Computers & Industrial Engineering 72 (2014) 12–23

Contents lists available at ScienceDirect

Computers & Industrial Engineering

journal homepage: www.elsevier.com/locate/caie

# A simulation optimization framework for ambulance deployment and relocation problems

Lu Zhen [a,*], Kai Wang [a], Hongtao Hu [b], Daofang Chang [b]

[a] School of Management, Shanghai University, 99 Shangda Road, Shanghai 200444, China
[b] Logistics Engineering College, Shanghai Maritime University, 1550 Haigang Avenue, Shanghai 201306, China

ARTICLE INFO

*Article history:*
Received 7 August 2013
Received in revised form 4 March 2014
Accepted 6 March 2014
Available online 15 March 2014

*Keywords:*
Ambulance deployment
Ambulance relocation
Simulation optimization
Genetic algorithm

ABSTRACT

This paper studies ambulance deployment and relocation problems, which are two of the core decisions faced by emergency medical service control centers in metropolis. The challenge in the problems is to estimate the operational performance of a deployment plan under stochastic environments. More specifically, the stochastic and dynamic nature of request arrivals, fulfillment processes, and complex traffic conditions as well as the time-dependent spatial patterns of some parameters complicate the decisions in the problems. This paper proposes a simulation optimization method that enables evaluating the operational performance of deployment plans through a detailed simulation model. For guiding the search process in the simulation optimization method, the genetic algorithm is employed in this study. On the basis of the deployment decisions, a mathematical model on ambulance relocation is also proposed for adapting to the dynamic changing environments along the time. To illustrate the proposed method's usage in practice, a demo example about its application in Shanghai is given. Some numerical experiments are also performed to validate the effectiveness and the efficiency of the proposed methods.


## 1. Introduction

Rising costs of medical equipments, increasing call volumes, and worsening traffic conditions in metropolis make emergency medical service control centers face increasing pressure so as to meet performance targets. The service control centers are supposed to locate a proper number of ambulances in some bases (waiting locations) so that medical service requesters can be reached in a time efficient manner. Uneven distribution of population in the city makes the ambulances should not be evenly deployed in the bases. A medical service control center needs to decide how many ambulances should be deployed in all the waiting locations, respectively. This decision making process is in a dynamic environment where the spatial distribution of potential requesters are changing along the time, and the spatial patterns of traffic situations in a city are also different in peak hours and off-peak hours. The ambulance deployment decision is also in a stochastic environment where the request calls arrive at the control center in a random manner; the travel time for a certain journey may contain randomness; the service time at the request calls' scenes and hospitals is also uncertain. The above mentioned dynamic and stochastic nature of the request arrivals and ambulance fulfillment processes as well as the environments complicates the ambulance deployment decision.

The challenge in ambulance deployment decision is to estimate the operational performance of a deployment plan. Simulation optimization method is a proper way that can enable assessing the operational performance of deployment plans through a detailed simulation model, which can capture multiple sources of uncertainties. Therefore this paper makes an explorative study on ambulance deployment problem by using a simulation optimization methodology.

Moreover, the ambulance deployment decision is not constant and static; it should be dynamically changing along the time because the input data for deployment decisions at different time intervals are varying. Thus, there are ambulance relocation processes between two consecutive time intervals. The relocation decision is based on the optimal deployment decisions in two consecutive time intervals. Therefore, besides the ambulance deployment problem, this paper also investigates the ambulance relocation problem, so that the results of this study can supply a potentially useful decision support tool on intelligent ambulance scheduling for emergency medical service providers.

The remainder of this paper is organized as follows. Section 2 is the literature review. Problem backgrounds of the ambulance deployment and relocation problems are elaborated in Section 3. Then the detailed introductions about the proposed simulation

* Corresponding author. Tel.: +86 21 66134237.
*E-mail address:* lzhen.sh@gmail.com (L. Zhen).

http://dx.doi.org/10.1016/j.cie.2014.03.008
0360-8352/

optimization framework and some key components are given in Section 4. Section 5 illustrates a demo example of the ambulance deployment and relocation problem applied in Shanghai. Some numerical experiments are performed in Section 6 for a further investigation on the proposed method. Closing remarks and conclusions are outlined in the last section.

## 2. Related works

Most studies on ambulance deployment or location problems are based on a minimal covering model (Toregas, Swain, ReVelle, & Bergman, 1971), which tries to minimize the number of ambulances necessary so as to cover all demand points, and a maximal covering model (Church & ReVelle, 1974), which tries to maximize the total demand covered given a feet of fixed size. Then based on the above models, some new models were proposed to consider the possibility that ambulances may be unavailable and some demand points may not be covered. For example, the double standard model (DSM) was proposed by Gendreau, Laporte, and Semet (1997); the probability model for unavailable ambulance was developed by Daskin (1983). The DSM model was applied to the data coming from the eight rural provinces in Austria (Doerner, Gutjahr, Hartl, Karall, & Reimann, 2005). Gendreau, Laporte, and Semet (2001) also extended their model into a dynamic environment to take advantage of the available time between consecutive calls by anticipating future decisions on the deployment of the fleet. The DSM was extended from single period to multiple periods (Schmid & Doerner, 2010). Some scholars used integer programming models to study the real-time ambulance redeployment problems (Brotcorne, Laporte, & Semet, 2003; Gendreau, Laporte, & Semet, 2006; Kolesar & Walker, 1974). The objectives of these integer programming models are mainly formulated from two perspectives, i.e., the backup coverage for future calls, and the relocation cost of ambulances. Solving these models is usually time-consuming as they need to solve an optimization sub-problem every time a decision is made. Therefore, decision makers usually resort to a parallel computing environment for implementing a real-time system. Shariat-Mohaymany, Babaei, Moadi, and Amiripour (2012) proposed two reliability-based linear models for optimal location of ambulances. Some studies considered the randomness in the system explicitly, either through a dynamic programming formulation or through heuristic approaches. Berman (1981a, 1981b, 1981c) proposed some frameworks by using the dynamic programming approaches for the ambulance redeployment problem. These papers follow an exact dynamic programming formulation, thus the formulation is tractable only in oversimplified versions of the problem with few vehicles and small transportation networks. Andersson and Vaerband (2007) investigated the ambulance deployment decision by using a preparedness function that essentially measures the capability of a certain ambulance configuration to cover future calls. The preparedness function is similar with the value function in a dynamic program, which assesses the impact of a current decision on the future evolution of the system. However, the way of the preparedness function belongs to some sort of heuristic methods in nature. The simulation optimization method was used in the personnel deployment at an emergency department healthcare unit (Ahmed & Alkhamis, 2009). Iannoni, Morabito, and Saydam (2009) studied how to optimize the locations of ambulance bases on highways. Zhang, Puterman, Nelson, and Atkins (2012) integrated the simulation optimization method with the demographic and survival analysis. A decision support system was developed for setting long-term care capacity planning. Underwood, Zhang, Denton, Shah, and Inman (2012) proposed a genetic algorithm based simulation optimization method to design PSA screening policies. Some interesting findings and policy recommendations were obtained from their study.

When compared with the above described models, our method provides some merits. In contrast to the models that are based on either integer programming or dynamic programming methodologies, our method can captures the more complex and random evolution of the system over time, and the stochastic nature of request arrivals, fulfillment processes, and complex traffic conditions as well as the time-dependent spatial patterns of some parameters, all of which complicate the decisions in the problem of this study. This paper proposes a simulation optimization method by involving the GA meta-heuristic and a simulator. Moreover, an ambulance relocation model is also proposed for adapting to the dynamic changing environments along the time.

## 3. Problem background

This paper studies the ambulance deployment problem and the ambulance relocation problem. (1) The ambulance deployment problem origins from the uneven spatial distribution of potential medical service requests. The problem is concerned with how to deploy a given amount of ambulances among the waiting stations in the city so as to optimize the service performance, which can be quantified by some indicators. (2) The ambulance relocation problem exists because the information about the request arrivals and traffic conditions is time dependant. This relocation problem is on the basis of the previous deployment problem. The optimal ambulance deployment decision depends on the parameter setting about the request distributions and traffic conditions, which are influenced by the time. Given two optimal deployment plans that belong to two consecutive time intervals, the ambulance relocation problem is concerned with how to relocate ambulances so as to minimize the total cost (distance) of relocation routes between the ambulance waiting stations in the city.

A request for medical service usually arrives by phone and is answered by a dispatcher, who inputs the information by asking some predefined questions, and determines the priority of the request calls. When a request becomes known to the dispatching system, the dispatcher checks with available ambulances and assigns the request to an ambulance. In this process, there are some possible criteria for the decision. For example, the nearest available ambulance may be chosen; or some other criteria can also be embedded in the decision support system for dispatching. The time span between the request arrival and the setting off of a chosen ambulance should be the shorter the better. This time span usually contains the time necessary for querying some information about the actual incident, negotiation with ambulance drivers, and the setup time for the crew to get ready. When the ambulance arrives at the patient location (i.e., the call's scene), the crew of the ambulance may need to take some first-aid measures for the patient. When the service is completed, the ambulance takes the patient to a hospital. Sometimes, the patient may appoint a hospital in case the incident is not very urgent; otherwise, the ambulance transport the patient to a hospital, which has the available resources for treating the patient and is the nearest from the call's scene. When the ambulance arrives at the hospital, the crew starts to unload the patient and deliver him (or her) to the corresponding department in the hospital. Then the whole service for the request is finally completed. The ambulance becomes idle (available) again and goes to a waiting location if there are not new requests assigned to it; otherwise, the ambulance needs to set off to the next call's scene immediately. Fig. 1 shows the above process.

The above elaborates the operational level process of ambulance dispatching. On the tactical level, the control center of ambulances needs to make a decision about how to deploy a given amount of ambulances among the waiting stations in the city

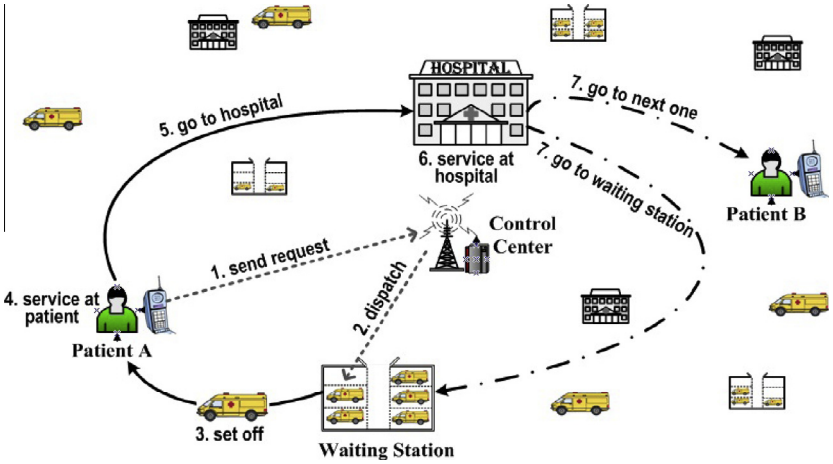



**Fig. 1.** The problem background of the ambulance scheduling.

so as to optimize the service performance, which can be evaluated by the average response time for all the requests, or some other indicators on the service performance. This decision is made on the basis of the above operational level ambulance traveling and serving processes, which is hard to be described by formulating some formulae because a lot of uncertainties contained in the processes. A simulation model is the best method to capture multiple sources of uncertainties and incorporate adaptive control policies for reflecting the above ambulance traveling and serving processes. This paper combines the simulation and optimization techniques and proposes a simulation optimization framework to determine how many ambulances should be deployed in each waiting station.

The optimal deployment of ambulances depends on the request arrival rates among different areas of the city, i.e., the distribution of the requests, which is changing along the time. Thus the deployment decision of ambulances should not be static. Another tactical level decision about relocating ambulance among their waiting stations also concerns the ambulance control center. So this paper proposes an optimization model for the ambulance relocation problem.

## 4. The simulation optimization framework

### 4.1. General framework

The general concept of a combined simulation optimization (SO) method is illustrated in Fig. 2. A simulator embedded in this SO framework takes a potential solution to the considered optimization problem (i.e., the ambulance deployment) and conducts a simulation to evaluate the solution's performance under a stochastic environment, in which the arrival time of requests, travelling time, servicing time are uncertain. Multiple simulation runs are performed to obtain results of statistical validity (for this study, 400 runs are performed in the simulation component). On the basis of the outcome of the above simulation process, the optimization component searches for alternative solutions and evaluates them again by the aforementioned simulator.

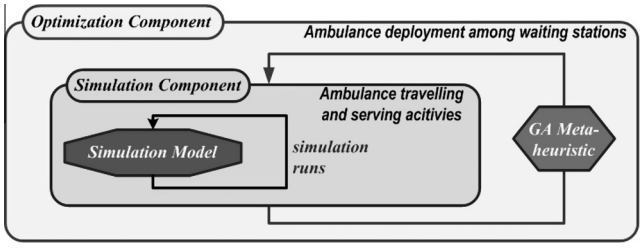



**Fig. 2.** The simulation optimization framework.

For the simulator, we employ a discrete event simulation that simulates the operational level process of ambulance travelling and serving activities in detail. It allows tracing the fulfillment processes of each request with respect to the random arrival time, random travel time from the waiting location to the call's scene, random service time at the call's scene, random traveling time to the hospital, and random service time at the hospital. The simulator can capture the operational level behavior of the ambulances under different dispatching policies. The simulation outputs a operational level performance measure $\widetilde{R}$, which denotes the average response time for all the requests.

For the optimizer, we employ the widely used genetic algorithm (GA) meta-heuristic. The GA allows searching solutions by imitating the natural evolution process, such as inheritance, mutation, selection, and crossover. The optimizer reconfigures a current deployment of ambulances by finding alternative values for decision variables $\boldsymbol{X} = \{x_1, x_2, \ldots, x_N\}$, which denote the numbers of ambulances that should be deployed in the $N$ waiting stations (bases), respectively.

Here we describe the flow the SO procedure for ambulance deployment as follows.

**SIMULATION OPTIMIZATION PROCEDURE:**
*//the outer framework is based on **GA**, objective evaluation for chromosomes is based on simulation*

**Initialize** a population of twenty chromosomes. *// every chromosome is an integer array, in which each item denotes the amount of ambulance deployed in a base. The size of the array equals to the number of bases in the problem case.*
**Evaluate** each chromosome's objective by a simulator **SIMU(X)**. *// X is a chromosome, which is denoted by an integer array $\boldsymbol{X} = \{x_1, x_2, \ldots, x_N\}$. The objective is the average of the results based on 400 randomly generated instances running on the simulator. The details of the simulator SIMU(X) will be elaborated in the next procedure.*
**While** (termination condition does not reach) *//the termination condition: the best objective has not been improved for 100 successive iterations.*

**{**
**Sort** the twenty chromosomes by the increasing order of their objectives.
The best 15% of chromosomes **evolve** to the next iteration without changes.
The worst 15% of chromosomes are handled by **mutation** operator.
The moderate 70% of chromosomes are handled by **crossover** operator.
For the newly obtained population, **evaluate** each chromosome's objective by **SIMU(*X*)**.
**Record** the objective of the best chromosome in the current iteration.
**Update** the incumbent best solution so far if necessary.
**}**

The simulator embedded in the above SO procedure is an important component for evaluating the objective for each chromosome (solution). The flowchart of the simulator is illustrated as follows.

**Simu(*X*)**: the simulator for evaluating objective of a chromosome *X*
***// Input**: the chromosome X. **Output**: the objective of the chromosome X, denoted by $OBJ_X$.*

**Input *X***. *// the integer array* $\boldsymbol{X} = \{x_1, x_2, \ldots, x_N\}$ *denotes numbers of ambulances deployed in N bases.*
**Initialize** environmental parameters for simulations. *// the parameters include: coordinates information for all the areas, probability density functions (PDFs) for request arrivals, PDFs for requests emerging in different areas, PDFs for service time, initial locations of ambulances, location of hospitals, bases, the settings on ambulance speed in different areas during different time intervals, etc.*
**Generate 400 instances randomly**. *// each instance includes a series of requests that are randomly generated according to the PDFs of requests arrival and distributions and the PDFs of service time for the requests.*
*idx* = 1. *// 'idx' denotes the instance index.*
**While** (*idx* <= 400) *//the simulation runs for 400 times*
{
**For** all the generated requests in an instance. *// according to the requests' chronological order.*
**{** *// the request in each iteration is denoted by r*
**Check** all the ambulances' **status** at the current time.
**Calculate** the time for all the available ambulances going to the request *r*'s location.
*// consider the ambulance speed varies in different areas during different time intervals.*
**Select** the ambulance with the shortest time, and assign it to the request *r*.
**Update** the assigned ambulance's status as 'occupied'.
The ambulance **prepares** for filling the task. *// setup time*.
The ambulance **sets out** from its current location, **arrives** at the request *r*'s location.
**Record** the response time ($t_r$) between the request ($r$) advent and the ambulance arrival at the request's location.
The ambulance **serves** for a time interval, which is randomly generated in the instance.
The ambulance goes to the nearest hospital. *// the travel time is also estimated by considering the ambulance speed varies in different areas during different time intervals.*
The ambulance **stays** for a time interval in the hospital, and then goes back to its base.
**Update** the ambulance's status as 'available'.
}
**Calculate** the average response time for all the requests. $T(idx) = \mathrm{Avg}_{\forall r} t_r$.
*idx* ++.
}
**Calculate** the average value of $T(idx)$ for all the instances, $OBJ_X = \frac{1}{400}\sum_{idx=1}^{400} T(idx)$.
**Return** $OBJ_X$.

As aforementioned in the above procedure, the simulation for evaluate each chromosome is based on 400 randomly generated instances. The number of the replications is mainly determined according to numerical experiments for the studied application problem. Nine series of experiments are performed by using the SO method. For each series, the number of instances run by the simulation is set as: 10, 50, 100, 200, 300, 400, 500, 600, and 700, respectively. For each series, the SO procedure is executed for ten times. For each series, the average (AVG) and the standard deviation (SD) of the ten obtained objectives are recorded and shown in Table 1. The average CPU time for running the SO procedure is also recorded. For each series, the ten obtained results (i.e., ambulance deployment plans) may be a uniform solution, or may have several different solutions. For each series, the number of identical solutions (#SLT) is also recorded.

From Table 1, it is observed that the standard deviation (SD) values decrease and the CPU time increases with the number of replications growing. When the number exceeds 400, the SD values do not decrease evidently, and their AVG values are very close to each other. For the number of identical solutions (#SLT), it is noted that the increasing of the number of simulation replications has no influence on the final decision of ambulance deployment when the number exceeds 400. It means that from the perspective of decision makers who make plan of ambulance deployment, 400 replications of simulation are enough for the whole SO procedure. With considering all of the above factors (e.g., SD of objectives, uniform solution, CPU time), 400 replications are a proper setting for the simulation model so as to ensure the SO procedure can derive good and unbiased results.

The following Sections 4.2 and 4.3 address the details on the simulation component, and optimization component, respectively.

### 4.2. Simulation component

Before performing a simulation, a map with coordinate plane should be defined for Shanghai. Hospitals and ambulance waiting stations are located in the coordinate plane in advance. In addition, we should define the speed ranges for ambulances in specific areas

**Table 1**
The results of SO under different numbers of simulation replications.

| Num of replications | AVG | SD | CPU time (s) | #SLT |
|---|---|---|---|---|
| 10 | 17.17 | 4.84 | 13 | 10 |
| 50 | 14.33 | 3.71 | 61 | 10 |
| 100 | 13.36 | 1.73 | 129 | 7 |
| 200 | 12.86 | 1.27 | 256 | 6 |
| 300 | 12.56 | 0.97 | 328 | 3 |
| 400 | 12.44 | 0.50 | 445 | 1 |
| 500 | 12.44 | 0.44 | 578 | 1 |
| 600 | 12.45 | 0.34 | 712 | 1 |
| 700 | 12.44 | 0.34 | 798 | 1 |

of the city, i.e., the spatial patterns of traffic situations in peak hours and off-peak hours.

The random requests generation follows the Poisson distribution. It is assumed that the average arrival rate is one request per minute. For the setup time of ambulances, the service time at requester calls' scenes, and the service time at hospitals, we assume they follow the negative exponential distribution with the average service time (or setup time) is two minutes, ten minutes, and three minutes, respectively.

The simulator generates about 1500 requests by following the Poisson distribution. Then the simulator needs to locate the generated requests in specific locations according to the distribution of requests densities among different areas in the city. It should be mentioned that the distribution varies with respect to time intervals (e.g., 0–6 am, 6–10 am, 10 am–16 pm, 16–19 pm, 19–24 pm, in the demo example in Section 5). When a request 'emerges' in a location, a search procedure on all the available ambulances will be triggered. An ambulance's status is availability if it is in a waiting station or is traveling from a hospital to a waiting station. If none of ambulances is available, the ambulance that will be available soon and can arrive at the call's scene in the earliest time will be assigned with the request.

In the simulation, the travel time is not assumed to follow probability distributions. When calculating the ambulance travel time between two given locations, the simulator first determines the areas (e.g., $A_1, A_2, \ldots, A_n$) that the journey between the two locations will pass, and finds the velocities (e.g., $v_1, v_2, \ldots, v_n$) that relate with the above obtained areas. Here we assume ambulances travel through each journey with uniform velocity, and the velocities in each area during each time interval are known in advance, which can be estimated according to the historical data of traffic conditions in the city. Then the travel time between the two locations (denoted by $t$) is calculated according to the lengths of the $n$ journey segments (e.g., $d_1, d_2, \ldots, d_n$); the formula for calculating $t$ is as follows: $t = d_1/v_1 + d_2/v_2 + \ldots + d_n/v_n$. This formula is used in the estimation of the travel times from the current location of an ambulance to calls' scenes, and then to hospitals; it is also used in the decision process on choosing an ambulance to fulfill the task.

For the hospital adoption, the chosen hospital (destination) should have the competency for treating the patient's illness and is the nearest to the call's scene, if the patient has not appointed a specific hospital; otherwise the ambulance will go to the hospital appointed by the patient.

When all the time points of events (e.g., requests emerge; ambulance set off; arrival at call's scene; arrival at hospital; etc.) are determined, the indicator of performance measures (e.g., response time of requests) can be calculated.

An interface of the simulation component is shown in Fig. 3. The simulator is developed by Visual Studio C# 2008. The random numbers are generated through the SPSS tool and then are imported to the simulation program in the C# 2008. The SPSS tool is widely used for generating true random numbers so as to derive unbiased results in the simulation runs.

### 4.3. Optimization component

In the optimization component, the widely used genetic algorithm (GA) is employed to search for alternative solutions which are assessed by the aforementioned simulator. The detailed description of the GA is as follows:

*Chromosome definition:* A chromosome is designed as an array. The amount of elements contained in the array equals to the amount of waiting stations (bases). The value of each element denotes the amount of ambulances deployed in each waiting station.

*Initialization of population:* The size of the population is twenty. The GA embedded in this optimization component imitates natural evolution of a population with twenty individual solutions, which are initially generated by allocating ambulances to waiting stations in a random manner.

*Fitness evaluations of chromosomes:* In each iteration of the evolution process, all the individual solutions (chromosomes) are evaluated by 400 runs of simulations. The detailed description of the simulator is addressed in Section 4.2.

The main evolution process of the chromosome population is shown in Fig. 4.

As shown in Fig. 4, the twenty chromosomes are sorted according to their fitness values. Then, the best fifteen percent of the chromosomes evolve to the next iteration without any change; the worst fifteen percent of the chromosomes are mutated; and others are crossed over by pairs of the chromosomes. The crossover and the mutation processes are as follows.

*Crossover:* The seventy percent of the chromosomes (i.e., the middle part of the population as shown in Fig. 4) are treated by crossover operator. The fourteen chromosomes are randomly divided into seven pairs. For each pair of the two chromosomes, the crossover operator randomly determines a crossover point within the two chromosomes. As shown in Fig. 5, two new chromosomes are generated after the crossover operation.

*Mutation:* The worst fifteen percent of the chromosomes are changed by mutation operator, which randomly choose an element of an individual's chromosome and change it.

It should be mentioned that there is another process ('Normalization' in Fig. 5) for making a proper modification on the newly generated but infeasible chromosomes after the crossover and the mutation operations. More specifically, if total number of ambulances is $N$, and the sum of the numbers in all the elements of a chromosome is $N'$, all the numbers in the chromosome are adjusted by multiplying a ratio $N/N'$ and then are rounded to integers such that the sum of all the elements equals to $N$ after the 'Normalization' process.

*Terminating condition:* The above mentioned evolution process is repeated until a termination condition has been reached. The terminating condition used in this optimization component is: the best individual's objective value has no longer been updated in one hundred successive iterations.

### 4.4. Model for ambulance relocation

By using the above mentioned SO procedure, we can obtain the ambulance deployment plans in several a certain time interval. During different time intervals, their environmental parameter settings will change. The setting includes the ambulance velocities among areas, the probability distributions of requests emerging among areas, etc. For one day, we assume it is divided into $T$ time intervals. We perform the SO procedure (by $T$ times) for the $T$ time intervals under their environmental parameter settings. For the two consecutive time intervals, their ambulance deployment plans (i.e., the results by the SO procedures) are different from each other. Here a mathematical model is proposed for decision on the ambulance relocation between the two consecutive time intervals.

In the time interval $t$, the numbers of ambulances deployed in the $N$ waiting stations are $n_i^t$, $i \in B$, here $B$ is the set of all the waiting stations (bases) of ambulances. The parameters $n_i^t$ are obtained by the above simulation optimization framework. The parameters $n_i^{t+1}$, $i \in B$, denote the deployment plan in the next time interval $t+1$; they are also obtained by the simulation optimization framework. Then, the number of ambulances which should be relocated from the waiting station $i$ to other sites is denoted by $(n_i^t - n_i^{t+1})^+$; and the number of ambulances which should be relocated into the waiting station $i$ from other sites is denoted by $(n_i^{t+1} - n_i^t)^+$. For the cost parameters, $c_{ij}$ is about the transportation cost for an

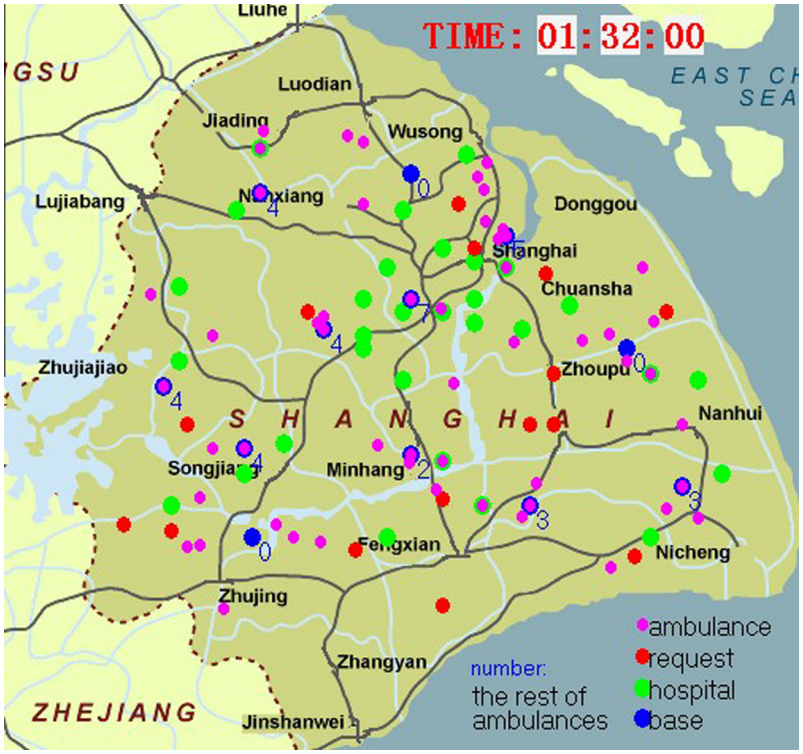



**Fig. 3.** An interface of the simulation component.

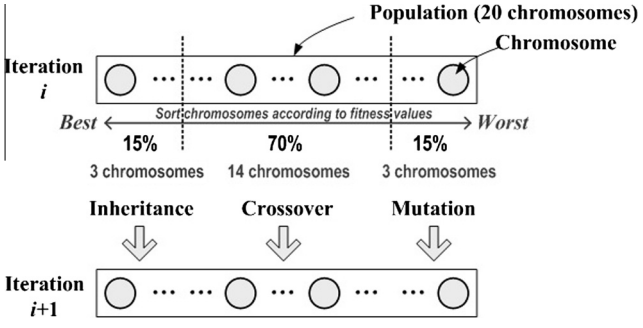



**Fig. 4.** The evolution process of the chromosome population.

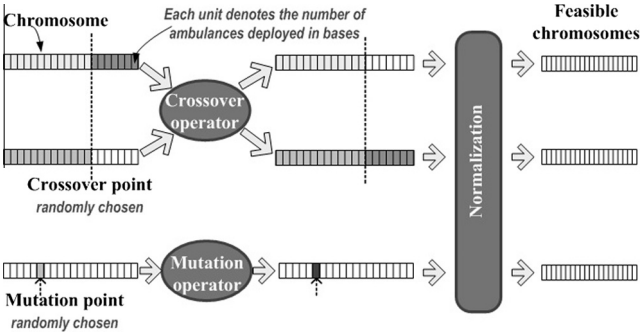



**Fig. 5.** The crossover and mutation processes.

ambulance relocating from the waiting station $i$ to the waiting station $j$; $f_{ij}$ is about the fixed cost for ambulance relocation between two waiting stations $i$ and $j$. The consideration on the fixed cost can avoid the situations as follows. For example, $c_{ac} = C_{ab} + C_{bc}$, if the fixed cost is waived, a relocation plan that relocates some ambulances from waiting station $a$ to $b$, and then relocates them to $c$, is the same as a relocation plan that relocates ambulances from $a$ to $c$ directly. Actually the latter plan is more convenient than the former plan in reality. Therefore the involvement of fixed cost $f_{ij}$ into the model can avoid the above unrealistic relocation plans.

The decision variables are denoted by $x_{ij}$, $i, j \in B, i \neq j$. They reflect the number of ambulances that are relocated from the waiting station $i$ to the waiting station $j$. Another decision variable is denoted by $y_{ij}$, $i, j \in B, i \neq j$. They are binary variables and reflect whether or not there are ambulances relocated between two waiting stations $i$ and $j$.

On the basis of the above definitions on the parameters and the variables, a mathematical model for ambulance relocation (**M_AR**) is formulated as follows:

$$(\mathbf{M\_AR}) \quad Min \quad Z = \sum_{i \in B} \sum_{i \in B, j \neq i} (c_{ij} x_{ij} + f_{ij} y_{ij}) \tag{1}$$

$$s.t. \quad \sum_{i \in B} x_{ij} = \left(n_j^{t+1} - n_j^t\right)^+ \quad \forall j \in B \tag{2}$$

$$\sum_{jinB} x_{ij} = \left(n_i^t - n_i^{t+1}\right)^+ \quad \forall i \in B \tag{3}$$

$$x_{ij} \leq y_{ij} \cdot M \quad \forall j \in B \tag{4}$$

$$x_{ij} \geq 0 \quad \forall j \in B \tag{5}$$

$$y_{ij} \in \{0, 1\} \quad \forall j \in B \tag{6}$$

In Constraint (4), $M$ is a sufficiently large positive number. Objective (1) minimizes all the transportation cost and the fixed cost incurred in the ambulance relocation process. Constraints (2) and (3) limit the number of ambulances that go out and go in each waiting station. Constraint (4) links the two decision variables $x_{ij}$ and $y_{ij}$. Constraints (5) and (6) define decision variables. The model can be solved by some commercial solver, e.g., LINDO, CPLEX, etc.

## 5. A demo example in Shanghai

This section gives a demo example for the application of the proposed method in Shanghai, which is the largest city by population in China. Shanghai has a population of over 23 million and a land area of about 6340 square kilometers. In such a megalopolis, the medical service call center usually receives a request and set off an ambulance every 1.2 min on average. Facing so many arriving medical service requests, a good decision on ambulance deployment and relocation is very necessary for reducing the average response time. For patients, the first few hours are the best time (golden hours) for giving them some proper treatments. Thus the average response time for all the requests reflects the service level of a city's medical service response system. It is also used as the criterion in the simulation optimization model of the proposed method. In this demo example, 30 hospitals, 12 ambulance bases (waiting stations), and 80 ambulances are considered, which are shown in Fig. 6.

The distributions of potential requests are changing along the time. For example, the request density may be much higher in the urban areas than the outskirts. In evening when people go back from their working or sight-seeing locations in downtown to residential districts in the suburbs, the distributions of potential requests may become different from the situations in the daytime. In this demo example, five time intervals (i.e., 0–6 am, 6–10 am, 10 am–16 pm, 16–19 pm, 19–24 pm) are considered for a day. The distributions of potential requests during these five time intervals are also shown in Fig. 6.

Moreover, the spatial patterns of the traffic situations are also different in peak hours and off-peak hours. Here the traffic situation mainly refers to the average speed of ambulances travelling in some areas. So this demo example considers the factor of time-dependant travelling speeds. According to the realistic situations in Shanghai, the traffic peak hours are usually two time intervals, i.e., 7:30–9:30 am, 16:30–18:30 pm; and the remainder time is the off-peak hours. Fig. 7 illustrates the spatial patterns of the traffic situations in peak and off-peak hours.

Based on the above input data, the proposed simulation optimization method can derive ambulance deployment plans in the five time intervals one by one. Fig. 8 illustrates the evolution process of solutions in the optimization component (i.e., GA). The curve in Fig. 8 is just an example which reflects the results in the time interval of '10 am–16 pm'. It validates the convergence of the GA method used in the simulation optimization framework. After the ninety generations, the result converges to 12.45 min, which is the average response time in the time interval of '10 am–16 pm'.

After the simulation optimization method performs for five times, each of which obtains an optimal deployment plan for one time interval. The results are shown in Table 2 as follows. From the table, it is observed that there are obvious differences between the optimal deployment plans in two consecutive time intervals. This result actually validates the needs for ambulance relocation processes.

Based on the above ambulance deployment plans in different time intervals, the ambulance relocation plan can be obtained by solving the relocation model (**M_AR**), which is elaborated in Section 4.4. The results of the relocation are illustrated in Table 3 as follows.

From the relocation plans in Table 3, we can make a statistic on the number of relocated ambulances in each shift of two consecutive time intervals, i.e., the shifts 'a', 'b', 'c', 'd', 'e', as noted in Table 3. The numbers of the relocated ambulances in these five shifts are 14, 24, 20, 15, and 11, respectively. It implies that the relocation activities are busier in the daytime than nighttime. In the daytime, more than one fourth of the ambulances need to be relocated among bases. Another finding is that the optimal relocation plan is related with all the ambulance bases rather than just focusing on some bases.

## 6. Numerical experiments

Some numerical experiments are performed in this section to investigate the influences of some parameters on the performances so that some managerial implications can be obtained.

### 6.1. Influences of parameters on the average response time

Four series of experiments are conducted to analyze the influence of some parameters on the average response time. The results are shown in Fig. 9.

From Fig. 9(1), it is observed that when the arrival rate exceeds one request per minute, the average response time increases evidently; and the average response time does not decease with the arrival rate deceasing when the rate is less than one request per minute. Thus it implies that one request per minute is a threshold value for the current system configuration, i.e., 80 ambulances in 12 bases.

From Fig. 9(2), it is observed that the number of ambulances influences the average response time obviously when the number is below a threshold. It also implies that given a certain arrival rate of requests, there is no need to maintain a fleet of as many ambulances as possible. For the example (i.e., one request per minute), 70–80 ambulances are enough.

Fig. 9(3) shows the similar implication as the above. The number of ambulance bases impacts the average response time when the number is below a threshold. For the example (i.e., one request per minute), 12 ambulance bases are enough because establishing more bases does not improve the response time evidently.

Fig. 9(4) illustrates the influence of the number of hospitals on the results. Although the calculation of the response time is not related with hospitals, the hospital amount also has significant influence on the average response time.

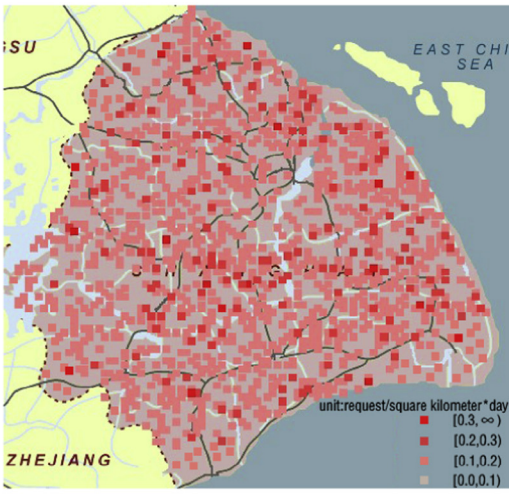
The above experiments also validate the system configuration setting for the demo example of Shanghai. '80 ambulances, 12 bases, 30 hospitals' are a proper combination for Shanghai with about one request every 1.2 min on average.

### 6.2. Influences of parameters on the ambulance deployment

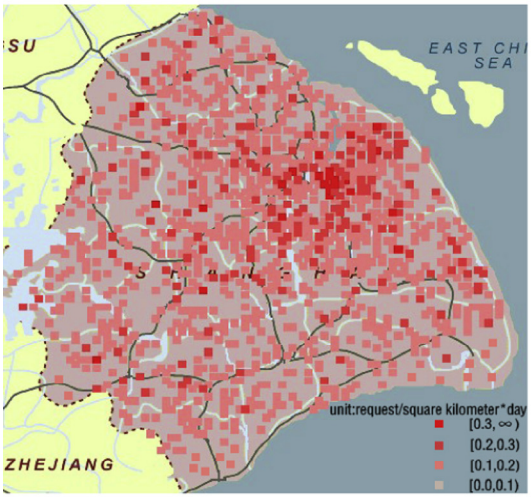
Some experiments are conducted to analyze the influence of parameters on the optimal ambulance deployment plans. Table 4 shows the impact of request arrival rate. In the figure, the standard deviation (SD) and the coefficient of variance (CV) are calculated according to the amounts of ambulances deployed in each base. In Fig. 3, it is observed that the optimal deployment plan becomes more and more uneven with the request arrival rate growing, which is reflected by the decreasing trend of the average time between two consecutive requests in Table 4. It implies that the simulation optimization (SO) method is especially necessary for making a good deployment decision in the situation with high frequent arriving request, rather than just deploying all the ambulances among the bases in a balanced manner. The reason may lie in that under an environment with higher frequent arriving requests, the uneven distributions of population and the uneven densities of traffic congestions among areas will have a more obvious influence on unbalanced deployment of ambulances among bases. Therefore, the proposed SO method is more necessary for the cases with shorter average time between two consecutive requests.

Table 5 shows the impact of ambulance amount on the deployment plan. As the total numbers of ambulances are different in test

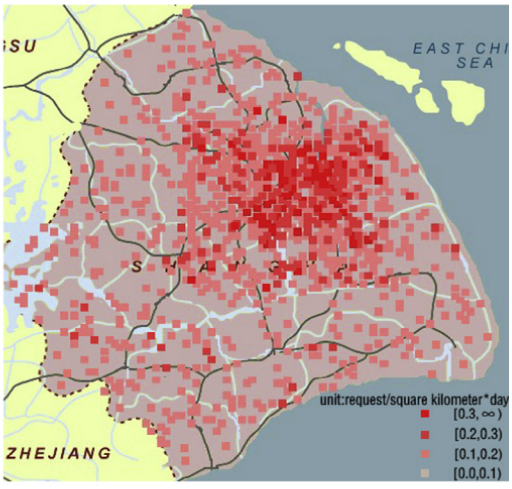

Locations of hospitals and bases



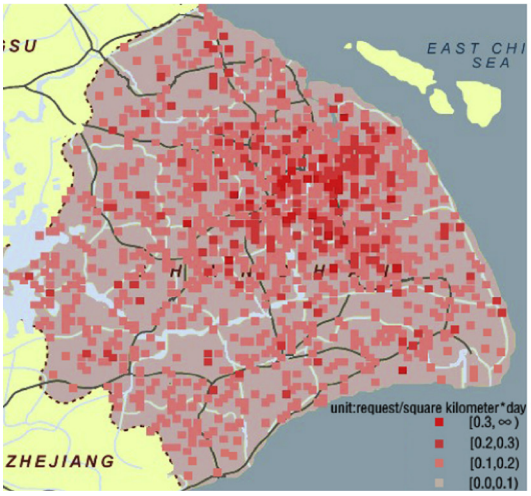

Request distribution in (**0** am – **6** am)



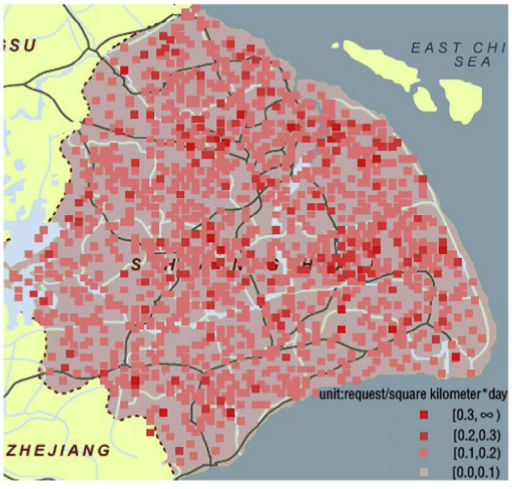

Request distribution in (**6** am – **10** am)




Request distribution in (**10** am – **16** pm)




Request distribution in (**16** pm – **19** pm)




Request distribution in (**19** pm – **24** pm)

**Fig. 6.** Request distributions during different time intervals in Shanghai.

cases, the CV is more proper than the SD to reflect the uneven status of the ambulance deployment among bases. It is observed that when the ambulance amount grows, ambulances are deployed among bases in a more and more balanced way, which is reflected by the decreasing trend of the CV values. The results imply that the SO method is especially necessary for making the deployment decision when there are a limited number of ambulances. The reason for this phenomenon is easily understood. The fewer are the ambulances, it is more necessary to use a well designed methodology to optimize the deployment of these scarce resources rather

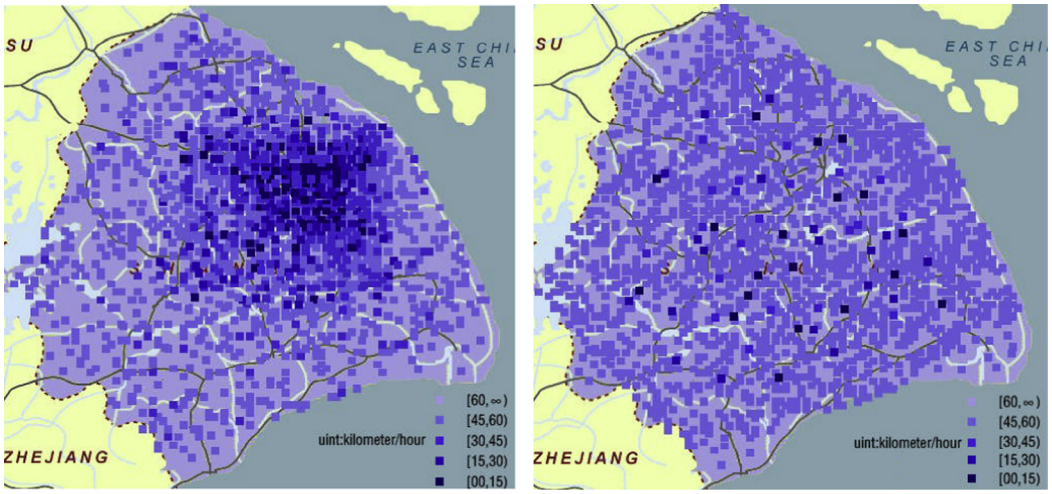



**Fig. 7.** Spatial patterns of the traffic situations in peak and off-peak hours.

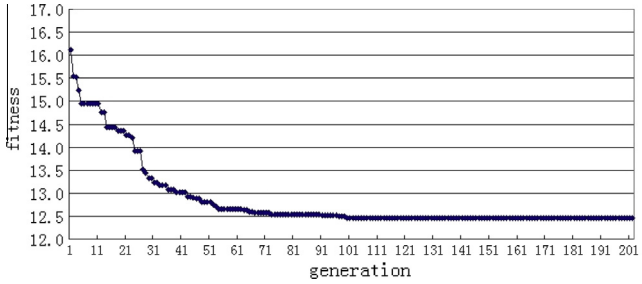



**Fig. 8.** The evolution process of solutions in the optimization component.

than allocating them in a balance manner. Therefore, the proposed SO method is more necessary for the cases with fewer ambulances.

Table 6 shows the impact of ambulance base amount on the deployment plan. It is observed that the base amount has no significant influence on the CV values. The results imply that establishing more bases will not make the ambulances be deployed in a more evenly balanced way. The reason may lie in that an uneven ambulance deployment decision is essentially incurred by the uneven distributions of population and the uneven densities of traffic congestions among areas. These external factors cannot be changed by setting more ambulance bases. In addition, from the CPU time in Tables 5 and 6, it is observed that the amounts of ambulances and bases influence the SO solving time in two opposite manners. The cases with more ambulances or less bases could be solved by SO in a shorter time.

Similar with the results in previous experiments, Table 7 shows the hospital amount also has no significant influence on the SD and CV values. The results imply that involving more hospitals into the system will not make the ambulances be deployed in a more evenly balanced way. The reason is similar as the above mentioned. These external factors (i.e., the uneven distributions of population and the uneven densities of traffic congestions among areas) cannot be changed by establishing more hospitals. Therefore, the uneven degree of ambulance deployment decision is not evidently impacted by the number of hospitals.

### 6.3. Experiments of the ambulance relocations

Some experiments are conducted to analyze how the variation of the request distributions influences the ambulance relocations. To capture and denote the variation degree of the request distributions in a quantitative way, we build a grid system by dividing the city into 53 squares with the size 10 km × 10 km. For each square, we calculate the standard deviation (SD) and the coefficient variance (CV) of the numbers of the requests that emerge in the square during the five time intervals. Then the average of the SD values (and the average of the CV values) for the 53 squares is obtained to reflect the aforementioned variation degree of the request distributions.

As shown in Table 8, experiments are performed under five different variation degrees. For each one of the five cases, the variation of the request distributions in time intervals, and the total movement distance of all the ambulances in their relocation activities are recorded in Table 8. The result validates that the variance degree of request distributions has a significant positive influence on the total movement distance of relocation activities.

**Table 2**
The optimal deployment plans in each time interval.

| Time intervals | Index of waiting | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| *Number stations of deployed (bases) ambulances* | | | | | | | | | | | | |
| 0–6 am | 3 | 4 | 9 | 7 | 3 | 3 | 6 | 9 | 7 | 10 | 8 | 11 |
| 6–10 am | 1 | 3 | 10 | 9 | 8 | 1 | 7 | 3 | 9 | 13 | 7 | 9 |
| 10 am–16 pm | 2 | 1 | 4 | 11 | 1 | 3 | 1 | 7 | 22 | 13 | 9 | 6 |
| 16–19 pm | 5 | 3 | 10 | 8 | 3 | 2 | 3 | 9 | 11 | 9 | 12 | 5 |
| 19–24 pm | 1 | 5 | 7 | 5 | 12 | 1 | 7 | 6 | 6 | 14 | 5 | 11 |

**Table 3**
The optimal relocation plans between two consecutive time intervals.

| From Ambulance Base Index (**origin** of a relocation route) | To Ambulance Base Index (**destination** of a relocation route) | | | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | 1 | 2 | 3 | 4 | 5 | 6 | 7 | 8 | 9 | 10 | 11 | 12 |
| 1 | | 1/d | | 1/a | | | | | | 1/a | | 1/d |
| 2 | | | 1/a<br>1/e | | | | | | | | 2/b | 1/e |
| 3 | | 2/d | | 2/b | 1/d | | | | 4/b | | | |
| 4 | | | 3/c | | 2/d | | | | | | | |
| 5 | | | 1/e | 1/e | | | 2/e | | 7/b | | | |
| 6 | | | | | 2/a | | 1/c | | | | | |
| 7 | | | | | | 2/b | | 4/b | | | | |
| 8 | | | | | 3/a | | 1/a | | 2/a | | | 2/d |
| 9 | | | 3/c | | 2/c<br>1/d | 1/e | 1/c<br>1/d | 2/c | | 1/d | 3/c | |
| 10 | 2/c | 2/c | | | | | | 1/e | | | | 2/e |
| 11 | | | | 1/a | | | | 1/e | | 3/d | | |
| 12 | 1/b<br>1/c | | | | | | | | 2/b | 2/a | | |

*Note*: (1) numbers in the table denote the amount of ambulances relocated between two bases. (2) letter 'a' means the ambulance relocation between the time intervals '0–6 am' and '6–10 am'; 'b': '6–10 am' to '10 am–16 pm'; 'c': '10 am–16 pm' to '16–19 pm'; 'd': '16–19 pm' to '19–24 pm'; and 'e': '19–24 pm' to '0–6 am'.

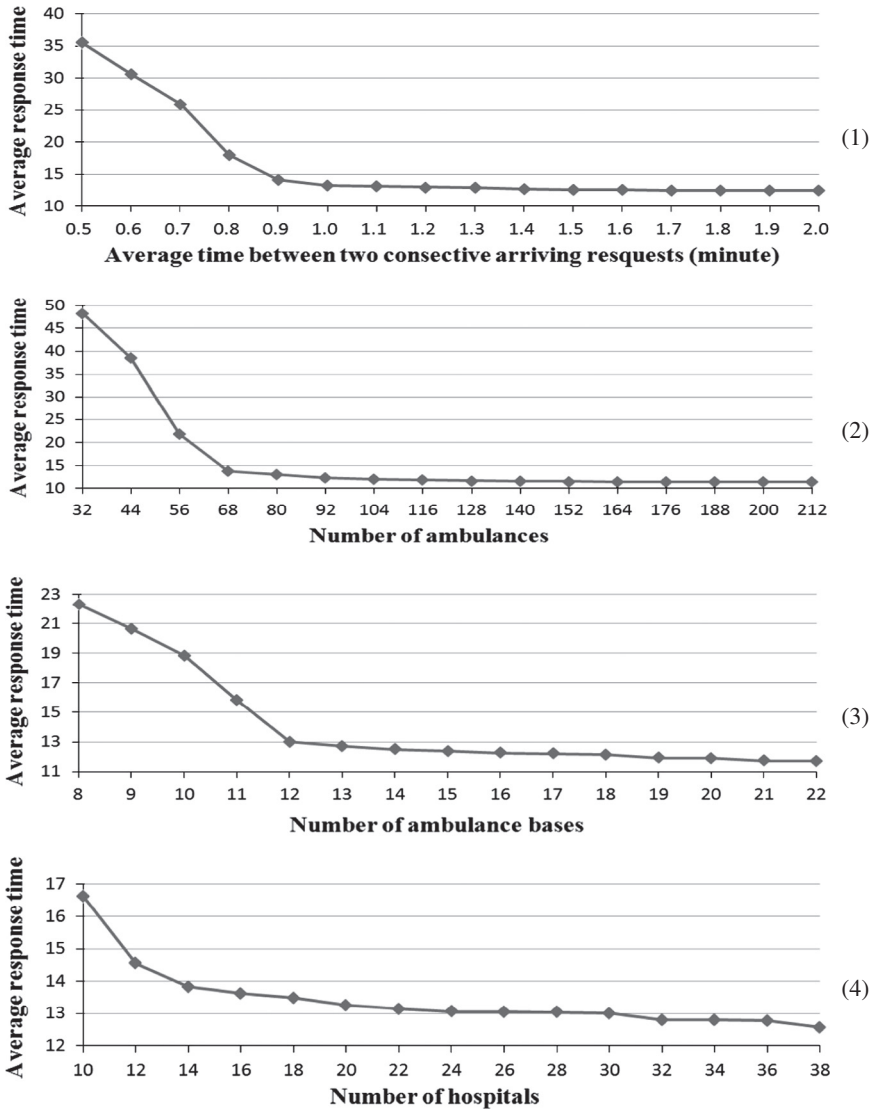



**Fig. 9.** The influences of some parameters on the average response time.

**Table 4**
The influence of request arrival rate on the optimal deployment plan.

| Avg. time between two requests (min) | The optimal deployment plan (num. of ambulances in each base) | Standard deviation | Coefficient of variance | CPU time (s) |
|---|---|---|---|---|
| 0.6 | (1, 0, 5, 5, 0, 0, 4, 6, 19, 19, 15, 6) | 6.79 | 1.02 | 691 |
| 0.7 | (0, 1, 3, 5, 0, 2, 2, 6, 25, 21, 8, 7) | 7.77 | 1.17 | 503 |
| 0.8 | (0, 1, 3, 4, 1, 4, 4, 6, 25, 17, 11, 4) | 7.15 | 1.07 | 478 |
| 0.9 | (1, 1, 3, 7, 2, 1, 3, 6, 24, 20, 8, 4) | 7.26 | 1.09 | 825 |
| 1.0 | (0, 4, 4, 5, 5, 2, 7, 5, 18, 19, 8, 3) | 5.66 | 0.85 | 442 |
| 1.1 | (2, 4, 4, 2, 4, 2, 8, 4, 17, 21, 8, 4) | 5.89 | 0.88 | 446 |
| 1.2 | (1, 3, 6, 3, 5, 4, 6, 5, 19, 17, 7, 4) | 5.31 | 0.80 | 426 |
| 1.3 | (1, 3, 5, 5, 4, 3, 7, 5, 18, 18, 8, 3) | 5.37 | 0.81 | 435 |
| 1.4 | (2, 5, 3, 6, 5, 5, 9, 6, 15, 13, 7, 4) | 3.73 | 0.56 | 307 |
| 1.5 | (4, 3, 6, 1, 8, 2, 6, 5, 19, 15, 7, 4) | 5.07 | 0.76 | 402 |
| 1.6 | (0, 4, 4, 11, 9, 4, 2, 6, 13, 12, 7, 8) | 3.90 | 0.59 | 351 |

**Table 5**
The influence of ambulance amount on the optimal deployment plan.

| Ambulance amount | The optimal deployment plan (num. of ambulances in each base) | Standard deviation | Coefficient of variance | CPU time (s) |
|---|---|---|---|---|
| 56 | (0, 1, 2, 4, 0, 3, 3, 4, 17, 14, 7, 1) | 5.23 | 1.12 | 700 |
| 68 | (0, 2, 3, 3, 4, 2, 5, 5, 20, 15, 7, 2) | 5.66 | 1.00 | 814 |
| 80 | (1, 3, 6, 3, 5, 4, 6, 5, 19, 17, 7, 4) | 5.31 | 0.80 | 426 |
| 92 | (6, 3, 4, 5, 9, 4, 6, 8, 19, 16, 9, 3) | 4.87 | 0.64 | 397 |
| 104 | (7, 9, 7, 5, 9, 7, 5, 7, 19, 16, 9, 4) | 4.29 | 0.49 | 367 |
| 116 | (7, 11, 7, 9, 8, 9, 7, 8, 19, 17, 7, 7) | 3.92 | 0.41 | 342 |
| 128 | (9, 11, 9, 9, 8, 8, 10, 8, 19, 20, 7, 10) | 4.09 | 0.38 | 318 |
| 140 | (9, 13, 11, 9, 10, 9, 12, 11, 19, 20, 7, 10) | 3.82 | 0.33 | 286 |
| 152 | (11, 9, 12, 11, 14, 9, 15, 12, 19, 17, 8, 15) | 3.25 | 0.26 | 265 |
| 164 | (11, 16, 11, 15, 13, 12, 13, 11, 19, 16, 14, 13) | 2.36 | 0.17 | 243 |
| 176 | (16, 13, 15, 13, 15, 12, 15, 12, 19, 18, 14, 14) | 2.09 | 0.14 | 219 |

**Table 6**
The influence of ambulance base amount on the optimal deployment plan.

| Base amount | The optimal deployment plan (num. of ambulances in each base) | Standard deviation | Coefficient of variance | CPU time (s) |
|---|---|---|---|---|
| 10 | (1, 2, 5, 13, 3, 7, 7, 7, 14, 21) | 5.93 | 0.74 | 247 |
| 11 | (1, 2, 4, 11, 3, 4, 7, 5, 21, 15, 7) | 5.83 | 0.80 | 289 |
| 12 | (1, 3, 6, 3, 5, 4, 6, 5, 19, 17, 7, 4) | 5.31 | 0.80 | 426 |
| 13 | (2, 2, 7, 4, 3, 2, 6, 5, 19, 16, 8, 5, 1) | 5.27 | 0.86 | 429 |
| 14 | (2, 6, 5, 5, 3, 3, 5, 5, 8, 16, 8, 3, 0, 11) | 3.92 | 0.69 | 497 |
| 15 | (2, 2, 4, 2, 8, 3, 6, 6, 8, 16, 8, 2, 1, 8, 4) | 3.79 | 0.71 | 587 |
| 16 | (4, 5, 4, 3, 3, 2, 5, 7, 7, 16, 4, 2, 2, 8, 4, 4) | 3.34 | 0.67 | 632 |
| 17 | (3, 3, 6, 3, 5, 2, 6, 4, 7, 14, 7, 4, 0, 7, 4, 3, 2) | 3.02 | 0.64 | 671 |
| 18 | (6, 2, 6, 5, 0, 4, 4, 7, 7, 12, 6, 4, 0, 7, 4, 3, 2, 1) | 2.89 | 0.65 | 698 |

**Table 7**
The influence of hospital amount on the optimal deployment plan.

| Hospital amount | The optimal deployment plan (num. of ambulances in each base) | Standard deviation | Coefficient of variance | CPU time (s) |
|---|---|---|---|---|
| 20 | (2, 2, 7, 8, 0, 4, 4, 9, 17, 15, 7, 5) | 4.90 | 0.74 | 322 |
| 22 | (3, 3, 0, 8, 3, 3, 9, 5, 18, 16, 7, 5) | 5.22 | 0.78 | 318 |
| 24 | (3, 3, 0, 6, 3, 3, 7, 7, 19, 15, 8, 6) | 5.19 | 0.78 | 424 |
| 26 | (2, 2, 0, 6, 8, 3, 6, 7, 18, 16, 9, 3) | 5.31 | 0.80 | 385 |
| 28 | (0, 4, 4, 5, 4, 2, 7, 7, 19, 17, 9, 2) | 5.60 | 0.84 | 392 |
| 30 | (1, 3, 6, 3, 5, 4, 6, 5, 19, 17, 7, 4) | 5.31 | 0.80 | 426 |
| 32 | (1, 4, 3, 4, 10, 2, 6, 5, 19, 16, 7, 3) | 5.39 | 0.81 | 391 |
| 34 | (3, 5, 3, 6, 5, 4, 6, 5, 18, 16, 7, 2) | 4.84 | 0.73 | 406 |
| 36 | (0, 2, 5, 4, 4, 3, 9, 7, 18, 17, 9, 2) | 5.51 | 0.83 | 430 |

**Table 8**
The influence of the request distribution variation on the relocations.

| Variation degree of request distributions in time intervals | | Total movement distance of all ambulances in their relocation activities (km) |
|---|---|---|
| Standard deviation | Coefficient of variance | |
| 2.45 | 0.50 | 2038 |
| 2.32 | 0.48 | 1945 |
| 2.24 | 0.47 | 1734 |
| 2.14 | 0.47 | 1686 |
| 2.04 | 0.45 | 1615 |

### 6.4. Managerial implications

From the above numerical experiments, some managerial implications are summarized as follows: (1) the request arrival rate does not always positively influence the average response time; when the rate is below a threshold, the average response time stays at a certain level. (2) For the medical resources, it is not the more resources the better performance. The number of resources (e.g., ambulances, bases, hospitals) also does not always positively influence the average response time; when the number of resources exceeds a certain level, the average response time stays at a value rather than deceasing gradually. For Shanghai with about one request every 1.2 min on average, a combination of 80 ambulances, 12 bases, and 30 hospitals is a proper setting for the current medical

service response system. (3) The ambulances should not be deployed among the bases in a balanced manner. The proposed simulation optimization (SO) method is especially necessary for making a good deployment decision in the situation with high frequent arriving request, in which case, the optimal deployment plan is in a very uneven pattern. (4) Establishing more ambulance bases or involving more hospitals into the medical service response system will not make the ambulances be deployed in a more evenly balanced way; but different ambulance base or hospital distributions and amounts can influence the specific deployment of ambulances. (5) The ambulance relocation is mainly impacted by the variance of the request distributions. The more variances are contained with respect to the request distributions during time intervals, the longer movement distance will be incurred in the ambulance relocation activities.

## 7. Conclusions

To facilitate the emergency medical service scheduling, this paper studies the ambulance deployment and relocation problems by using the simulation optimization methodology. A demo example of its application in Shanghai is illustrated. Some numerical experiments are also conducted for further investigations on ambulance deployment and relocation decisions.

By comparing with the literature on the related topics, the contribution of this paper is mainly as follows. Most related studies on ambulance deployment problems use the dynamic programming methods or some integer programming models. The former one is usually limited by its problem scales; the latter one may be not easy to consider more detailed and comprehensive random factors in the ambulance scheduling and fulfillment activities. This study employs the simulation and optimization methodology, which can capture stochastic and dynamic nature of request arrivals, fulfillment processes, and complex traffic conditions as well as the time-dependent spatial patterns of some parameters. The GA embedded in the simulation and optimization framework can guide the search process for finding a near optimal solution (i.e., ambulance deployment decision) in an efficient manner. Moreover, on the basis of the deployment decisions, a mathematical model on ambulance relocation is also proposed for adapting to the dynamic changing environments along the time. This study can pave a way for developing some comprehensive decision support tools on medical service scheduling in the future.

There are also limitations in this study. The current model uses the dispatching rule that the nearest available ambulance from a call's scene is assigned with the task for fulfilling the call's request. This 'the nearest distance' dispatching rule may not be the best rule in some situations (Schmid, 2012). In addition, the heterogeneous ambulances with respect to their drivers' experiences, equipments, crew capabilities, etc. (López, Innocenti, & Busquets, 2008), have not been considered in this study. All of these issues will be investigated in our future researches.

## Acknowledgements

This research is supported by the Excellent Young Faculty Research Program in Shanghai University under the 085 Project 'Smart City and Metropolis Development', the Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning, National Natural Science Foundation of China (Grant No. 71101087).

## References

Ahmed, M. A., & Alkhamis, T. M. (2009). Simulation optimization for an emergency department healthcare unit in Kuwait. *European Journal of Operational Research, 198*(3), 936–942.
Andersson, T., & Vaerband, P. (2007). Decision support tools for ambulance dispatch and relocation. *Journal of the Operational Research Society, 58*(2), 195–201.
Berman, O. (1981a). Dynamic repositioning of indistinguishable service units on transportation networks. *Transportation Science, 15*(2), 115–136.
Berman, O. (1981b). Repositioning of distinguishable urban service units on networks. *Computers and Operations Research, 8*(2), 105–118.
Berman, O. (1981c). Repositioning of two distinguishable service vehicles on networks. *IEEE Transactions on Systems, Man, Cybernetics SMC, 11*(3), 187–193.
Brotcorne, L., Laporte, G., & Semet, F. (2003). Ambulance location and relocation models. *European Journal of Operational Research, 147*(3), 451–463.
Church, R., & ReVelle, C. (1974). The maximal covering location problem. *Papers in Regional Science, 32*(1), 101–118.
Daskin, M. S. (1983). A maximum expected covering location model: Formulation, properties and heuristic solution. *Transportation Science, 17*(1), 48–70.
Doerner, K. F., Gutjahr, W. J., Hartl, R. F., Karall, M., & Reimann, M. (2005). Heuristic solution of an extended double-coverage ambulance location problem for Austria. *Central European Journal of Operations Research, 13*(4), 325–340.
Gendreau, M., Laporte, G., & Semet, F. (1997). Solving an ambulance location model by Tabu search. *Location Science, 5*(2), 75–88.
Gendreau, M., Laporte, G., & Semet, F. (2001). A dynamic model and parallel Tabu search heuristic for real-time ambulance relocation. *Parallel Computing, 27*(12), 1641–1653.
Gendreau, M., Laporte, G., & Semet, S. (2006). The maximal expected coverage relocation problem for emergency vehicles. *Journal of the Operational Research Society, 57*(1), 22–28.
Iannoni, A. P., Morabito, R., & Saydam, C. (2009). An optimization approach for ambulance location and the districting of the response segments on highways. *European Journal of Operational Research, 195*(2), 528–542.
Kolesar, P., & Walker, W. E. (1974). An algorithm for the dynamic relocation of fire companies. *Operations Research, 22*(2), 249–274.
López, B., Innocenti, B., & Busquets, D. (2008). A multiagent system for coordinating ambulances for emergency medical services. *IEEE Intelligent Systems, 9–10*, 50–57.
Schmid, V. (2012). Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming. *European Journal of Operational Research, 219*(3), 611–621.
Schmid, V., & Doerner, K. F. (2010). Ambulance location and relocation problems with time-dependent travel times. *European Journal of Operational Research, 207*(3), 1293–1303.
Shariat-Mohaymany, A., Babaei, M., Moadi, S., & Amiripour, S. M. (2012). Linear upper-bound unavailability set covering models for locating ambulances: Application to Tehran rural roads. *European Journal of Operational Research, 221*(1), 263–272.
Toregas, C., Swain, R., ReVelle, C., & Bergman, L. (1971). The location of emergency service facilities. *Operations Research, 19*(6), 1363–1373.
Underwood, D. J., Zhang, J. Y., Denton, B. T., Shah, N. D., & Inman, B. A. (2012). Simulation optimization of PSA-threshold based prostate cancer screening policies. *Health Care Management Science, 15*(4), 293–309.
Zhang, Y., Puterman, M. L., Nelson, M., & Atkins, D. (2012). A simulation optimization approach to long-term care capacity planning. *Operations Research, 60*(2), 249–261.