Journal of Multidisciplinary Engineering Science and Technology (JMEST)
ISSN: 2458-9403
Vol. 5 Issue 10, October - 2018

An Optimization Model for Emergency Medical Service

Carlos Escobar
Dept. of Industrial Engineering,
New Mexico State University
Las Cruces, NM 88003, USA
caed@nmsu.edu

Alireza Moghimi
Applied Economics and Management,
Cornell University,
Ithaca, NY 14853, USA
am2393@cornell.edu

German Reyes
Dept. of Industrial Engineering,
New Mexico State University
Las Cruces, NM 88003, USA
greyes@nmsu.edu

Han-suk Sohn
Dept. of Industrial Engineering,
New Mexico State University
Las Cruces, NM 88003, USA
hsohn@nmsu.edu

***Abstract*—This paper presents a new hybrid algorithm and its application to a classical ambulance deployment problem. In the proposed algorithm, a discrete event simulation model is embedded into a heuristic algorithm, in order to determine a high fractile-performance ambulance deployment strategy. The applicability of the proposed algorithm is demonstrated in the case study of Doña Ana County's emergency medical service system. The results demonstrate that the proposed algorithm is a practical and flexible tool in solving realistic emergency medical service deployment problems.**

Keywords— Emergency Medical Service; discrete event simulation; heuristic algorithm; ANOVA; coefficient of variation

I. INTRODUCTION

A. Background

With the rising number of retired baby boomers and a constantly increasing trend in emergency medical service (EMS) demand, ambulance service providers are facing the most critical challenges of their history [1]. In addition, the EMS systems are facing another issue, i.e., the increment of the EMS system operating costs, which are also being impacted by the continuous rising costs of fuel, professional labor, and technology improvement. To overcome these challenges, ambulance service providers must find a way to continuously optimize resources. There are four attributes for the EMS system to be nationally considered as high performance: clinical excellence, response-time reliability, economic efficiency, and customer satisfaction within budget constraints [2]. Among them, the response-time reliability takes a major role in the patients' outcome. Therefore, the response-time reliability is the main driver of this research, and the goal of this research is to improve the ambulance deployment strategy by identifying good locations for the base station of the ambulances and its fleet size so that more accurate deployments of resources can be accomplished. This would reduce operating costs while improving the EMS system reliability.

B. Emergency Medical Service (EMS) System

EMS system consists of the organizations, individuals, facilities, and equipment where participation is required to ensure timely and medically appropriate responses to each request for out-of-hospital care and medical transportation [3]. Every response to an EMS request proceeds in stages from receipt of the request through delivery of the patient and return of the responding units to available status. Each ambulance is placed at a pre-determined base station and waits for an EMS request. When an EMS request call arrives, the control room dispatcher evaluates the system status and determines which ambulance to send to the scene. Decisions of dispatching and ambulance location are critical factors in EMS system success.

C. Ambulance Response Time (RT) and system Reliability

The ambulance response time is the interval between when the system first gains enough information to initiate a response and the time a properly equipped and staffed ambulance arrives at the scene [4]. About half the systems start the response time clock when the call is received, and some systems are at the receipt of dispatch. Yet, others start the clock when the ambulance wheels begin to roll. The national standard set by the Commission on Accreditation of Ambulance Services (CAAS), 90% of EMS request calls within an urban area should be responded to within 8-minutes, whereas the target RT for suburban and rural areas are 15:59 and 20:59, respectively. As a measure of response time performance, we use fractile response time. The fractile response time is the percent of responses within a defined time limit,

and is the preferred method to measuring system performance [2].

II. Literature Review

A. Discrete Event Simulation (DES) for Emergency Medical Service (EMS) Systems

A simulation can be defined as numerically exercising a model for the inputs in question to see how they affect the output measures of performance [6]. A computer simulation can be seen as a way to conduct such thought experiments leading to prediction, proof, and discovery [7]. Unlike the closed form of analytic models, which are often capable of being solved exactly due to a principal understanding of system relationships, complex models that require simulation are often stochastic in nature, preventing any definitive prediction. One of the earliest applications of the simulation modeling to EMS system was done by Savas [8]. He conducted a case study of the district of New York, and concluded that the dispersion of the ambulance base-location improves the efficiency of the EMS systems. Later, it was supported by Monroe who conducted a similar study for Madison, Wisconsin [9]. Fitzsimmons combined the queueing theory and simulation model to analyze the EMS system of San Fernando Valley area in Los Angeles, California [10]. A similar approach was proposed by Franci in Italy [11]. Both of them concluded that the systematic distribution of ambulances throughout the region is more important than simply increasing the total number of ambulances in the system. They also described the importance of the dimension of coverage area as well as the volume of calls generated in that specific area. Lubicz and Mielczarez described that, after the EMS system reached a certain level, simply adding more ambulances to the system will not help improve system performance [12]. They also pointed out that increasing ambulance utilization could lead the system to an unacceptable level of the system performance.

B. ANOVA

Single-factor ANOVA is used to test if the means from more than two populations or groups are equal. By using it, we are able to group the data and find an improved ambulance deployment strategy for each group. The relevant null hypothesis is “population means are equal” whereas, the alternative hypothesis is “at least two of the population means are different” [12]. The basic assumptions required to use the ANOVA include: (1) each sample is an independent random sample, (2) homoscedasticity, and (3) the distribution of the response variable follows a normal distribution. ANOVA tests the given null hypothesis versus the alternative hypothesis by using the F distribution. The F distribution is a continuous probability distribution that is most widely used with ANOVA. The post-hoc tests (i.e., usually all pairwise comparisons) in ANOVA is the follow up analysis when the null hypothesis has been rejected and additional exploration of the differences among means is needed to determine which means are significantly different from each other. There are several multi-comparison techniques with different agendas, however, the investigator should be careful to choose the most appropriate method for the particular goals of the study. Typical post-hoc tests are: LSD, Duncan, Tukey, Bonferronni, and Scheffe [12]. Due to the economic implications of adding or eliminating ambulances to the system, we use Tukey post-hoc test in this study.

III. Methodology

This section describes how various concepts and techniques of statistical data analyses, clustering algorithm, simulation modeling, and heuristic algorithm can be utilized to identify good locations of the ambulance base station and the fleet size of the ambulance. The applicability of these techniques is demonstrated in the case study of Doña Ana County’s EMS system.

A. Review of Data

Doña Ana County occupies a total area of 3,815 square miles in south-central New Mexico and is the home to more than 210,000 residents. The County is divided into three distinct ambulance zones, described below as Central, Northern, and Southern Counties (Dona Ana County Ambulance RFP, 2014). The Central Doña Ana County is primarily urban and includes many different fire districts. The City of Las Cruces is the county seat and is located in the center of the county, which is the largest user of county ambulance services. The Northern Doña Ana County is primarily rural and consists of only a few fire districts with a population of less than 2400 in each fire district (2012 Census Population). The Southern Doña Ana County is primarily suburban. In this study, each of these ambulance zones have been subdivided into smaller response time compliance areas, where there are three different response time divisions (i.e., urban, suburban, and rural) based on population density.

The County currently considers two types of reporting categories, i.e., Code1 and Code3. The Code1 is an emergency perceived as a non-life threatening situation. This type of emergency is responded without use of lights and sirens. The Code3 is an emergency perceived as a life threatening situation. This type of emergency requires the immediate dispatch of an ambulance with use of lights and sirens.

Emergency response data was collected by Mesilla Valley Regional Dispatch Authority (MVRDA) and provided by American Medical Response (AMR). The original data includes 31,614 emergency call records that cover the entire Dona Ana County for a 2 year period from January 1, 2012 to December 31, 2013. We visualize the emergency call locations on the map using ArcMaps software (see Fig. 1). The current study includes 27 potential candidate site stations for ambulance vehicles, which were identified by the

Dona Ana County and AMR. The actual locations are also depicted on the map in Fig. 1.

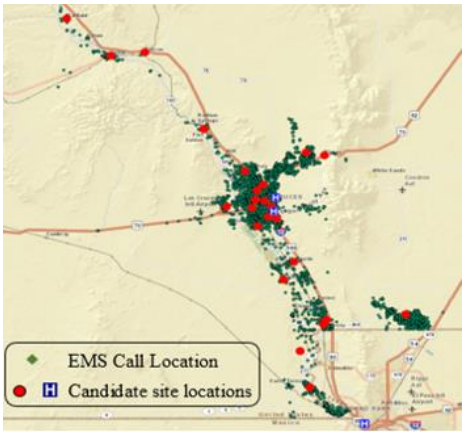

Fig. 1. Emergency call locations of Doña Ana County (2012-2013) and Initial 27 candidate site locations including three hospitals

B. Descriptive analysis

According to literature surveys, there are some critical factors for driving response time performances and reliabilities. They are (1) abilities to understand and predict call volumes based on the time of the day, (2) abilities to understand and predict call volumes based on the days of the week, (3) abilities to understand and predict call volumes based on the month of the year, and (4) abilities to understand and predict geographical locations of the calls [3]. Therefore, we reviewed the County's emergency call volumes and trends based on different times of the day, different days of the week and different month of the year, and summarize as follow. About 65% of the calls occurred from 10am to 10pm, and the highest system demand occurred during the lunch time (12pm-1pm) with an hourly average of 2.4 calls. The weekly peak demand occurs on Friday. However, there are no significant differences on the call volume among the seven days. It is also observed that relatively high demands occurred from January through May. Note that more than 55% of the calls were generated in the County during these 5 months. Next, we have conducted statistical analyses to test if all these patterns are significant, which will provide more accurate information for simulation modeling later.

C. Tests for the equality of means and coefficient of variation

On-way ANOVA was performed to test for population equal means considering time (e.g. months) as the only factor. Note that, when significant omnibus F-test was obtained, Tukey's post-hoc tests was followed to define month-groups. The data was analyzed by months to determine if emergency demand is sensitive to month of the year. According to the one-way ANOVA results, there are significant variations in the call volumes among 12 months of the year with a p-value < 0.0001. The Shapiro-Wilk test fail to reject the null hypothesis, therefore it is concluded that data is normal. Result from the Tukey's test indicates that year demand should be broken down into three groups (i.e., January through May, June, and July through December). The coefficient of variation analysis also shows that grouping the EMS request calls into peak hours (10am-10pm) and off-peak hours (10pm-10am) seems reasonable. In modeling the computer simulation as well as conducting the analysis, therefore, the EMS request call data are grouped into three distinct periods, namely, January through May, June, and July through December, and then these three periods are subdivided into peak hours (10am-10pm) and non-peak hours (10pm-10am).

D. K-means data clustering algorithm

In EMS simulation models, locating each single demand separately is computationally impractical because of considerably large amount of calls received by EMS systems. Therefore, the emergency response data were grouped together based on the similarities of their geographical location. For this purpose, we have applied *K-means* clustering algorithm to the data set to find the partition of emergency call locations so that the distances between the call locations within the same cluster are minimized. In the *K-means* algorithm, initial centroids are often chosen randomly, and the centroid is typically the mean of the points in the cluster. It is also important to note that different initializations in the *K-means* algorithm can lead to different clustering since the algorithm does not guarantee to find the global optima. In order to overcome this shortcoming, for a given number of clusters *K*, we have run the *K-means* algorithm with several different initial partitions, and we chose the one with the smallest value of the sum of squared errors (SSE). We have also run the algorithm independently for different values of *K*, and the partition that appears the most meaningful to the system is selected. Note that a cluster with a heavy weight is highly likely to generate more emergency calls than a cluster with a lighter weight in our simulation model. In general, increasing the value of *K* will reduce the sum of squared error (SSE). However, a good clustering with smaller value of *K* can lead to have a lower SSE than a poor clustering with higher value of *K*. With these in mind, we have carefully chosen *K*=18 as the desired number of clusters for the Dona Ana County's emergency call data. The *K-means* clustering algorithm applied to the County's historical emergency call data produces a set of centroids capturing patterns corresponding to the error function. They are mapped in Fig. 2, where different clusters are represented by different color and 18 centroids are denoted by asterisk.

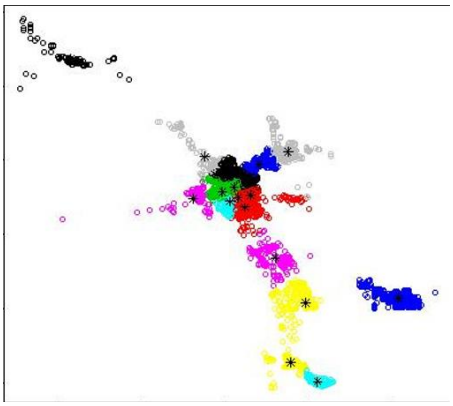

Fig. 2. 18 clusters and their centroid locations for the County's EMS request call data.

All of these 18 centroid locations are specified on the map in Fig. 3, which helps us for a better understanding of where these points are physically located in the Doña Ana County.

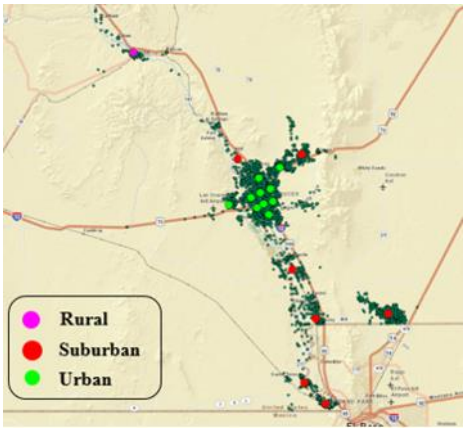

Fig. 3. Classification of the 18 centroids.

Among these 18 service locations, 10 centroids are identified as Urban areas that are located in the center of the county, which are represented by light-green dots, and we use the Standard RT of 0:08:59. Only one of them is identified as a Rural area which applies to the Standard RT of 0:20:59. This rural area is represented by pink dots. The other 7 centroids were classified as Suburban area, which are represented by red dots and we use 0:15:59 as the Standard RT. Five of them are located in the Southern Doña Ana County and two of them are in the Central Doña Ana County. All of this classification information will be used to model the geographical distribution of emergencies in our simulation model later.

E. Discrete Event Simulation

Discrete event simulation (DES) is one of the major paradigms in computational simulation modeling, which built around events and the related states of the system. The DES provides a proven venue for representing confounded systems in a traceable and rigorous manner that is particularly useful for gaining insight into complex problems. We used the SIMIO computer simulation package as our architecture to build the simulation. Our simulation model is composed of three primary modules, ambulance, patient, and control center modules (see Fig. 4).

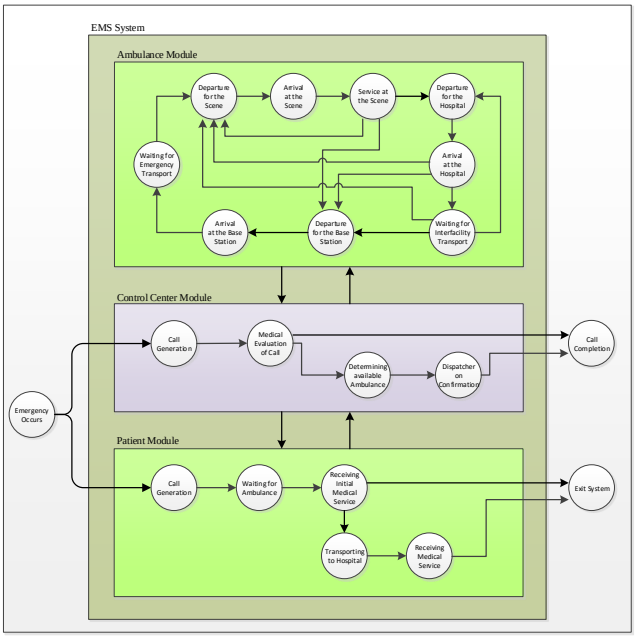

Fig. 4. Event diagram for the basic EMS system module (Source: [14]).

Interaction between modules is essential as the dispatching decision is dependent on the availability of the ambulance and the medical evaluation status of the patient.

Each ambulance is placed at a pre-determined base station and waits for an EMS request. The ambulance starts its task upon request from the control center. First, it arrives at the scene. Then, it provides initial medical treatment. After the initial treatment at the scene, the ambulance may or may not provide transportation to a hospital. If transportation is not required, the ambulance is assigned to another call or returns directly to its base from the scene. Otherwise, the ambulance transports the patient to a hospital, and then either provides inter-facility transportation service, responds to another call, or returns to its base station.

Each patient enters the system by generating an EMS request call. After a certain period of time waiting, the patient receives initial treatment at the scene. If the patient requires a higher level of medical treatment, then he/she is transported to a hospital. Otherwise, the patient exits the EMS system.

The dispatcher at the control center is the decision core of the entire EMS system. When an EMS request call arrives, the dispatcher evaluates the system status and determines the appropriate ambulance to send to the scene. The decisions of dispatching and ambulance location are critical factors in the success of the EMS system. The standard dispatch strategy is to send the closest ambulance that is either idle at its base station or returning from a previous assignment.

F. *Hybrid DES/Heuristic Algorithm*

The EMS system configuration resulted from the simulation run may be feasible or infeasible with respect to satisfying the target fractile response time. In this section, we present a hybrid algorithm, where the discrete event simulation (DES) model is imbedded into a greedy like heuristic algorithm. The heuristic algorithm aims to modify the system configuration resulted from the simulation run in order to meet the targeted fractile value of the response time. The basic idea is that the new system configuration is obtained by moving a set of ambulance vehicles between potential location sites or adding a set of ambulances on potential site locations. The complete details of the hybrid algorithm are depicted in Fig. 5.

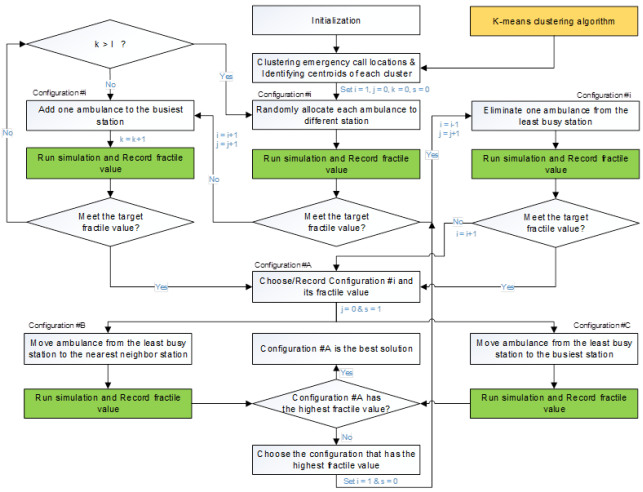

Fig. 5. Hybrid DES/Heuristic framework (Source: [13] and [14]).

IV. Results and Discussion

The hybrid DES/Heuristic algorithm was implemented on randomly generated data, and we have obtained the best system configuration for each of the six different time frames. The results are summarized in Tables I through IV, where their system performances are compared to the ones from the County's current system configuration. Note that the proposed system arrangement for the non-peak hours are the same as each other throughout the whole year.

TABLE I. Performance Comparison (January – May during peak-hours)

System Configuration	# of Ambulances Required	Fractile of RT (%)					Average RT (min)	
		Urban	Suburban	Rural	Overall Average	95% CI	Overall Average	95% CI
SYS current	10	78.35	80.33	98.79	79.14	78.25-80.03	9.52	9.27-9.76
SYS proposed_10	10	92.91	87.54	98.11	91.94	91.56-92.32	6.50	6.42-6.58
SYS proposed_best	11	93.12	91.13	97.9	92.7	92.34-93.60	6.39	6.27-6.50

TABLE II. Performance Comparison (June during peak-hours)

System Configuration	# of Ambulances Required	Fractile of RT (%)					Average RT (min)	
		Urban	Suburban	Rural	Overall Average	95% CI	Overall Average	95% CI
SYS current	10	80.83	80.63	98	81.13	78.84-83.42	8.85	8.2-9.51
SYS proposed_best	9	92.42	93.8	1	92.88	92.13-93.63	6.35	6.16-6.54
SYS proposed_10	10	93.92	93.92	99.17	93.97	93.45-94.49	5.98	5.83-6.14

TABLE III. Performance Comparison (July – December during peak-hours)

System Configuration	# of Ambulances Required	Fractile of RT (%)					Average RT (min)	
		Urban	Suburban	Rural	Overall Average	95% CI	Overall Average	95% CI
SYS current	10	81.2	85.83	1	82.73	79.59-85.87	8.25	7.69-8.76
SYS proposed_best	9	93.17	94.63	1	93.63	92.97-93.96	6.45	6.25-6.64
SYS proposed_10	10	95.35	95.47	99.52	95.45	95.09-95.81	5.74	5.63-5.86

TABLE IV. Performance Comparison (January – May during nonpeak-hours)

System Configuration	# of Ambulances Required	Fractile of RT (%)					Average RT (min)	
		Urban	Suburban	Rural	Overall Average	95% CI	Overall Average	95% CI
SYS current	7	57.13	92.8	98.57	66.4	65.23-67.57	10.72	10.3-11.0
SYS proposed_best	6	91.52	94.28	97.78	92.1	91.64-92.56	6.01	5.86-6.16
SYS proposed_7	7	94.69	94.36	97.98	94.67	93.99-95.35	5.82	5.62-6.03

For example, for the case of system configuration during peak-hours from January to May, the proposed simulation optimization model was able to find a feasible solution (i.e., *SYS proposed_best*), which satisfies the minimum compliance rate for all three service areas (see Table I). It recommends operation of 11 ambulances and records the fractile rates of 93.12%, 91.13% and 97.9% for Urban, Suburban, and Rural areas, respectively, and overall 92.7%, as well. The corresponding 95% confidence interval (CI) is between 92.34% and 93.60%. Since the County currently operates only 10 ambulances, for the comparison purpose, we also provide a system arrangement, which utilizes only 10 ambulances (i.e., *SYS proposed_10*). With this modification (i.e., *SYS proposed_10*), the proposed system did not meet the minimum compliance rate of 90% in the Suburban area. However, the recorded values of the fractile of RTs in the Urban and Suburban areas are a lot higher than the ones from the current system.

V. Conclusion

The primary criterion for responsiveness of EMS is the response time. Although the response time is not the only measure of performance, it takes a major role in the patients' outcome. Therefore, our research was focused on improving the response-time reliability by identifying good locations for the base station of the ambulances and its fleet size. We performed statistical data analysis for homogeneity tests of data and *K*-means data clustering algorithm for the natural groupings of the EMS request call locations. We also proposed a hybrid DES/Heuristic algorithm, in which a DES model is embed into a greedy like heuristic algorithm to identify and evaluate a set of base locations for ambulances and its fleet size.

The proposed DES/Heuristic algorithm was able to find the best system configuration for each of the six different time frames. The results demonstrate that the proposed algorithm is a practical and flexible tool in solving realistic EMS deployment problems. By using it, the EMS provider can enhance the delivery of emergency services to the residents, which has a promising potential that will benefit the local community.

ACKNOWLEDGMENT

This work was supported by the Consejo Nacional de Ciencia y Tecnologia (CONACYT under grant #404325/ 215143). This work was also partially supported by the US Department of Agriculture (USDA) under grants (# 2011-38422-30803 and # 2015-38422-24112).

REFERENCES

[1] American Ambulance Association. (2008). EMS Structured for Quality: Best Practices in Designing, Managing and Contracting for Emergency Ambulance Service.

[2] J. Overton and J. Stout (2002). System design. In A. E. Kuehl, Prehospital Systems and Medical Oversight: National Association of EMS Physicians (Third ed., pp. 114-131). Dubuque, IA: Kendall/Hunt.

[3] J. Stout, (1987, September). Measuring response time performance. Journal of Emergency Medical Services.

[4] A.M. Law, (2007). Simulation Modeling and Analysis. Mc Graw Hill.

[5] E.S. Savas, (1969). Simulation and Cost-Effectiveness Analysis of New York's Emergency Ambulance Service. Management Science, 15(12), 608-627.

[6] C. Monroe, (1980). A simulation model for planning emergency response systems. Social Science & Medicine. Part D: Medical Geography, 71-77.

[7] J.A. Fitzsimmons, (1973). A Methodology for Emergency Ambulance Deployment. *Management Science, 19*(6), 627-636.

[8] B.M. Marek Lubicz, (1987, May). Simulation modelling of emergency medical services. European Journal of Operational Research, 29(2), 178-185

[9] J.L. Devore, (2012). Probability and Statistics for Engineering and the Sciences (8th ed.). Boston: MA: Brooks/Cole, Cengage Learning.

[10] J.P. Stevens, (1999). Intermediate Statistics: A Modern Approach (2nd ed.). Hillsdale, NJ: Lawrence Erlbaum.

[11] R.O. Kuehl, (2000). Design of Experiments: Statistical Principles of research Design and Analysis. Belmont, CA: Brooks/Cole CENAGE Learning.

[12] Z. Zhu, M. McKnew, and J. Lee, (1992). Effects of time-varied arrival rates: an investigation in emergency ambulance service systems. 1992 WSC.

[13] H. Sohn, G. Reyes, and C. Escobar, (2014) Optimizing the Performance of the Dona Ana County Ambulance Service, *Technical Report*, Dona Ana County Health & Human Services Department, New Mexico.

[14] C. Escobar, G. Reyes, M. Alireza, and H. Sohn, A hybrid algorithm for emergency ambulance service, (*unpublished, 2018)*.