Math Meth Oper Res (2011) 74:281–310
DOI 10.1007/s00186-011-0363-4

ORIGINAL ARTICLE

Covering models and optimization techniques for emergency response facility location and planning: a review

Xueping Li · Zhaoxia Zhao · Xiaoyan Zhu · Tami Wyatt

Received: 3 September 2010 / Accepted: 4 June 2011 / Published online: 22 July 2011

Abstract With emergencies being, unfortunately, part of our lives, it is crucial to efficiently plan and allocate emergency response facilities that deliver effective and timely relief to people most in need. Emergency Medical Services (EMS) allocation problems deal with locating EMS facilities among potential sites to provide efficient and effective services over a wide area with spatially distributed demands. It is often problematic due to the intrinsic complexity of these problems. This paper reviews covering models and optimization techniques for emergency response facility location and planning in the literature from the past few decades, while emphasizing recent developments. We introduce several typical covering models and their extensions ordered from simple to complex, including Location Set Covering Problem (LSCP), Maximal Covering Location Problem (MCLP), Double Standard Model (DSM), Maximum Expected Covering Location Problem (MEXCLP), and Maximum Availability Location Problem (MALP) models. In addition, recent developments on hypercube queuing models, dynamic allocation models, gradual covering models, and cooperative covering models are also presented in this paper. The corresponding optimization

X. Li (✉) · Z. Zhao · X. Zhu
Department of Industrial and Information Engineering, University of Tennessee,
416 East Stadium Hall, Knoxville, TN 37919, USA
e-mail: Xueping.Li@utk.edu

Z. Zhao
e-mail: zzhao8@utk.edu

X. Zhu
e-mail: xzhu5@utk.edu

T. Wyatt
College of Nursing, University of Tennessee, 200 Volunteer Boulevard,
Knoxville, TN 37996-4180, USA
e-mail: twaytt@utk.edu

techniques to solve these models, including heuristic algorithms, simulation, and exact methods, are summarized.

Keywords Emergency facility location · Modeling and optimization · Mathematical modeling · Covering model · Genetic algorithm · Tabu search · Simulation

1 Introduction

In the past few decades, the Emergency Medical Services (EMS) systems have drawn a great deal of attention from researchers. In the EMS systems, response time is a critical factor when making decisions on the system configurations that affect life or death care. If the emergency response system cannot provide service promptly, people's lives are jeopardized. While the public expects availability of EMS facilities to provide timely services, this expectation is hard to realize due to limited available resources and other factors such as stringent governmental budget. Therefore, efficiently locating available emergency response facilities becomes an important issue.

Most recently, researchers proposed more realistic models for locating and planning emergency response facilities, including hypercube queuing models, dynamic models, gradual covering models, and cooperative covering models. However, little work has been done to summarize the optimization techniques to solve these models. This motivates us to review covering models, with emphasis on recent developments, and optimization techniques for emergency facility location and planning from the mathematical methods and operations research perspective.

Traditionally, emergency facility location problems deal with decisions from two aspects: which sites should be selected as depots for facilities and how many facilities should be placed in each depot, given demand points and potential facility sites. Plenty of models have been developed to solve facility location problems. Most of these models simplify the facility location problems by treating emergency calls generated from discrete demand points. These models can be divided into three broad groups: (1) covering models, which emphasize providing coverage for emergency calls within a predefined distance standard; (2) *p*-median models, which minimize the total or average service distance for all demand points; and (3) *p*-center models, which aim to minimize the maximum service distance for all demand points. Covering models are concern with covering demands, and in most covering models, demand is said to be covered when it can be reached within a predefined distance standard by at least one facility. However, *p*-center and *p*-median models place stress on the distance between demand points and their nearest facilities. Among the three groups, the covering models are most prevalent and thus are intensively reviewed in this paper.

According to the literature, the first emergency facility location covering model was the Location Set Covering Problem (LSCP), proposed by Toregas et al. (1971). The LSCP is a mandatory covering model and its objective is to find the minimum number of facilities to cover all demand points. However, full coverage is hard to achieve in reality due to the limited resources. For a demand point far away from the others, it probably cannot be covered within the predefined distance standard. A few years later, the first maximum deterministic covering model was proposed by

Church and ReVelle (1974), named the Maximal Covering Location Problem (MCLP). Given a limited number of facilities, this model aims to maximize the demand coverage. The LSCP and MCLP have a common shortcoming; once a facility is called for service, demand points under its coverage are not covered by it any more. In the literature, there are two strands of research on overcoming this drawback. One is to provide multiple coverage, such as the Double Standard Model (DSM), proposed by Gendreau et al. (1997). The DSM aims to allocate facilities among potential sites to provide full coverage within a longer distance standard while maximizing coverage within a shorter distance standard. The other strand is to explicitly consider the busy probabilities and reliabilities of facilities as represented by the Maximum Expected Covering Location Problem (MEXCLP) and the Maximum Availability Location Problem (MALP), proposed by Daskin (1983) and ReVelle and Hogan (1989), respectively.

The strong assumptions in the MEXCLP, such as independent facilities and same busy probability for all facilities, are relaxed in hypercube queuing models. Hypercube queuing models provide more accurate presentation of real systems. Along with the development of information technologies, the dynamic allocation models are proposed to solve real time facility location and allocation problems as represented by the Dynamic Double Standard Model ($DDSM^t$) and the Dynamically Available Coverage Location (DACL) model, proposed by Gendreau et al. (2001) and Rajagopalan et al. (2008), respectively.

Recently, researchers proposed gradual covering models to relax the assumption that a demand point is covered when it can be reached within a predefined distance standard by at least one facility. In the literature, various mathematical functions are proposed to model the gradual decline of coverage along with the increase of the distance. In some of the covering models, only one facility, namely the nearest one, determines whether a demand point is covered. The cooperative coverage models, which study the cooperative behavior of different facilities, are proposed to relax this "individual coverage" assumption.

Various optimization techniques were developed to solve the proposed models. This paper reviews three main categories of these techniques: heuristic algorithms, simulation, and exact methods. The common heuristic algorithms used are Genetic Algorithm (GA), Tabu Search (TS), Lagrangian Relaxation (LR), Simulated Annealing (SA), Ant Colony Optimization (ACO), and Local Search heuristics. Simulation is usually used to evaluate the performance of a system or combine it with heuristic algorithms to provide near optimal solutions. Ambulance location models are mostly formulated as integer programming problems, and Branch and Bound (B&B) algorithms can be applied to obtain optimal solutions.

This paper concentrates on the mathematical models and their recent extensions for EMS facility location and planning, as well as main optimization techniques applied to these models. For past surveys on facility location problems, one may refer to the following literatures: Daskin et al. (1988), ReVelle (1989), ReVelle (1991), Schilling et al. (1993), Daskin (1995), Marianov and ReVelle (1995), Owen and Daskin (1998), Marianov and Serra (2002), Berman and Krass (2002), Brotcorne et al. (2003), Green and Kolesar (2004), Goldberg (2004) and Cordeau et al. (2007).

The rest of this paper is organized as follows: covering models are introduced in detail in Sects. 2 and 3 reviews the optimization techniques used to solve these models, and Sect. 4 concludes the review with a discussion of directions for future research.

2 Covering models for emergency facility location and planning

In most of the covering models, a demand is covered if at least one EMS facility can serve the emergency call within a predefined distance standard. The standard was specified in the EMS Act of 1973, which required that in urban areas 95% of requests should be reached within 10 min and in rural areas, calls should be reached within 30 min or less. Because of this Act, covering models are widely used and have been studied for several decades. Most of these models are well known. We use the following notations to describe the models.

- V the set of demand points;
- i the index for demand points;
- W the set of potential facility sites;
- j the index for the potential facility sites;
- t_{ij} the distance from demand point i to the facility at site j;
- r the distance threshold for a demand point to be considered as being covered;
- W_i the set of the facility sites covering demand point i, i.e. $\{j \in W_i | t_{ij} \leq r\}$
- d_i the population size of demand point i;
- p the total number of available facilities;
- y_i binary variable, equal to 1 if and only if demand point i is covered at least once;
- x_j binary variable, equal to 1 if and only if a facility is located at site j;

2.1 Location set covering problem

The Location set covering problem (LSCP) is probably the first emergency facility location covering model, proposed by Toregas et al. (1971). This mandatory covering model can be formulated as follows:

LSCP:

$$\min \sum_{j \in W} x_j \tag{1}$$

$$\text{subject to} \sum_{j \in W_i} x_j \geq 1, \quad i \in V \tag{2}$$

$$x_j \in \{0, 1\}, \quad j \in W. \tag{3}$$

In the formulation above, the objective function (1) minimizes the total number of facilities required. Constraint (2) specifies that all the demand points must be covered by at least one facility. Figure 1 illustrates this model with a feasible solution. There are four potential facility sites and twelve demand points. A demand point is covered

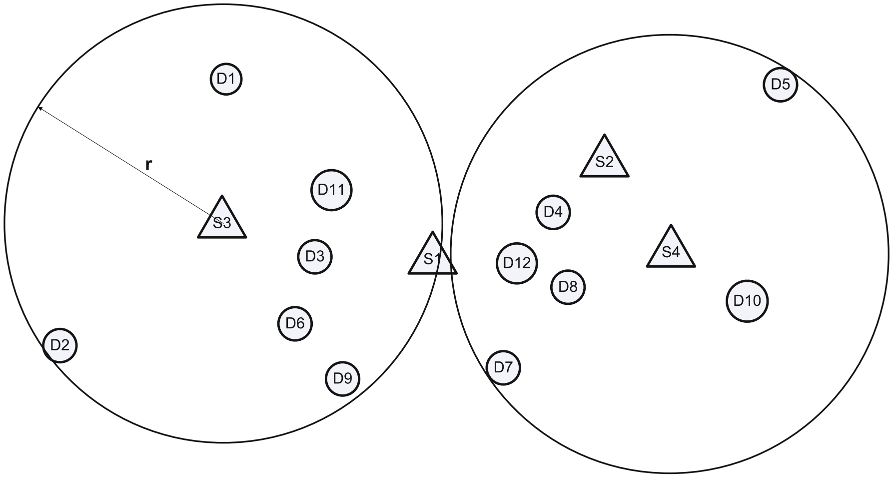

Fig. 1 An illustration of the LSCP. D_1 to D_{12} are demand points while S_1 to S_4 are EMS facilities. A demand point is covered by a facility if the distance between the demand point and the facility is within r. A feasible solution is given in which S_3, S_4 cover all the demand points

by a facility as long as that facility can reach the demand point within the distance standard r. In this illustration, locating facilities at site S_3, S_4 can cover all the demand points, hence it is a feasible solution.

The LSCP simplifies the real world EMS facility location problem by treating the system as static and deterministic. Specifically, the resource is considered unlimited and it is assumed that a facility can serve all EMS requests from its assigned demand points. However, it is useful on the strategic level to determine the minimum number of facilities needed to provide full coverage. A variety of models are proposed to relax one or some of the strong assumptions in the LSCP. Aly and White (1978) studied the problem with the assumption that emergency calls are generated in a continuous region instead of discrete points. They formulated a model to account for stochastic response time. Daskin and Stern (1981) proposed a hierarchical version of the LSCP (HOSC), with the objective of minimizing the number of facilities providing full coverage within a distance standard first and then maximizing the number of demand points with multiple coverage. ReVelle and Hogan (1989) formulated a probabilistic version of the LSCP which required all the demand points to be covered with α reliability level. Ball and Lin (1993) established a new version of probabilistic LSCP. In their model, the uncovered probability of each demand point must be below a preset value. Queuing Probabilistic Location Set Covering Problem (QPLSCP) formulated by Marianov and ReVelle (1994) relaxed the assumption that servers were operated independently. Shiah and Chen (2007) introduced the Ambulance Allocation Capacity Model (AACM), which integrated the concept of ambulance service capacity into the LSCP and considered the road condition and the population distribution.

2.2 Maximal covering location problem

The maximal covering location model was introduced by Church and ReVelle (1974), and is known as Maximal covering location problem (MCLP). This model is presented as follows:

MCLP:

$$\max \sum_{i \in V} d_i y_i \tag{4}$$

$$\text{subject to} \sum_{j \in W_i} x_j \geq y_i, \quad i \in V \tag{5}$$

$$\sum_{j \in W} x_j = p \tag{6}$$

$$x_j, \ y_i \in \{0, 1\}, \quad j \in W, \ i \in V. \tag{7}$$

The objective (4) is to maximize the demand coverage. Constraint (5) guarantees that demand point i is covered only if one or more facilities are placed within the distance standard, and constraint (6) specifies that the total number of available facilities equals p. This model considers the demand size and uses it as the weight of each demand point in the objective function, which makes the model more realistic. The MCLP aims to make the best possible use of available limited resources (Brotcorne et al. 2003).

Eaton et al. (1981, 1985, 1986) successfully applied MCLP to solve the practical EMS vehicle location problems in Colombia, United States, and Dominican Republic. Chung (1986) examined the applications of MCLP on other subjects, including data abstraction, statistical classification, cognitive process modeling, etc.

There are various extensions of the MCLP. Dessouky (2006), Jia et al. (2007a) and Jia et al. (2007b) studied multiple quality levels and multiple quantities of the facilities at each quality level for demand points in large scale EMS systems. Their model can be formulated as:

$$\max \sum_{k} \sum_{i \in V} c^k d_i y_i^k \tag{8}$$

$$\text{subject to} \sum_{j \in W} x_j \leq p \tag{9}$$

$$\sum_{j \in W_i^k} x_j \geq Q_i^k y_i^k \quad i \in V, \quad k = 1, \ldots, q \tag{10}$$

$$x_j, \quad y_i^k \in \{0, 1\} \quad j \in W, \quad i \in V, \quad k = 1, \ldots, q, \tag{11}$$

where c^k is the importance weighting factor of demand points having the quality level k. y_i^k is a binary variable, equal to 1 if and only if demand point i is covered at quality level k. W_i^k represents the set of the facility sites that can cover demand point i at quality level k and Q_i^k denotes the minimum number of facilities that must be allocated to demand point i to achieve k quality level coverage. Jia et al. (2007b) suggested the number of quality levels and Q_i^k should be determined by population, the weighting factor, and the emergency occurrence likelihood at each demand point, due to the complexity of emergency incidents. The objective of this model is to maximize demands covered at different quality levels. Constraint (9) specifies the total number of facilities is less than p. Constraint (10) enforces y_i^k to be 0 if there are less than Q_i^k facilities that can cover demand point i at quality level k.

Other extensions of the MCLP found in the literature are summarized as follows. Several models studied systems in which service is provided by two distinct types of servers, such as Tandem Equipment Allocation Model (TEAM) (Schilling et al. 1979) and Backup Double Covering Model (BDCM) (Basar et al. 2008). The models proposed by Hogan and ReVelle (1986) are to maximize the population coverage with more than two facilities while forcing all demand points to be covered once, known as Backup Coverage Problem (*BACOP1* and *BACOP2*). Schilling (1980) developed a dynamic version of MCLP in multi-period context and used a multi-objective approach to achieve near optimal solution. Alsalloum and Rand (2003) and Alsalloum and Rand (2006) extended MCLP and developed Goal Programming models. They first determined locations of facilities to maximize expected demand coverage and then adjusted the capacity of each station while meeting the minimum performance requirements. Marianov and Serra (1998) introduced a queuing version of MCLP named the Maximal Covering Location-Allocation Problem (MCLAP) with constraints on waiting time in queue. Erkut et al. (2007) incorporated a survival function into the covering model and formulated the Maximum Survival Location Problem (MSLP). The survival function is a monotonic decreasing function, mapping response time to survival rate. They tested their model by out-of-hospital cardiac arrest emergency.

The LSCP and MCLP have a common drawback. When a facility is called for emergency services for one demand point, other demand points within its coverage area will not be covered. The literature devoted to this problem can be divided into two strands (Daskin et al. 1988): one is to provide additional coverage, such as HOSC, BACOP1, BACOP2, and DSM (which is to be itroduced next), and the other strand is to explicitly take into account the busy probabilities of facilities, such as QPLSCP, MCLAP, MEXCLP and MALP. The last two models will be introduced in Sects. 2.4 and 2.6.

2.3 Double standard model (DSM)

The double standard model (DSM), proposed by Gendreau et al. (1997), uses two distance standards r_1 and r_2, $(r_1 < r_2)$.

DSM:

$$\max \sum_{i \in V} d_i y_{i2} \tag{12}$$

$$\text{subject to} \quad \sum_{j \in W_{i2}} x_j \geq 1, \quad i \in V \tag{13}$$

$$\sum_{i \in V} d_i y_{i1} \geq \alpha \sum_{i \in V} d_i \tag{14}$$

$$y_{i2} \leq y_{i1}, \quad i \in V \tag{15}$$

$$\sum_{j \in W_{i1}} x_j \geq y_{i1} + y_{i2}, \quad i \in V \tag{16}$$

$$\sum_{j \in W} x_j = p \tag{17}$$

$$x_j \leq p_j, \quad j \in W \tag{18}$$

$$y_{i1}, \ y_{i2} \in \{0, 1\}, \quad i \in V \tag{19}$$

$$x_j \quad \text{integer}, \quad j \in W, \tag{20}$$

where y_{i1} and y_{i2} are binary variables, equal to 1 if and only if demand point i is covered at least once and twice within r_1, respectively; W_{i1} and W_{i2} represent the sets of facility sites that can cover demand point i within r_1 and r_2, respectively. The objective of the DSM is to maximize the demands that are covered at least twice within r_1. Constraint (13) and (14) express the coverage requirements that all demands must be covered within r_2 and a proportion α of the total demands that must be covered within r_1. Constraint (15) imposes that a demand point cannot be covered twice if it is not covered at least once. Constraint (15) and (16) together ensure the demand point i is covered twice only if there are two or more facilities within r_1. Constraint (17) and (18) limit the number of facilities at each facility site. DSM considers the demand size at each demand point and relaxes the assumption that only one facility can be sited at each facility site. Figure 2 illustrates the DSM and a feasible solution of this model. The positions of demand points and potential sites are the same as those in Fig. 1. To maximize demands covered twice within r_1, an additional EMS facility is placed at site S_1.

Doerner et al. (2005) and Doerner and Hartl (2008) developed their models based on the DSM, augmenting penalty terms to the objective function to avoid unmet coverage requirements and uneven workload. Specifically, for each demand point i, the workload w_i per EMS facility assigned within r_2 is computed as $w_i = \frac{d_i}{\sum_{j \in W_{i2}} x_j}$. If the ratio exceeds a given standard w_0, a penalty term $M(w_i - w_0)$ is subtracted from the objective function. The model can be formulated as follows:

$$\max \sum_{i \in V} d_i y_{i2} - M_1 f_1 - M_2 f_2 - M_3 f_3 \tag{21}$$

subject to (15)–(20),

$$f_1 = \left| \left\{ i \in V : \sum_{j \in W_{i2}} x_j = 0 \right\} \right| \tag{22}$$

$$f_2 = \alpha - min \left(\alpha, \frac{\sum_{i \in V} d_i y_{i1}}{\sum_{i \in V} d_i} \right) \tag{23}$$

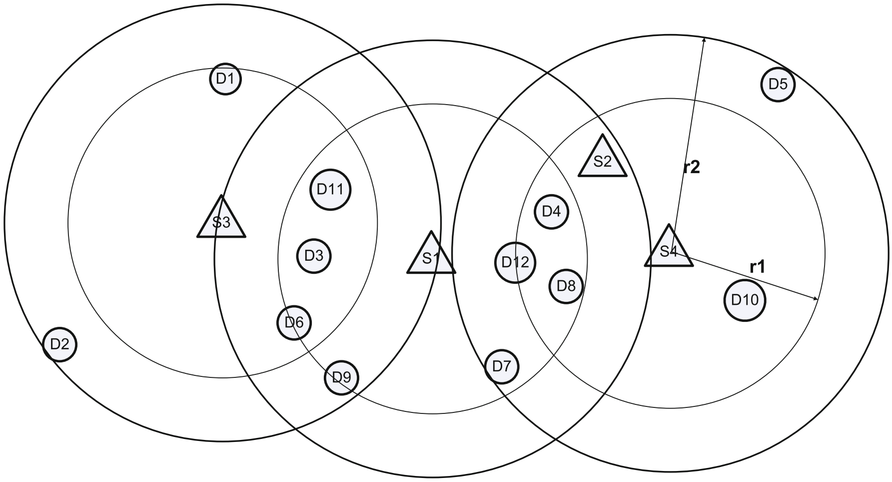

Fig. 2 The illustration of the Double Coverage Facility Location Problem. The positions of demand points and potential sites are the same as those in Fig. 1. To maximize demands covered twice within r_1, additional EMS facility is placed at site S_1

$$f_3 = \sum_{i \in V} (w_i - w_0)^+ \tag{24}$$

f_1 is a modified Eq. 13 to compute the number of uncovered demand points within the large distance standard r_2. f_2 is a modified Eq. 14 to compute the differences between the actual coverage within r_1 and the predetermined proportion α. f_3 is the penalty term for facilities workload exceeding a given standard. By assigning different values of M_1, M_2 and M_3, decision makers can determine the relative importance of these soft constraints.

2.4 Maximum expected covering location problem (MEXCLP)

In the MEXCLP introduced by Daskin (1983), all facilities are assumed to have the same busy probability and operate independently. The model can be formulated as follows:

MEXCLP:

$$\max \sum_{i \in V} \sum_{k=1}^{p} d_i (1-q) q^{k-1} y_{ik} \tag{25}$$

$$\text{subject to} \quad \sum_{j \in W_i} x_j \geq \sum_{k=1}^{p} y_{ik}, \quad i \in V \tag{26}$$

$$\sum_{j \in W} x_j \leq p \tag{27}$$

$$y_{ik} \in \{0, 1\}, \quad i \in V, k = 1, \ldots, p \tag{28}$$

$$x_j \quad \text{integer}, \quad j \in W, \tag{29}$$

where y_{ik} is a binary variable, equal to 1 if and only if demand point i is covered by at least k facilities. The objective of this model is to maximize the expected coverage given a limited number of facilities. The left-hand side of constratint (26) represents the total number of facilities covering demand point i within r, while the right-hand side is the number of times that demand point i is covered. Since the objective is for maximization, constraint (26) and (27) will be satisfied as equalities (Brotcorne et al. 2003). As with the DSM, the MEXCLP also allows for more than one facility at one site. The MEXCLP has two strong assumptions: independent facilities and the same busy probabilities of facilities. However, in reality, demands are not evenly distributed temporally and spatially, and thus the busy probability varies from facility to facility. In the following years, many researchers attempted to relax these assumptions and built more practical models.

Repede and Bernardo (1994) made some extensions to the MEXCLP to incorporate temporal varying demands into the model named the TIMEXCLP. Their model, together with a simulation model, were used to evaluate alternative plans for ambulance deployment in Louisville, Kentucky, and yielded increased coverage from 84% to 95% and 36% decrease in response time compared to the previous system. The TIMEXCLP can be formulated as follows:

TIMEXCLP:

$$\max \quad \sum_{t}^{T} \sum_{i \in V} \sum_{k=1}^{p_t} (d_{i,t})(1 - q_t) \left(q_t^{k-1}\right) (y_{ik,t}) \tag{30}$$

$$\text{subject to} \quad \sum_{j \in W_{i,t}} x_{j,t} \geq \sum_{k=1}^{p_t} y_{ik,t}, \quad i \in V, \; t \leq T \tag{31}$$

$$\sum_{j \in W} x_{j,t} \leq p_t \quad t \leq T \tag{32}$$

$$y_{ik,t} \in \{0, 1\}, \quad i \in V, \; k = 1, \ldots, p, \; t \leq T \tag{33}$$

$$x_{j,t} \quad \text{integer}, \quad j \in W, \; t \leq T, \tag{34}$$

where p_t is the number of available ambulances at time period t. q_t is the system-wide busy fraction at time period t for a fleet size of p_t. $d_{i,t}$ is the demand size generated at demand point i during time period t. $y_{ik,t}$ is a decision variable, equal to 1 if and only if demand point i is covered by at least k ambulances during time period t covers demand point i. $x_{j,t}$ is the number of ambulance located at site j during the time period t. $W_{i,t}$ is the set of facilities that cover demand point i during the time period t within r. Equatios (30)–(34) have similar meaning with Eqs. 25–29, respectively.

There are other extensions of the MEXCLP found in the literature. Saydam and McKnew (1985) used a separable programming approach to reformulate the MEXCLP into a nonlinear form. Fujiwara et al. (1987, 1988) applied the MEXCLP to locate EMS in Bangkok and the near optimal solutions obtained from the model were further analyzed by simulation. Rajagopalan et al. (2007) developed several heuristic algorithms, including GA, TS, SA and hybridized hill climbing, to optimize the MEXCLP. The performance of these algorithms was compared and analyzed using ANOVA. Sorensen and Church (2010) formulated Local Reliability based MEXCLP (LR-MEXCLP) with the same objective as the MEXCLP, while incorporating local reliability estimation.

The assumptions in the MEXCLP make models easier to build and solve. However, they lack an accurate estimation of the expected coverage. Some researchers resorted to hypercube queuing models to obtain better estimation of the expected coverage.

2.5 Hypercube queuing model

The hypercube queuing model is designed for analyzing the behaviors of a multi-server queuing system with distinguishable servers. The first hypercube queuing model was introduced by Larson (1974). In this model, the region under investigation is assumed to be partitioned into several cells or geographical atoms with a certain fraction of region-wide workload. In addition, it is assumed the nearest available facility is selected to dispatch when demands arise. The state of the system is described by facilities status idle (0) or busy (1). By building equilibrium equations of the steady states, the probability of each state can be calculated. Thus, performance measures of the system, such as the loss probability, fraction of dispatches and mean system travel time, can be easily computed. This decision process is time consuming, especially when the problem size is large. Larson (1975) proposed an approximation procedure to calculate performance measures. They claimed that the hypercube model is a special $M/M/p$ queuing system with a more finely structured state space. The correction factor proposed in this paper has been widely used. The correction factor is developed in the $M/M/p/\infty$ system. Let $q = \lambda/(p\mu) < 1$, and P_j denote the steady-state probability that exactly j severs are busy:

$$
\begin{aligned}
P_j &= p^j q^j / j! P_0, \quad j = 1, 2, \ldots, p-1, \\
P_p &= p^p q^p / p!(1-q) P_0, \\
P_0 &= 1 \Bigg/ \left[\sum_{i=0}^{p-1} p^i q^i / i! + p^p q^p / p!(1-q) \right].
\end{aligned}
$$

They randomly selected the facilities without replacement, and let S_j be the event that exact j servers are busy, B_k denote the event that the kth selected facility is busy, and F_k denote the event that the kth selected facility is free. The probability that the $(k+1)$th selected server is the first available one is:

$$
\begin{aligned}
P\{B_1 B_2 \ldots B_k F_{k+1}\} &= \sum_{j=0}^{j=p} P\left\{B_1 B_2 \ldots B_k F_{k+1} | S_j\right\} P_j \\
&= \sum_{j=0}^{j=p} P\left\{F_{k+1} | B_1 B_2 \ldots B_k S_j\right\} P\left\{B_k | B_1 B_2 \ldots B_{k-1} S_j\right\}
\end{aligned}
$$

$$
\begin{aligned}
&\dots P\left\{B_1|S_j\right\} P_j \\
&= \sum_{j=0}^{j=p} \frac{p-j}{p-k} \frac{j-(k-1)}{p-(k-1)} \dots \frac{j-1}{p-1} \frac{j}{p} P_j \\
&= \sum_{j=k}^{j=p-1} \frac{p-j}{p-k} \frac{j-(k-1)}{p-(k-1)} \dots \frac{j-1}{p-1} \frac{j}{p} \frac{p^j q^j}{j!} P_0 \\
&= \left[\sum_{j=k}^{j=p-1} \frac{(p-k-1)!(p-j)}{(j-k)!} \frac{p^j}{p!} q^{j-k} \right] \frac{P_0}{1-q} q^k (1-q) \\
&= Q(p,q,k) q^k (1-q)
\end{aligned}
$$

where

$$
Q(p,q,k) = \frac{\sum_{j=k}^{p-1} \{(p-k-1)!(p-j)/(j-k)!\} (p^j/p!) q^{j-k}}{(1-q)\left(\sum_{i=0}^{p-1} (p^i/i!) q^i\right) + p^p q^p / p!}, \quad k = 0, 1, \dots, p-1 \tag{35}
$$

The busy probabilities of these facilities are the same, equal to q. Assuming the facilities are independent, then the probability that the first available facility is the $(k+1)$th is $q^k(1-q)$. Therefore, $Q(p,q,k)$ can be interpreted as a factor which corrects the independent argument to obtain the exact result.

After the MEXCLP was proposed, researchers endeavored to improve the model in different directions. One of the directions is to incorporate hypercube queuing theory since it is suitable for describing multi-server queuing systems and providing abundant statistics data about the servers and demand points. Batta et al. (1989) proposed the Adjusted MEXCLP (AMEXCLP), which embedded the hypercube queuing theory into the MEXCLP and relaxed the assumption of independent busy probability by using the correction factor derived in Larson (1975). The objective function of the AMEXCLP can be written as follows:

$$
\max \sum_{i \in V} \sum_{k=1}^{p} Q(p,q,k-1) d_i (1-q) q^{k-1} y_{ik}, \tag{36}
$$

note that when the correction factors are equal to 1, the AMEXCLP reduces to the MEXCLP.

Saydam and Aytug (2003) changed the objective function of the MEXCLP into a nonlinear form which can be formulated as follows:

$$
\max \quad \sum_{i \in V} d_i (1-q)^{y_i} \tag{37}
$$

$$
\text{subject to} \quad \sum_{j \in W_i} x_j = y_i, \quad i \in V \tag{38}
$$

$$\sum_{j \in W} x_j \leq p \tag{39}$$

$$x_j, \ y_i \quad \text{integer}, \quad j \in W, \ i \in V, \tag{40}$$

where y_i denotes the number of times demand point i is covered. Hypercube methodology was incorporated into fitness evaluations in a GA. They applied the approximation of the hypercube model in Jarvis (1985) and used Jarvis' procedure to determine the server-location specific busy probabilities. By accomplishing this, Saydam and Aytug changed the Eq.37 into:

$$\sum_{i \in V} d_i \left(1 - \prod_{j \in W_i} q^j \right). \tag{41}$$

Compared with the solutions from Daskin (1983), this hybrid GA can yield better solutions in reasonable computational time. In addition, they pointed out the empirical 'optimal' solutions in Daskin (1983) tended to over or underestimate the coverage rate.

Besides providing accurate estimation of expected coverage, hypercube queuing theory was adopted to locate EMS facilities under different situations. Takeda et al. (2007) used hypercube queuing models to analyze the effects of ambulance decentralization in Campinas, Brazil. They made some assumptions, including independent Poisson arrivals, independent severs, and fixed dispatching preference. Specifically, they divided the whole city into five non-overlapping areas and each area was treated as two atoms for generating basic and advanced requests, respectively. And there were ten ambulances in the system among which two were advanced ambulances. Demand rate of each atom and service time were derived from historical data. For each atom, the dispatching preference of ambulances was fixed. The results of their experiments showed that as a large number of ambulances were decentralized, the performance measures of the system, such as the response time and workload, were improved. However, the total decentralization did not yield satisfactory results. McLay (2009) proposed the *MEXCLP2* to efficiently deploy two types of medical units to serve multiple types of customers. A hypercube queuing model is developed to analyze the dependencies between facilities within the same type and dependencies of facilities among different types.

The first study to investigate the application of hypercube queuing models in the deployment of EMS facilities on highway is Mendonca and Morabito (2001). They adopted the model in Larson (1974) and made some changes. Upon an emergency, the center dispatches the first preference facility if it is available, otherwise the next facility on this list is dispatched. If the first preference facility and the backup facility are busy, the call is transferred into other systems. In other words, the call is lost. Their model successfully reduced the unbalanced workloads. Based on their previous work, Iannoni and Morabito (2007) considered more complex situation where the calls and servers were of different types. According to the nature of the calls, different types and numbers of facilities were dispatched. After that, they embedded their model into a GA to optimize response areas of each facility in Iannoni et al. (2008). Iannoni et al. (2009) optimized not only the districting coverage areas of facilities but also the facility locations along the highway.

Geroliminis et al. (2004) combined the MCLP and the hypercube queuing theory and formulated the Stochastic Hybrid Queuing Location Model (SHQLM). By placing p servers among potential sites, this model aims to minimize mean response time while forcing a minimum predefined level of coverage to be satisfied. Geroliminis et al. (2006) developed a generalized hypercube model (GHM) in which demands were temporal and spatial varying and the service time was server and emergency specified. Other studies on hypercube queuing models for locating EMS facilities include Larson and Odoni (1981), McKnew (1983), Brandeau and Larson (1986), Burwell (1986), ReVelle (1989), Goldberg et al. (1990b), Goldberg and Szidarovszky (1991), Burwell et al. (1993), Zaki et al. (1997), and Chan (2001). For a detailed review of these models, one may refer to Goldberg (2004).

2.6 Maximum availability location problem (MALP)

Another probabilistic covering location model named the MALP was proposed by ReVelle and Hogan (1989). There are two versions of the MALP. The MALP-I assumed that the facilities had the same busy fraction q. However, in the MALP-II, the busy fraction q_i associated with demand point i was computed as the ratio of the total duration of all calls generated from demand point i to the total availability of all facilities in W_i. The objective of MALP-I is to maximize the population covered by a facility within r at α reliability level. Demand point i is covered with reliability α only if $1 - q^{\sum_{j \in W_i} x_j} \geq \alpha$. Therefore to achieve α reliability coverage, the number of facilities covering demand point i must satisfy $\sum_{j \in W_i} x_j \geq \lceil \frac{\log(1-\alpha)}{\log q} \rceil = b$ The MALP-I can be formulated as follows:

MALP-I:

$$\max \sum_{i \in V} d_i y_{ib} \tag{42}$$

$$\text{subject to} \quad \sum_{j \in W_i} x_j \geq \sum_{k=1}^{b} y_{ik}, \quad i \in V \tag{43}$$

$$y_{i,k+1} \leq y_{ik}, \quad i \in V, k = 1, \ldots, b-1 \tag{44}$$

$$\sum_{j \in W} x_j = p \tag{45}$$

$$x_j, y_{ik} \in \{0, 1\}, \quad j \in W, \ i \in V, \ k = 1, \ldots, p \tag{46}$$

The left-hand side of constraint (43) is the total number of facilities than can cover demand point i within r, and the right-hand side represents the times of demand point i being covered and it is less than b. Because the concavity property in the MEXCLP no longer holds in this model, constraint (44) is required (Brotcorne et al. 2003). Constraint (45) specifies the total number of facilities equal to p.

To relax the assumption in the MALP-I that the probability of different servers being busy are independent, Marianov and ReVelle (1996) proposed the queuing maximal availability location problem (Q-MALP). In their model, the arrival and service activities in the neighborhood around i were treated as a M/G/s-loss queuing system. The Q-MALP used region-specific busy fractions, and the dependence between busy fractions at a local neighborhood level was allowed. They calculated b_i, the smallest number of facilities that must be located to cover demand point i with reliability α, by the queuing theory. Let s be the number of servers in the neighborhood, state k of the system be k servers being busy, λ_i be the arrival rate in region i, and $(1/\mu_i)$ be the mean of the service time for a single server. According to the queuing theory, P[getting into state k]-P[getting out of state k]= 0, that is: $[P_{k-1}\lambda_i + (k+1)\mu_i P_{k+1}] - [P_k\lambda_i + k\mu_i P_k] = 0$ for states 1,2,3,…,s, and for state 0 $\mu_i P_1 - P_0\lambda_i = 0$. Let $\rho_i = \lambda_i/\mu_i$, then the probability of all servers being busy is:

$$P_s = \frac{(1/s!)\rho_i^s}{1 + \rho_i + (1+2!)\rho_i^2 + \cdots + (1/s!)\rho_i^s}.$$

By determining $P_1,\ P_2, \ldots P_s$, and P_{s+1} in sequence, b_i is chosen as the smallest value of k that satisfies $P_k \leq 1-\alpha$.

Galvao et al. (2005) developed the EMALP by integrating hypercube queuing model into the MALP. The identical servers assumption in the MALP was relaxed. They suggested that it was necessary to identify which server was located at which site. Thus they changed the decision variable x_j into x_{jk}, which is equal to 1 if and only if facility k is located at site j. In addition, y_i is redefined as a binary variable, equal to 1 if and only if the demand point i is covered with α reliability. This model can be formulated as follows:

EMALP:

$$\max \sum_{i\in V} d_i y_i \tag{47}$$

$$\text{subject to} \quad \left[1 - \prod_{k=1}^{p} q_k^{\sum_{j\in W_i} x_{jk}} Q\left(p, q, \sum_{j\in W_i}\sum_{k=1}^{p} x_{jk} - 1\right) - \alpha\right] y_i \geq 0 \quad i \in V \tag{48}$$

$$\sum_{j\in W}\sum_{k=1}^{p} x_{jk} = p \tag{49}$$

$$y_i, x_{jk} \in \{0, 1\}, \quad i \in V,\ j \in W,\ k = 1, \ldots, p \tag{50}$$

Just as MALP, the objective of EMALP is to maximize the population covered with α reliability. In constraint (48), q_k is the workload of facility k; $\sum_{j\in W_i} x_{jk}$ equals 1 if facility k covers demand point k, otherwise 0; Q is the correction factor with the form of Eq. 35; $\sum_{j\in W_i}\sum_{k=1}^{p} xjk$ is the total number of facilities that can cover demand point i. This constraint ensures that y_i equal 0 if demand point i cannot be covered with α reliability. Constraint (49) imposes the total number of facilities sited equal to p.

2.7 Dynamic allocation and relocation models

Static models are useful in the strategic level but lack the flexibility in the operational level. While demands vary spatially and temporally, to maximize the coverage of emergency calls, idle EMS facilities siting in low demand areas are needed to move to busier areas. In other words, decision makers need to re-deploy facilities to provide better coverage. As early as the 1970s, Scott (1971) and Wesolowsky and Truscott (1976) studied dynamic location-allocation facility problems. Later, some researchers incorporated temporal and spatial varying demands into their dynamic location models. Recently, the real time facility redeployment problem was thoroughly studied and various models were proposed. Maxwell et al. (2009a) classified research on dynamic allocation problems into three categories: (1) solving the model in real-time (see Gendreau et al. 2001; Rajagopalan et al. 2005; Rajagopalan and Saydam 2005; Nair and Miller-Hooks 2006; Rajagopalan et al. 2008), (2) using the optimal facility position computed in advance (see Gendreau et al. 2006), (3) incorporating system randomness into the model, either by modeling the problem as a Markov decision process (see Berman 1981a, b, c; Zhang et al. 2008; Restrepo 2008; Maxwell et al. 2009a, b) or make decisions under particular system configurations (see Andersson and Vaerband 2007; Andersson 2005).

Gendreau et al. (2001) appears to be the first to consider real time EMS facility redeployment. Based on their previous work in the DSM, they introduced new variables and parameters to reflect the dynamic nature of the new model, named DDSM^t. x_{jk} is a binary variable, equal to 1 if and only if facility k is located at site j. M^t_{jk} represents the penalty related to the relocation of facility k from its current site to site j at time t. The problem is solved at each instant t when a request for facility arises. This model can be written as:

DDSM^t:

$$\max \sum_{i\in V} d_i y_{i2} - \sum_{j\in W}\sum_{k=1}^{p} M^t_{jk} x_{jk} \tag{51}$$

$$\text{subject to} \quad \sum_{j\in W_{i2}}\sum_{k=1}^{p} x_{jk} \geq 1, \quad i \in V \tag{52}$$

$$\sum_{i\in V} d_i y_{i1} \geq \alpha \sum_{i\in V} d_i \tag{53}$$

$$y_{i2} \leq y_{i1}, \quad i \in V \tag{54}$$

$$\sum_{j\in W_{i1}}\sum_{k=1}^{p} x_{jk} \geq y_{i1} + y_{i2}, \quad i \in V \tag{55}$$

$$\sum_{j\in W} x_{jk} = 1, \quad k = 1, \ldots, p \tag{56}$$

$$\sum_{k=1}^{p} x_{jk} \leq p_j, \quad j \in W \tag{57}$$

$$y_{i1},\ y_{i2} \in \{0, 1\}, \quad i \in V \tag{58}$$

$$x_{jk} \in \{0, 1\}, \quad j \in W, k = 1, \ldots, p. \tag{59}$$

The objective of DDSM^t is to maximize the demand covered twice within r_1, minus facilities redeployment cost at time t. Constraint (52) and (53) express the coverage requirement that all demands must be covered within r_2 and a proportion α of the total demands that must be covered within r_1. Constraint (54) imposes that a demand point cannot be covered twice if it is not covered at least once. Constraint (54) and (55) together ensure the demand point i is covered twice only if there are more than two facilities within r_1. Constraint (56) ensures that each facility must be placed to one site. Constraint (57) limited the number of facilities at each site. Note the penalty coefficients are changed over time. By assigning different values, the second item in the objective function can restrict the movement of the same facilicy for round or long trips.

Rajagopalan et al. (2008) developed the Dynamically Available Coverage Location (DACL) model for dynamic redeploying facilities to time-varied demands. They divided the time horizon into clusters based on significant change of demands. The model incorporates the hypercube theory, wit facilities working independently with different busy probabilities. In this model, $x_{jk,t}$ is the binary variable, equal to 1 if and only if facility k is placed at site j at time t; $y_{i,t}$ is the binary variable, equal to 1 if and only if demand point i is covered with α_t reliability at time t; $W_{i,t}$ is the set of facility sites that cover demand point i at time t; p_t is the number of available facilities at time t ; $q_{k,t}$ is the busy fraction of facility k at time t; q_t is the system-wide busy fraction at time t; and d_t is the demand size at demand point i at time t. The DACL model is presented as follows:

DACL:

$$\min \sum_{t=1}^{T} \sum_{j \in W} \sum_{k=1}^{p_t} x_{jk,t} \tag{60}$$

$$\text{subject to} \left[1 - \prod_{k=1}^{p_t} q_{k,t}^{\sum_{j \in W_{i,t}} x_{jk,t}} Q\left(p_t, q_t, \sum_{j \in W_{i,t}} \sum_{k=1}^{p_t} x_{jk,t} - 1\right) - \alpha_t\right] y_{i,t} \geq 0, \quad \forall i, t \tag{61}$$

$$\sum_{i \in V} d_{i,t} y_{i,t} \geq c_t, \quad \forall t \tag{62}$$

$$x_{jk,t}, y_{i,t} \in \{0, 1\}, \quad \forall i, j, k, t. \tag{63}$$

Objective (60) minimizes the number of facilities needed. Similar with Eq. 48, constraint (61) ensures that $y_{i,t}$ equals 0 if demand point i cannot be covered with α_t reliability at time t. Constraint (62) specifies the lower bound of the demands covered with α_t reliability at time t.

Gendreau et al. (2006) proposed a model named the Maximal Expected Coverage Relocation Problem (MECRP) and provided a dynamic relocation strategy for idle EMS facilities siting in low demand areas. The objective is to maximize the expected demand coverage with the number of relocated facilities not exceeding a predefined value. They tested their model in a system where the number of emergency vehicles was relatively small. The solutions obtained from CPLEX were verified by simulation.

2.8 Gradual coverage model

Most of the covering location models are built based on an assumption that a demand point is said to be covered if it can be reached by at least one facility within a predefined threshold. According to this assumption if two demand points A and B are located right inside and outside the boundaries specified by radius r, respectively, point A is fully covered while point b is not covered at all, which is not reasonable. The gradual covering models are proposed to relax this "all or nothing" assumption by using mathematical functions to model the gradual decline of coverage along the increase of the distance. Figure 3 presents four common coverage decay functions. For a review on coverage decay functions, one may refer to Eiselt and Marianov (2009).

Karasakal and Karasakal (2004) developed a partial coverage version of MCLP (MCLP-P) and applied LR to optimize it. They relaxed the "all or nothing" assumption in MCLP by using a sigmoid function (as the one in Fig. 3d) to model the gradual decline of coverage along with the distance increase. They changed a single distance threshold into a distance range within which the coverage changed from "covered" to "not covered". The intermediate coverage level between full coverage and none is called partial coverage. The model is formulated based on the classical p-median

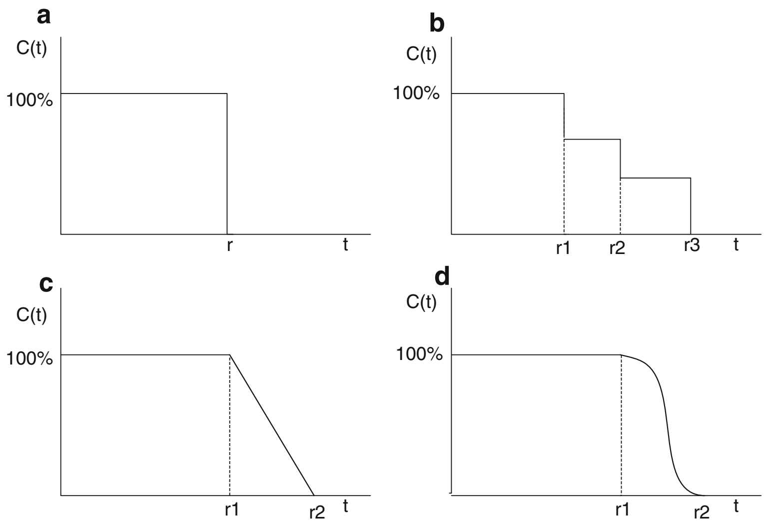

Fig. 3 The coverage decay functions

model. The objective is to maximize the demand coverage instead of minimizing the total distance.The model can be written as follows:

MCLP-P:

$$\max \sum_{i \in V} \sum_{j \in W_i} C_{ij} z_{ij} \tag{64}$$

$$\text{subject to } \sum_{j \in W} x_j = p \tag{65}$$

$$\sum_{j \in W_i} z_{ij} \leq 1, \quad i \in V \tag{66}$$

$$z_{ij} \leq x_j, \quad i \in V, j \in W_i \tag{67}$$

$$x_j \in \{0, 1\}, \quad j \in W \tag{68}$$

$$z_{ij} \in \{0, 1\}, \quad i \in V, j \in W_i, \tag{69}$$

where z_{ij} equals to 1 if and only if demand point i is partially or fully covered by an facility at site j, W_i is the set of the facilities sites that can either fully or partially cover demand point i. If $r_1 < t_{ij} < r_2$, $C_{ij} = f(t_{ij})$; if $t_{ij} < r_1$, $C_{ij} = 1$; otherwise, $C_{ij} = 0$. $f(t_{ij})$ is a sigmoid function, and r_1 and r_2 are the lower bound and the upper bound of the distance range, respectively. The objective of MCLP-P is to maximize the coverage level. Constraint (65) specifies that total p facilities are to be placed. Constraint (66) requires that a demand point will be covered at most once. If there is more than one facility site that can cover demand point i, the one that can provide the maximal coverage will be chosen. This constraint is derived from the p-median model, in which the closest facility is selected to provide service. Constraint (67) forces all $z'_{ij}s$ associated with facility site j to be 0 if no facility is placed at site j. MCLP-P does not take the population size of the demand point into consideration.

Drezner et al. (2010) pointed out that the previous gradual coverage model may not be the correct approach in many situations. They proposed a stochastic gradual coverage model in which the short and long distance standards, i.e., r_1 and r_2, are random variables. They claimed that this generalization is suitable to address settings where customers are heterogeneous and their sensitivity for the distance standard r_1 and r_2 are different. It is assumed that distributions of r_1 and r_2 are known, denoted as ϕ_{r_1} and ϕ_{r_2} and the coverage declined linearly along with the increase of the distance (as in Fig. 3c). They developed the expected coverage at distance t as:

$$c(t) = Pr(r_1 \geq t) + \int_0^t \int_t^\infty \frac{z - t}{z - y} \phi_{r_1}(y) \phi_{r_2}(z) dz dy$$

The objective of the model is to maximize the coverage, which is formulated as $\sum_{i \in W} d_i c(t_i)$, where t_i is the distance between demand point i and its closet facility.

2.9 Cooperative coverage model

In some of the covering models discussed above, only one facility (namely the closest one) determines whether a demand point is covered or not. Berman et al. (2010) pointed out that this individual assumption might lead to solutions that require more facilities to cover the same amount of demands. They proposed the following coverage mechanism: each facility at site j emits a "signal" that decays over distance according to a function $\phi(t)$ (which is similar to the decay function in the gradual covering models). A demand point i receives signals from all facilities and is covered only if the "signal" it receives exceeds a certain threshold, i.e., $\sum_{j \in W} \phi(t_{ij}) \geq A$. The summation is referred to as the aggregation operator in their paper, and researchers should choose appropriate aggregation operators to represent the cooperative behavior of facilities in different systems. Berman et al. (2010) formulated the Cooperative Location Set Covering Problem (CLSCP) and the Cooperative Maximum Covering Location Problem (CMCLP). The objective functions and constraints in the two models have similar meaning to that in the LSCP and the MCLP, respectively. Therefore we present the two models without further explanations.

CLSCP:

$$\min \sum_{j \in W} x_j \tag{70}$$

$$\text{subject to} \sum_{j \in W} \phi(t_{ij}) x_j \geq A, \quad i \in V \tag{71}$$

$$x_j \in \{0, 1\}, \quad j \in W. \tag{72}$$

CMCLP:

$$\max \sum_{i \in V} d_i y_i \tag{73}$$

$$\text{subject to} \sum_{j \in W} \phi(t_{ij}) x_j \geq A y_i, \quad i \in V \tag{74}$$

$$\sum_{j \in W} x_j = p \tag{75}$$

$$x_j, \ y_i \in \{0, 1\}, \quad j \in W, \ i \in V. \tag{76}$$

3 Optimization techniques

In this section, the heuristic algorithms, simulation techniques, and exact methods used to solve EMS facility location models are presented. Heuristic algorithms are broadly used to solve large scale problems. Simulation models are developed either to validate solutions obtained from heuristic algorithms or to combine with other techniques to improve the quality of solutions. B&B is efficient on solving small size problems since most of the problems are formulated as integer programming. Section 3.1 describes some main heuristic algorithms, including GA, TS, LR and other heuristics.

Section 3.2 presents the application of simulation. Section 3.3 briefly reviews the exact methods developed in literature.

3.1 Heuristics

Table 1 summarizes the main heuristic algorithms for various EMS facility location models.

Table 1 Heuristic methods for EMS facility location models

Models	Optimization techniques	References
LSCP	GA	Aickelin (2002)
Non-unicost LSCP	GA	Beasley and Chu (1996)
MCLP	LR	Galvao and ReVelle (1996)
MCLP	TS	Diaz and Rodriguez (1997)
MCLP with partial coverage	LR	Karasakal and Karasakal (2004)
MCLP with multiple quantity	GA	Jia et al. (2007b)
Requirement at different	LR	
Quality level	Locate-allocate	
Extension of MCLP	Goal programming	Alsalloum and Rand (2003)
Extension of MCLP	Goal programming	Alsalloum and Rand (2006)
DSM	TS	Gendreau et al. (1997)
DSM with penalty term to the objective function	TS	Doerner et al. (2005)
MEXCLP	Hypercube embedded GA	Saydam and Aytug (2003)
MEXCLP	GA with local search	Aytug and Saydam (2002)
MEXCLP	GA	Rajagopalan et al. (2007)
	TS	
	SA	
	Hybrid hill-climbing	
Extension of MEXCLP and MALP	SA	Galvao et al. (2005)
Deployment on highway with partial backup and multi-dispatching	Hypercube embedded GA	Iannoni and Morabito (2007)
Same model with above	Hypercube embedded GA with local search	Iannoni et al. (2009)
$DDSM^t$	Parallel TS	Gendreau et al. (2001)
DACL	Reactive TS	Rajagopalan et al. (2008)
Multi-objective covering-based vehicle location model	Fuzzy goal programming	Araz et al. (2007)
Maximum covering model	GA	Jaramillo et al. (2002)

3.1.1 GA

GA is one of the most widely used heuristic approaches. Known as an intelligent probabilistic search algorithm, GA has been applied to a wide range of optimization problems. According to the principles of natural selection and survival of the fittest, genes from the highly fit individuals will pass on to an increasing number of individuals in each successive generation, which results in more fit offspring. In other words, species evolve to become better adapted to their environment (Beasley and Chu 1996).

Beasley and Chu (1996) seem to be the first to apply GA for covering model. Several techniques, including a crossover-fusion operator, variable mutation rate, and a heuristic feasibility operator, were proposed to improve the performance of their algorithm. For small size instances, their algorithm can generate optimal solutions. For large size instances, it can provide near optimal solutions.

Aickelin (2002) used GA to solve the set covering problem. The algorithm splits search into three distinct phases. First, it specifies permutation and other parameters. Second, it uses diversity information to improve the solution. Finally, solutions are post-optimized using a hill-climber heuristic method.

Inspired by Beasley and Chu (1996) and Aytug and Saydam (2002) applied GA to solve the MEXCLP. The performance of their algorithm was compared with that of CPLEX and Daskin's heuristics. The results showed that GA with Local Search heuristics was robust and yielded satisfactory solutions in a reasonable computational time. They provided two techniques for generating the initial population and five crossover operators, then conducted experiments to compare their performance. It demonstrated that GA with ambulance swap crossover operator and random initialization overrode others. The experimental results showed that at least one of their GAs provided optimal or near optimal solutions. Their GA was improved by embedded hypercube queuing model in Saydam and Aytug (2003).

Jaramillo et al. (2002) built a GA similar to the one in Beasley and Chu (1996). Their algorithm was tested on two standard data sets. The solutions obtained from their GA were slightly better, compared with a Lagrangian heuristic followed by a substitution procedure. Jia et al. (2007b) used GA to solve their multi-quantity-quality model. In order to generate good-quality solutions and expedite the convergence of their GA, they applied a greedy process in the individual initialization phase and the crossover operation. Iannoni and Morabito (2007) embedded hypercube queuing model into GA to determine the response district for each ambulance in a highway segment. The hypercube model was applied to evaluate the fitness of each individual in each generation. Iannoni et al. (2009) introduced a local search procedure into their algorithm. By applying their hybrid GA in sequence, the location of ambulance along the highway and the response segments of each ambulance were determined.

3.1.2 TS

TS is a local search method with a unique feature – using a memory or tabu list to record the solution recently examined. This list avoids testing the same solution in a given time period. At each iteration, the solution is moved to the best one among its neighborhood regardless of whether the overall objective value is improved or not.

By accepting worse solutions, TS can effectively escape from local optima. Arostegui et al. (2006) and Rajagopalan et al. (2007) conducted studies to compare the performance of main heuristics algorithms including TS, SA and GA in solving facility location problems. Both of their studies showed that TS could find satisfactory solutions faster with relatively small variability.

Diaz and Rodriguez (1997) developed a simple version of TS to optimize the MCLP. Gendreau et al. (1997) proposed a TS to optimize the DSM. They combined constraints in Eqs. 13 and 14 with the objective function in a hierarchical fashion, in case of the infeasibility of the original problem. The initial population was based on rounded solutions of linear relaxation of the DSM. Gendreau et al. (2001) proposed the DDSM^t and applied parallel TS to provide ambulance deployment schedules in advance for different scenarios. Rajagopalan et al. (2008) formulated the DACL and developed a reactive TS with a *look-ahead* procedure to calculate the number of servers required to satisfy the coverage constraints quickly. Doerner et al. (2005) modified the objective function of the DSM, and implemented the TS in Gendreau et al. (2001) to obtain near optimal solutions.

3.1.3 Other heuristic methods

Unlike the popular GA and TS, other heuristic methods appear to be less commonly investigated for EMS facility location problems. By applying LR, researchers reformulate models into tractable models. The most common application of LR found in the literature is to solve the MCLP or its extensions. Galvao and ReVelle (1996) developed a Lagrangian heuristic for the MCLP. In their algorithm, the objective of MCLP was changed to be minimizing the minus total coverage. They applied a vertex addition and substitution heuristics to obtain the upper bound, and built a sub-gradient optimization algorithm to produce the lower bound. Karasakal and Karasakal (2004) formulated the MCLP in the presence of partial coverage and developed a solution procedure based on LR. Jia et al. (2007b) formulated the MCLP with multiple facility quantity-of-coverage and quality-of-coverage requirements and developed a LR to solve the model.

The *LocAlloc* heuristic was proposed by Cooper (1964) to solve location problems. The procedure of the *LocAlloc* heuristic is to : (1) choose an initial location for each ambulance, (2) determine the response area of each ambulance given the locations, (3) divide the demand points into several groups with at least one ambulance in each group and find the best ambulance location in each group, and (4) if any of the locations has changed, repeat this allocation-location process. This heuristics method uses the property that separating phases of EMS facility location problems makes it more tractable than combining the phases. Jia et al. (2007b) adapted the LocAlloc heuristic to solve their location problem with multiple facility quantity-of-coverage and quality-of-coverage requirement. The initial location was obtained from a greedy process. Because the demand points need to be served at multiple coverage quality, each demand point may belong to several groups. Within each group, the demand points are divided into several subgroups according to the coverage quality they received from the facility in that group. When relocating the facilities, the corresponding weights for demand points in each subgroup are considered.

Galvao et al. (2005) applied SA to optimize the extended MEXCLP and MALP. The solutions obtained via SA are better than those obtained from vertex substitution. In Arostegui et al. (2006) and Rajagopalan et al. (2007), simple version of SA is implemented to solve the MECXLP. Doerner et al. (2005) applied ACO to solve the extended double coverage model in Eqs. 21–24.

3.2 Simulation

As noted by Marianov and ReVelle (1995), the development of EMS facility location and relocation models is likely to be parallel with the growth of the information technologies, and therefore, simulation becomes a more powerful tool to study complex systems. There are three applications of simulation: (1) provide insight into the implementation of policies derived from optimal or near optimal solutions, (2) evaluate and compare performances of different optimization techniques, and (3) combine with other methods to yield better or faster solutions.

Since the assumptions of existing queuing models usually cannot be satisfied in practice, some researchers resorted to simulation (Repede and Bernardo 1994). In their research, the TIMEXCLP was built and incorporated into a decision support system to assist the planners in deploying ambulances in decision periods. Their system was implemented as follows: first, the TIMEXCLP was applied based on the initial information to achieve a certain service level; then by using the output of the TIMEXCLP as input data, a simulation model was implemented to obtain coverage and response time for each demand point. If the planner was satisfied with the result this location scheme would be implemented; otherwise, the TIMEXCLP would be reformulated and this procedure would repeat until satisfactory results were achieved.

Some researchers integrated simulation techniques with other methods. Maxwell et al. (2009a) developed a simulation model to analyze the performance of a current redeployment policy and incorporated this simulation model into an approximate dynamic programming procedure to obtain allocation policies. Harewood (2002) created a simulation model to verify the optimal solutions, conducted sensitive analysis of different system configurations on performance measures, and compared the current deployment policy and the one derived from the optimal solution of the model.

Some simulation models were developed to support decisions on facility location and allocation. Fujiwara et al. (1987, 1988) utilized simulation techniques to develop effective policies for ambulance allocation in Bankok. Liu and Lee (1988) used simulation to analyze a hospital EMS system in Taipei. Goldberg et al. (1990a) developed a simulation model to compare two alternative sets of ambulance locations in Tucson. Zaki et al. (1997) presented a simulation model to optimize the resource allocation and management policy in Richmond. Henderson and Mason (2004) used the analysis software named BARTSIM to integrate geographic information systems (GIS) into their simulation model, and provided support for deployment schedule of ambulances. GIS within BARTSIM provided spatial visualization of different policies.

3.3 Exact methods

Ambulance location models are mostly formulated as integer programming. The B&B was usually applied to obtain optimal solutions for small size instances. Swoveland et al. (1973) combined the B&B technique with simulation to optimize policies for ambulance services. The outputs of a simulation model were used as an initial solution in the B&B procedure. Near optimal solutions obtained by the B&B were further validated by simulation models. Mannino and Sassano (1995) used the B&B to solve set covering problems. Marianov and ReVelle (1994) proposed the QPLSCP and relaxed it as a linear programming which is solved by the B&B. Similarly, Marianov and ReVelle (1996) introduced the Q-MALP and used B&B to solve the linear programming relaxation of this model.

4 Conclusion and discussion

With emergencies unfortunately being part of our lives, it is crucial to efficiently plan and allocate emergency response facilities to deliver effective and timely relief to people most in need. This paper reviewed covering models and optimization techniques for emergency response facility location and planning, from the perspective of mathematical models and operations research. The earlier studies of the covering models are represented by the Location Set Covering Problem (LSCP) and the Maximal Coverage Location Problem (MCLP). The LSCP is a mandatory covering model in which all the demand points are covered at least once, while the MCLP attempts to maximize coverage, given limited resources. Both the LSCP and MCLP have a common drawback in that once an EMS facility is dispatched to serve an emergency call, other demands in its coverage area are not covered by it any more. In this context, multiple coverage concept was introduced to handle excess demands in some locations, such as the recent Double standard Model (DSM) that was proposed to remedy this situation by allocating facilities among potential sites to provide full coverage within a longer distance standard, and to maximize coverage within a shorter distance standard. Another strand of research designed to overcome this drawback is to explicitly model the busy probabilities and reliabilities of facilities. The most frequently used probabilistic models are the Maximum Expected Covering Location Problem (MEXCLP) and the Maximum Availability Location Problem (MALP). Further research developed hypercube queuing models to relax the strong assumptions in previous probabilistic models and obtain a more accurate analysis of the system. Dynamic models were proposed to allocate facilities in real time and provide better coverage for demands. Recently, the commonly used "all or nothing" assumption about the coverage of demands is relaxed by gradual covering models that model the gradual decline of coverage along with the increase of distance by mathematical functions. In addition, the individual assumption that only the nearest facility determines whether a demand point is covered or not is relaxed by cooperative covering models.

This paper further summarizes optimization techniques used to solve the above proposed models, including heuristics, simulation, and exact methods. Genetic Algorithm (GA) and Tabu Search (TS) are the most popular among all heuristics algorithms.

Heuristic methods are usually applied to solve large scale problems. Although sometimes they cannot provide optimal solutions, they usually yield high quality solutions in reasonable computational time. Some research developed hybrid versions of heuristic methods by combining some of them. Simulation is used to either evaluate the performance of polices derived from the solutions obtained via heuristic methods or combined with heuristic methods to provide solutions with better quality. For small size instances, optimal or near-optimal solutions can be obtained by exact methods.

By tracing the development of the above various models, we deem the following research directions are worthy of further research endeavors:

- In previous research, emergency calls or demands were treated as discrete points. All demands in an area are generated from the weighted center of this area. This approximation may result in inaccurate representation of real world situations. It is noteworthy to investigate the possibility of using a continuous area instead of discrete points for demands generation.
- In reality, emergency calls may have different priorities that require different types and/or numbers of emergency services/vehicles. It is of interest to integrate the concept of quality levels as well as priorities in the models.
- Most of the covering models for the EMS facility location and allocation problems use coverage of emergency calls as an objective. Other criteria such as survival rate may serve as a better objective that directly reflects the effectiveness of emergency responses.
- The efficiency of the EMS facility location has been investigated thoughtfully but the equity on the facility distribution has been overlooked. The public expect fair access to EMS facilities. Building models to analyze the trade-off between the two conflicting aspects become necessary.
- The recent advanced geographic information systems (GIS) techniques can be incorporated into the models, which creates the possibility of using visualization to evaluate relief efforts and serve as an effective tool for decision makers. The impacts from other modern techniques/tools such as smart phones and text messaging on emergency relief efforts are yet to be explored.

References

Aickelin U (2002) An indirect genetic algorithm for set covering problems. J Oper Res Soc 53:1118–1126

Alsalloum OI, Rand GK (2003) A goal programming model applied to the ems system at riyadh city, saudi arabia, working Paper

Alsalloum OI, Rand GK (2006) Extensions to emergency vehicle location models. Comput Oper Res 33:2725–2743

Aly AA, White JA (1978) Probabilistic formulation of the emergency service location problem. J Oper Res Soc 29:1167–1179

Andersson T (2005) Decision support tools for dynamic fleet management. Ph.D. thesis, Department of Science and Technology, Linkoepings Universitet, Norrkoeping, Sweden

Andersson T, Vaerband P (2007) Decision support tools for ambulance dispatch and relocation. J Oper Res Soc 58:195–201

Araz C, Selim H, Ozkarahan I (2007) A fuzzy multi-objective covering-based vehicle location model for emergency services. Comput Oper Res 34:705–726

Arostegui MA, Kadipasaoglu SN, Khumawala BM (2006) An empirical comparison of tabu search, simulated annealing, and genetic algorithms for facilities location problems. Int J Prod Econ 103:742–754

Aytug H, Saydam C (2002) Solving large-scale maximum expected covering location problems by genetic algorithms: a comparative study. Eur J Oper Res 141:480–494

Ball M, Lin F (1993) A reliability model applied to emergency service vehicle location. Oper Res 41:18–36

Basar A, Catay B, Unluyurt T (2008) A new model and tabu search approach for planning the emergency service stations. In: Operations research proceedings

Batta R, Dolan JM, Krishnamurthy NN (1989) The maximal expected covering location problem: revisited. Transport Sci 23:277–287

Beasley J, Chu P (1996) A genetic algorithm for the set covering problem. Eur J Oper Res 94:392–404

Berman O (1981) Dynamic repositioning of indistinguishable service units on transportation networks. Transport Sci 15:115–136

Berman O (1981) Repositioning of distinguishable urban service units on networks. Comput Oper Res 8:105–118

Berman O (1981) Repositioning of two distinguishable service vehicles on networks. IEEE Trans Syst Man Cybernet 11:187–193

Berman O, Krass D (2002) Facility location problems with stochastic demands and congestion. In: Facility locations: application and theory. Springer, Berlin, pp 329–371

Berman O, Drezner Z, Krass D (2010) Discrete cooperative covering problems. J Oper Res Soc. Advance online publicatin 15 December 2010

Brandeau M, Larson R (1986) Extending and applying the hypercube queueing model to deploy ambulances in Boston. TIMS Stud Manage Sci 22:121–153

Brotcorne L, Laporte G, Semet F (2002) Fast heuristics for large scale covering-location problems. Comput Oper Res 29:651–665

Brotcorne L, Laporte G, Semet F (2003) Ambulance location and relocation models. Eur J Oper Res 147(4):451–463

Burwell T (1986) A spatially distributed queuing model for ambulance systems. Ph.D. thesis, Clemson University, Clemson

Burwell T, Jarvis J, McKnew M (1993) Modeling co-located servers and dispatch ties in the hypercube model. Comput Oper Res 20:113–119

Ceria S, Nobili P, Sassano A (1998) A lagrangian-based heuristic for large-scale set covering problems. Math Program 81:215–228

Chan Y (2001) Location theory and decision analysis. South Western College Publishing, Cincinnati

Chung C (1986) Recent applications of the maximal covering location planning (M.C.L.P.) model. J Oper Res Soc 37:735–746

Church RL, ReVelle CS (1974) The maximum covering location problem. Papers Reg Sci Assoc 32: 101–118

Cooper L (1964) Heuristic methods for location-allocation problems. Soc Indus Appl Math 6:37–53

Cordeau J, Laporte G, Potvin J, Salvesbergh M (2007) Transportation on demand. In: Barnhart C, Laporte G (eds) Transportation, handbooks in operations research and management science. Elsevier, Amsterdam pp 429–466

Coskun N, Erol R (2010) An optimization model for locating and sizing emergency medical service stations. J Med Syst 34:43–49

Daskin M, Stern E (1981) A hierarchical objective set covermg model for emergency medical service deployment. Tansport Sci 15:137–152

Daskin MS (1983) A maximum expected covering location model: Formulation, properties and heuristic solution. Trans Sci 17:48–68

Daskin M, Hogan K, ReVelle C (1988) Integration of multiple, excess, backup and expected covering models. Environ Plan B 15:13–35

Daskin M (1995) Network and discrete location. Wiley, New York

Dessouky M (2006) Rapid distribution of medical supplies. In: Patient flow: reducing delay in healthcare delivery. Springer, USA, pp 309–339

Diaz B, Rodriguez F (1997) A simple search heuristic for the mclp: Application to the location of ambulance based in a rural region. Int J Manage Sci 25:181–187

Doerner K, Gutjahr W, Hartl R, Karall M, Reimann M (2005) Heuristic solution of an extended double-coverage ambulance location problem for austria. Cent Eur J Oper Res 13:325–340

Doerner KF, Hartl RF (2008) Health care logistics, emergency preparedness, and disaster relief: New challenges for routing problems with a focus on the austrian situation. In: The vehicle rounting problem: lastest Advances and New Challenges. Springer, USA, pp 527–550

Drezner T, Drezner Z, Goldstein Z (2010) A stochastic gradual cover location problem. Naval Res Logist 57:367–372

Eaton D, Church R, Bennett V, Hamon B, Lopez L (1981) On deployment of health resources in rural valle del cauca, colombia. TIMS Stud Manage Sci 17:331–359

Eaton D, Daskin M, Simmons D, Bulloch B, Jansma G (1985) Determining emergency medical service vehicle deployment in austin, texas. Interfaces 15:96–108

Eaton D, Sanchez H, Lantigua R, Morgan J (1986) Determining ambulance deployment in santo domingo, dominican republic. J Oper Res Soc 37:113–126

Eiselt HA, Marianov V (2009) Gradual location set covering with service quality. Socio Econ Plan Sci 43:121–130

Erkut E, Ingolfsson A, Erdogan G (2007) Ambulance location for maximum survival. Naval Res Logist 55:42–58

Fujiwara O, Makjamroen T, Gupta K (1987) Ambulance deployment analysis: a case study of Bangkok. Eur J Oper Res 31(1):9–18

Fujiwara O, Kachenchai K, Makjamroen T, Gupta K (1988) An efficient scheme for deployment of ambulances in metropolitan Bangkok. In: Rand GK (ed) Operational research '87, pp 730–741

Galvao RD, ReVelle C (1996) A lagrangean heuristic for the maximal covering location problem. Eur J Oper Res 88:114–123

Galvao RD, Chiyoshi FY, Morabito R (2005) Towards unified formulations and extensions of two classical probabilistic location models. Comput Oper Res 32:15–33

Gendreau M, Laporte G, Semet F (1997) Solving an ambulance location model by tabu search. Location Sci 5(2):77–88

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

Gendreau M, Laporte G, Semet F (2006) The maximal expected coverage relocation problem for emergency vehicles. J Oper Res Soc 57:22–28

Geroliminis N, Karlaftis M, Skabardonis A (2006) A generalized hypercube queueing model for locating emergency response vehicles in urban transportation networks. TRB 2006 Annual Meeting CD-ROM

Geroliminis N, Karlaftis M, Stathopoulos A, Kepaptsoglou K (2004) A districting and location model using spatial queues. TRB 2004 Annual Meeting CD-ROM

Goldberg J (2004) Operations research models for the deployment of emergency services vehicles. EMS Manage J 1:20–39

Goldberg J, Dietrich R, Chen J, Mitwasi M, Valenzuela T, Criss E (1990) A simulation model for valuating a set of emergency vehicle base location: development, validation, and usage. Socio Econ Plan Sci 24:125–141

Goldberg J, Dietrich R, Chen J, Mitwasi M, Valenzuela T, Criss E (1990) Validating and applying a model for locating emergency medical vehicles in tucson, az. Eur J Oper Res 49:308–324

Goldberg J, Szidarovszky F (1991) Methods for solving nonlinear equations used in evaluating vehicle busy probabilities. Oper Res 39:903–916

Green L, Kolesar P (2004) Improving emergency responsiveness with management science. Manage Sci 50:1001–1014

Harewood S (2002) Emergency ambulance deployment in barbados: a multi-objective approach. J Oper Res Soc 53:185–192

Henderson S, Mason A (2004) Ambulance service planning: simulation and data visualization. In: Operations research and health care: a handbook of methods and applications. Kluwer, Boston, pp 77–102

Hogan K, ReVelle C (1986) Concepts and applications of backup coverage. Manage Sci 32:1434–1444

Iannoni AP, Morabito R (2007) A multiple dispatch and partial backup hypercube queuing model to analyze emergency medical systems on highways. Trans Res Part E 43:755–771

Iannoni AP, Morabito R, Saydam C (2008) A hypercube queueing model embedded into a genetic algorithm for ambulance depolyment on highways. Ann Oper Res 157:207–224

Iannoni AP, Morabito R, Saydam C (2009) An optimization approach for ambulance location and the districting of the response segments on highways. Eur J Oper Res 195:528–542

Jaramillo J, Bhadury J, Batta R (2002) On the use of genetic algorithms to solve location problems. Comput Oper Res 29:761–779

Jarvis JP (1985) Approximating the equilibrium behavior of multi-server loss systems. Manage Sci 32: 235–239

Jia H, Ordonez F, Dessouky MM (2007) A modeling framework for facility location of medical service for large-scale emergency. IIE Trans 39(1):35–41

Jia H, Ordonez F, Dessouky MM (2007) Solution approaches for facility location of medical supplies for large-scale emergecies. Comput Indus Eng 52(1):257–276

Karasakal O, Karasakal EK (2004) A maximal covering location model in the presence of partial coverage. Comput Oper Res 31:1515–1526

Laporte G, Louveaux FV, Semet F, Thirion A (2009) Application of the double standard model for ambulance location. In: Innovations in distribution logistics. Springer, Berlin, pp 235–249

Larson R, Odoni A (1981) Urban operations research. Prentice-Hall, Englewood Cliffs

Larson RC (1974) A hypercube queuing model for facility location and redistricting in urban emergency services. Comput Oper Res 1:67–95

Larson RC (1975) Approximating the performance of urban emergency service systems. Oper Res 23: 845–867

Liu M, Lee J (1988) A simulation of a hospital emergency call system using slam ii. Simulation 51:216–221

Mannino C, Sassano A (1995) Solving hard set covering problem. Oper Res Lett 18:1–5

Marianov V, ReVelle C (1994) The queuing probabilistic location set covering problem and some extensions. Socio Econ Plan Sci 28:167–178

Marianov V, ReVelle C (1995) Sitting emergency services. In: Facility location: a survey of appication and methods. Springer, New York, pp 199–223

Marianov V, ReVelle C (1996) The queueing maximal availability locationm problem: a model for the siting of emergency vehicles. Eur J Oper Res 93(1):110–120

Marianov V, Serra D (1998) Probabilistic maximal covering location allocation models for congested systems. J Reg Sci 38:401–424

Marianov V, Serra D (2002) Location problems in the public sector. In: Facility locations: application and theory. Springer, Berlin, pp 119–150

Maxwell MS, Henderson SG, Topalogu H (2009a) Ambulance redeployment: an approximate dynamic programming approach. In: Rossetti MD, Hill RR, Johansson B, Dunkin A, Ingalls R (eds) Proceedings of 2009 winter simulation conference

Maxwell MS, Restrepo M, Henderson SG, Topaloglu H (2009b) Approximate dynamic programming for ambulance redeployment (to appear)

McKnew MA (1983) An approximation to the hypercube model with patrol initiated activities: an application to police. Decis Sci 14:408–418

McLay LA (2009) A maximum expected covering location model with two types of servers. IIE Trans 41:730–741

Mendonca FC, Morabito R (2001) Analysing emergency medical service ambulance deployment on a brazilian highway using the hypercube model. J Oper Res Soc 52:261–270

Nair R, Miller-Hooks E (2006) A case study of ambulance location and relocation. Presentation in INFORMS Annual Meeting, Pittsburgh Pennsylvania

Ohlsson M, Peterson C, Soderberg B (2001) An efficient mean field approach to the set covering problem. Eur J Oper Res 133:583–595

Owen S, Daskin M (1998) Strategic facility location: a review. Eur J Oper Res 111:423–447

Rajagopalan HK, Saydam C (2005) An effective and accurate hybrid meta heuristic for a probabilistic coverage location problem for dynamic deployment. In: Proceedings of 35th international coference on computers and industrial engineering

Rajagopalan HK, Saydam C, Xiao J (2005) A multiperiod expected covering location model for dynamic redeployment of ambulances. In: Proceedings of the joint conference-10th EWGT meeting and 16th Mini-EURO conference, Poznan, Poland

Rajagopalan HK, Saydam C, Xiao J (2008) A multiperiod set covering location model for dynamic redeployment of ambulances. Comput Oper Res 35:814–826

Rajagopalan HK, Vergara FE, Saydam C, Xiao J (2007) Developing effective meta-heuristics for a probabilistic location model via experimental design. Eur J Oper Res 177:83–101

Repede JF, Bernardo JJ (1994) Developing and validating a decision support system for locating emergenct medical vehicles in louisville kentucky. Eur J Oper Res 75(5):567–581

Restrepo M (2008) Computational methods for static allocation and real-time redeployment of ambulances. Ph.D. thesis, Cornell University, Ithaca, New York

ReVelle C (1989) Review, extension and prediction in emergency siting models. Eur J Oper Res 40:58–69

ReVelle C (1991) Siting ambulances and fire companies: new tools for planners. J Am Plan Assoc 57: 471–484
ReVelle C, Hogan K (1989) The maximum availability location problem. Trans Sci 23:192–200
ReVelle C, Hogan K (1989) The maximum reliability location problem and alpha reliable p-center problems: derivatives of the probabilistic location set covering problem. Ann Oper Res 18:155–174
Saydam AC, McKnew MA (1985) A separable programming approach to expected coverage: an application to ambulance location. Decis Sci 16:381–398
Saydam C, Aytug H (2003) Accurate estimation of expected coverage: revisited. Socio Econ Plan Sci 37:69–80
Schilling D (1980) Dynamic location modeling for public-sector facilities: a multicriteria approach. Decis Sci 11:714–724
Schilling D, Elzinga D, Cohon J, Church RL, ReVelle C (1979) The teem/fleet models for simultaneous facility and equipment sitting,. Trans Sci 13:163–175
Schilling D, Jayaraman V, Barkhi R (1993) A review of covering problems in facility location. Locat Sci 1:25–55
Scott A (1971) Dynamic location-allocation systems: some basic planning strategies. Environ Plan 3:73–82
Shiah D-M, Chen S-W (2007) Ambulance allocation capacity model. In: e-Health networking, applications and services, 2007 9th international conference, Taipei,Taiwan
Sorensen P, Church R (2010) Integrating expected coverage and local reliability for emergency medical services location problems. Socio Econ Plan Sci 44:8–18
Swoveland C, Uyeno D, Vertinsky I, Vickson R (1973) A simulation-based methodology for optimization of ambulance service policies. Socio Econ Plan Sci 7:697–703
Takeda RA, Widmer JA, Morabito R (2007) Analysis of ambulance decentralization in an urban emergency medical service using the hypercube queueing model. Comput Oper Res 34:727–741
Toregas C, Swain R, ReVelle C, Bergman L (1971) The location of emergency service facilities. Oper Res 19:1363–1373
Wesolowsky G, Truscott W (1976) The multiperiod location-allocation problem with relocation of facilities. Manage Sci 22:57–65
Zaki A, Cheng H, Parker B (1997) A simulation model for the analysis and management of an emergency service system. Socio Econ Plan Sci 31:173–189
Zhang O, Mason AJ, Philpott AB (2008) Simulation and optimisation for ambulance logistics and relocation. Presentation in INFORMS Annual Meeting, Washington, DC