

## Research Article

# Analysis of MCLP, Q-MALP, and MQ-MALP with Travel Time Uncertainty Using Monte Carlo Simulation

Noraida Abdul Ghani<sup>1</sup> and Norazura Ahmad<sup>2</sup>

<sup>1</sup>*School of Distance Education, Universiti Sains Malaysia (USM), 11800 Gelugor, Penang, Malaysia*

<sup>2</sup>*School of Quantitative Sciences, UUM College of Arts and Sciences, Universiti Utara Malaysia (UUM), 06010 Sintok, Kedah, Malaysia*

Correspondence should be addressed to Norazura Ahmad; [norazura@uum.edu.my](mailto:norazura@uum.edu.my)

Received 8 March 2017; Accepted 12 June 2017; Published 30 July 2017

Academic Editor: Fu-Yun Zhao

Copyright © 2017 Noraida Abdul Ghani and Norazura Ahmad. This is an open access article distributed under the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

This paper compares the application of the Monte Carlo simulation in incorporating travel time uncertainties in ambulance location problem using three models: Maximum Covering Location Problem (MCLP), Queuing Maximum Availability Location Problem (Q-MALP), and Multiserver Queuing Maximum Availability Location Problem (MQ-MALP). A heuristic method is developed to site the ambulances. The models are applied to the 33-node problem representing Austin, Texas, and the 55-node problem. For the 33-node problem, the results show that the servers are less spatially distributed in Q-MALP and MQ-MALP when the uncertainty of server availability is considered using either the independent or dependent travel time. On the other hand, for the 55-node problem, the spatial distribution of the servers obtained by locating a server to the highest hit node location is more dispersed in MCLP and Q-MALP. The implications of the new model for the ambulance services system design are discussed as well as the limitations of the modeling approach.

## 1. Introduction

In emergency medical services, a responsive and well-managed ambulance service is one of the key factors that can reduce fatality and suffering in patients. For emergency patients, each minute passed by will increase the severity of illnesses or injuries. Therefore, to provide a quick response, ambulances must be positioned at appropriate locations that can cover emergency demand within an acceptable response time. The response time is the most common measure of ambulance location performance, generally defined as the time between the dispatch of medical personnel and arrival of the personnel at the scene. It consists of the dispatch delay time plus the travel time to the scene. Yet, from the literature, some have also considered using the travel time as a surrogate for response time in alleviating ambulance location problem [1, 2]. This is a common practice and the most meaningful measure because the only component that is affected by changing the location of the ambulance is the travel time.

Traditional models have studied ambulance location model using constant travel time, with the assumption that

travel time does not vary with time as there is no difference in velocity and speed [3–5]. However, in reality, there are variations in travel times and neglecting the variations could lead to an inaccurate estimation of the optimal ambulance fleet size. Variations in travel times from run to run would still occur even on an ambulance that traveled from a particular facility site using the same route repeatedly under essentially constant conditions such as same vehicle, driver, weather, and time of day. Thus, the uncertainty of ambulance travel time must be considered when doing analysis using travel time as a performance measure.

Studies on ambulance location problems have a long history in operations research. Among the many types of the ambulance location models in the literature, most of the models can be classified under the category of covering model. There are several types of covering model ranging from a simple to a more complex model such as the Location Set Covering Problem (LSCP) [6], Maximal Covering Location Problem (MCLP) [7, 8], Double Standard Model (DSM) [9], Maximum Expected Covering Location Problem (MEXCLP) [10], Maximum Availability Location

Problem (MALP) [11, 12], Queuing Maximum Availability Location Problem (Q-MALP) [10], and Multiserver Queuing Maximum Availability Location Problem (MQ-MALP) [12] models. In all covering models, the aim is to find a set of optimal locations of facilities, such as ambulances, that can cover all or a maximum number of demand points, where coverage is defined similarly in some models while differently in others.

In this paper, with travel time used as a proxy to the response time, we consider the application of the Monte Carlo approach to incorporating travel times uncertainty in the Maximum Covering Location Problem (MCLP), Queuing Maximum Availability Location Problem (Q-MALP), and Multiserver Queuing Maximum Availability Location Problem (MQ-MALP) models. All three models are applied to the 33-node problem representing Austin, Texas, and the 55-node problem.

The remainder of the paper is structured as follows. Firstly, some pertinent literatures on ambulance location models are presented. Next, a brief review of the MCLP, MALP, and the MQ-MALP models is provided. This is then followed by the Monte Carlo simulation of the MCLP, Q-MALP, and MQ-MALP with travel time uncertainty and the method used to site the ambulances. Analysis and discussion on the obtained results are in the subsequent section. Finally, a brief conclusion concludes the paper.

## 2. Ambulance Location Models

Ambulance location models in the literature may be classified into two categories. The first, descriptive in nature, are the stochastic models which include queuing and simulation studies. They are generally used to determine the performance of a specified allocation and as such, it is often useful in addressing the tactical issues of the deployment of the vehicles. These approaches provide a richness of detail but they are also significantly more complex and require huge data.

The most well-known and comprehensive queuing model is the hypercube model created by Larson [13, 14]. In attempting to get a more accurate estimate for the probability of a vehicle being unavailable, Larson [13] developed a hypercube queuing model that considers the dependency among ambulances. Later, to reduce computational loads of the original model, Larson [14] created an approximate hypercube queuing model. The model can be used to assist planning by describing the consequences of a proposed change in terms of performance measures. Larson's work was followed by many studies that were aimed at incorporating more realistic model features and integrating them into a probabilistic location model [15–17].

In terms of simulation model, one of the earliest applications to the ambulance location problem was by Savas [18]. Other simulation models were then developed in the context of medical facility or resource location problems by Fitzsimmons [19], Swoveland et al. [20], Berlin and Lieberman [21], Uyeno and Seeberg [22], Goldberg et al. [23], Henderson and Mason [24], Gunes and Szechtman [25], and Aringhieri et al. [26]. These simulation models were commonly applied

to evaluate the performance of solutions that were obtained from mathematical models. Unlike the analytical models, simulation models can evaluate rather complicated deployment strategies that do not require the restrictive assumptions frequently needed in analytical models. However, there are some practical challenges in simulation optimization such as the difficulty of evaluating optimality of obtained solutions and possibly long computation time that require considerable computer resources.

The second category of relevant literature is the family of deterministic location models based on a network. Prescriptive in nature, these models require less data and less computer time than the descriptive models. In this paper, the deterministic models are intensively reviewed since our implicit goal is to present a network based model that will fill a gap between the stochastic and deterministic models. The historical starting point of deterministic location model is the  $p$ -median problem proposed by Hakimi [27]. The two main properties of the  $p$ -median problem are the coverage distance that is unrestricted and the number of resources to be located is already identified [28]. The model seeks to locate facilities in such a way that the average response time is minimized. A major problem of this formulation is that solutions to these problems tended to burden some portion of the region with unacceptably long response times [29]. This problem led to the notion of “demand covering.”

There are two types of covering problem. One is to minimize the number of facilities to cover all demand points, while the other type aims at locating a limited number of resources that maximize the number of covered demand points. The first type is originally proposed by Toregas et al. [3], known as the Location Set Covering Problem (LSCP) which sought to position the minimum number of servers in such a way that each and every demand point on the network had at least one server initially positioned some distance or some standard time. However, as the nearest facility may not always be available at the time a demand call is received, the LSCP has been extended to locate a limited number of resources that maximize the number of covered demand points. Church and ReVelle [4] introduced the first maximum covering model known as the Maximal Covering Location Problem (MCLP).

The MCLP seeks the placement of a fixed number of servers so that the population or calls for service that have a server positioned within the standard would be maximized. There are a few extensions of MCLP that were applied to solve problems related to ambulance location. Some successful MCLP applications are the work done by Eaton et al. [30], who managed to solve ambulance location problem in Dominican Republic, and Dessouky et al. [31, 32] who studied large scale MCLP focusing on multiple quality levels and multiple quantities of the vehicles at the demand points. Other extensions of MCLP were also successfully applied on other areas such as gradual covering models by Berman et al. [33] and integrated GIS with MCLP by Alexandris and Giannikos [34].

The LSCP and MCLP, however, have a drawback; once a resource is requested for service, other demand points under its coverage are ignored and need to be catered by

other resources. The drawback leads to the extensions and modifications of the basic model formulations. Two types of model were proposed to overcome the drawback. One type of the covering model aims to provide multiple coverage demand points by using more than one resource such as the Double Standard Model (DSM), introduced by Gendreau [5]. The DSM attempted to allot resources among potential sites using at least two vehicles to provide full coverage for longer distance standard while at the same time maximizing coverage within a shorter distance standard [35]. In the basic DSM model, demand is assumed equal for all nodes. However, Doerner et al. [36] modified and extended the basic DSM model by applying different capacity at the demand nodes.

The other type of the models attempted to take explicit account of the probabilities of servers being busy to compute the amount of redundancy actually needed, such as the MEXCLP by Daskin [37] and MALP by ReVelle and Hogan [38]. Daskin [37] extends the MCLP to account for the possibility that a server may be unable to respond to new demands. The objective of Daskin's siting model was to maximize the expected population covered given a limited number of ambulances to be deployed. Later, ReVelle and Hogan [38] enhanced the MEXCLP by introducing the local estimate of the busy fraction, in the coverage area around node. Instead of maximizing expected coverage as in Daskin, they constrained the level of server availability to be greater than or equal to a preset value, while minimizing the total number of servers. The model, known as the Probabilistic Location Set Covering Problem (PLSCP), was essentially a version of the LSCP with a probabilistic constraint.

Before long realizing that solution to the PLSCP could lead to a large number of servers, potentially larger than what available funds could achieve, ReVelle and Hogan [38] formulated the Maximum Availability Location Problem (MALP). It sought to maximize the population which had a service available within a stated travel time with a specified reliability, given that only servers are to be located. The number of servers needed for  $\alpha$ -reliable coverage of node is computed using the same reasoning as in PLSCP. Other researchers also modified and enhanced the MEXCLP and PLSCP to tackle other EMS location problems such as MOFLEET [39], AMEXCLP [40], and TIMEXCLP [41].

The preceding models discussed so far made an assumption that the probabilities of two vehicles being busy within the same region are independent. Batta et al. [40] relaxed the independence assumption through the use of Larson's approximated hypercube while still maintaining the system-wide busy fraction. Marianov and ReVelle [42], on the other hand, use the region-specific busy fraction in their model. They formulated the Queuing-MALP (Q-MALP) in which the assumption of independence of server availability is relaxed and modeled the behavior of each region as an M/M/s/s queuing system (Poisson arrivals, exponentially distributed service time, servers, and up to calls being serviced at the same time). They later modified the model to include the general service time distribution [43]. Noraida [12] later extended the Q-MALP model by developing the MQ-MALP that considers two types of demand (critical and noncritical) with two types of servers, that is, advanced life support (ALS)

and basic life support (BLS). Though the BLS is not equipped with ALS capabilities, this unit acts as a backup for the ALS in providing coverage to critical calls. In addition, her model takes into account the stochastic nature of the travel times.

Other studies of ambulance location problem were also aware of the importance of taking into account the uncertainty in the travel times, as it will significantly influence the quality of the achieved solution. Mirchandani and Odoni [44] extended the  $p$ -median problem to account for stochastic travel times. Their model assumes travel times to be known when a demand for service arises; however, the state of the system (as described by the link travel times) changes over time according to a Markov process. Daskin [45] formulated a multiobjective model that simultaneously determines the number of vehicles to deploy and their locations and identifies the appropriate dispatch policy and the routes vehicles should use in traveling to emergency locations. His model accounts for the effects of travel time variability on the system performance measures.

### 3. MCLP, Q-MALP, and MQ-MALP Models

In covering models, a demand is considered covered if at least one vehicle can serve the emergency call within a standard distance. As stipulated in the EMS Act of 1973, the standard distance required that 95% of demand calls in urban areas should be reached within 10 minutes, whereas in rural areas the calls should be reached not more than 30 minutes. Adhering to the Act, many studies of ambulance location problem had applied covering models. In this section, we provide further description on the MCLP, Q-MALP, and the MQ-MALP as the direction of this paper is based on these three models.

**3.1. MCLP Model Formulation.** Church and ReVelle [4] introduced the MCLP to modify the assumption of unlimited resources to serve the demand as considered in the LSCP. The MCLP is formulated to maximize the number of demands that can be covered within the desired service distance or time given a limited number of facilities. A demand is considered covered if there is at least one ambulance stationed within some distance away. The MCLP does not consider server availability; hence, it is assumed that once stationed, the ambulance is always available. The model can be stated as follows:

$$\max \quad Z = \sum_{i \in I} f_i y_i \quad (1)$$

$$\begin{aligned} \text{Subject to} \quad & y_i \leq \sum_{j \in N_i} x_j \quad \forall i \in I, \\ & \sum_{j \in J} x_j = P, \\ & x_j, y_i = 0, 1 \quad \forall j \in J, i \in I, \end{aligned} \quad (2)$$

where

$I$  is set of demand nodes (indexed by  $i$ ),

$J$  is set of eligible facility sites (indexed by  $j$ ),

$t_{ij}$  is shortest time from potential facility site  $j$  to demand node  $i$ ,

$S$  is time or distance standard for coverage,

$f_i$  is demand at node  $i$ .

$N_i = \{j \mid t_{ij} \leq S\}$ ; that is,  $N_i$  is the set of nodes  $j$  located within the time or distance standard of demand node  $i$  or the neighborhood of  $i$ . If a call for service originating at this node is answered by available servers stationed inside this neighborhood, it will be answered within the time or distance standard.

$$x_j = \begin{cases} 1 & \text{if a facility is located at node } j, \\ 0, & \text{otherwise,} \end{cases}$$

$$y_i = \begin{cases} 1 & \text{if 1 or more facilities are within } S \text{ of demand node } i, \\ 0, & \text{otherwise.} \end{cases} \quad (3)$$

**3.2. Q-MALP Model Formulation.** Marianov and ReVelle [43] introduced the queuing maximal availability location problem (Q-MALP) that relaxes the assumption that the probability of different servers being busy is independent. In the model, the arrival and servers activities within the neighborhood were treated as M/G/s-loss queuing system. The model allows dependency busy fraction between different servers at a local neighborhood. Using queuing theory,  $b_i$ , the smallest number of vehicles that must be located to cover demand point  $i$  with reliability coverage  $\alpha$  can be computed. In the Q-MALP model, a demand is considered covered if there is at least one ambulance stationed within some distance away. However, once stationed, the ambulance is not always available to respond as it might be servicing other demands or it might be down or under repair and server availability is dependent on other servers in the system. The model can be formulated as follows:

$$\begin{aligned} \max \quad & Z = \sum_{i \in I} f_i y_{ib_i} \\ \text{subject to} \quad & \sum_{k=1}^{b_i} y_{ik} \leq \sum_{j \in N_i} \sum_{k=1}^{c_j} x_{kj} \quad \forall i \in I, \\ & y_{ik} \leq y_{i(k-1)} \quad \forall i, k = 2, 3, \dots, b_i \\ & \sum_{j \in J} \sum_{k=1}^{c_j} x_{kj} = p, \\ & x_{kj}, y_{ik} = 0, 1 \quad \forall i, j, k, \end{aligned} \quad (4)$$

where

$I$  is set of demand nodes (indexed by  $i$ ),

$J$  is set of eligible facility sites (indexed by  $j$ ),

$t_{ij}$  is shortest time from potential facility site to demand node  $i$ ,

$S$  is time or distance standard for coverage,

$a$  is reliability of a server,

$M_i = \{k \mid t_{ik} \leq S, k \in I\}$ ; that is, the set of demand nodes located within  $S$  of node  $i$  or the neighborhood of  $i$ ,

$N_i = \{j \mid t_{ij} \leq S, j \in J\}$ ; that is, the set of potential facility sites  $j$  located within  $S$  of node  $i$  or the neighborhood of  $i$ ,

$\lambda_i$  is arrival rate (calls/day),

$\mu_i$  is service rate (calls/day),

$p_i = \lambda_i / \mu_i$ ,

$b_i$  is the smallest integer satisfying  $(1/b_i!) p_i^{b_i} / (1 + p_i + (1/2!) p_i^2 + \dots + (1/b_i!) p_i^{b_i}) \leq 1 - \alpha$

$$x_j = \begin{cases} m & \text{if } m \text{ servers are positioned at location } j, \\ 0, & \text{otherwise,} \end{cases}$$

$$y_{ik} = \begin{cases} 1 & \text{if } k \text{ or more servers are within } S \text{ of demand area } i, \\ 0, & \text{otherwise.} \end{cases} \quad (5)$$

**3.3. MQ-MALP Model Formulation.** Noraida [12] developed the MQ-MALP to reflect two categories of ambulance, the advanced life support (ALS) and basic life support (BLS). ALS ambulances are manned by paramedic and are equipped to handle life-threatening demands such as cardiac resuscitation and airway management. BLS ambulances provide services for noncritical problems that are managed by emergency medical technician (EMT). The BLS also plays a role as a "backup" for the ALS in providing coverage to critical calls. A call of noncritical nature is considered covered if there is at least one ALS/BLS unit stationed within time standard  $S$  with  $\alpha$ -reliability. Likewise, a call of critical nature is considered covered if there is at least one ALS stationed within time standard  $T$  with probability  $\alpha$ . The concept of coverage in the MQ-MALP is discussed in detail by Noraida [12].

The model can be stated as follows:

$$\begin{aligned} \max \quad & Z = w_a \sum_{i \in I} f_{ai} y_{ai} b_{ai} + w_b \sum_{i \in I} f_{bi} y_{bi} b_{bi} \\ \text{subject to} \quad & \sum_{k=1}^{b_{ai}} y_{aik} \leq \sum_{j \in N_{si}} x_{aj} + \sum_{j \in N_{si}} x_{bj} \quad \forall i \in I, \\ & \sum_{k=1}^{b_{bi}} y_{bik} \leq \sum_{j \in N_{Ti}} x_{bj} \quad \forall i \in I, \\ & y_{aik} \leq y_{ai(k-1)} \quad \forall k = 2, 3, \dots, b_{ai}, \\ & y_{bik} \leq y_{bi(k-1)} \quad \forall k = 2, 3, \dots, b_{bi}, \\ & \sum_{j \in J} x_{aj} = p^a \\ & \sum_{j \in J} x_{bj} = p^b \end{aligned}$$

$$\begin{aligned}
x_{aj} + x_{bj} &\leq c_j \quad \forall j \in J, \\
x_{aj} + x_{bj} &= \text{integer} \leq c_j \quad \forall j \in J, \\
y_{aik} + y_{bik} &= 0, 1 \quad \forall i, k, \\
N_{si} &= \{j \mid t_{ij} \leq s, j \in J\}, \\
N_{Ti} &= \{j \mid t_{ij} \leq T, j \in J\},
\end{aligned} \tag{6}$$

where

$I$  is set of demand nodes (indexed by  $i$ ),  
 $J$  is set of eligible facility sites (indexed by  $j$ ),  
 $t_{ij}$  is shortest time from potential facility site  $j$  to demand node  $i$ ,  
 $S$  is time standard for coverage of critical calls,  
 $T$  is time standard for coverage of noncritical calls,  
 $f_{ai}$  is demand of critical nature at node  $i$  (number of calls per day),  
 $f_{bi}$  is demand of noncritical nature at node  $i$  (number of calls per day),

$a$  is reliability of a server,

$b_{ai}$  is the minimum number of EMS (ALS or BLS) units which must be located within  $S$  unit of node  $i$  for node  $i$  to be covered with reliability  $a$ , precomputed using

$$\frac{(1/b_{ai}!) p_i^{b_{ai}}}{1 + p_{ai} + (1/2!) p_{ai}^2 + \dots + (1/b_{ai}!) p_{ai}^{b_{ai}}} \leq 1 - a, \tag{7}$$

$b_{bi}$  is the minimum number of BLS units which must be located within  $T$  unit of node  $i$  for node  $i$  to be covered with reliability, precomputed using

$$\frac{(1/b_{bi}!) p_i^{b_{bi}}}{1 + p_{bi} + (1/2!) p_{bi}^2 + \dots + (1/b_{bi}!) p_{bi}^{b_{bi}}} \leq 1 - a, \tag{8}$$

$p_a$  is number of available ALS units to locate,

$p_b$  is number of available BLS units to locate,

$c_j$  is capacity of site  $j$ .

$w_a, w_b \geq 0$  are the weights associated with each objective:

$$\begin{aligned}
y_{aik} &= \begin{cases} 1 & \text{if at least } k \text{ EMS (AL or BLS) units are within } S \text{ of demand node } i, \\ 0, & \text{otherwise,} \end{cases} \\
y_{bik} &= \begin{cases} 1 & \text{if at least } k \text{ BLS units are within } T \text{ of demand node } i, \\ 0, & \text{otherwise,} \end{cases} \\
y_{aj} &= \begin{cases} m & \text{if at least } m \text{ ALS servers are located at side } j, \\ 0, & \text{otherwise,} \end{cases} \\
y_{bj} &= \begin{cases} m & \text{if at least } m \text{ BLS servers are located at side } j, \\ 0, & \text{otherwise.} \end{cases}
\end{aligned} \tag{9}$$

To account for the stochastic nature of the travel times, the above MQ-MALP model formulation was extended by incorporating a probabilistic measure in  $N_{Si}$  and  $N_{Ti}$  as explained in Section 4.

It is better to calculate the average duration of critical call originating from demand node  $i$ . However, as data is not collected at that level,  $\bar{t}_a$  is used as an estimate instead. This is where the assumption [see Marianov and ReVelle [43]] of treating the demand neighborhoods as self-containing comes into play.

As indicated by Figure 1, this assumption may not be consistent as it is possible that facility node  $m$  is part of the response area of demand node  $i$  as well as of demand node  $j$ . Averaging with respect to critical calls to demand node  $i$  (i.e., calculate  $\bar{t}_{ai}$ ) would improve the apparent inconsistency as this average would take into account the fact that responses to critical calls originating at demand node  $i$  are delayed because

servers at facility node  $m$  are also responding to demand node  $j$ . However, even with this adaptation the minimum number of servers required calculated in our model would still be a lower bound on the number that is actually needed to arrive at this reliability due to the self-containment assumption of demand node  $i$ .

#### 4. Monte Carlo Simulation of MCLP, Q-MALP, and MQ-MALP Models with Travel Time Uncertainty

In the formulations of MCLP, Q-MALP, and MQ-MALP, the response time/travel times along the arcs of the network are assumed to be deterministic. In other words, the probability distribution of the response time,  $T_{ij}$ , is degenerate. Suppose



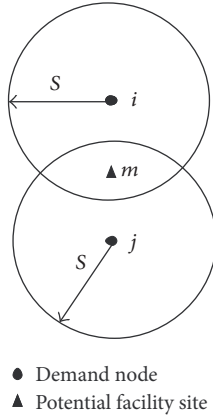


FIGURE 1: Example of facility site within  $S$  neighborhood of node  $i$  and node  $j$ .

now that the response times  $T_{ij}$  are nondegenerate random quantities with probability distribution,  $F_{T_{ij}}$  (Figure 2).

By treating the response times as random quantities, an improvement was made by Noraida [12] in the way  $N_{s_i}$  and  $N_{T_i}$  are computed. This was done by choosing a neighborhood of each node in such a way that, if a call for service originating at this node is answered by an available server located within the neighborhood, it will be answered within time standards with probability  $\gamma$ .  $N_{s_i}$  is redefined as

$$\begin{aligned} N_{s_i} &= \{j \mid p(T_{ij} \leq S) \geq \gamma\} = \{j \mid F_{T_{ij}}(S) \geq \gamma\} \\ &= \{j \mid S \geq F_{T_{ij}}^{-1}(\gamma)\} = \{j \mid F_{T_{ij}}^{-1}(\gamma) \geq S\}. \end{aligned} \quad (10)$$

Similarly, the same can be done with  $N_{T_i}$  and is redefined as

$$N_{T_i} \{j \mid F_{T_{ij}}^{-1}(\gamma) \geq T\}. \quad (11)$$

However, the above are only applicable if and only if  $F_{T_{ij}}^{-1}$  exists. Instead of redefining  $N_{s_i}$  and  $N_{T_i}$  in the formulation, another approach towards taking travel time uncertainty into account is to use a Monte Carlo regimen. Here, one would sample the travel times from their respective distribution and solve the MCLP, Q-MALP, and MQ-MALP model described in Sections 3.1, 3.2, and 3.3, respectively. This constitutes one iteration, resulting in an optimal solution for the sampled travel times.

By following this approach repeatedly, a whole series of optimal solutions can be generated. One would hope that particular nodes would be members of optimal solutions more frequently than other nodes. Those nodes that are hit the most should be considered for placement of EMS (ALS or BLS) units.

## 5. Results and Analysis

This section considers the application of the Monte Carlo approach to incorporating travel times uncertainty in the MCLP, Q-MALP, and MQ-MALP models. We will use this approach to the 33-node problem and the 55-node problem.

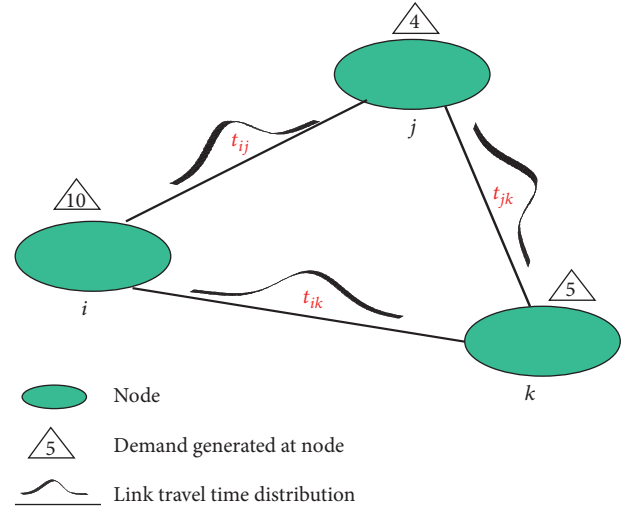


FIGURE 2: A network with travel time distribution.

The 33-node problem from Daskin [46] represents Austin, Texas, at the census tract level. Interzonal travel times are given by travel matrix with intrazonal times taken to be one minute. The weights associated with each zones are the number of calls for ambulance services recorded in the census tract during the five-month period for which the data were available.

The 55-node problem is the test problem proposed by Toregas [3]. The data is drawn from a common base composed of 55 points identified by their coordinates and their  $(x, y)$  user population. The travel distance matrix is generated by

$$d_{ij} = \sqrt{(x_i - x_j)^2 + (y_i - y_j)^2}. \quad (12)$$

However, for the purpose of analysis, the following modifications were made similar to Noraida [12]. First, for each network, the population concentration at each node was multiplied by a constant factor such that the resulting average calls per day over the entire network are 0.4 calls per day. These are used as estimates of the number of calls per node per day. Second, we used the finding from Eaton et al. [30] (only 20–25 percent of calls for ambulance service in Austin require the advanced skills of paramedics) in order to get estimates for the breakdown of these calls into critical and noncritical calls. Thus, the population concentration at each node was multiplied by a constant factor such that the resulting average calls for ALS services per node per day over the entire network are 0.1 calls per day. These are used as estimates of the number of critical calls per node per day. The differences between these two estimates are used as estimates of the number of noncritical calls per node per day. Third, an average duration of a single service of 3/4 of an hour was used [43]. This figure was estimated based on the average of three cases: firstly, the ambulance goes to the site of the call, stays there for some time, and then goes back to the facility site; secondly, the ambulance reaches the emergency site, takes the

patient to a hospital, and returns to its assigned facility site; finally, there is the case of possibility of a false alarm or the event that the emergency is over when the ambulance reaches the alarm site.

The distributions of the travel time were assumed to be distributed as Weibull with scale and shape parameters,  $\alpha$  and  $\beta$ , chosen such that the means equaled their counterparts for the deterministic travel times. As for the variances, a more realistic model would be to use the variances of the travel time distributions itself. However, in the absence of such data, one could use a standard coefficient of variation to reflect the proportionality of the variances with the means. However, for simplicity we will use a constant variance of 4 (a standard deviation of 2) for all travel times. Reliability for server availability was set at 0.95 and 0.9 for the 33-node problem and the 55-node problem. The number of ALS and BLS servers to locate was set at 3 and 5, respectively. A response time of 10 minutes was used in conformance with stipulations of the EMS Act for urban areas for all servers with the exception of 8 minutes for the ALS servers.

At each node, two sets of 1000 travel time samples each were generated using Delphi version 4.0. In generating the travel time ( $X_i$ ), a random number generator was used to first generate  $U_i \sim U(0, 1)$ ,  $i = 1, \dots, 1000$ . By letting  $X_i = \beta[-\ln(1 - U_i)]^{1/\alpha}$ , the resulting  $X_i \sim \text{Weibull}(\alpha, \beta)$ ,  $i = 1, \dots, 1000$ . The first set was generated assuming complete independence of travel time between each pair of nodes, while the second set was generated assuming complete dependence of travel time between each pair of nodes. These assumptions represent the two extremes of the relationship of travel time between each pair of nodes. A different seed number was used in the random number generator for each pair of nodes in generating the complete independence of travel time between each pair of nodes, while the same seed number was used in the random number generator for all pairs of nodes in generating the complete dependence of travel time between each pair of nodes. The generated travel times were used as inputs into the optimization problem MQ-MALP and solved using Xpress-MP version 13.1.

Figures 3-4 display the percentage and cumulative percentage of EMS vehicle node location of the optimal solutions of MCLP and Q-MALP models for the 33-node problem, while Figures 5-6 show the percentage and cumulative percentage of ALS and BLS vehicle node location of the optimal solutions of MQ-MALP model for the 55-node problem. In overall, Figures 3-6 show that the percentage distribution of EMS vehicle node location of the optimal solutions of the MCLP, Q-MALP, and the MQ-MALP models appears steeper when dependent travel times between each node were used. For example, in the 33-node problem, the cumulative percentage of the three highest ranking nodes (as identified by the optimal solution of the MCLP using independent travel times) equals 31.1 percent, while the cumulative percentage of the three highest ranking nodes (as identified by the optimal solution of the MCLP using dependent travel times) equals 44.8 percent. On the other hand, the difference in the cumulative percentages of the

three highest ranking nodes is more pronounced in the Q-MALP model (48% when using independent travel times as opposed to 76.1% when using dependent travel times).

As for the MQ-MALP model, the cumulative percentage of the three highest ranking nodes as identified by the optimal solution of the model stands higher at approximately 60 percent for ALS node location and 80 percent for the BLS node location when using independent travel times, while the cumulative distribution of the three highest ranking nodes as identified by the optimal solution of the model stands at approximately 77 percent for ALS node location and 89 percent for the BLS node location when using dependent travel times.

Likewise, in the 55-node problem, the cumulative percentage of the three highest ranking nodes as identified by the optimal solution of the MCLP using independent travel times approximately equals 25 percent, while the cumulative percentage of the three highest ranking nodes as identified by the optimal solution of the MCLP using dependent travel times approximately equals 35 percent. The difference in cumulative percentages, however, is less pronounced when Q-MALP model was employed (29.6% when using independent travel times as opposed to only 33% when using dependent travel times). As for the MQ-MALP model, the cumulative distribution of the three highest ranking nodes as identified by the optimal solution of the model stands at approximately 27 percent for ALS node location and 52 percent for the BLS node location when using independent travel times, while the cumulative distribution of the three highest ranking nodes as identified by the optimal solution of the model stands at approximately 33 percent for ALS node location and 56 percent for the BLS node location when using dependent travel times.

In comparing the three models, the cumulative distribution of EMS vehicle node location appears to be the steepest in the MQ-MALP model followed by the Q-MALP model and the MCLP model. This can be partially explained by the fact that the Q-MALP and the MQ-MALP allow for the placement of more than one EMS vehicle at a particular node to reflect the capacity of each node to site the vehicles while the MCLP does not. In addition, the incorporation of server availability in the definition of coverage for the Q-MALP and MQ-MALP model at times calls for the placement of more than one server at a particular node in its effort to maximize coverage.

Tables 1 and 2 show the distribution of the servers for the 33-node problem and the 55-node problem, respectively, when the servers are located at the highest hit node location as identified by the optimal solutions of the MCLP, Q-MALP, and MQ-MALP models. Spatial distribution of the servers corresponding to Tables 1 and 2 is as shown in Figures 7 and 8. From Tables 1 and 2, one can see that there is not any node in common between the solutions obtained by all 6 models in terms of server location. However, within each model employed, the results obtained using independent travel times and dependent travel times show that there are some nodes that are in common in terms of server location. For example, in the 33-node problem, only one node (node 25) is in common when the MCLP model is employed, while two

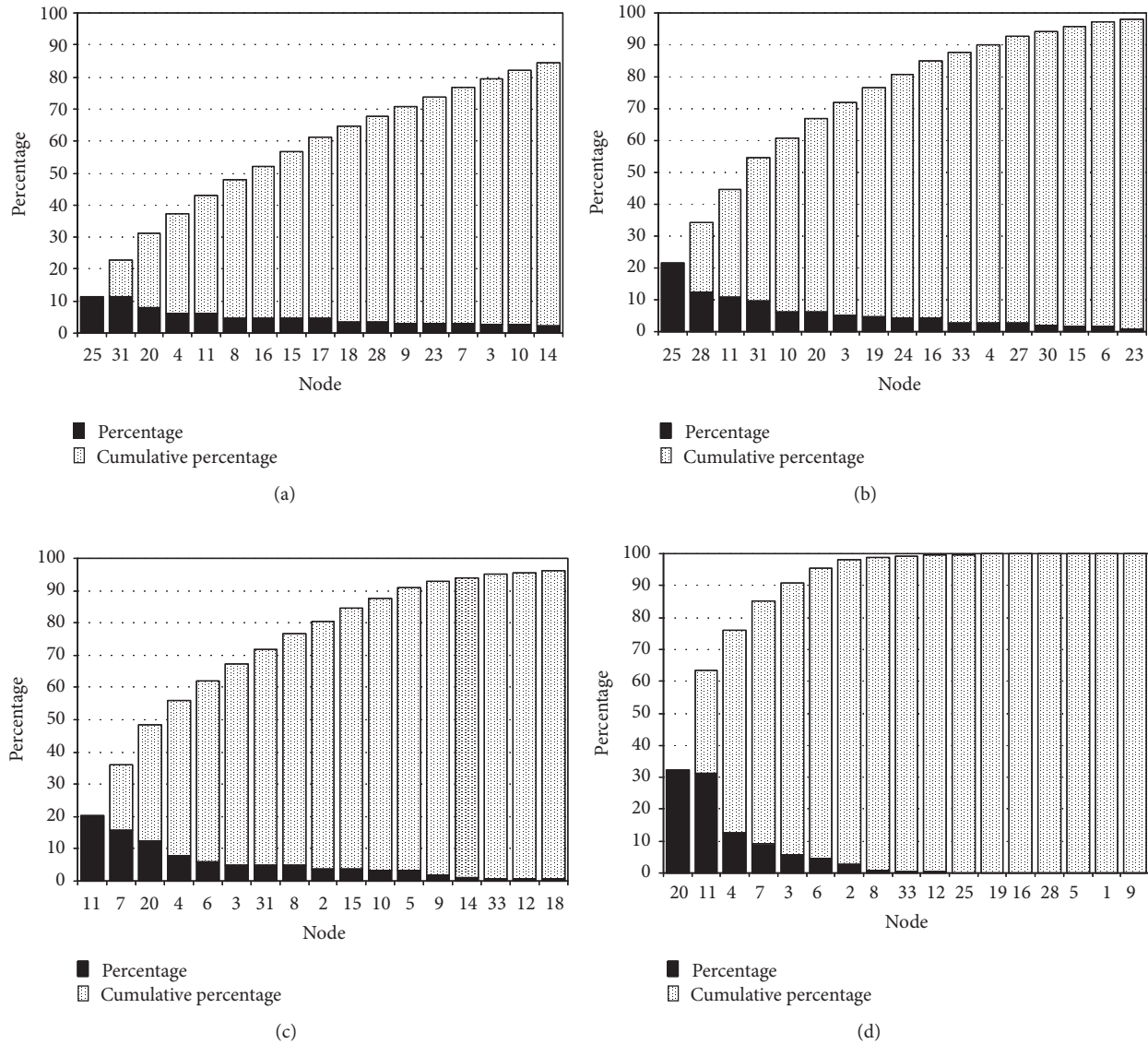


FIGURE 3: Bar graphs of the percentage and cumulative percentage of EMS vehicle node location of optimal solutions of MCLP and Q-MALP with independent and dependent travel times applied to the 33-node problem. (a) MCLP: EMS vehicle node location with independent travel times. (b) MCLP: EMS vehicle node location with dependent travel times. (c) Q-MALP: EMS vehicle node location with independent travel times. (d) Q-MALP: EMS vehicle node location with dependent travel times.

nodes (nodes 11 and 20) are in common when the Q-MALP model is employed. On the other hand, server placements are identical when the MQ-MALP model is employed. The same, however, cannot be said of the MQ-MALP model when applied to the 55-node problem.

Figures 7 and 8 allow a comparison between the six models employed in terms of the spatial pattern of the server nodes chosen. As can be seen from these figures, whether one is using the independent or dependent travel time distribution, the servers are less spatially distributed in the latter two models, that is, Q-MALP and MQ-MALP, when the uncertainty of server availability is taken into account. While the servers are more spatially distributed in the MCLP model,

this model guarantee no more than the initial placement of a server within the coverage standard and does not account for the possibility that call volume and other factors can frequently make the servers unavailable at the time of a call.

For the 33-node problem, in general, the spatial distribution of the servers obtained by locating a server to the highest hit node location as identified by the optimal solutions of the models employed is comparable to those by Noraida [12], an approach in which the uncertainty of the travel time in terms of a probability measure is incorporated into the model formulation itself. If the placements of two servers are close to each other, then from an economic point of view, it makes sense to place both servers at only one of the locations



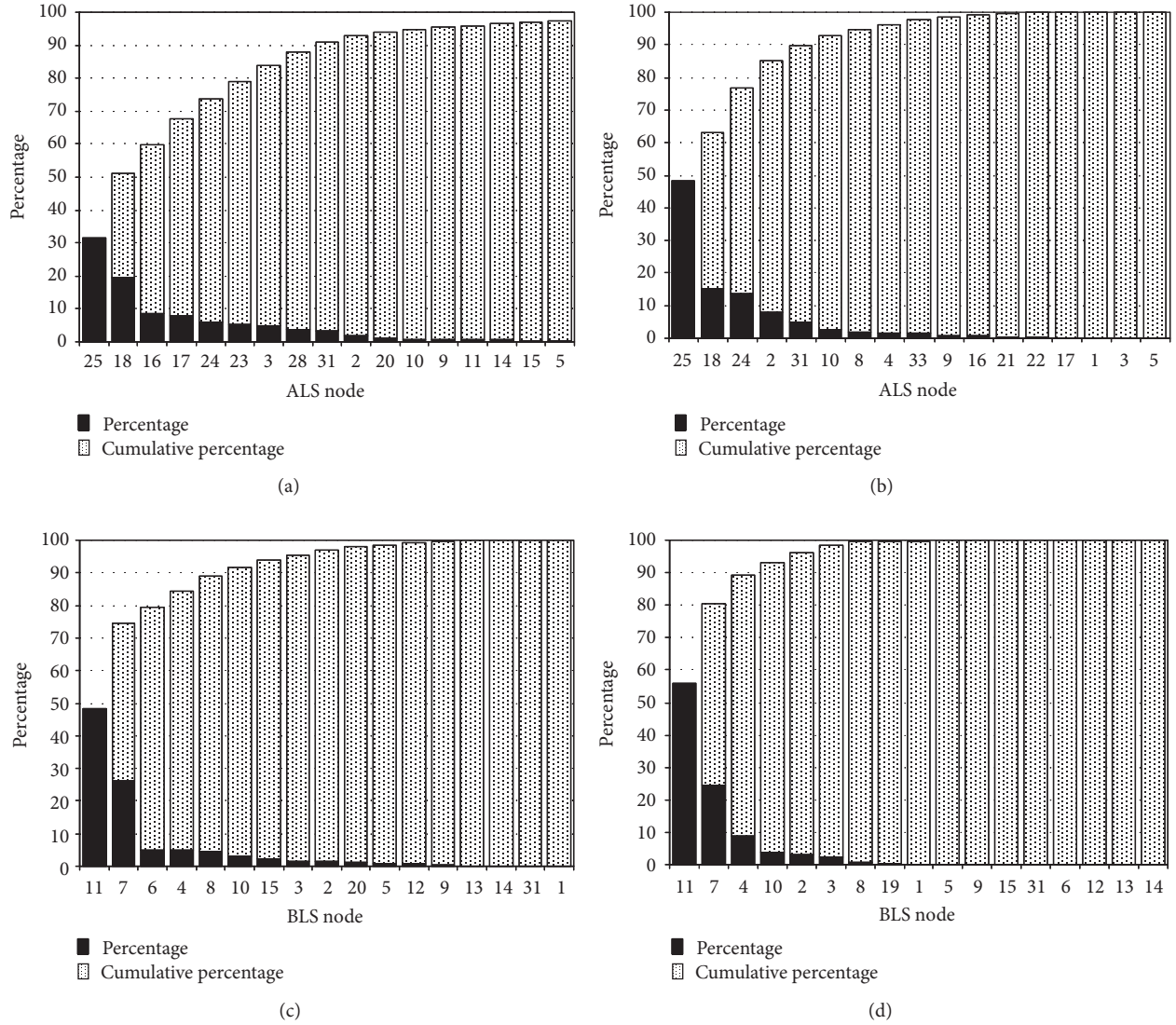


FIGURE 4: Bar graphs of the percentage and cumulative percentage of ALS and BLS vehicle node location of optimal solutions of MQ-MALP with independent and dependent travel times applied to the 33-node problem. (a) MQ-MALP: ALS vehicle node location with independent travel times. (b) MQ-MALP: ALS vehicle node location with dependent travel times. (c) MQ-MALP: BLS vehicle node location with independent travel times. (d) MQ-MALP: BLS vehicle node location with dependent travel times.

as one would have to build only one facility site instead of two.

Likewise, for the 55-node problem, in general, the spatial distribution of the servers obtained by locating a server to the highest hit node location as identified by the optimal solutions of the MCLP and Q-MALP model is comparable to those obtained by Noraida [12]. The same, however, cannot be said of the MQ-MALP model.

## 6. Conclusion

From the literature, covering models are the most prevalent location models to alleviate the ambulance location problem for long term planning. The concept of coverage stated that a demand point is considered covered by an ambulance

location if the travel time between the ambulance location and demand point is less than or equal to a standard arrival time. However, estimating the expected coverage in reality is not easy as it depends on the uncertainty of many elements such as travel time, server availability, and demand. Thus, incorporating these uncertainties into a covering model is essential in obtaining a precise estimation of the coverage. This paper proposed the application of the Monte Carlo approach to incorporating travel times uncertainty in the MCLP, Q-MALP, and MQ-MALP models. The MCLP does not take into account server availability unlike the formulation of Q-MALP and MQ-MALP. Further, MCLP and Q-MALP assumed a single type of ambulance, while the MQ-MALP model considered two types of ambulance, that is, the ALS and BLS. The formulation of multiobjective

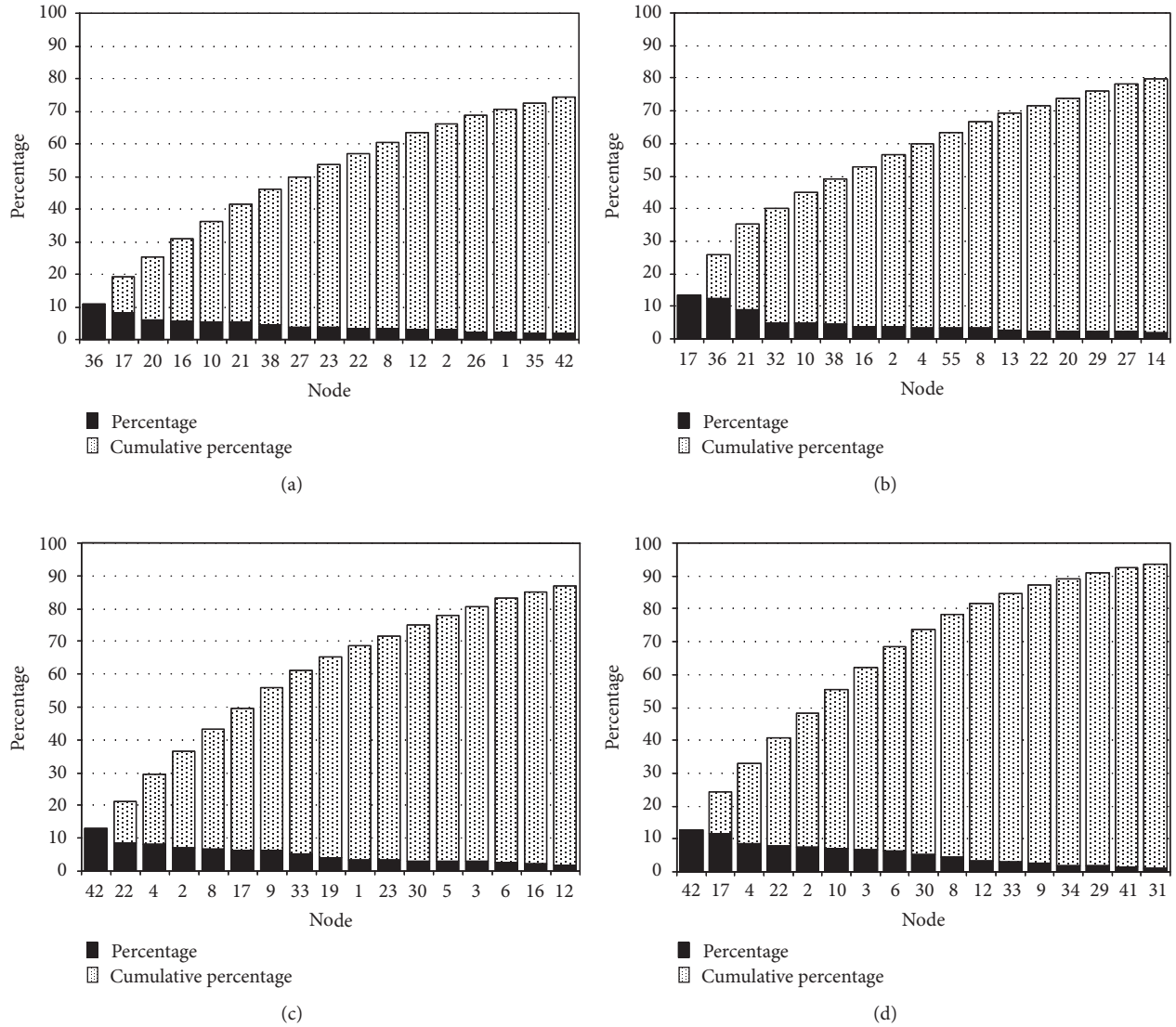


FIGURE 5: Bar graphs of the percentage and cumulative percentage of EMS vehicle node location of optimal solutions of MCLP and Q-MALP with independent and dependent travel times applied to the 55-node problem. (a) MCLP: EMS vehicle node location with independent travel times. (b) MCLP: EMS vehicle node location with dependent travel times. (c) Q-MALP: EMS vehicle node location with independent travel times. (d) Q-MALP: EMS vehicle node location with dependent travel times.

function provides several alternatives for the decision maker in determining a good location for ambulances.

Results from the models show that the cumulative distribution of EMS vehicle node location for MQ-MALP model is the steepest followed by Q-MALP model. This may be possibly due to the fact that the Q-MALP and the MQ-MALP allow the position of more than one ambulance as allowed by the capacity at a particular node to site the vehicles, while the MCLP does not. In addition, the incorporation of server availability in the definition of coverage for the Q-MALP and MQ-MALP model at times calls for the placement or more than one server at a particular node in its effort to maximize coverage.

In terms of server location, results show that the solutions obtained by all six models do not have any common nodes.

However, within each model employed, the results obtained using independent travel times and dependent travel times show that there are some nodes that are in common in terms of server location. On the other hand, server placements are identical when the MQ-MALP model is employed. The same, however, cannot be said of the MQ-MALP model when applied to the 55-node problem. From an economic point of view, if the placements of two servers are close to each other using our approach, it makes sense to place both servers at only one of the locations as one would have to build only one facility site instead of two.

While the model formulations progress the state of the art of the emergency services modeling, there are a few limitations which need to be considered in future works. First, the models do not explicitly consider other stochastic

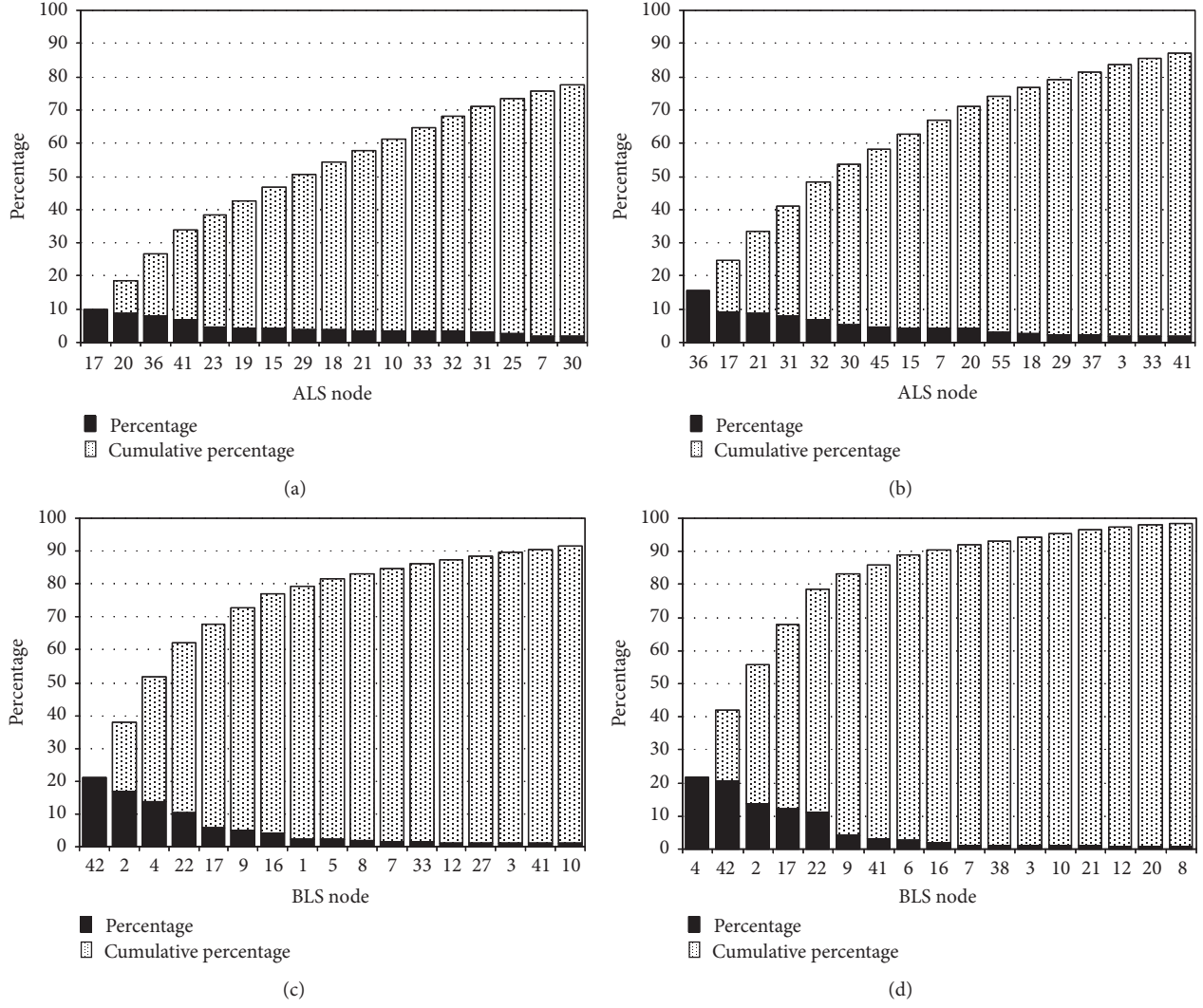


FIGURE 6: Bar graphs of the percentage and cumulative percentage of ALS and BLS vehicle node location of optimal solutions of MQ-MALP with independent and dependent travel times applied to the 55-node problem. (a) MQ-MALP: ALS vehicle node location with independent travel times. (b) MQ-MALP: ALS vehicle node location with dependent travel times. (c) MQ-MALP: BLS vehicle node location with independent travel times. (d) MQ-MALP: BLS vehicle node location with dependent travel times.

TABLE 1: Server location of optimal solution of MCLP, Q-MALP, and MQ-MALP with independent and dependent travel times applied to the 33-node problem.

Model	Travel time	Node location		
		ALS	BLS	EMS
MCLP	Independent			20, 25, 31
MCLP	Dependent			11, 25, 28
Q-MALP	Independent			7, 11, 20
Q-MALP	Dependent			4, 11, 20
MQ-MALP	Independent	25	7, 11	
MQ-MALP	Dependent	25	7, 11	

natures that are often important in designing emergency service such as demand for services. Hence, incorporating the stochastic demand into the formulation would improve the credibility of the optimized placement solution. Second, the Monte Carlo uncertainty analysis in the model is performed

by assigning ambulances to the facility nodes that are “hit” the most as optimal solutions. Such a solution may in fact not satisfy the feasibility constraints of the original model. Perhaps a solution could be suggested that maximizes the percentage of times the nodes are optimal while still

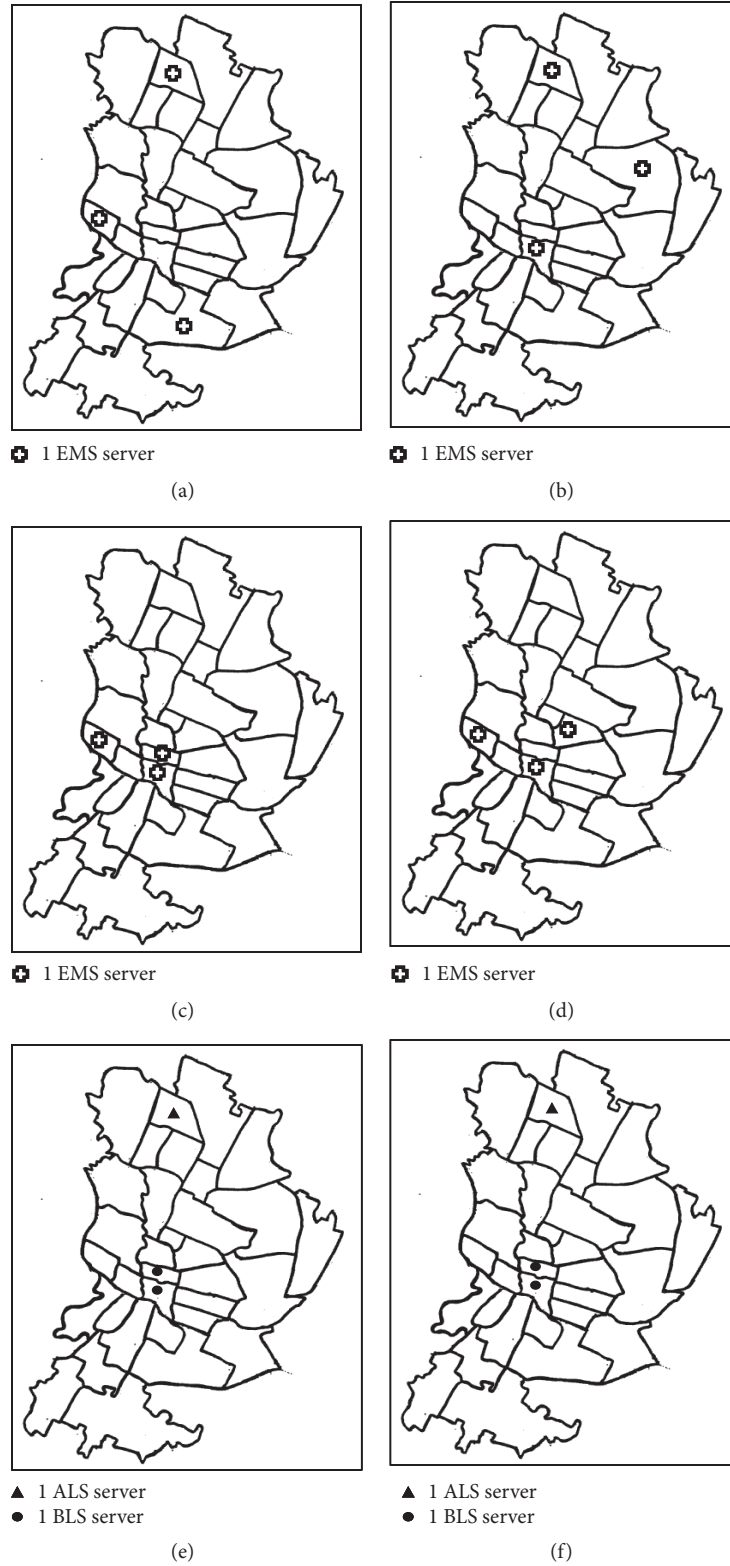


FIGURE 7: Spatial distribution of server node location of optimal solutions of MCLP, Q-MALP, and MQ-MALP with independent and dependent travel times applied to the 33-node problem. (a) MCLP: EMS vehicle node location with independent travel times. (b) MCLP: EMS vehicle node location with dependent travel times. (c) Q-MALP: EMS vehicle node location with independent travel times. (d) Q-MALP: EMS vehicle node location with dependent travel times. (e) MQ-MALP: ALS and BLS vehicle node location with independent travel times. (f) MQ-MALP: ALS and BLS vehicle node location with dependent travel times.

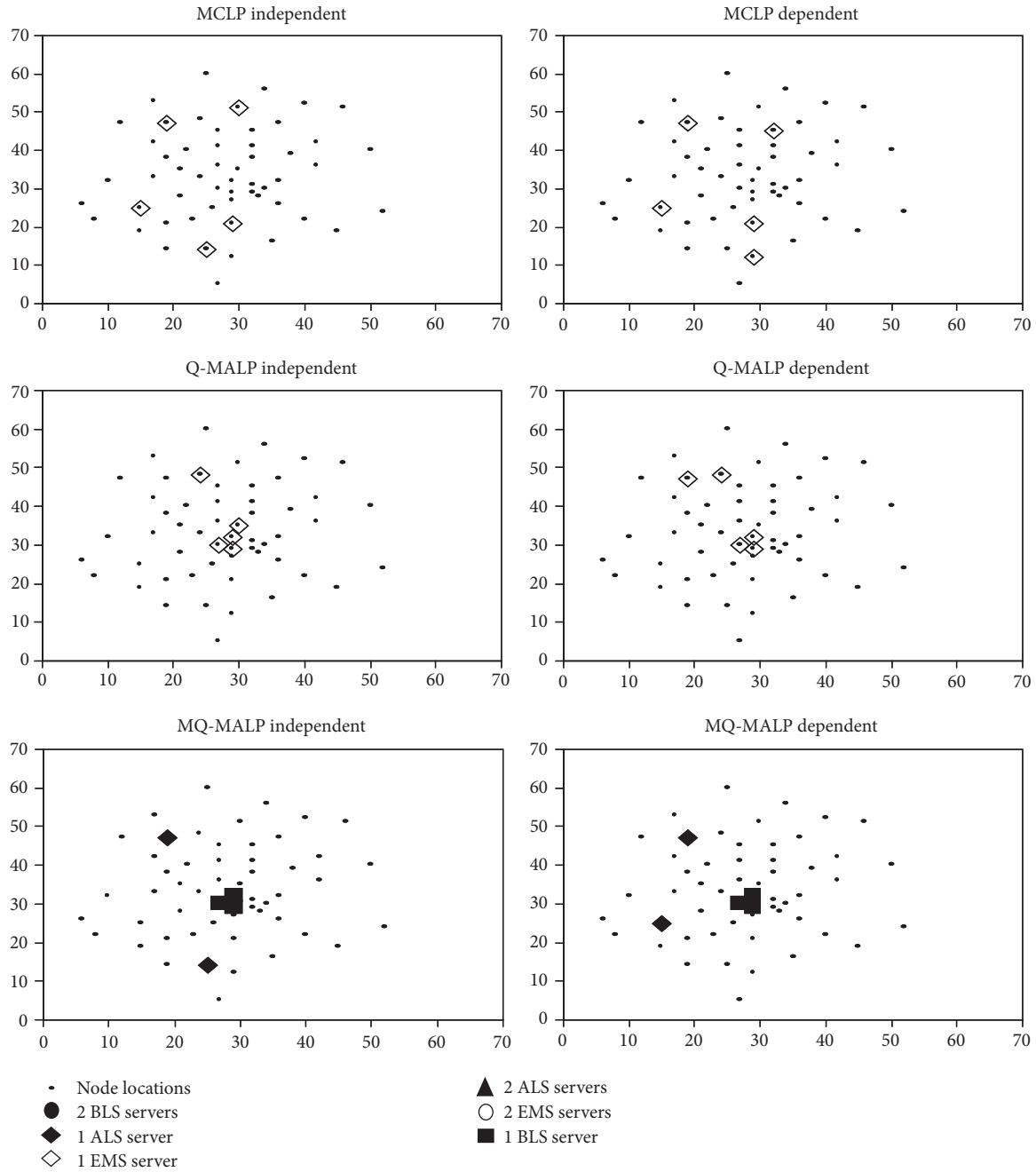


FIGURE 8: Spatial Distribution of Server Node Location of Optimal Solutions of MCLP, Q-MALP and MQ-MALP with Independent and dependent travel times Applied to the 55-Node Problem.

TABLE 2: Server location of optimal solution of MCLP, Q-MALP, and MQ-MALP with independent and dependent travel times applied to the 55-node problem.

Model	Travel time	Node location		
		ALS	BLS	EMS
MCLP	Independent			10, 16, 17, 20, 36
MCLP	Dependent			10, 17, 21, 32, 36
Q-MALP	Independent			2, 4, 8, 22, 42
Q-MALP	Dependent			2, 4, 17, 22, 42
MQ-MALP	Independent	17, 20	2, 4, 42	
MQ-MALP	Dependent	17, 36	2, 4, 42	



satisfying some efficiency and equity requirements. Although the implementation of such a regimen is straightforward given the original formulation, computations time may not allow for a great number of iterations. Therefore, a second tier could be added following the Monte Carlo approach that takes both the feasibility of a suggested solution into account as well as the probability distribution of facility nodes being “hit” as optimal. Finally, the importance of some areas may be different than other, such as area where schools are located. To address the importance of those nodes perhaps in the formulation, certain weightage can be adapted to the nodes.

## Conflicts of Interest

The authors declare that there are no conflicts of interest regarding the publication of this paper.

## References

- [1] M. J. Burt and J. S. Dyer, “Estimation of travel times in multiple mode systems,” *Operational Research Quarterly*, vol. 22, no. 2, pp. 155–163, 1974.
- [2] C. Kenneth and J. P. Jarvis, “Estimating the probability distribution of travel times for urban emergency service systems,” *Operations Research*, vol. 27, no. 1, pp. 199–203, 1979.
- [3] C. Toregas, R. Swain, C. ReVelle, and L. Bergman, “The location of emergency service facilities,” *Operations Research*, vol. 19, no. 6, pp. 1363–1373, 1971.
- [4] R. L. Church and C. S. ReVelle, “The maximal covering location problem,” *Papers of the Regional Science Association*, vol. 32, pp. 101–118, 1974.
- [5] M. Gendreau, “Solving an ambulance location model by tabu search,” *Location Science*, vol. 5, no. 2, pp. 75–88, 1997.
- [6] F. Boras and J. T. Pastor, “The ex-post evaluation of the minimum local reliability level: an enhanced probabilistic location set covering model,” *Annals of Operation Research*, vol. 111, no. 2002, pp. 51–74, 2002.
- [7] S. A. Snyder and R. G. Haight, “Application of the maximal covering location problem to habitat reserve site selection: a review,” *International Regional Science Review*, vol. 39, no. 1, pp. 28–47, 2016.
- [8] A. T. Murray, “Maximal coverage location problem: impacts, significance, and evolution,” *International Regional Science Review*, vol. 39, no. 1, pp. 5–27, 2016.
- [9] G. Laporte, F. X. Louveaux, F. Semet, and A. Thirion, “Application of the double standard model for ambulance location,” in *Innovations in Distribution Logistics*, vol. 619 of *Lecture Notes in Economics and Mathematical Systems*, pp. 235–249, Springer, Berlin, Germany, 2009.
- [10] X. Li, Z. Zhao, X. Zhu, and T. Wyatt, “Covering models and optimization techniques for emergency response facility location and planning: a review,” *Mathematical Methods of Operations Research*, vol. 74, no. 3, pp. 281–310, 2011.
- [11] E. Erkut, A. Ingolfsson, T. Sim, and G. Erdoğan, “Computational comparison of five maximal covering models for locating ambulances,” *Geographical Analysis*, vol. 41, no. 1, pp. 43–65, 2009.
- [12] A. G. Noraida, “Multi-server queuing maximum availability location problem with stochastic travel times,” in *Proceedings of the World Congress on Engineering 2012*, pp. 137–143, 2012.
- [13] R. C. Larson, “A hypercube queuing model for facility location and redistricting in urban emergency services,” *Computers & Operations Research*, vol. 1, no. 1, pp. 67–95, 1974.
- [14] R. C. Larson, “Approximating the performance of urban emergency service systems,” *Operations Research*, vol. 23, no. 5, pp. 845–868, 1975.
- [15] J. P. Jarvis, “Approximating the equilibrium behavior of multi-server loss systems,” *Management Science*, vol. 32, no. 2, pp. 235–239, 1985.
- [16] J. Goldberg and F. Szidarovszky, “Methods for solving nonlinear equations used in evaluating emergency vehicle busy probabilities,” *Operations Research*, vol. 39, no. 6, pp. 903–916, 1991.
- [17] A. Ingolfsson, S. Budge, and E. Erkut, “Optimal ambulance location with random delays and travel times,” *Health Care Management Science*, vol. 11, no. 3, pp. 262–274, 2008.
- [18] E. S. Savas, “Simulation and cost-effectiveness analysis of New York’s emergency ambulance service,” *Management Science*, vol. 15, no. 12, pp. 608–627, 1969.
- [19] J. A. Fitzsimmons, “A methodology for emergency ambulance deployment,” *Management Science*, vol. 19, no. 6, pp. 627–636, 1973.
- [20] C. Swoveland, D. Uyeno, I. Vertinsky, and R. Vickson, “Ambulance location: a probabilistic enumeration approach,” *Management Science*, vol. 20, no. 4, pp. 686–698, 1973.
- [21] G. N. Berlin and J. C. Liebman, “Mathematical analysis of emergency ambulance location,” *Socio-Economic Planning Sciences*, vol. 8, no. 6, pp. 323–328, 1974.
- [22] D. H. Uyeno and C. Seeberg, “A practical methodology for ambulance location,” *Simulation*, vol. 43, no. 2, pp. 79–87, 1984.
- [23] J. Goldberg, R. Dietrich, J. Chen, T. Valenzuela, and E. Criss, “A simulation model for evaluating a set of emergency vehicle base location: deployment, validation and usage,” *Socio-Economic Planning Sciences*, vol. 24, pp. 125–141, 1990.
- [24] S. Henderson and A. Mason, “Ambulance service planning: simulation and data visualisation,” in *Operations Research and Health Care: A Handbook of Methods and Applications*, vol. 70, pp. 77–102, Springer, 2004.
- [25] E. Gunes and R. Szechtman, “A simulation model of a helicopter ambulance service,” in *Proceedings of the Winter Simulation Conference*, pp. 951–957, December 2005.
- [26] R. Aringhieri, G. Carello, and D. Morale, “Ambulance location through optimization and simulation?: the case of Milano urban area,” in *Proceedings of the the 38th Annual Conference of the Italian Operation Research Society Optimization and Decision Sciences*, pp. 1–29, 2007.
- [27] S. L. Hakimi, “Optimum locations of switching centers and the absolute centers and medians of a graph,” *Operations Research*, vol. 12, no. 3, pp. 450–459, 1964.
- [28] M. Dzator and J. Dzator, “An effective heuristic for the P-median problem with application to ambulance location,” *OPSEARCH*, vol. 50, no. 1, pp. 60–74, 2013.
- [29] C. ReVelle, D. Bigman, D. Schilling, J. Cohon, and R. Church, “Facility location: a review of context-free and EMS models,” *Health Services Research*, vol. 12, no. 2, pp. 129–146, 1977.
- [30] D. Eaton, H. Sanchez, R. Lantigua, and J. Morgan, “Determining ambulance deployment in Santo Domingo, Dominican republic,” *Journal of Operational Research Society*, vol. 37, pp. 113–126, 1986.
- [31] M. Dessouky, “Rapid distribution of medical supplies,” in *Patient Flow: Reducing Delay in Healthcare Delivery*, vol. 91, pp. 309–339, Springer, Boston, Mass, USA, 2006.

- [32] H. Jia, F. Ordonez, and M. M. Dessouky, "A modeling framework for facility location of medical service for large-scale emergency," *IIE Transactions*, vol. 39, no. 1, pp. 35–41, 2007.
- [33] O. Berman, Z. Drezner, and D. Krass, "Discrete cooperative covering problems," *Journal of the Operational Research Society*, vol. 62, no. 11, pp. 2002–2012, 2011.
- [34] G. Alexandris and I. Giannikos, "A new model for maximal coverage exploiting GIS capabilities," *European Journal of Operational Research*, vol. 202, no. 2, pp. 328–338, 2010.
- [35] V. Schmid and K. F. Doerner, "Ambulance location and relocation problems with time-dependent travel times," *European Journal of Operational Research*, vol. 207, no. 3, pp. 1293–1303, 2010.
- [36] K. F. Doerner, W. J. Gutjahr, R. F. Hartl, M. Karall, and M. Reimann, "Heuristic solution of an extended double-coverage ambulance location problem for Austria," *Central European Journal of Operations Research*, vol. 13, no. 4, pp. 325–340, 2005.
- [37] M. S. Daskin, "A maximum expected covering location model: formulation, properties and heuristic solution," *Transportation Science*, vol. 17, no. 1, pp. 48–70, 1983.
- [38] C. ReVelle and K. Hogan, "The maximum availability location problem," *Transportation Science*, vol. 23, no. 3, pp. 192–200, 1989.
- [39] G. Bianchi and R. L. Church, "A hybrid fleet model for emergency medical service system design," *Social Science & Medicine*, vol. 26, no. 1, pp. 163–171, 1988.
- [40] R. Batta, J. M. Dolan, and N. N. Krishnamurthy, "Maximal expected covering location problem. Revisited," *Transportation Science*, vol. 23, no. 4, pp. 277–287, 1989.
- [41] J. F. Repede and J. J. Bernardo, "Developing and validating a decision support system for locating emergency medical vehicles in Louisville, Kentucky," *European Journal of Operational Research*, vol. 75, no. 3, pp. 567–581, 1994.
- [42] V. Marianov and C. ReVelle, "The queueing probabilistic location set covering problem and some extensions," *Socio-Economic Planning Sciences*, vol. 28, pp. 167–178, 1994.
- [43] V. Marianov and C. ReVelle, "The queueing maximal availability location problem: a model for the siting of emergency vehicles," *European Journal of Operational Research*, vol. 93, no. 1, pp. 110–120, 1996.
- [44] P. B. Mirchandani and A. R. Odoni, "Locations of medians on stochastic networks," *Transportation Science*, vol. 13, no. 2, pp. 85–97, 1979.
- [45] M. S. Daskin, "Location, dispatching and routing models for emergency services with stochastic travel times," in *Spatial Analysis and Location-Allocation Models*, A. Ghosh and G. Rushton, Eds., pp. 224–265, Van Nostrand Reinhold Co., New York, NY, USA, 1987.
- [46] M. S. Daskin, "Application of an expected covering model to EMS system design," *Decision Sciences*, vol. 13, no. 3, pp. 416–439, 1982.

