecting the right value for the parameters needed to determine an adequate approximation of the value function. When such a good approximation is found, the optimal policy can be identified by enumerating each possible decision and evaluating the corresponding expected value using Monte Carlo simulation. Results of this study show that the optimal policy allows for an improvement of approximately 4 % over a myopic policy, i.e. returning the vehicle to its home base. In this case, the data used comes from Edmonton, Canada. Moreover, they show that the system's performances can be improved by considering more frequent relocations and involving more vehicles in the relocation process as it was done in previous models. Consequently, computational time would increase significantly.

Schmid (2012) also used dynamic programming to formulate the dynamic ambulance relocation problem.

As in the case of Maxwell et al. (2009), relocations decisions are considered when a vehicle completes its mission and the only vehicle involved in the relocation process is the newly idle one. In the particular context of the study, relocating idle/free vehicles is prohibited by law. The problem addressed in Schmid (2012) has been developed to support both relocation and dispatching decisions. Its objective is to minimize the average response time over a finite planning horizon, while considering the variation of the travel times and demand density with respect to time. The Bellman equation associated with this problem is then formulated as follows:

$$V_t(S_t) = \min_{x_t}(c(S_t, x_t) + E\{V_{t+1}(S_{t+1}(S_t, x_t, W_{t+1}))\}), \tag{68}$$

where $S_t$ denotes the state of the system at time $t$, i.e. the demand's state and the vehicle state, $x_t$, a decision taken at time $t$, $W_t$, the information that is revealed for time $t-1$ to $t$, and $c(S_t, x_t)$, the contribution of a decision $x_t$ taken when the system's state is $S_t$ in the response time computation. Decisions are made according to a policy $X_t^\pi(S_t)$ that returns a decision vector $x_t$ that is feasible at state $S_t$. The optimal policy is thus minimized, given a discounted factor $\gamma$, the sum of the expected response times over the planning horizon $T$ using:

$$\min_{\pi \in \Pi} E \sum_{t=0}^{T} \gamma^t c_t(S_t, X_t^\pi(S_t)). \tag{69}$$

As in Maxwell et al. (2009), approximate dynamic programming has been considered to determine the optimal policy. Results obtained based on the data of Vienna, Austria, showed that the optimal policy allows for a 13% average response time improvement over myopic policies, which consist in always dispatching the nearest vehicle to a call and returning the vehicle to its home base.

Naoum-Sawaya and Elhedhli (2013) addressed the dynamic ambulance relocation problem by means of a two-stage stochastic programming approach. They considered the first stage decisions to be the ones concerned with vehicle location, and the second stage decisions as the ones concerned with the assignment of vehicles to emergency demands, upon actual reception of the emergency demands. To formulate the problem, the authors used $c_k$, the vehicle $k$ relocation cost, $\lambda$, the cost incurred when a demand is not served within the prescribed delay, $p_s$, the probability associated with scenario $s \in S$, the set of scenarios, $|S| = s_{max}$, $\alpha$, the percentage of calls that should be reached within the prescribed delay, $P_j$, the maximum number of vehicles that can be located in $j$, $D_{tot}$, the total number of demands and $U(t)$, the set of periods $t' \in T$ where a vehicle is considered to be unavailable after its assignment to a call, and denote $\delta_{kj}$, a parameter equal to 1 if, and only if, the location of vehicle $k$ to $j$ implies a relocation, 0, otherwise, $a_{jts}$, a parameter equal to 1 if, and only if, a demand received in $t$ of scenario $s$ can be reached within the prescribed delay from station $j$, 0, otherwise, $r_{kjt}$, a parameter equal to 1 if, and only if, a vehicle $k$ can reach $j$ before the beginning of period $t$, 0, otherwise, and $\gamma_{ts}$, a parameter equal to 1 if, and only if, a demand is received at the period $t$ of scenario $s$, 0, otherwise. They also use the following variables: $y_{kj}$ is a binary variable equal to 1 if, and only if, vehicle $k$ is located in $j$, and $x_{kts}$, a binary variable equal to 1 if, and only if, a vehicle $k$ is dispatched to a demand placed in $t$ of scenario $s$. The problem is then formulated as follows:

$$\min \sum_{k=1}^{P} \sum_{j=1}^{m} c_k \delta_{kj} y_{kj} + \lambda \sum_{s=1}^{s_{max}} p_s \sum_{t=1}^{T} (1 - \sum_{k=1}^{P} x_{kts}) \tag{70}$$

subject to:

$$x_{kts} - \sum_{j=1}^{m} (r_{kjt} a_{jts} y_{kj}) \leq 0, \ k = 1, ..., P, \ t = 1, ..., T, \ s = 1, ..., s_{max}, \tag{71}$$

$$x_{kts} + \sum_{t' \in U(t)} x_{kt's} \leq 1, \ k = 1, ..., P, \ t = 1, ..., T, \ s = 1, ..., s_{max}, \tag{72}$$

$$\sum_{k=1}^{P} x_{kts} \leq \gamma_{ts}, \ t = 1, ..., T, \ s = 1, ..., s_{max}, \tag{73}$$

$$\sum_{k=1}^{P} y_{kj} \leq P_j, \ j = 1, ..., m, \tag{74}$$

$$\sum_{k=1}^{P} \sum_{t=1}^{T} \sum_{s=1}^{s_{max}} x_{kts} \geq \alpha D_{tot}, \tag{75}$$

$$y_{kj} \in \{0, 1\}, x_{kts} \in \{0, 1\}, \ k = 1, ..., P, \ j = 1, ..., m, \ t = 1, ..., T, \ s = 1, ..., s_{max}. \tag{76}$$

The objective function of the two-stage program consists firstly in minimizing the number of relocated vehicles and secondly, the number of demands that cannot be served within the prescribed delay (70). A vehicle is assumed to be able to adequately serve a demand if, and only if, it is waiting at a standby site located within a given time frame of the demand (71). After being dispatched to a call, a vehicle is assumed to remain unavailable for a given time period (72). In addition, the model ensures that at most one vehicle is dispatched to each demand (73), that the capacity of each site is satisfied (74), and that at least a given percentage of the demands is served within a prescribed time frame (75). The model proposed by Naoum-Sawaya and Elhedhli (2013) has been applied to the case of the EMS operating in Waterloo, Canada. It was solved with CPLEX, considering 50 scenarios and a planning horizon of 2 hours divided into 120 one-minute periods. In this case, the computation time is relatively short, around 40 seconds for the instances considered.

Mason (2013) presented a dynamic ambulance relocation problem that is in many ways similar to the one proposed by Gendreau et al. (2001). This model, called the real-time multi-view generalized-cover repositioning model (RtMvGcRM), is implemented within an EMS management software called Optima Live. This software provides EMS managers with real-time relocation recommendations. The RtMvGcRM aims to determine the location of available vehicles such that the service quality is maximized and relocation costs (i.e. round-trips as well as too frequent and long trips) are minimized. In a similar way as it was done in Gendreau et al. (2001), a penalty term has been integrated into the objective function to minimize the relocations costs. No solution strategy or application context is explicitly given in Mason (2013).

Finally, Jagtenberg et al. (2015) proposed a dynamic version of the MEXCLP (Daskin, 1983) to address a real-time relocation problem with the goal of minimizing the expected fraction of late arrivals. However, such as in Maxwell et al. (2009) and Schmid (2012), they assumed that an vehicle is only allowed to be relocated at the end of a mission. Results obtained for the region of Utrecht, the Netherlands, showed that, despite its simplicity, the proposed relocation policy results in a significant decrease in the fraction of late arrivals, from 9.5% to 7.9%, when compared to the static policy where each vehicle always returns to its home base.

The relocation of vehicles seeks to maintain an adequate service level considering the evolution of the system over time, which was not taken into account when dealing with the static location. Clearly, relocation decisions are dynamic responses to the actual realization of the demand, which could not be tackled by static models. However, relocation generates movements that produce undesirable consequences from both economical and managerial standpoints. To better assess the behaviour of the system in pseudo-realistic contexts, relocation strategies can be evaluated using simulation. In fact, this was done for most of the models presented in this section, as well as in Bélanger et al. (2014), where an empirical comparison of several relocation strategies has been conducted. More recently, Aboueljinane et al. (2014) used simulation in an optimization scheme to determine the best possible multi-period relocation plans in the context of a French department emergency medical service. Following their experiments, many authors concluded that relocation can help maintaining an adequate service level and that the established mechanisms allowed to achieve it with limited efforts. Besides, most authors also pointed out the necessity of developing efficient solution approaches to support real-time decision-making, which was not such an important issue in the static case.

## 4. Dispatching decisions

Dispatching rules are used to determine which vehicle to assign to an emergency call. Clearly, dispatching decisions can have a significant impact on response times, and thus greatly affect the system's performances. They can also impact the system's capacity to adequately serve future demands. Indeed, a degradation of the coverage in a given region can be expected when a vehicle located in that particular region is dispatched to a call. On the other hand, the relocation of vehicles can help reduce such a coverage degradation. Dispatching rules and relocation strategies are thus closely related. For these reasons, we present some of the works related to dispatching decisions we deem to be the most relevant to the purpose of this survey.

When an emergency call is received, a vehicle must be dispatched. Such a decision must be made as quickly as possible to avoid unnecessary delay, and to ensure that the vehicle will arrive at the emergency scene in a timely fashion. In most contexts, the nearest vehicle is assigned to the emergency. Indeed, this strategy is greatly accepted in the practice for several reasons. First of all, it is easy to implement, but more importantly, it ensures a rapid intervention for the most urgent calls. In most EMS, the nearest vehicle strategy is adopted for the most urgent cases, but also for the less prioritized ones. However, although this strategy intends to minimize the response time to reach a call, it does not consider the impact of vehicles' non-availability on the capacity of the system to adequately serve future demands. Based on this observation, some authors proposed that other dispatching strategies should be considered for less prioritized calls (but still needing an immediate service) taking into account the impact on future system's performances while ensuring that each call can be served within a proper time limit.

Within their integrated management tool, Gendreau et al. (2001) proposed to select among all available vehicles that can reach the emergency scene within a prescribed time frame, the one whose dispatch will minimize relocation costs. In this case, they proposed to compute the relocation plan corresponding to each possible dispatching decisions using the time elapsed between the reception of two consecutive calls. Then, when an emergency call is received, the vehicle to assign to the call is selected considering the relocation plan associated with each possible decisions. Andersson and Värbrand (2007) also studied dispatching decisions. They suggested to select among all available vehicles that can reach the call within a prescribed time frame, the one whose dispatch will cause the smallest preparedness degradation. A simple heuristic has been developed to select the best vehicle to assign to a call following this rule. Schmid (2012) was able to show that the *nearest vehicle* strategy may not be the best to adopt when the level of priority allows to deviate from this rule. Indeed, according to the results reported in Schmid (2012) based on the dynamic programming model she proposed, better dispatching rules coupled with good repositioning strategies can lead to the improvement of system performances. Finally, Toro-Diaz et al. (2013) proposed a model that combines location and dispatching decisions. In this case, dispatching decisions are taken according to fixed preference lists. Their results showed that, given the context under study, both the minimization of response time and the maximization of coverage can be achieved using the *nearest vehicle* strategy. Nevertheless, joint location and dispatching decisions can find better solutions when other performance indicators are considered, e.g. solution fairness.

In our opinion, dispatching decisions could also be determined considering other objectives, e.g. to balance workload, or practical issues such as lunch and coffee breaks, or ends of shifts. In this way, when several vehicles are available within the time frame, the vehicle with the smallest workload or the one that does not have any break or end of shift planned in a near future can be selected. It seems that such considerations are taken into account in practice in some sort of way. However, to do so, the information or at least an estimation of information related to the workload, the planned activities and the location of vehicles must be known in real-time.

## 5. Conclusion and perspectives

Ambulance location and relocation decisions can have a significant impact on the service provided to the population. As it was possible to note, different modelling and solution approaches have been proposed to support such decisions. Over the years, location models have evolved to represent the context under study more accurately and, more importantly, the different sources of uncertainty. Approaches from the integration of multiple coverage to the use of stochastic programming have been proposed for this purpose. Since the early 1990s, some researchers have also been interested in the multi-period and dynamic relocation of emergency vehicles, which consider the evolution of the system over time. Different strategies have thus been developed to integrate changes in the system, but also to limit relocation costs. Several dispatching rules have also been proposed in order to better account for the system capacity to serve future demands.

Although ongoing efforts have been deployed over the years to solve both location and relocation problems, different research avenues still need to be explored. Firstly, even if researchers have demonstrated clear interest in capturing the uncertainty and dynamism inherent to these problems, a limited number of contributions has been made to this day, and there is still room left for the development of more sophisticated dynamic and stochastic approaches. This therefore confirms predictions made in Brotcorne et al. (2003) who anticipated a growing interest in dynamic models and predicted the use of stochastic programming with recourse for the development of such models. This leads us to believe that these research avenues are still relevant. Secondly, the growing size of the problems under study, as well as the consideration of stochastic programming, also make us believe that future efforts need to be devoted to the development of more efficient solving methods. Thirdly, the increasing presence of new technologies allows for the use of real-time information about the system that may be more closely considered in the decision-making process, such as workload or the time elapsed since the beginning of an intervention. It is worth noting that, in practice, dispatchers seem to consider such information when dealing with dispatching or relocation decisions, but this is not reflected in any of the previous models. Certainly, this may be due to the difficulty of handling the required data in real-time. However, we believe that it should be considered more formally in decision support tool proposal. Finally, we think that dispatching rules and relocations strategies are closely related. This relationship should be more carefully analyzed, through simulation for instance.

More importantly, from this analysis we observe a growing collaboration between researchers and practitioners. Indeed, many recent studies have been conducted based on real-life cases. This collaboration allows practitioners to benefit from the theoretical and methodological knowledge of researchers on the one hand and, on the other hand, researchers to better understand the context, main concerns, difficulties and limitations from a practical perspective. In our opinion, this new trend, where a closer relation between researchers and practitioners is the core of the solution search process, needs to be continued and reinforced. Despite the challenging aspects, this will ensure that implementable solutions taking into account managerial aspects other than only the response time or coverage measure will be found. In the end, everyone should benefit from such a collaboration, including, of course, the population.

It is clear from this analysis that a lot of work has been done to support the complex decision-making process faced by EMS, mainly from the location and relocation standpoint, but in our opinion, it remains to be done in this register. For this reason, we strongly believe that EMS management still lead to really interesting and relevant research opportunities, as it is the case for most problems related to the health systems management.

## References

Aboueljinane, L., Sahin, E., Jemai, Z., 2013. A review of simulation models applied to emergency medical service operations. Computers & Industrial Engineering 66, 734–750.

Aboueljinane, L., Sahin, E., Jemai, Z., Marty, J., 2014. A simulation study to improve the performance of an emergency medical service: Application to the French Val-de-Marne department. Simulation Modelling Practice and Theory 47, 46–59.

Alsalloum, O. I., Rand, G. K., 2006. Extensions to emergency vehicle location model. Computers & Operations Research 33, 2725–2743.

Andersson, T., Värbrand, P., 2007. Decision support tools for ambulance dispatch and relocation. Journal of the Operational Research Society 58, 195–201.

Aytug, H., Saydam, C., 2002. Solving large-scale maximum expected covering location problem by genetic algorithms : A comparative study. European Journal of Operational Research 141, 480–494.

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, 627–637.

Başar, A., Çatay, B., Ünlüyurt, T., 2012. A taxonomy for emergency service station location problem. Operations Research Letters 6, 1147–1160.

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

Batta, R., Dolan, J. M., Krishnamurty, N. N., 1989. The maximal expected covering location problem : Revisited. Transportation Science 23, 277–287.

Bélanger, V., Kergosien, Y., Ruiz, A., Soriano, P., 2014. Empirical comparison of relocation strategies in real-time ambulance fleet management. Tech. Rep. CIRRELT-2014-73, CIRRELT.

Bélanger, V., Ruiz, A., Soriano, P., 2012. Déploiement et redéploiement des véhicules ambulanciers dans la gestion des services préhospitaliers d'urgence. INFOR 50, 1–30.

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

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

Bianchi, G., Church, R. L., 1988. A hybrid fleet model for emergency medical service system design. Social Sciences & Medicine 26, 163–171.

Borràs, F., Pastor, J. T., 2002. The ex-post evalulation of the minimum location reliability : An enhanced probabilistic location set covering model. Annals of Operations Research 111, 51–74.

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

Church, R. L., ReVelle, C. S., 1974. The maximal covering location problem. Papers of Regional Science Association 32, 101–118.

Daskin, M. S., 1982. Application of an expected covering model to emergency medical service design. Decision Sciences 13, 416–439.

Daskin, M. S., 1983. A maximum expected location problem : Formulation, properties and heuristic solution.

Transportation Science 17, 416–439.

Daskin, M. S., 1987. Location, dispatching, and routing models for emergency services with stochastic travel times. In: Ghosh, A., Rushton, G. (Eds.), Spatial Analysis and Location-Allocation Models. Van Nostrand Reinhold, New York, N.Y., pp. 224–265.

Daskin, M. S., Hogan, K., ReVelle, C. S., 1988. Integration of multiple, excess, backup and expected covering models. Environment and Planning B 15, 15–35.

Daskin, M. S., Stern, E. H., 1981. A hierarchical objective set covering model for emergency medical service vehicle deployment. Transportation Science 15, 137–152.

Doerner, K. F., Gutjahr, W. J., Hartl, R. F., Karall, M., Reimann, M., 2005. Heuristic solution of an extended double-coverage ambulance location problem for austria. Central European Journal of Operations Research 13, 325–340.

Eaton, D. J., Daskin, M. S., Simmons, D., Bulloch, B., Jansma, G., 1985. Determining emergency medical deployment in Austin, Texas. Interfaces 15, 96–108.

Eaton, D. J., Sanchez, H. M. U., Lantigua, R. R., Morgan, J., 1986. Determining ambulance deployment in Santo Domingo, Dominican Republic. Journal of the Operational Research Society 37, 113–126.

Erkut, E., Ingolfsson, A., Sim, T., Erdogan, G., 2009. Computational comparison of five maximal covering models for locating ambulances. Geographical Analysis 41, 43–65.

Fujiwara, O., Makjamroen, T., Gupta, K. K., 1987. Ambulance deployment analysis : A study case of Bangkok. European Journal of Operational Research 31, 9–18.

Galvão, R. D., Chiyoshi, F. Y., Morabito, R., 2005. Towards unified formulations and extensions of two classical probabilistic location model. Computers & Operations Research 32, 15–33.

Galvão, R. D., ReVelle, C. S., 1996. A lagrangean heuristic for the maximal covering location problem. European Journal of Operational Research 88, 114–123.

Gendreau, M., Laporte, G., Semet, F., 1997. Solving an ambulance location model by tabu search. Location Science 5, 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, 1641–1653.

Gendreau, M., Laporte, G., Semet, F., 2006. The maximal expected relocation problem for emergency vehicles. Journal of the Operational Research Society 57, 22–28.

Goldberg, J., 2004. Operations research models for the deployment of emergency services vehicle. EMS Management Journal 1, 20–39.

Goldberg, J., Dietrich, R., Chen, J. M., Mitwasi, M. G., Valenzuela, T., Criss, E., 1990. Validating and applying a model for locating emergency medical services in Tucson, AZ. European Journal of Operational Research 49, 308–324.

Goldberg, J., Paz, L., 1991. Locating emergency vehicle bases when service time depends on call location. Transportation Science 25, 264–280.

Harewood, S. I., 2002. Emergency ambulance deployment in Barbados : a multi-objective approach. Journal of the Operational Research Society 53, 185–192.

Hogan, K., ReVelle, C. S., 1986. Concepts and application of backup coverage. Management Science 34, 1434–1444.

Ingolfsson, A., Budge, S., Erkut, E., 2008. Optimal ambulance location with random delays and travel times. Health Care Management Science 11, 262–274.

Jagtenberg, C. J., Bhulai, S., van der Mei, R. D., 2015. An efficient heuristic for real-time ambulance redeployment. Operations Research for Health Care 4, 27–35.

Laporte, G., Louveaux, F. V., Semet, F., Thirion, A., 2009. Applications of the double standard model for ambulance location. In: Bertazzi, L., Speranza, M. G., van Nunen, J. (Eds.), Innovations in Distribution Logistics. Springer, Berlin, pp. 235–249.

Larson, R. C., 1974. A hypercube queuing model for facility location and redistricting in urban emergency services. Computers & Operations Research 1, 67–85.

Larson, R. C., 1975. Approximating the performance of urban emergency service systems. Operations Research 23, 845–868.

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

Mandell, M. B., 1988. Covering models for two-tiered emergency medical services system. Location Science 6, 355–368.

Marianov, V., ReVelle, C. S., 1994. The queuing probabilistic location set covering problem and some extensions. Socio-Economic Planning Sciences 28, 167–178.

Marianov, V., ReVelle, C. S., 1995. Siting emergency services. In: Drezner, Z. (Ed.), Facility Location. A survey of Applications and Methods. Springer, New York, N.Y., pp. 119–223.

Marianov, V., ReVelle, C. S., 1996. The queuing maximal availability location problem : A model for the siting of emergency vehicles. European Journal of Operational Research 93, 110–120.

Mason, A. J., 2013. Simulation and real-time optimised relocation for improving ambulance operations. In: Denton, B. (Ed.), Handbook of Healthcare Operations: Methods and Applications. Springer, New York, N.Y., pp. 289–317.

Maxwell, M. S., Restepo, M., Henderson, S. G., Topaloglu, H., 2009. Approximate dynamic programming for ambulance redeployment. INFORMS Journal on Computing 22, 266–281.

Moeini, M., Jemai, Z., Sahin, E., September 2013. An integer programming model for the dynamic location and relocation of emergency vehicles: A case study. In: Csendes, T., Stim, L. L., Zerovnik, J. (Eds.), Proceedings of the 12th International Symposium on Operational Research (SOR'2013). Slovenia, pp. 343–350.

Nair, R., Miller-Hooks, E., 2009. Evaluation of relocation strategies for emergency medical service vehicles. Journal of the Transportation Research Board 2137, 63–73.

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

Rajagopalan, H. K., Saydam, C., Xiao, J., 2008. A multiperiod set covering location model for dynamic redeployment of ambulances. Computers & Operations Research 35, 814–826.

Repede, J. F., Bernardo, J. J., 1994. Developping and validating a decision support system for location emergency medical vehicles in Louisville, Kentucky. European Journal of Operational Research 75, 567–581.

ReVelle, C. S., 1989. Review, extension and prediction in emergency services siting models. European Journal of Operational Research 40, 58–69.

ReVelle, C. S., Hogan, K., 1988. A reliability constrained siting model with local estimates of busy fractions. Environment and Planning B 15, 143–152.

ReVelle, C. S., Hogan, K., 1989. The maximum availability location problem. Transportation Science 23, 192–200.

ReVelle, C. S., Marianov, V., 1991. A probabilistic FLEET model with individual reliability requirements. European Journal of Operational Research 53, 93–105.

Saydam, C., Rajagopalan, H. K., Sharer, E., Lawrimore-Belanger, K., 2013. The dynamic redeployment coverage location model. Health Systems 2, 103–119.

Schilling, D. A., Elzinga, D. J., Cohon, J., Church, R. L., ReVelle, C. S., 1979. The TEAM/FLEET models for simultaneous facility and equipment setting. Transportation Science 13, 163–175.

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

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

Shariat-Mohaymany, A., Babaei, M., Moadi, S., Amiripour, S. M., 2012. Linear upper-bound unavailability set covering models for locating ambulances: Application to Tehran rural roads. European Journal of Operational Research 221, 263–272.

Storbeck, J., 1982. Slack, natural slack and location covering. Socio-Economic Planning Sciences 16, 99–105.

Su, Q., Luo, Q., Huang, H., 2015. Cost-effective analyses for emergency medical services deployment: A case study in shanghai. International Journal of Production Economics 163, 112–123.

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

Toro-Diaz, H., Mayorga, M. E., Chanta, S., McLay, L. A., 2013. Joint location and dispatching decisions for emergency medical services. Computers & Industrial Engineering 64, 917–928.

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

Zhang, Z., Jiang, H., 2014. A robust counterpart approach to the bi-objective medical service design problem.

Applied Mathematics Modelling 38, 1033–1040.

Zhen, L., Wang, K., Hu, H., Chang, D., 2014. A simulation optimization framework for ambulance deployment and relocation problems. Computers & Industrial Engineering 72, 12–23.

| | Toregas et al. (1971) | Church and ReVelle (1974) | Schilling et al. (1979) | Daskin and Stern (1981) | Storbeck (1982) | Eaton et al. (1985) | Eaton et al. (1986) | Hogan and ReVelle (1986) | Galvão and ReVelle (1996) | Gendreau et al. (1997) | Doerner et al. (2005) | Su et al. (2015) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| **A. Objectives** | | | | | | | | | | | | |
| A.1 Min. number of vehicles | X | | | X | | | X | | | | | |
| A.2 Max. demand covered once | | X | X | | X | X | | X | X | | | |
| A.3 Max. demand covered more than once $\alpha$ | | | | X | X | | X | | | | | |
| A.4 Max. demand covered twice | | | | | | | | X | | X | X | |
| A.5 Min. delayed services and operational cost | | | | | | | | | | | | X |
| **B. Covering constraints** | | | | | | | | | | | | |
| B.1 Each demand (zone) at least once within $S$ | X | | | X | | | X | X | | | | |
| B.2 Each demand (zone) at least once within $S'$ | | | | | | | | | | X | X | X |
| B.2 $\alpha$ of demand (zone) at least once within $S$ | | | | | | | | | | X | X | X |
| **C. Standby site constraints** | | | | | | | | | | | | |
| C.1 At most one per site | X | X | X | X | X | X | X | X | X | | | |
| C.2 At most $p_j$ per site | | | | | | | | | | X | X | X |
| C.4 Tandem location | | | X | | | | | | | | | |
| C.3 Limited number of sites to use | | | | | | | | | | | | |
| **D. Ambulances** | | | | | | | | | | | | |
| D.1 One type of vehicle | X | X | | X | X | X | X | X | X | X | X | |
| D.2 Two types of vehicles | | | X | | | | | | | | | |
| D.3 Given number of vehicles | | X | X | | X | X | | | X | X | X | |
| D.4 Limited number of demands per vehicle | | | | | | | | | | | X | |
| D.5 Limited workload per vehicle | | | | | | | | | | | | X |
| **E. Solution strategy** | | | | | | | | | | | | |
| F.1 Branch and bound | | X | X | | | | | X | | | | |
| F.2 Branch and cut | X | | | X | | | | | | | | |
| F.2 Greedy heuristic | | X | | | | | | | | | | |
| F.3 Lagrangean heuristic | | | | | | | | | X | | | |
| F.4 Heuristic method | | | | | | | X | | | | | |
| F.5 Tabu search | | | | | | | | | | X | X | |
| F.6 Ant colony algorithm | | | | | | | | | | | X | X |
| F.7 Not presented | | | | | X | | | | | | | |
| **F. Region of interest** | | | | | | | | | | | | |
| G.1 New York City, NY | X | | | | | | | | | | | |
| G.2 Austin, TX | | | | X | | X | | | | | | |
| G.3 Baltimore City, MD | | | X | | | | | | | | | |
| G.4 Dominican Republic | | | | | | | X | | | | | |
| G.5 Montreal, Canada | | | | | | | | | | X | | |
| G.6 Austrian provinces | | | | | | | | | | | X | |
| G.7 Shanghai, China | | | | | | | | | | | | X |

Table A1: Summary of deterministic models (single and multiple coverage)

| | **Daskin (1982, 1983)** | **Fujiwara et al. (1987)** | **Bianchi and Church (1988)** | **Batta et al. (1989)** | **Goldberg et al. (1990)** | **Goldberg and Paz (1991)** | **Mandell (1988)** | **Aytug and Saydam (2002)** | **Galvão et al. (2005)** |
|---|---|---|---|---|---|---|---|---|---|
| **A. Objectives** | | | | | | | | | |
| A.1 Min. number of vehicles | | | | | | | | | |
| A.2 Max. expected covered demand | X | X | | X | X | X | X | X | X |
| A.3 Max. demand covered with reliability $\alpha$ | | | | | | | | | |
| A.4 Min. uncovered demand | | | X | | | | | | |
| A.5 Min. costs | | | | | | | | | |
| **B. Covering constraints** | | | | | | | | | |
| B.1 Each demand (zone) with a reliability $\alpha$ | | | | | | | | | |
| B.2 Each demand (zone) at least once | | | | | | | | | |
| **C. Standby site constraints** | | | | | | | | | |
| C.1 At most one per site | | | | | | | X | | |
| C.2 At most $p_j$ per site | | | | | | | | | |
| C.3 Limited number of sites to use | | | X | | | | | | |
| **D. Ambulances** | | | | | | | | | |
| D.1 One type of vehicle | X | X | X | X | X | X | | X | X |
| D.2 Two types of vehicles | | | | | | | X | | |
| D.3 Given number of vehicles | X | X | X | X | X | X | X | X | X |
| D.4 Lower bound computed for each zone | | | | | | | | | |
| **E. Uncertainty** | | | | | | | | | |
| **E.1 Vehicle availability** | | | | | | | | | |
| *E.1.1 System-wide busy fraction* | X | X | X | | | | | X | |
| *E.1.2 Zone-specific busy fraction* | | | | | | | | | |
| *E.1.3 Considered using queuing theory* | | | | X | X | X | X | | X |
| **E.2 Travel time** | | | | | X | X | | | |
| **E.3 Demand realization** | | | | | | | | | |
| **F. Solution strategy** | | | | | | | | | |
| F.1 Branch and bound | X | | X | | | | | | |
| F.2 Heuristic method | X | | X | X | | | | | |
| F.3 Descent method | | | | X | | | | | |
| F.4 Iterative method | | | | | X | | | | |
| F.5 Exchange-based heuristic | | | | | | X | | | |
| F.6 Genetic algorithm | | | | | | | | X | |
| F.7 Simulated annealing | | | | | | | | | X |
| F.8 CPLEX | | | | | | | X | | |
| **G. Region of interest** | | | | | | | | | |
| G.1 Bangkok, Thailand | | X | | | | | | | |
| G.2 Tucson, AZ | | | | | X | X | | | |

Table A2: Summary of probabilistic models (expected coverage models)

| | ReVelle and Hogan (1988) | ReVelle (1989) | ReVelle and Marianov (1991) | Ball and Lin (1993) | Marianov and ReVelle (1994) | Marianov and ReVelle (1996) | Harewood (2002) | Galvão et al. (2005) | Shariat-Mohaymany et al. (2012) |
|---|---|---|---|---|---|---|---|---|---|
| **A. Objectives** | | | | | | | | | |
| A.1 Min. number of vehicles | X | | | | X | | | | |
| A.2 Max. expected covered demand | | | | | | | | | |
| A.3 Max. demand covered with reliability $\alpha$ | | X | X | | | X | X | X | |
| A.4 Min. uncovered demand | | | | | | | | | |
| A.5 Min. costs | | | | X | | | X | | X |
| **B. Covering constraints** | | | | | | | | | |
| B.1 Each demand (zone) with a reliability $\alpha$ | X | | | | X | | | | X |
| B.2 Each demand (zone) at least once | | | | X | | | | | |
| **C. Standby site constraints** | | | | | | | | | |
| C.1 At most one per site | | X | X | | | | X | | |
| C.2 At most $p_j$ per site | | | | X | | | | | |
| C.3 Limited number of sites to use | | | X | | | | | | |
| **D. Ambulances** | | | | | | | | | |
| D.1 One type of vehicle | X | X | | X | X | X | X | X | X |
| D.2 Two types of vehicles | | | X | | | | | | |
| D.3 Given number of vehicles | | X | X | | | X | X | X | |
| D.4 Lower bound computed for each zone | X | X | X | X | X | X | X | X | |
| **E. Uncertainty** | | | | | | | | | |
| **E.1 Vehicle availability** | | | | | | | | | |
| *E.1.1 System-wide busy fraction* | | X | | | | | | | |
| *E.1.2 Zone-specific busy fraction* | X | X | X | | | | | | |
| *E.1.3 Considered using queuing theory* | | | | | X | X | X | | |
| *E.1.4 Corrective factor in constraints* | | | | | | | | X | |
| *E.1.5 Limit on the resulting busy fraction* | | | | | | | | | X |
| **E.2 Travel time** | | | | | | | | | |
| **E.3 Demand realization** | | | | X | | | | | |
| **F. Solution strategy** | | | | | | | | | |
| F.1 Branch and bound | | X | X | X | X | X | | | |
| F.2 LINDO | | | | | | | X | | |
| F.3 Heuristic method | | X | | | | | | | |
| F.4 Simulated annealing | | | | | | | | X | |
| **G. Region of interest** | | | | | | | | | |
| G.1 Barbados | | | | | | | | X | |
| G.2 Tehran, Iran | | | | | | | | | X |

Table A3: Summary of probabilistic models (Chance-constrained models)

| | Beraldi et al. (2004) | Alsalloum and Rand (2006) | Beraldi and Bruni (2009) | Zhang and Jiang (2014) |
|---|---|---|---|---|
| **A. Objectives** | | | | |
| A.1 Min. number of vehicles | | X | | |
| A.2 Max. expected covered demand | | | | |
| A.3 Max. demand covered with reliability $\alpha$ | | | | |
| A.4 Min. uncovered demand | | X | | X |
| A.5 Min. costs | X | | X | X |
| **B. Covering constraints** | | | | |
| B.1 Each demand (zone) with a reliability $\alpha$ | X | | | |
| B.2 Each demand (zone) at least once | | | X | |
| **C. Standby site constraints** | | | | |
| C.1 At most one per site | | | | |
| C.2 At most $p_j$ per site | X | X | X | X |
| C.3 Limited number of sites to use | | X | | |
| **D. Ambulances** | | | | |
| D.1 One type of vehicle | X | X | X | |
| D.2 Two types of vehicles | | | | |
| D.3 Given number of vehicles | | | | |
| D.4 Lower bound computed for each zone | | X | | |
| **E. Uncertainty** | | | | |
| **E.1 Vehicle availability** | | | | |
| *E.1.1 System-wide busy fraction* | | | | |
| *E.1.2 Zone-specific busy fraction* | | | | |
| *E.1.3 Considered using queuing theory* | | X | | |
| *E.1.4 Corrective factor* | | | | |
| **E.2 Travel time** | | | | |
| **E.3 Demand realization** | X | | X | X |
| **F. Solution strategy** | | | | |
| F.1 CPLEX | X | | | X |
| F.2 Exact method | | | X | |
| F.3 Heuristic method | | | X | |
| F.4 Not presented | | X | | |
| **G. Region of interest** | | | | |
| G.1 Riyadh, Saudi Arabia | | X | | |
| G.2 A city in China | | X | | X |

Table A4: Summary of new approaches

| | Repede and Bernardo (1994) | Rajagopalan et al. (2008) | Schmid and Doerner (2010) | Başar et al. (2011) | Saydam et al. (2013) | van den Berg and Aardal (2015) |
|---|---|---|---|---|---|---|
| **A. Objectives** | | | | | | |
| A.1 Min. number of sites/vehicles | | X | | | X | |
| A.2 Max. expected covered demand | X | | | | | X |
| A.3 Max. demand covered twice | | | X | X | | |
| A.4 Min. number of relocated vehicles | | | X | | X | |
| A.5 Min. relocation costs | | | | | | X |
| **B. Covering constraints** | | | | | | |
| B.1 Each demand (zone) with a reliability $\alpha$ | | | | | | |
| B.2 Each demand (zone) at least once | | X | | | X | |
| B.3 $\alpha$ of demand (zone) at least once | | | | | | |
| **C. Standby site constraints** | | | | | | |
| C.1 At most one per site | | | | X | | |
| C.2 At most $p_j$ per site | | | X | | | |
| C.3 Limited number of sites to use | | | | X | | |
| **D. Ambulances** | | | | | | |
| D.1 One type of vehicle | X | X | X | X | X | X |
| D.2 Given number of vehicles | X | | X | X | | X |
| D.3 Limited number of demands per vehicle | | | X | | | |
| D.4 Lower bound computed for each zone | | X | | | X | |
| **E. Time-dependent** | | | | | | |
| E.1 Demand | X | | | | | X |
| E.2 Travel time | X | X | X | | X | X |
| E.3 Number of vehicles/sites | X | | | X | | X |
| E.4 Busy fraction | X | X | | | X | X |
| **F. Solution strategy** | | | | | | |
| F.1 Tabu Search | | X | | X | X | |
| F.2 Variable neighborhood search | | | X | | | |
| F.3 CPLEX | | | | | | X |
| F.4 Not presented | X | | | | | |
| **G. Region of interest** | | | | | | |
| G.1 Louisville, KY | X | | | | | |
| G.2 Mecklenburg County, NC | | X | | | X | |
| G.3 Istanbul, Turkey | | | | X | | |
| G.4 Vienna, Austria | | | X | | | |
| G.5 Amsterdam, The Netherlands | | | | | | X |

Table A5: Summary of multi-period models

| | Gendreau et al. (2001) | Gendreau et al. (2006) | Andersson and Värbrand (2007) | Nair and Miller-Hooks (2009) | Maxwell et al. (2009) | Schmid (2012) | Naoum-Sawaya and Elhedhli (2013) | Moeini et al. (2013) | Mason (2013) | Jagtenberg et al. (2015) |
|---|---|---|---|---|---|---|---|---|---|---|
| **A. Objectives** | | | | | | | | | | |
| A.1 Max. expected covered demand | | X | | | | | | | | |
| A.2 Max. demand covered twice | X | | | X | | | | X | X | |
| A.3 Min. number of relocated vehicles | | | | | | | X | | | |
| A.4 Min. relocation costs | X | | | X | | | | X | X | |
| A.5 Min. travel time to perform relocation | | | X | | | | | | | |
| A.6 Min. number of calls that cannot be reached in time | | | | | X | | X | X | | |
| A.7 Min. average response time | | | | | | X | | | | |
| **B. Covering constraints** | | | | | | | | | | |
| B.1 Each demand (zone) at least once in $S$ or $S'$ | X | | | | | | | X | | |
| B.2 $\alpha$ of demand (zone) at least once in $S$ | X | | | | | | X | X | | |
| **C. Standby site constraints** | | | | | | | | | | |
| C.1 At most one per site | | X | | | | | X | | | |
| C.2 At most $p_j$ per site | X | | | X | | | X | X | | |
| **D. Ambulances** | | | | | | | | | | |
| D.1 One type of vehicle | X | X | X | X | X | X | X | X | X | |
| D.2 Given number of vehicles | X | X | X | X | X | X | X | X | X | |
| **E. Relocation constraints** | | | | | | | | | | |
| E.1 At most one vehicle relocated | | | | | X | X | | | | |
| E.2 At most $r_s$ vehicles relocated | | X | | X | | | | | | |
| E.3 To reach a given service level | | | X | | | | | | | |
| **F. State-dependent** | | | | | | | | | | |
| F.1 Demand or call rate | | | | X | X | X | | X | X | |
| F.2 Travel time | | | | X | | X | X | | | |
| F.3 Number of vehicles/sites | X | X | X | X | X | X | X | | | |
| F.4 Relocation costs/impact | X | | | | | | | | | |
| **G. Solution strategy** | | | | | | | | | | |
| F.1 Tree-search heuristic | | | X | | | | | | | |
| F.2 Tabu Search | X | | | | | | | | | |
| F.3 Approximate dynamic programming | | | | | X | X | | | | |
| F.4 Monte Carlo simulation | | | | | X | | | | | |
| F.5 CPLEX | | X | | | | | X | X | | |
| F.6 Not presented | | | | X | | | | | X | |
| **H. Region of interest** | | | | | | | | | | |
| G.1 Montreal, Canada | X | X | | X | | | | | | |
| G.2 Stockholm, Sweden | | | X | | | | | | | |
| G.3 Edmonton, Canada | | | | | X | | | | | |
| G.4 Vienna, Austria | | | | | | X | | | | |
| G.5 Waterloo, Canada | | | | | | | X | | | |
| G.6 Val-de-Marne county, France | | | | | | | | X | | |
| G.7 Utrecht region, The Netherlands | | | | | | | | | | X |

Table A6: Summary of dynamic models