

An interactive decision support system for real-time ambulance relocation with priority guidelines

Mahdi Hajiali^a, Ebrahim Teimoury^a, Meysam Rabiee^b, Dursun Delen^{c,d,*}

^a School of Industrial Engineering, Iran University of Science and Technology, Narmak, Tehran, Iran

^b Business School, University of Colorado Denver, Denver, CO 80202, USA

^c Spears School of Business, Oklahoma State University, Stillwater, OK, USA

^d School of Business, Ibn Haldun University, Istanbul, Turkey

ARTICLE INFO

Keywords:

Emergency management system
Ambulance redeployment
Response time
Optimization modeling
DSS
Health care logistics

ABSTRACT

Changes in demand patterns and unexpected events are the two primary sources of delays in healthcare emergency operations. To mitigate such delays, researchers proposed the movement of idle ambulances between emergency bases as one of the effective ways to improve the areal coverage of future demands. In this study, we have developed a model-driven decision support system that simultaneously seeks to maximize demand coverage while minimizing travel time by optimally relocating emergency response vehicles. The developed mathematical model partitions and prioritizes demand into four categories and continuously updates them over time. Furthermore, it dynamically calculates the number of coverages in different regions based on the current location of idle ambulances. Also, we developed a real-time risk assessment DSS for recommended relocations, which could be utilized as a reference by the EMS user while implementing suggested relocation decisions. A real case study is used to validate the proposed DSS, and its final output is compared to the existing operational policy. The findings show that the average workload added to each ambulance due to relocations has significantly improved the response time and coverage ratio. Compared to the existing operational policy, the developed decision support system decreased the time to respond to calls, which was deemed to be more than to offsets the increase in travel time due to relocation. Furthermore, the system also reduced the total working time of all ambulances by about 9% per shift.

1. Introduction

Emergency Medical Services (EMS) organizations are among the most critical systems for providing community health services. They are expected to make effective decisions and implement them rapidly while providing services to their customers in real-time. Determining the type and severity of incoming calls and selecting the most appropriate ambulance to send are life-saving and time-sensitive decisions [4,19,41,49]. These systems are under constant pressure from contractual obligations and management objectives to meet and exceed certain levels of performance expectations. For example, according to some emergency system standards [42], the maximum response time to emergency calls must be within 10 min in urban areas, and 30 min in rural areas [6].

Decisions in the emergency system can be divided into three levels: strategic, tactical, and operational. Strategic decisions include the location of ambulance stations and specifications for ambulances. Tactical-level decisions involve standby sites location, scheduling crews, and determining fleet policy [1,14]. Indeed, fleet management strategies, such as static or dynamic policies, are decided at the tactical level. When there is a static policy at the strategic or tactical level, the ambulances' base stations and standby sites are set for various goals, such as providing maximum coverage for a single operation, and each ambulance returns to its home base after each mission [5]. Static models do not provide the necessary operational flexibility [37]. At the operational level, the state of the system is continuously changing, so a decision about an ambulance's base stations needs to be modified dynamically [15]. The state of the system can be changed by factors such

* Corresponding author at: Journal of Business Analytics, Spears Chair in Business Administration, Patterson Foundation Chair in Business Analytics, Director of Research, Center for Health Systems Innovation, Department of Management Science and Information Systems, Spears School of Business, Oklahoma State University, USA.

E-mail addresses: teimoury@iust.ac.ir (E. Teimoury), meysamr@uoregon.edu (M. Rabiee), dursun.delen@okstate.edu (D. Delen).

<https://doi.org/10.1016/j.dss.2021.113712>

Received 18 January 2021; Received in revised form 7 December 2021; Accepted 9 December 2021

Available online 14 December 2021

0167-9236/© 2021 Elsevier B.V. All rights reserved.

as changes in demand patterns, the occurrence of unforeseen events, and changes in the number of available ambulances.¹ In real-time, when the state of the system changes, the primary concern of the emergency system is to cover the current demand points and minimize the response time for future calls. The movement of idle ambulances between emergency bases is one of the effective ways to improve the coverage of future calls [29,33,35,53–55,59].

There are two ways to approach the dynamic relocation problem: offline and online. In the offline approach, the system state is determined by the number of available ambulances and changes when that number changes. Each state of the system has a compliance table that designates the bases that must have at least one ambulance. In contrast, the online (or real-time) approach relocates ambulances in real-time and takes into account more dimensions of the problem than the offline approach. In the online approach, several factors can each be considered as a criterion for changing the system state, such as the number of available ambulances, the demand or call rate, the relocation cost, and the workload of ambulances. The essential component in the online approach is making a proper decision in the shortest amount of time, taking into account all available information. Since relocation is done to improve coverage, this movement may be undesirable for employees and may incur additional costs to the system. Therefore, while improving coverage, the online approach should minimize the relocation costs and the added workload.

In this study, we present a decision support system that solves the ambulance relocation problem based on a mathematical model with an online approach. This system is updated based on time and changes in the system's state, such as relocation workload and the number of coverages.² The decision support system determines the new location of idle ambulances, maximizing demand coverage while taking into account the workload limitation³ and the cost of relocating. The proposed mathematical model is a two-objective model: the first goal is to maximize demand coverage, and the second goal is to minimize travel time. When maximizing demand coverage, three time-ranges⁴ and four types of demand are considered. Demand is categorized based on the time required and the degree of urgency. Maximizing the coverage of demand points is based on the weights of their demands during the decision-making period, and the weights of demand points are updated at each period. When the emergency system needs relocation of ambulances, users can enter some of the required data and rapidly decide on a new location for ambulances. The remainder of the paper is organized as follows. Section 2 reviews the literature related to the current problem. Section 3 provides an overview of the problem and the system's events. Section 4 describes the proposed decision support system and its subsystems, such as the mathematical model, database, and user interface. Section 5 compares the current policy in the eastern part of Tehran to that proposed by our DSS using real data. Section 6 concludes the paper and suggests future research directions.

2. Literature review

Over the past twenty years, much attention has been paid to mathematical models with a dynamic approach to relocating ambulances in an emergency system. Deciding on the relocation of ambulances is an operational decision that is fraught with uncertainty. The requirements for advanced strategies to address this uncertainty and to consider social

goals such as clinical outputs have posed new challenges in this field. To respond to this uncertainty, many models have turned to dynamic approaches. Some researchers also divide a working day into several time periods to manage fluctuations in demand and consider a location plan for each time period. Ambulances perform relocation based on different location plans [11,16,23,47,48,53]. There are two main ways of solving dynamic optimization problems: offline and online. The offline technique is a posteriori in nature, requiring the solution of mathematical programs to generate compliance tables for each potential state. The online approach facilitates real-time decision-making to determine the optimal relocation strategy given the system's current state. Also, the relocation problem can be divided into different categories.

There are two major streams in the extant literature that aim to improve emergency management system services: One focuses on the demand aspect in order to maximize demand coverage. The other focuses on the costs associated with ambulance relocation. Both of these streams employ different objective functions and constraints in mathematical models to achieve their goals. In the context of demand areas, the majority of studies have had a primary goal of improved coverage; different models have used different approaches to reach this goal. In Table 1, we describe the reviewed studies by their approach (real-time or offline) and the dimensions they use to address the relocation problem.

2.1. Online approach

The first redeployment model to use an online approach was proposed by Gendreau et al. (a) [28]. They presented a model that uses two time-ranges. Two types of constraints are considered in this model: the absolute covering constraint, which seeks to cover all demand points by at least one ambulance within a longer time range; and the relative covering constraint, which seeks to cover a proportion α of the total demand within a shorter time range. The model tries to maximize the coverage of demand in the shortest time by at least two ambulances. It then seeks to minimize the cost of the system due to relocation.

Anderson and Varbrand [3] developed an algorithm for ambulance relocation and automatic dispatching to minimize travel time. Their algorithm was based on a new quantitative preparedness measure that determines the system's ability to serve patients by available ambulances at present and future. They used a tree-search heuristic to solve the model.

Naoum-Sawaya and Elhedhli [46] developed a two-stage stochastic optimization model for the relocation problem. In the first stage, ambulances are assigned to bases to minimize the number of relocations. Then, in the second stage, ambulances are assigned to calls. This step minimizes the number of calls that are not served within an appropriate time. Similarly, Mason [39] provided a real-time multi-view generalized-cover repositioning model (RMvGcRM) by proposing an objective function with two parts: the first part seeks to maximize the overall coverage obtained from the redeployment of ambulances; the second part, as in [28], seeks to minimize the penalty for redeploying an ambulance from its current base to a new base. They have developed their model in a real-time online optimization system called Optima Live.

As mentioned earlier, the challenge for most relocation models is to gain an accurate understanding of all the details of the system and to have an algorithm that takes a short time to determine further actions. To address this challenge, Jagtenberg et al. [30] proposed a polynomial-time heuristic algorithm. They presented the model proposed in [51] with a dynamic approach that aims to minimize the expected fraction of late arrivals. Then, they compared their obtained results to the common method (a static solution) of the emergency management system.

Moieni et al. [43] developed a model based on [28] by dividing the demand points into points that require a single ambulance and points that require two ambulances. In some cases, a demand area needs to be covered by more than one vehicle. They classified high-emergency

¹ Ambulances that are idle in bases and ambulances that have completed their missions.

² The number of ambulances deployed at bases that cover the demand area.

³ Ambulances make additional moves to better respond to future calls. The additional burden on the ambulance's employees due to relocation is workload limitation.

⁴ The maximum response time to a call by an emergency base is referred to as the time range.

Table 1

Classification of recent literature on dynamic relocation, based on problem characteristics and approaches.

										Demand side		EMS side			
Author	Year	Real-time	Offline	Real-Time Aspects						Coverage improvement	Demand prioritization		EMS cost	Workload limitation	Interactive DSS
				RC	NV	CR	AW	NC	RRARR		#of Levels ^a	Type ^b			
Gendreau et al. (a)	2001	✓		✓	✓					✓			✓	✓	
Gendreau et al. (b)	2006		✓							✓					
Andersson and Varbrand	2007	✓			✓								✓		
Nair and Miller-Hooks	2009		✓							✓			✓		
Naoum-Sawaya and Elhedhli	2013	✓			✓								✓	✓	
Mason	2013	✓				✓				✓			✓	✓	
Maleki et al.	2014		✓										✓		
Jagtenberg et al.	2015	✓			✓					✓					
Moeini et al.	2015	✓		✓	✓					✓	2	Static	✓	✓	
van Barneveld et al. (a)	2015	✓			✓								✓		
Bélanger et al.	2016	✓			✓					✓			✓	✓	
Sudtachat et al. (a)	2016		✓							✓					
van Barneveld et al. (b)	2017		✓							✓			✓		
Enayati et al. (a)	2018	✓					✓			✓			✓	✓	
Enayati et al. (b)	2018	✓					✓			✓				✓	
van Barneveld et al. (c)	2018		✓							✓					
Carvalho et al.	2020	✓		✓						✓			✓		
Sudtachat et al. (b)	2020		✓							✓					
This study	2021	✓			✓	✓	✓	✓	✓	✓	4	Dynamic	✓	✓	✓

^a Number of levels, determined by the number of ambulances necessary and the severity of the emergency.^b The demand is either constant (i.e., static) or changes over time over time (dynamic).

demand into two categories depending on the number of ambulances necessary. Additionally, their model's demand is static and does not change over time. They implemented their approach in a sparsely populated region of France. Additionally, our algorithm categorizes demand into four groups based on the amount of time required, the number of ambulances required, and the severity of the emergency. Our model's demand is dynamic and changes over time. In another study, Van Barneveld et al. (a) [7] proposed a dynamic relocation problem, focusing on rural areas and considering a limited number of vehicles. Their model is formulated based on a discrete-time Markov decision process; at each time step, a relocation policy decides on the movement of all idle ambulances.

Bélanger et al. [13] compared four strategies for fleet management. All strategies were based on the double standard model proposed by Gendreau et al. [27]. They classified strategies based on permission to relocate (moving idle ambulances between bases)- multi-period or dynamic planning and repositioning plan (determining the location after completing the mission). They concluded that the best performance in service delivery is provided if a dynamic approach is used and ambulances are allowed to have relocation and repositioning. Enayati et al. (a) [24] presented a two-step model for real-time relocation by taking into account the workload limitation. In the first step, it seeks to maximize demand coverage; in the second step, it minimizes the total travel time. Enayati et al. (b) [25] also presented a two-stage stochastic integer model to integrate dispatching and relocation policies. In the first stage, ambulances are assigned to new locations; in the second stage, they respond to a different call-in scenario. Carvalho et al. [18] developed a

mixed-integer programming model to integrate relocation and ambulance dispatching decisions. Their approach uses a time-preparedness measure to evaluate the system's ability to manage emergencies. This criterion is taken from the criterion introduced by [34], with a time dimension t added to it.

2.2. Offline approach

Gendreau et al. (b) [26] mentioned that executing online models after each dispatch of an ambulance may be time-consuming. They presented the Maximal Expected Coverage Relocation Problem model (MECRP) as the first offline approach model by taking advantage of the Maximal Covering Location Problem Model (MCLP) introduced by [22]. This model is a dynamic probabilistic model that seeks to maximize coverage in all system states. The criterion for changing the state of the system is the number of available ambulances, and in each state, the location of ambulances is determined.

To evaluate the benefits of relocation, Nair and Miller-Hooks [45] proposed a multi-objective, probabilistic, integer model, based on [52]. After the ambulance bases have been identified for each state of the system, the next issue is specifying the movement of the ambulances. Some ambulances become free and need to go to a new base. Some idle ambulances also need to change their bases due to system changes. After presenting the MECRP model, Gendreau et al. (b) [26] proposed the use of transportation planning models to solve this problem. After determining the location of ambulances using the MECRP model, Maleki et al. [38] proposed a "generalized assignment model" and "generalized

bottleneck assignment model” to solve the movement problem. The first model minimizes the total travel time; the second model minimizes the maximum travel time. Sudtachat et al. (a) [58] present a model with nested compliance tables as an integer programming model to maximize coverage. They perform their policy based on the Markov-chain model proposed by [2]. One of the challenges in the offline approach is the number of types of vehicles. If there is more than one type of vehicle in the problem, the problem becomes complicated. Van Barneveld et al. (b) [9] considered a system that includes two types of vehicles. They used an integer linear programming model to determine the location of ambulances; the inputs of this model use the outputs of a Hypercube model. Their model is a developed model of [26,40].

Van Barneveld et al. (c) [10] combined the methodology developed by [8,30] to provide an operational analysis of the dynamic ambulance relocation process. In this study, they examined the effect of the frequency of redeployment-decision moments and the performance criterion on the quality of the distribution strategy. Sudtachat et al. (b) [57] presented the maximum realized covering the relocation and districting problem. In this case, they first converted the regions into districts and then considered an independent relocation strategy for each district. An optimal nested-compliance model was applied to determine the offline relocation program of each district. They compared their approach with the model provided by Batta et al. [12].

2.3. The contribution of the study

In this study, we have extended the online approach to the ambulance relocation problem in the following directions:

- Since ambulance relocation is an operational decision and may be repeated several times during a work shift, in this study, we have proposed an online decision support system that makes suggestions based on a mathematical model.
- When covering constraints are formulated in the mathematical models, three questions must be considered:
 - a. What is the current number of coverages of demand points in different ranges?
 - b. How many coverages will be increased by relocations to demand points?
 - c. How many coverages will be decreased by relocations away from demand points?

All of the models proposed in previous studies answer only question b in their coverage constraints. They are technically capable of responding to all of them. However, they did not do so since they had no intention of using those numbers in future analysis. In our study, the number of coverages in various time ranges serves as the primary input to our final DSS, which is used to quantify the risk associated with demand zones (see Table 4).

- We developed a DSS for use in real-time risk assessment of recommended relocations (RRARR). The important features for predicting distinct types of demand are found in RRARR. The risk value of each zone is then determined in each feature. Zones are classified from very high to very low risk. Finally, relocation decisions are ranked according to their risk. That is, after running the model, the user is offered with the priority of picking any recommended relocation.

Table 1 provides a multi-dimensional classification of the recent literature on the dynamic relocation approach, providing specifics on the problem characteristics and solution approaches. The very last row in Table 1 illustrates the proposed study and its coverage as it is compared to the extant literature. The real-time aspects column in Table 1 attempts to differentiate these papers based on the parameters that continuously change in response to changing system states. These parameters include relocation costs (RC) - number of vehicles (NV) - call

rate (CR) - ambulance workload (AW) - number of coverages (NC) and real-time risk assessment for recommended relocations (RRARR).

3. Problem overview

The eastern region is one of the high-demand areas in Tehran, the capital city in Iran. It covers an area of 170 km² and has a population of about 3 million. With 30 ambulances and 30 base stations, this area should cover an average of 19 calls per hour. The most important problem of the emergency system in the eastern area is the inability to cover all the demands within their required time ranges. The approach used by the emergency system in this area is a static approach. To tackle this problem, we offer an online relocation decision support system that is model-driven. In this context, these three events change the state of the system: the arrival of a call and the assignment of an ambulance to it; a busy ambulance completes its mission; and idle ambulances are relocated. We have defined four types of demand based on the required time range to respond to calls and the degree of urgency as follows (in the following categories, it is assumed that the $r_1 < r_2 < r_3$):

1. Most Urgent - Case 1: Calls that must be covered by one ambulance within a maximum of r_1 minutes.
2. Most Urgent - Case 2: Calls that must be covered by two ambulances within a maximum of r_1 minutes.
3. Urgent - Calls that must be covered by an ambulance within a maximum of r_2 minutes.
4. Not Urgent - Calls that must be covered by an ambulance within a maximum of r_3 minutes.

When a call arrives, it is first assessed in terms of urgency and then assigned to one of the four defined types of demand. After assigning ambulances to calls and changing the system state (which results in a reduction in the number of available ambulances), the ambulances are relocated. In this paper, relocation decisions are made through a model-driven decision support system. This is a description of the events presented in this model to relocate the ambulances:

1. These events change the system state. The state of the system can change due to one or more of the following events:
 - A call arrives, and an ambulance is assigned to it, which decreases the number of available ambulances.
 - A busy ambulance completes its mission, which increases the number of available ambulances.
 - Idle ambulances are relocated.
2. The system is updated based on each event.
 - When a call arrives, the required covering time and the degree of urgency are determined, and the call is placed in one of the four demand categories. When calculating which ambulance to dispatch to the scene, all idle ambulances at their bases and all ambulances that have completed their missions and are returning to their bases are considered. An ambulance is dispatched to the scene based on the closest available ambulance within the required time range of the call. If there is no ambulance in the required time range, the closest available ambulance in the next range is considered. After dispatching, dynamic parameters of the problem are updated, such as the number of coverages of demand points at different time ranges. These parameters are used in the criteria for the relocation of ambulances in the system.
 - After an ambulance completes its mission, if the system needs to relocate ambulances according to the relocation criteria, the ambulance will first be assigned to the closest base as an idle ambulance, and then the relocation will be performed.
 - In case of relocation of ambulances, first, the ambulances will be deployed at their new bases, and then the dynamic parameters will be updated.

3. If relocation of ambulances is needed, determine relocation based on the mathematical model.

If either of these two criteria is met, the system will relocate ambulances:

- At least four demand points in r_1 and at least two demand points in r_2 have no coverage, or
- At least 60 min have elapsed since the last relocation.

Following the case of Bélanger et al. [13], we set out these criteria for relocation. The first criterion is defined based on the different system conditions such as fleet capacity and the volume of demand. This is to ensure that the increase in the number of movements can be justified by the obtained improvement. The second criterion guarantees that the number of uncovered areas will always be at an acceptable level. That is, the movements should not be so low that there is no improvement in demand coverage.

4. If there is system relocation, execute the ambulance relocation, and update the system.

The following assumptions are also considered in the designed model:

- The demand region of the problem is divided into sub-zones with approximately equal areas. This method has been utilized to simplify the problem, mathematical modeling, and finally discover a solution in a fair amount of time.
- The basis of coverage by base stations in the required time ranges is the fastest path from the base to the demand point.
- Each demand point can create four types of calls, and demands that require more than one ambulance are considered as two-ambulance demands. The two-ambulance demands are defined only in radius r_1 .
- The demands in the mathematical model are considered based on historical data. To be more specific, the EMS user sets the initial values of parameters at the start of each shift based on the predicted values obtained using time series and machine learning approaches. Then, during each shift, the demand parameter will be incrementally modified to account for the unexpected percentage of demand.
- The working hours of the emergency system are divided into two 12-h shifts.
- All demands are covered by one type of vehicle.

In the next section, we provide an explanation of the proposed DSS architecture along with the specifics of the mathematical formulation and the underlying workflow of the process model.

4. Proposed DSS architecture

This section describes the architecture of the DSS and its components: the mathematical model, the user interface, and the database.

Implementing and operationalizing this system involved three main steps: mathematical model design, database design, and user interface design. We prototyped our system using *Excel* and the final implementation was done using *C++* on the *.NET* framework. Also, while developing decision support systems, the choice of programming language and platforms largely depends on the type of mathematical model solver. These solvers use the Application Programming Interface (API) libraries to interface with programming languages. API libraries provide a platform for stand-alone and web-based interactions. To solve the mathematical model and communicate with the *.NET* framework, we used GAMS 25.1.2 and the CPLEX solver. Fig. 1 shows the association between system components and the DSS architecture.

4.1. Mathematical model

We have developed an integer linear programming model for the relocation problem. This model is multi-objective; in addition to maximizing the coverage of demand points in different time ranges, the model minimizes travel time due to relocation. This model also prevents movements of ambulances that would violate workload. Table 3 shows the notation for the mathematical model's indices, parameters, and decision variables.

4.1.1. Objective functions and constraints

To cover the demand points by the base stations, three time-ranges of 7 min, 15 min, and 25 min have been considered. Each base station covers several demand points in different ranges. During relocation, the new locations are determined in such a way that the coverage of demand points by the bases is maximized. According to historical data, the proportional call volume of demand zones varies throughout the day, and coverage maximization should be tailored to the areas' demand priorities. For example, during a shift period, the priority of some points may be covered within a 7-min time range. But some areas may need more coverage within a 15-min range. Therefore, the objective function (1) is defined to maximize the coverage of demand points according to their coverage priority. The objective function (2) seeks to move the available ambulances in the shortest possible time so that the move has the least cost for the emergency system.

$$\max \sum_{i \in N} (d_{ii}^1 X_i^1 + d_{ii}^2 X_i^2 + d_{ii}^3 Z_i + d_{ii}^4 W_i) \quad (1)$$

$$\min \sum_{k \in K} \sum_{f \in M} \sum_{j \neq f \in M} Y_{jf}^k B_{jf} C_{jf} \quad (2)$$

subject to:

$$O_{ii} + \sum_{k \in K} \sum_{f \in M} \sum_{j \neq f \in M} Y_{jf}^k L_{kf}^t C_{jf} H_{ij} - \sum_{k \in K} \sum_{f \in M} \sum_{j \neq f \in M} Y_{jf}^k L_{kf}^t C_{jf} H_{if} \geq X_i^1 + X_i^2 \quad \forall i \in N \quad (3)$$

$$E_{ii} + \sum_{k \in K} \sum_{f \in M} \sum_{j \neq f \in M} Y_{jf}^k L_{kf}^t C_{jf} S_{ij} - \sum_{k \in K} \sum_{f \in M} \sum_{j \neq f \in M} Y_{jf}^k L_{kf}^t C_{jf} S_{if} \geq Z_i \quad \forall i \in N \quad (4)$$

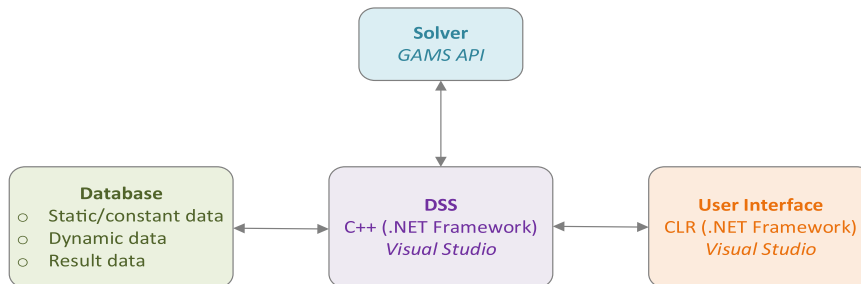


Fig. 1. High-level architecture of the proposed DSS.

$$Q_{it} + \sum_{k \in K} \sum_{f \in M} \sum_{j \neq f \in M} Y_{fj}^k L_{kf}^t C_{fj} R_{ij} - \sum_{k \in K} \sum_{f \in M} \sum_{j \neq f \in M} Y_{fj}^k L_{kf}^t C_{fj} R_{if} \geq W_i \quad \forall i \in N \quad (5)$$

$$X_i^1 \geq X_i^2 \quad \forall i \in N \quad (6)$$

$$V_{jt} + \sum_{f \in M} \sum_{k \in K} Y_{fj}^k C_{fj} - \sum_{f \in M} \sum_{k \in K} Y_{jf}^k C_{fj} \leq P_j \quad \forall j \in M \quad (7)$$

$$\beta_k^t + \sum_{f \in M} \sum_{j \neq f \in M} Y_{fj}^k B_{fj} C_{fj} L_{kf}^t \leq U_t \quad \forall k \in K, t \in (0, T) \quad (8)$$

$$\left(\sum_{k \in K} Y_{fj}^k C_{fj} L_{kf}^t \right) \left(\sum_{k \in K} Y_{jf}^k C_{fj} L_{kf}^t \right) = 0 \quad \forall f, j \neq f \in M, t \in (0, T) \quad (9)$$

$$L_{kf}^t \geq Y_{fj}^k C_{fj} \quad \forall k \in K, f, j \neq f \in M, t \in (0, T) \quad (10)$$

$$Z_i + W_i \geq 1 - X_i^1 \quad \forall i \in N \quad (11)$$

$$\sum_{j \in M} L_{fj}^t Y_{fj}^k \leq 1 \quad \forall k \in K, f \in M, t \in (0, T) \quad (12)$$

$$X_i^1, X_i^2, Z_i, W_i \in \{0, 1\}, \quad \forall i \in M \quad (13)$$

$$Y_{fj}^k \in \{0, 1\}, \forall k \in K, f, j \neq f \in M \quad (14)$$

Constraints (3), (4), and (5) are coverage constraints that are formulated based on the six steps described in Section 2.3. These constraints respectively consider the number of coverages in r_1 , r_2 , and r_3 minutes at time t , the number of coverages increased, and the number of coverages decreased due to relocation, simultaneously. In constraint (3), if $(X_i^1 = X_i^2 = 1)$, ambulance relocation should be performed in such a way that at least two ambulances cover the desired demand point in r_1 minutes. Also, in constraints (4) and (5), if the variables Z and W are one, the outcome of the coverage should be such that at least one ambulance in the r_2 and r_3 minutes covers the demand areas. Constraint (6) ensures that at least two ambulances are assigned to a demand point if at least one ambulance is assigned to it.

Constraint (7) considers the number of ambulances located at each base, the number of ambulances entering it, and the number of ambulances leaving it; this constraint prevents a violation of the maximum number of ambulances in each base. Constraint (8) refers to the relocation workload of ambulances in work shifts. Parameter β represents the accumulated relocation time of each ambulance at the current time t . This constraint ensures that the relocation time of ambulances will not exceed the maximum allowed accumulated relocation time. The parameter (U_t) will be computed dynamically according to Eq. (15):

$$U_t = \frac{\gamma \cdot t}{T} \quad (15)$$

Constraint (9) avoids round trips between two bases. This constraint is a nonlinear inequality; to linearize it, we add the three inequalities below.

$$\eta_{fj}^k \leq Y_{fj}^k \quad \forall k \in K, f \neq j \in M \quad (16)$$

$$\eta_{fj}^k \leq Y_{jf}^k \quad \forall k \in K, f \neq j \in M \quad (17)$$

$$\eta_{fj}^k \geq Y_{fj}^k + Y_{jf}^k - 1 \quad \forall k \in K, f \neq j \in M \quad (18)$$

$$\eta_{fj}^k = Y_{fj}^k \times Y_{jf}^k \quad \forall k \in K, f \neq j \in M \quad (19)$$

Inequalities (16) and (17) ensure that η_{fj}^k will be zero if either Y_{fj}^k or Y_{jf}^k are zero. Inequality (18) ensures that η_{fj}^k will take value one if both binary variables are set to 1. Constraint (10) ensures that an ambulance can be moved from a base if it is located at that base. In some critical situations, the number of available ambulances may be low, and it may

Table 3

Summary of notation for the mathematical model.

Sets & indices	
$i \in \{1, 2, \dots, N\}$	Demand Zones
$j, f \in \{1, 2, \dots, M\}$	Base Stations
$k \in \{1, 2, \dots, K\}$	Available Ambulances
$t \in (0, T)$	Current time in the shift
Constant parameters	
P_j	Maximum number of ambulances that can be located at base station j
r_1, r_2, r_3	Time ranges to cover demand zones ($r_1 < r_2 < r_3$)
γ	Maximum allowed relocation workload for each ambulance in a shift (hours)
H_{ij}	Binary parameter: 1 if demand point i is accessible from base station j in r_1 minutes
S_{ij}	Binary parameter: 1 if demand point i is accessible from base station j in r_2 minutes
R_{ij}	Binary parameter: 1 if demand point i is accessible from base station j in r_3 minutes
B_{fj}	Travel time between base stations f and j in the shortest path
C_{fj}	Binary parameter: 1 if travel from base station f to j is allowed
λ_1, λ_2	Weights of objective functions 1 and 2
Dynamic parameters	
d_{it}^1	Proportional call volume of the most urgent demands at the point i for one ambulance at time t
d_{it}^2	Proportional call volume of the most urgent demands at the point i for two ambulances at time t
d_{it}^3	Proportional call volume of the urgent demands at the point i for one ambulance at time t
d_{it}^4	Proportional call volume of the non-urgent demands at the point i for one ambulance at time t
V_{jt}	Number of ambulances located at base station j at time t
O_{it}	Number of ambulances covering point i in r_1 minutes at time t
E_{it}	Number of ambulances covering point i in r_2 minutes at time t
Q_{it}	Number of ambulances covering point i in r_3 minutes at time t
L_{kf}^t	Binary parameter: 1 if ambulance k is located at base station f at time t
U_t	Maximum allowed accumulated relocation time of each ambulance at time t
β_k^t	Accumulated relocation time of ambulance k at time t
Decision variables	
Y_{fj}^k	Binary variable: 1 if ambulance k moves from base station f to j , 0 otherwise
X_i^1	Binary variable: 1 if demand point i is covered at least one time in r_1 minutes, 0 otherwise
X_i^2	Binary variable: 1 if demand point i is covered at least two times in r_1 minutes, 0 otherwise
Z_i	Binary variable: 1 if demand point i is covered at least one time in r_2 minutes, 0 otherwise
W_i	Binary variable: 1 if demand point i is covered at least one time in r_3 minutes, 0 otherwise
η_{fj}^k	Binary variable: 1 if the product of Y_{fj}^k and Y_{jf}^k is equal to 1; 0 if at least one of them is equal to 0

not be possible to provide coverage in the minimum time for some demand points. Under these conditions, constraint (11) forces the model to allocate at least one ambulance in r_2 or r_3 minutes. Constraint (12) prevents an ambulance from being assigned to more than one base station. Finally, constraints (13) and (14) state types of the decision variables must all be binary.

4.2. DSS process model

To implement the mathematical model in the decision support system, we need to design a process model. This model is the basis of the DSS database design and user interface design. The developed database consists of three kinds of data: constant data, which does not change during model execution; dynamic data, which changes as the system state changes; and result data, which is generated from the combination of constant and dynamic data. In fact, result data is a kind of dynamic

data. These three types of data are related to each other, based on the developed process model shown in Fig. 2.

The user interface is designed based on the process model. The user first enters the working hours, then the system work shift is determined. According to the current working hour and shift, the parameter U_t is calculated, then it is called along with the demand weights. Also, if the model has the first run in the current work shift, ambulances have zero relocation workload ($\beta_k^t = 0$). Otherwise, their relocation workload will be called from the last run.

In addition to working hours, the user must assign available ambulances to base stations in the user interface environment according to emergency system information. By determining the location of ambulances, the parameter L_{kf}^t is called, and according to Eq. (20), the parameter V_{jt} is calculated.

$$\sum_{k \in K} L_{kf}^t = V_{jt} \quad \forall f = j \in M, t \in (0, T) \quad (20)$$

By calculating the total number of ambulances at each base station and calling the coverage parameters in each time range, the number of each demand point's coverages is calculated based on Eq. (21), (22), and (23).

$$\begin{aligned} a) \text{ if } H_{ij} = 1 \text{ then } O_{ij} = V_{jt} \quad \forall i \in N, j \in M, t \in (0, T) \\ b) \sum_{j \in M} O_{ij} = O_{it} \quad \forall i \in N, t \in (0, T) \end{aligned} \quad (21)$$

$$\begin{aligned} a) \text{ if } S_{ij} = 1 \text{ then } E_{ij} = V_{jt} \quad \forall i \in N, j \in M, t \in (0, T) \\ b) \sum_{j \in M} E_{ij} = E_{it} \quad \forall i \in N, t \in (0, T) \end{aligned} \quad (22)$$

$$\begin{aligned} a) \text{ if } R_{ij} = 1 \text{ then } Q_{ij} = V_{jt} \quad \forall i \in N, j \in M, t \in (0, T) \\ b) \sum_{j \in M} Q_{ij} = Q_{it} \quad \forall i \in N, t \in (0, T) \end{aligned} \quad (23)$$

The parameters O_{ij} , E_{ij} , and Q_{ij} are virtual parameters that represent the number of coverages created by base station j for the demand point i in r_1 , r_2 , and r_3 minutes, respectively. After that, the constant parameters of the model are called and the model will be executed. By displaying the results, the user can select the desired movements and apply them to the system. If the user wants to continue running the model in the current shift, the relocation workload of ambulances that have been relocated will be updated according to Eq. (24). Otherwise, if the user wants to continue in the next shift, the parameter β_k^t will be zero.

$$\begin{aligned} a) \text{ if } Y_{fj}^k = 1 \text{ then } \quad \forall k \in K, f \neq j \in M \\ b) \beta_k^{t \text{ next run}} = \beta_k^{t \text{ last run}} + B_{fj} \quad \forall k \in K, f \neq j \in M, t \in (0, T) \end{aligned} \quad (24)$$

The user interface environment is displayed in Fig. 3. Although we have used the C++ programming language in the .Net framework to develop the decision support system, this language does not allow us to build the user interface. Therefore, in the .Net framework in the Visual Studio environment, the Common Language Runtime (CLR) tool has been used to design the user interface.

Based on the designed user interface, the user must first specify the working hours and location of available ambulances at the base stations. After running the model, the user can select some of the results provided by the model and apply them to the system. After applying the results, the user will continue to run the model either in the current shift or in the next shift. To select each of these states, the required options in the user interface are considered.

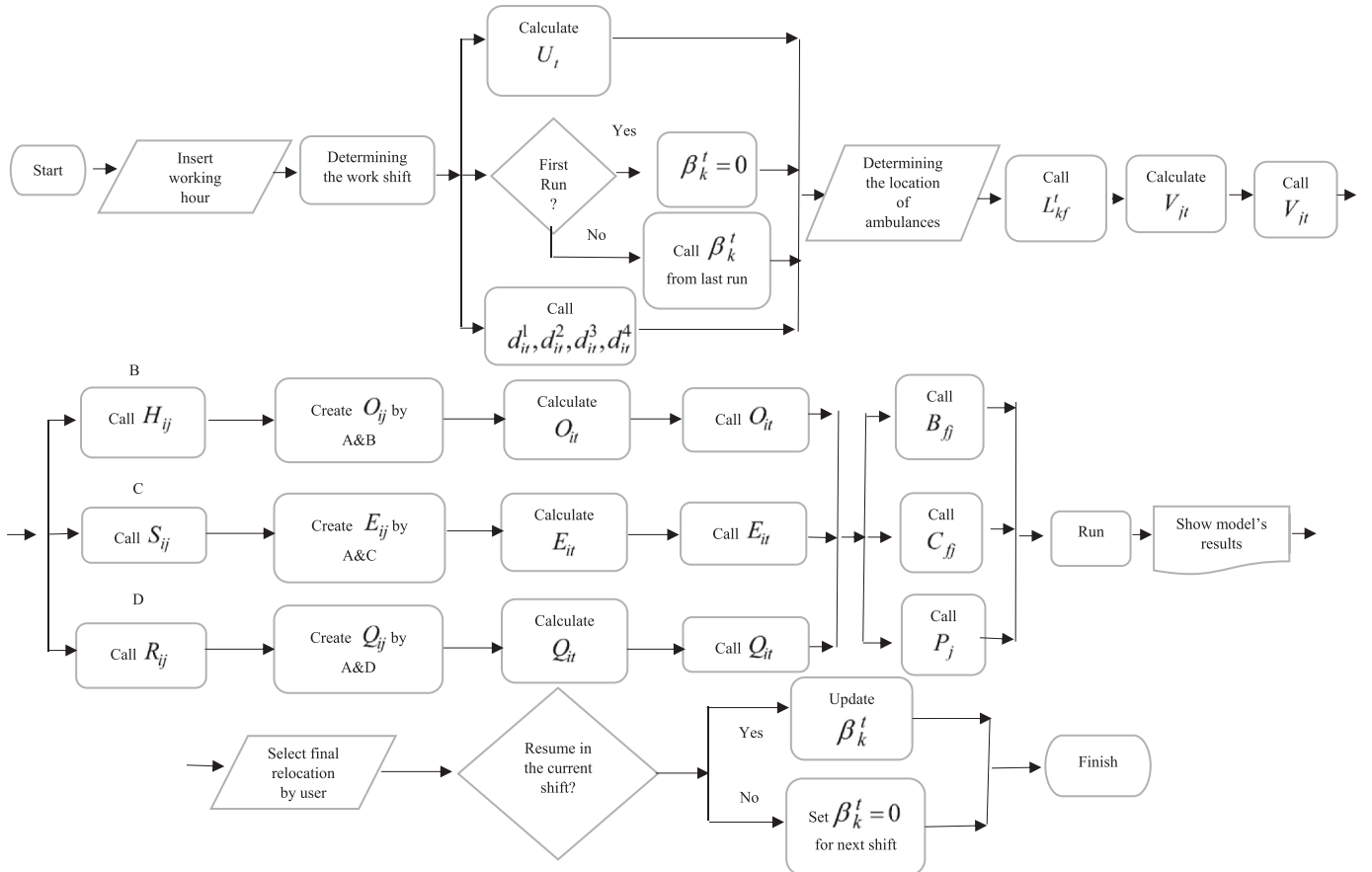


Fig. 3. Screenshot of the graphical user interface.

Fig. 2. A process model of the proposed DSS.

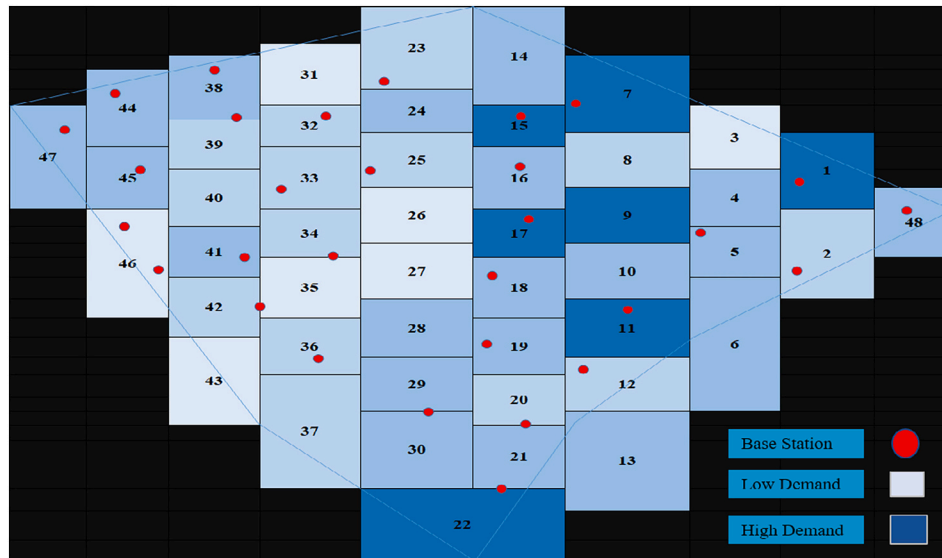


Fig. 4. Spatial demand distribution for the 48 demand zones in the eastern area of Tehran

4.3. Real-time risk assessment for relocation decisions

The user must implement all, some, or none of the recommended relocations each time the model is run. We designed the real-time risk assessment DSS to serve as a clear and broad guidance for this decision. RRARR identifies critical characteristics for forecasting various demand types (i.e., Table 4). The notations for the RRARR indices and parameters are shown in Table 5. The RRARR is developed in the following manner:

Step 1. Select one or more of the characteristics shown in Table 4 ($i = 1, \dots, |I|$).

Step 2: Determine the importance weight (w_i) of selected features

using Best-Worst-Method [50].

Step 3: Subdivide the study area into sub-zones ($j = 1, \dots, |J|$).

Step 4. Assign the selected features to one of the following categories: 1) a feature of historical data in which we do not have access to earlier statistics but only to the current value. For instance, imagine "Pollution level" is one of our selected features and we lack access to historical data for this value.

2) A feature of historical data that makes prior statistics accessible 3) DSS output is available in real-time. If the chosen feature comes within category 2 of the preceding step, utilize the following formulas to determine the variance of each element and then the normalized risk. Otherwise, go to Step 6. It is important to provide the time unit for the

Table 4
List of important features for in risk determination of different zones.

Feature name	Feature type (Obj/ Subj)	Reference
Historical data	Month/Season	Time (Obj)
	Day of week	Time (Obj)
	Time of the day	Time (Obj)
	Holiday	Time (Obj)
	Special events	Event (Subj)
	# Weekdays population	Socioeconomic (Obj)
	#Holiday and night population	Socioeconomic (Obj)
	% of senior residents (65 and older)	Socioeconomic (Obj)
	Special place	Location (Subj)
	Traffic flow	Location (Obj)
	% Road crash	Location (Obj)
	Pollution level	Location (Obj)
	Unusual call volume	Random (Subj)
	# Coverage by idle ambulances in r_1 range for each zone	Model output (Obj)
	# Coverage by idle ambulances in r_2 range for each zone	Model output (Obj)
	# Coverage by idle ambulances in r_3 range for each zone	Model output (Obj)
DSS data		

Table 5
Notations for the RRARR.

Sets & indices	$i \in 1, \dots, I $	Features
	$j \in 1, \dots, J $	Zones
Parameters	W_i	Weight of selected feature i
	Var_{ij}	Variance of feature i for zone j
	NR_{ij}	Normalized value of variance of feature i for zone j
	TR_j	Total risk of zone j

features prior to computing the variance. For instance, we can use the month as the time unit to determine the degree of pollution in each zone. As a result, we now know the monthly variance of the pollution index for each zone.

Step 5. If the selected feature falls into category 2 of the previous step, calculate the variance of each element and then calculate the normalized risk using the following formulas. Otherwise, go to Step 6.

Note that the time unit for the features should be specified before calculating the variance. For example, we can consider month as a time unit to calculate the Pollution level for each zone. So, in this case, we have the variance value of the pollution index for each zone per month.

$$MaxVar_i = \underbrace{Max(Var_{ij})}_{\text{between all } j \in J} \quad \forall i \in I \quad (25)$$

$$NR_{ij} = \frac{Var_{ij}}{MaxVar_i} \quad \forall i \in I, j \in J \quad (26)$$

Step 6. Use the corresponding formulas to find the normalized value for the features belong to categories 1 and 3. Positive criteria are the ones that the higher value is the better in terms of minimizing the risk such as the number of coverages by idle ambulances in r_1 range for each zone. However, the negative criteria are the ones that the lower value is equivalent to less risk on the demand side such as pollution level or percentage of senior residents who are 65 and older in any zone.

Positive Criteria → Negative Criteria (27)

$$\frac{Min^{Criterion}}{Value^{Criterion}} \rightarrow \frac{Value^{Criterion}}{Max^{Criterion}}$$

Step 7. Calculate the overall risk for each zone by multiplying the

normalized risk of the selected feature (or elements of that feature if it has more than one dimension) by the following formula:

$$TR_j = \prod_{i=1}^{|I|} w_i NR_{ij} \quad \forall j \in J \quad (28)$$

Step 8. Calculate the maximum and minimum risk scores for each zone. Then, determine the range of possible scores. We utilize this range to determine the incremental value required to create four distinct levels of urgency as a guideline.

$$MaxTR = Max(TR_1, TR_2, \dots, TR_{J-1}, TR_J) \quad (29)$$

$$MinTR = Min(TR_1, TR_2, \dots, TR_{J-1}, TR_J) \quad (30)$$

$$Range = MaxTR - MinTR \quad (31)$$

$$\alpha = \frac{Range}{4} \quad (32)$$

Each zone will be assigned one of the following levels of urgency based on their total risk score:

- I. j^{th} zone is a “Low risk” zone if $TR_j \in [MinTR, MinTR + \alpha]$
- II. j^{th} zone is a “Medium risk” zone if $TR_j \in [MinTR + \alpha, MinTR + 2\alpha]$
- III. j^{th} zone is a “High risk” zone if $TR_j \in [MinTR + 2\alpha, MinTR + 3\alpha]$
- IV. j^{th} zone is a “Very High risk” zone if $TR_j \in [MinTR + 3\alpha, MaxTR]$

Step 9. Finally, suppose the DSS runs the model and makes some recommendations for improving the coverage rat by relocations. We propose the following guidelines for implementing the recommended relocation decisions (i.e., Table 6). We classify actions in Table 6 into five categories: Highest Priority (HTP), High Priority (HP), Medium Priority (MP), Low Priority (LP), and Lowest Priority (LP) (LTP).

5. Computational experiments

The case study is related to the eastern part of Tehran in Iran, with a population of 3 million people and an area of 170 km². Thirty ambulances are stationed at thirty bases to answer calls. We have divided this area into 48 areas with approximately equal ranges, according to Fig. 4. The data includes 14,000 calls over a one-month period. The statistical summary of the collected data is shown in Table 7. The model is executed on a laptop with an Intel Core i5-4200M 2.50 GHz processor with 6 GB RAM running the operating system Windows 10 Enterprise (64-bit). The maximum run time is 40 s.

We simulated our proposed DSS (RRP) for one week using real data to quantify the potential benefits of policies recommended by the DSS over the static policy (SP). The results of our model and the static policy will be compared using five performance criteria: average number of ambulances available to cover calls (ANAA); average coverage rate (ACR); average response time (ART); average workload of each ambulance in one shift (AWA); and total working time of all ambulances in one shift (TWAA).

Table 6
Priority guidelines for relocation decisions.

		Destination zone			
		“Low risk”	“Medium risk”	“High risk”	“Very High risk”
Start Zone	“Low risk”	MP	HP	HTP	HTP
	“Medium risk”	LP	MP	HP	HTP
	“High risk”	LTP	LP	MP	HP
	“Very High risk”	LTP	LTP	LP	MP

Table 7

Summary statistics of the real dataset of EMS in the eastern part of Tehran, Iran.

	d_{it}^1	d_{it}^2	d_{it}^3	d_{it}^4	All Calls
Average number of calls per shift	179	11	30	16	236
Call arrival rate per hour	14	1	3	1	19
Dispatch to hospital	93%	100%	39%	1%	81%

5.1. Proposed approach and obtained results

The model implementation process is based on the four steps described in Section 3. Some data, such as call arrival time, call type, and call request area are fixed in both our approach and the static approach. Calls enter the system at specific times and from specific demand areas. After entering the call, the number of ambulances in the required time range is determined by the parameters O_{it} , E_{it} and Q_{it} . If the number of available ambulances covers the number of required ambulances, it means that the call will be responded to within the required range. Otherwise, the nearest ambulances will respond to the call in the next time range. Also, in this process, the dependence of demand coverage on previous relocations is considered. That is, calls that we could not cover if we had not relocated are different from other calls. After dispatch, the ambulance has been working at the scene for some time. This time is based on actual call data. Ambulances may then move patients to the hospital. According to the real data, 81% of the calls have been sent to the hospital. Also, the time of dispatch from the scene to the hospital is considered based on the actual call data. We take the total of ambulances' work time from the moment of dispatch to completion of the mission at the scene or hospital. Therefore, since the duration of stay on scene and the duration of dispatch from the scene to the hospital for each call are the same in both policies, the call response time, which is different in our approach and the static policy (SP), is the difference in total activity time. As stated in Section 3, we have considered two criteria for relocating ambulances in the system. If either of these criteria is met, the required information will be entered in the user interface, and the new locations of the ambulances will be determined. The movement time of ambulances in each relocation will be considered as their relocation activity. We show this time by β_k^t in the mathematical model. Also, while displaying the results, the user can select some of the provided results and apply them to the system. In this case, the β_k^t parameter will be updated only for selected ambulances. The weights of objective functions (1) and (2) are calculated to be 0.64 and 0.36, respectively. We used the fuzzy AHP method to calculate the weight of objectives. This approach is based on the method developed by Chang [20]. Also, due to the non-homogeneity of the objective functions, we first normalize the objective functions, and by assigning the calculated weights to them, we turn the model into one with a single objective function using the *weighted sum* approach. This function seeks to minimize the respective gaps between the two original objective functions and their ideal values. Eq. (33) shows the normalized single-objective model. After model implementation, performance criteria were extracted and compared with the static policy (SP). Real data was executed for a period of 7 days.

We performed the simulation in all work shifts (14 work shifts in total) and calculated the performance measures in all shifts.

Table 8

Comparison of the results for the ANAA and the ACR for each call.

Call type	ANAA			ACR		
	RRP*	SP	%Improvement	RRP*	SP	%Improvement
d_{it}^1	1.32	0.84	57%	88%	61%	44%
d_{it}^2	2	1.53	31%	82%	55%	49%
d_{it}^3	2.4	1.91	26%	90%	73%	23%
d_{it}^4	2.93	2.46	19%	100%	88%	14%
All Calls	1.6	1.12	54%	89%	73%	22%

* The compared results against the existing system are statistically significant.

The results displayed are the average performance measures calculated in all shifts. The first performance measure is the average number of ambulances available for each type of call (ANAA). Using the parameters O_{it} , E_{it} , and Q_{it} , we have calculated the ANAA each time a call is made. Table 8 shows the results for all four types of calls. The next performance measure is the average coverage rate (ACR). If a call's response time is within the required time range, it means that the call is covered. In Table 8, we categorize the ACR by demand type and compare them to the static policy (SP). In the SP, the ACR is calculated based on the call response time. If the response time is within the required range, it means the call is covered.

As Table 8 shows, real-time relocation policy has led to the optimal performance of resources, and the availability of ambulances at each call has improved over the static policy. In addition, 89% of calls were covered within their required range, and 11% of calls were covered by the closest ambulances in the next time range. This is while the ACR in the static policy is 73%. The average response time (ART) in both policies is given in Table 9. Our policy has decreased response time by an average of 4.4 min. According to Hakon Leknes et al. [36] and Andreas Bürger et al. [17], reducing reaction time improves survival and hospital release rates. In current study, the reported improvement in response times is expected to also positively impact the discharge rate.

One of the important aspects is to pay attention to the workload limitation due to relocation. The results show that executing the relocation policy in a work shift creates an average of 13.8 min of extra workload for each ambulance. However, by decreasing the response time, this policy reduces the average activity of each ambulance by 24 min per shift.

This implies that implementing the relocation policy will have a significant impact on reducing response time, hence reducing travel time and, as a result, reducing ambulance activity. We compare the key elements of ambulance activity in relocation policy with static policy in Table 10 to make it clearer.

Ambulances undertake an average of 13.8 min of relocation per shift, according to relocation policy. However, as Table 9 demonstrates, our strategy has resulted in a 35% reduction in ambulance response time. Indeed, as shown in Table 11, our approach has resulted in a 9% reduction in the average overall activity of an ambulance during a single shift when compared to the static policy. Also, this improvement can be recognized in the total working time of all ambulances in one shift (TWAA).

For more clarity, we provide a numerical example from our case study: Suppose ambulance x is placed at base 1 and the EMS uses the

$$\min \lambda_1 \left(\frac{\max_1 - \left(\sum_{i \in N} (d_{it}^1 X_i^1 + d_{it}^2 X_i^2 + d_{it}^3 Z_i + d_{it}^4 W_i) \right)}{\max_1 - \min_1} \right) + \lambda_2 \left(\frac{\left(\sum_{k \in K} \sum_{j \in M} \sum_{j' \in M} Y_{jj'}^k B_{jj'} C_{jj'} \right) - \min_2}{\max_2 - \min_2} \right) \quad (33)$$

Table 9

ART (in min) for static policy and our approach, for each type of call.

Call type	RRP*	SP	%Improvement
	ART	ART	
d_{it}^1	6.5	9.2	29%
d_{it}^2	6.5	8.8	26%
d_{it}^3	12.6	17.2	27%
d_{it}^4	19.4	27.5	29%
All Calls	8.2	12.6	35%

* The compared results against the existing system are statistically significant.

Table 10

Activity components' comparison between static and relocation policy from EMS's perspective.

Activity components	Response time	On scene care	Transfer to hospital	Back to base	Relocation
SP	✓	✓	✓	✓	
RRP	✓	✓	✓	*✓	✓

* When a mission is over, the ambulance would either go to another base or come back to its base in RRP.

Table 11

Comparison of results for the working time of ambulances (in minutes).

	AWA	TWAA
RRP*	248	7442
SP	272	8174
% Improvement	9%	9%

* The compared results against the existing system are statistically significant.

static policy. Imagine a call from demand point 4 enters the system and ambulance x will be sent to the area to respond to the call. After the mission is completed, the ambulance returns to the home base (base 1). Another call is entered from an area relatively far from the first call. Because there is no ambulance at the closest base to the demand area, ambulance x must respond to it. But the response time is longer than the previous call. The total activity of ambulance x to respond to two calls in static policy is 128 min as follows (See Table 12):

We also review the response to these calls in our real-time relocation policy. In RRP, after completing the mission in the first call, the system decides to relocate and ambulance x is relocated to base 2. After the second call arrives, the closest base to respond to it is base 2, and ambulance x is dispatched to the scene at a shorter response time. The total activity of ambulance x to respond to these calls is as follows (see Table 13):

The total activity of ambulance x to respond to two calls in real-time relocation policy is 99 min. Note that in this case, ambulance x is relocated after completing the mission. In some cases, ambulances return to home bases after completing the mission, and then the decision to relocate the ambulance may be made later.

Table 12Activity time of ambulance x in static policy (in minutes).

Static policy	Response time	On scene care	Transfer to hospital	Back to base	Total
First Call	8	15	14	15	52
Second Call	19	15	14	28	76
Total Activity	27	30	28	43	128

Table 13Activity time of ambulance x in real-time relocation policy (in minutes).

Real-time relocation policy	Response time	On scene care	Transfer to hospital	Back to base (Relocation)	Total
First Call	8	15	14	13	50
Second Call	6	15	14	14	49
Total Activity	14	30	28	27	99

6. Summary and conclusion

In this study, we designed and developed a mathematical model-driven decision support system for the real-time ambulance relocation problem. The system relocates ambulances in real-time based on a two-objective mathematical model. In addition to seeking maximum demand coverage, this model takes into account the relocation costs incurred by the emergency system. On the demand side, our model defines four types of demand; using real data, we extract the demand pattern of demand points in different time periods, and we maximize coverage based on the requirements of demand points in each time period. On the EMS side, we attempt to avoid long trips and decrease the costs of the system by defining the objective function of minimizing travel time due to relocation. We developed a DSS that is used to analyze the risk associated with recommended relocations in real-time. That is, when the model has been run, the user is provided with the option of selecting any recommended relocation and our developed DSS offer some practical guidelines for the user in the implementation phase of relocation.

We formulated the coverage constraints based on the number of coverages of demand points at the time of the decision, the number of coverages increased due to relocation, and the number of coverages decreased due to relocation. In our approach, due to the variation of the system state, the number of coverages of demand points is calculated in real-time and called in the model. We have implemented the developed system in the eastern part of Tehran in Iran, using real data, and compared it with the existing static approach. The results showed that the proposed system has led to better management of limited resources by improving the coverage rate and the response time. Also, the proposed approach adds an average of 13.8 min of workload due to relocation per shift for each ambulance. However, by decreasing the response time, our policy reduces the travel time of ambulances and consequently reduces their overall workload. Thus, compared to the static policy, our policy results in a 9% decrease in the average total working time of all ambulances per shift.

In future research, different types of ambulances can be considered. For example, motor lances have an essential role in the initial and temporary responses to demands. In Tehran, their temporary location is determined based on the experience of emergency users. Our policy can be applied to determine the temporary locations of motor lances with a dynamic approach in future research. It can also be useful to apply different dispatch policies and integrate dispatch decisions with relocation decisions. Sensitivity analysis of relocation time criteria and investigation of changes in performance metrics are two recommendations for future research that we did not pursue. Finally, in future research, the use of a robust optimization approach is recommended in order to take into account the demand uncertainty.

Credit author statement

All authors have contributed collectively and equally to the original and revised version of this manuscript in terms of conceptualization, implementation, and paper writing.

References

- [1] L. Aboueljinnane, E. Sahin, Z. Jemai, A review on simulation models applied to emergency medical service operations, *Comput. Ind. Eng.* 66 (4) (2013) 734–750.

- [2] R. Alanis, A. Ingolfsson, B. Kolfal, A Markov chain model for an EMS system with repositioning, *Prod. Oper. Manag.* 22 (1) (2013) 216–231.
- [3] T. Andersson, P. Värbrand, Decision support tools for ambulance dispatch and relocation, *J. Oper. Res. Soc.* 58 (2) (2007) 195–201.
- [4] R. Aringhieri, M.E. Bruni, S. Khodaparasti, J.T. van Essen, Emergency medical services and beyond: addressing new challenges through a wide literature review [Internet], *Comput. Oper. Res. Els.* 78 (2017) 349–368.
- [5] R. Aringhieri, G. Carello, D. Morale, Supporting decision making to improve the performance of an Italian emergency medical service, *Ann. Oper. Res.* 236 (1) (2016) 131–148.
- [6] M.O. Ball, F.L. Lin, Reliability model applied to emergency service vehicle location, *Oper. Res.* 41 (1) (1993) 18–36.
- [7] T.C. van Barneveld, S. Bhulai, R.D. van der Mei, A dynamic ambulance management model for rural areas, *Health Care Manag. Sci.* 20 (2) (2017) 165–186.
- [8] T.C. Van Barneveld, S. Bhulai, R.D. Van Der Mei, The effect of ambulance relocations on the performance of ambulance service providers, *Eur. J. Oper. Res.* 252 (1) (2016) 257–269.
- [9] T.C. van Barneveld, R.D. van der Mei, S. Bhulai, Compliance tables for an EMS system with two types of medical response units, *Comput. Oper. Res.* 80 (2017) 68–81.
- [10] T. van Barneveld, C. Jagtenberg, S. Bhulai, R. van der Mei, Real-time ambulance relocation: assessing real-time redeployment strategies for ambulance relocation, *Socio Econ. Plan. Sci.* 2018 (62) (2016) 129–142. February.
- [11] A. Başar, B. Çatay, T. Ünliuyurt, A multi-period double coverage approach for locating the emergency medical service stations in Istanbul, *J. Oper. Res. Soc.* 62 (4) (2011) 627–637.
- [12] R. Batta, J.M. Dolan, N.N. Krishnamurthy, Maximal expected covering location problem, Revisited *Transp. Sci.* 23 (4) (1989) 277–287.
- [13] V. Bélanger, Y. Kergosien, A. Ruiz, P. Soriano, An empirical comparison of relocation strategies in real-time ambulance fleet management, *Comput. Ind. Eng.* 94 (2016) 216–229.
- [14] V. Bélanger, A. Ruiz, P. Soriano, Recent optimization models and trends in location, relocation, and dispatching of emergency medical vehicles, *Eur. J. Oper. Res.* 272 (1) (2019) 1–23.
- [15] V. Bélanger, E. Lanzarone, P. Soriano, E. Lanzarone, A. Ruiz, P. Soriano, The ambulance relocation and dispatching problem, in: *Tech Rep CIRRELT-2015-59*, CIRRELT, 2015 (November).
- [16] P.L. Van Den Berg, K. Aardal, Time-dependent MEXCLP with start-up and relocation cost, *Eur. J. Oper. Res.* 242 (2) (2015) 383–389.
- [17] A. Bürger, J. Wnent, A. Bohn, T. Jantzen, S. Brenner, R. Lefering, et al., The effect of ambulance response time on survival following out-of-hospital cardiac arrest—an analysis from the German resuscitation registry, *Dtsch. Arztebl. Int.* 115 (33–34) (2018) 541–548.
- [18] A.S. Carvalho, M.E. Captivo, I. Marques, Integrating the ambulance dispatching and relocation problems to maximize system's preparedness, *Eur. J. Oper. Res.* 283 (3) (2020) 1064–1080.
- [19] A.M. Caunhye, X. Nie, S. Pokharel, Optimization models in emergency logistics: a literature review, *Socio Econ. Plan. Sci.* 46 (1) (2012) 4–13.
- [20] D.Y. Chang, Applications of the extent analysis method on fuzzy AHP, *Eur. J. Oper. Res.* 95 (3) (1996) 649–655.
- [21] N. Channouf, P. L'Ecuyer, A. Ingolfsson, A.N. Avramidis, The application of forecasting techniques to modeling emergency medical system calls in Calgary, Alberta, *Health Care Manag. Sci.* 10 (1) (2007) 25–45.
- [22] R.L. Church, C. Revell, The maximal covering location problem, *Pap. Reg. Sci.* 71 (3) (1974) 199–215.
- [23] D. Degel, L. Wiesche, S. Rachuba, B. Werners, Time-dependent ambulance allocation considering data-driven empirically required coverage, *Health Care Manag. Sci.* 18 (4) (2015) 444–458.
- [24] S. Enayati, M.E. Mayorga, H.K. Rajagopalan, C. Saydam, Real-time ambulance redeployment approach to improve service coverage with fair and restricted workload for EMS providers, *Omega (United Kingdom)* 79 (2018) 67–80.
- [25] S. Enayati, O.Y. Özalpın, M.E. Mayorga, C. Saydam, Ambulance redeployment and dispatching under uncertainty with personnel workload limitations, *IIE Trans.* 50 (9) (2018) 777–788.
- [26] M. Gendreau, G. Laporte, F. Semet, The maximal expected coverage relocation problem for emergency vehicles, *J. Oper. Res. Soc.* 57 (1) (2006) 22–28.
- [27] M. Gendreau, Solving an ambulance location model by tabu search, *Locat. Sci.* 5 (2) (1997) 75–88.
- [28] M. Gendreau, G. Laporte, F. Semet, A dynamic model and parallel tabu search heuristic for real-time ambulance relocation, *Parallel Comput.* 27 (12) (2001) 1641–1653.
- [29] S. Ibrı, M. Nourelfath, H. Drias, A multi-agent approach for integrated emergency vehicle dispatching and covering problem, *Eng. Appl. Artif. Intell.* 25 (3) (2012) 554–565.
- [30] C.J. Jagtenberg, S. Bhulai, R.D. van der Mei, An efficient heuristic for real-time ambulance redeployment, *Oper. Res. Health Care* 4 (2015) 27–35.
- [31] R.D. Kamenetzky, L.J. Shuman, H. Wolfe, Estimating need and demand for prehospital care, *Oper. Res.* 30 (6) (1982) 1148–1167.
- [32] T.O. Kvalseth, J.M. Deems, Statistical models of the demand for emergency medical services in an urban area, *Am. J. Public Health* 69 (3) (1979) 250–255.
- [33] S.S.W. Lam, C.B.L. Ng, F.N.H.L. Nguyen, Y.Y. Ng, M.E.H. Ong, Simulation-based decision support framework for dynamic ambulance redeployment in Singapore, *Int. J. Med. Inform.* 106 (February) (2017) 37–47.
- [34] S. Lee, A new preparedness policy for EMS logistics, *Health Care Manag. Sci.* 20 (1) (2017) 105–114.
- [35] C. Lei, W. Lin, L. Miao, A stochastic emergency vehicle redeployment model for an effective response to traffic incidents, *IEEE Trans. Intell. Transp. Syst.* 16 (2) (2015) 898–909.
- [36] H. Leknes, E.S. Aartun, H. Andersson, M. Christiansen, T.A. Granberg, Strategic ambulance location for heterogeneous regions, *Eur. J. Oper. Res.* 260 (1) (2017) 122–133.
- [37] X. Li, Z. Zhao, X. Zhu, T. Wyatt, Covering models and optimization techniques for emergency response facility location and planning: a review, *Math. Meth. Oper. Res.* 74 (3) (2011) 281–310.
- [38] M. Maleki, N. Majlesinasab, Sepehri M. Mehdi, Two new models for redeployment of ambulances, *Comput. Ind. Eng.* 78 (2014) 271–284.
- [39] A.J. Mason, Simulation and real-time optimised relocation for improving ambulance operations, *Handb. Healthc. Oper. Manag.* 184 (2015) 1–18.
- [40] L.A. McLay, A maximum expected covering location model with two types of servers, *IIE Trans. (Inst. Ind. Eng.)* 41 (8) (2009) 730–741.
- [41] L.A. McLay, M.E. Mayorga, Evaluating emergency medical service performance measures, *Health Care Manag. Sci.* 13 (2) (2010) 124–136.
- [42] H.K. Mell, S.N. Mumma, B. Hiestand, B.G. Carr, T. Holland, J. Stopyra, Emergency medical services response times in rural, suburban, and urban areas, *JAMA Surg.* 152 (10) (2017) 983–984.
- [43] M. Moeni, Z. Jemai, E. Sahin, Location and relocation problems in the context of the emergency medical service systems: a case study, *Cent. Eur. J. Oper. Res.* 23 (3) (2015) 641–658.
- [44] S.S. Mohri, H. Haghsheenas, An ambulance location problem for covering inherently rare and random road crashes, *Comput. Ind. Eng.* 151 (October) (2021), 106937.
- [45] R. Nair, E. Miller-Hooks, Evaluation of relocation strategies for emergency medical service vehicles, *Transp. Res. Rec.* 2137 (2009) 63–73.
- [46] J. Naoum-Sawaya, S. Elhedhli, A stochastic optimization model for real-time ambulance redeployment, *Comput. Oper. Res.* 40 (8) (2013) 1972–1978.
- [47] H.K. Rajagopalan, C. Saydam, J. Xiao, A multiperiod set covering location model for dynamic redeployment of ambulances, *Comput. Oper. Res.* 35 (3) (2008) 814–826.
- [48] J.F. Repede, J.J. Bernardo, Developing and validating a decision support system for locating emergency medical vehicles in Louisville, Kentucky, *Eur. J. Oper. Res.* 75 (3) (1994) 567–581.
- [49] M. Reuter-Oppermann, P.L. van den Berg, J.L. Vile, Logistics for emergency medical service systems, *Health Syst.* 6 (3) (2017) 187–208.
- [50] J. Rezaei, Best-worst multi-criteria decision-making method, *Omega (United Kingdom)* 53 (2015) 49–57.
- [51] M.S. Daskin, A maximum expected covering location model, *Transp. Sci.* 17 (1) (1983) 47–70 (August 2015).
- [52] A. Sathe, E. Miller-Hooks, Optimizing location and relocation of response units in guarding critical facilities, *Transp. Res. Rec.* 1923 (2005) 127–136.
- [53] C. Saydam, H.K. Rajagopalan, E. Sharer, K. Lawrimore-Belanger, The dynamic redeployment coverage location model, *Health Syst.* 2 (2) (2013) 103–119.
- [54] V. Schmid, Solving the dynamic ambulance relocation and dispatching problem using approximate dynamic programming, *Eur. J. Oper. Res.* 219 (3) (2012) 611–621.
- [55] V. Schmid, K.F. Doerner, Ambulance location and relocation problems with time-dependent travel times, *Eur. J. Oper. Res.* 207 (3) (2010) 1293–1303.
- [56] H. Setzler, C. Saydam, S. Park, EMS call volume predictions: a comparative study, *Comput. Oper. Res.* 36 (6) (2009) 1843–1851.
- [57] K. Sudtachat, M.E. Mayorga, S. Chanta, L.A. Albert, Joint relocation and districting using a nested compliance model for EMS systems, *Comput. Ind. Eng.* 2020 (142) (2019) 106327. July.
- [58] K. Sudtachat, M.E. Mayorga, L.A. Mclay, A nested-compliance table policy for emergency medical service systems under relocation, *Omega (United Kingdom)* 58 (2016) 154–168.
- [59] L. Trujillo, G. Álvarez-Hernández, Y. Maldonado, C. Vera, Comparative analysis of relocation strategies for ambulances in the city of Tijuana, Mexico, *Comput. Biol. Med.* 116 (2020).
- [60] J.L. Vile, J.W. Gillard, P.R. Harper, V.A. Knight, Predicting ambulance demand using singular spectrum analysis, *J. Oper. Res. Soc.* 63 (11) (2012) 1556–1565.

Mahdi Hajiali studied Industrial Engineering at Bu-Ali Sina University and received his Master's degree in Logistics and Supply Chain Management from Iran University of Science and Technology in 2020. His current research vision is concentrated mainly on healthcare logistics and healthcare network design. He has been employing mathematical models to design and develop healthcare-focused decision support systems.

Ebrahim Teimoury is an Associate Professor of Industrial Engineering at Iran University of Science and Technology. He received his Ph.D. from Iran University of Science and Technology in 2000 and initiated his work as a faculty member at SIE since 2001. His research vision is concentrated mainly on Supply Chain Management. He teaches Supply Chain Management, E-supply Chain Management, Socio-economic Systems Modeling, Systems Engineering, Queuing Theory, Probability Theory, and Mathematical Statistics.

Meysam Rabiee is a PhD student in Operations and Business Analytics at the University of Oregon. He has served as a Faculty Member and the Program Director of the Industrial Engineering Department at the Bu-Ali Sina University in Iran prior his current appointment at University of Oregon. His research has appeared in major journals including International Journal of Production Economics and International of Production Research, among others. His research interests include Decision Support Systems, Sustainable Supply Chain, HealthCare Delivery Network Design, Multi-Criteria Decision Making, Multi-Objective Optimization, Data Analytics and Scheduling.

Dursun Delen is the holder of Spears and Patterson Endowed Chairs in Business Analytics, Director of Research for the Center for Health Systems Innovation, and Regents Professor of Management Science and Information Systems in the Spears School of Business at Oklahoma State University. He has authored/co-authored more than 120 journal papers and numerous peer-reviewed conference proceeding articles. His research has appeared in major journals including Decision Sciences, Journal of Production Operations Management, Decision Support Systems, Communications of the ACM, Computers and Operations Research, Computers in Industry, Journal of the American Medical Informatics Association, Artificial Intelligence in Medicine, International Journal of Medical Informatics,

Health Informatics Journal, among others. He has published 10 books/textbooks in the broad area of Business Intelligence and Business Analytics. He is often invited to national and international conferences and symposiums for keynote addresses, and companies and government agencies for consultancy/education projects on data science and analytics related topics. Dr. Delen served as the general co-chair for the 4th International Conference on Network Computing and Advanced Information Management (held in Seoul, South Korea), and regularly chairs tracks and minitracks at various information systems and analytics conferences. He is currently serving as the editor-in-chief, senior editor, associate editor and editorial board member of more than a dozen academic journals.