An improved emergency medical service system simulation-optimization model with Poisson mixture distribution

Nuşin Uncu* and Melik Koyuncu**

*Department of Industrial Engineering, Adana Alparslan Turkes Science and Technology University, Adana, Turkey

**Department of Industrial Engineering, Çukurova University, Adana, Turkey,

* Corresponding Author: nuncu@atu.edu.tr

Submitted **:** 01-03-2022

Revised **:** 02-09-2022

Accepted **:** 13-09-2022

ABSTRACT

A discrete event system (DES) simulation and optimization model is proposed for ambulance allocation in emergency medical service (EMS) systems. The accuracy of the proposed model is improved by estimating the input parameters using the Poisson mixture model and the expectation-maximization algorithm. The results are confirmed using the Mann–Whitney U and Kolmogorov–Smirnov tests. The DES model is executed using the OptQuest program in the Arena software. The response times of cases in urban and rural regions improved by 40% and 45%, respectively, and the station success rates in those regions increased to 95% and 96%, respectively. Our results indicate that the number of ambulances must be revised, and their allocation to stations in the actual system must be reorganized to achieve the required EMS system performance standards.

Keywords: Allocation model; emergency medical service systems; expectation maximization; optimization; Poisson mixture distribution.

INTRODUCTION

The emergence of COVID-19 has highlighted the importance of health systems in human life. During the pandemic, not only the entire healthcare system, but also emergency medical service (EMS) systems, which are vital to the healthcare system, are in high demand. The main purpose of EMS systems is to minimize loss of life while increasing service quality and performance. Sánchez-Mangas et al. (2010) suggested that reducing the response time by 10 min (from 25 to 15 min) resulted in a one-third reduction in the likelihood of death in both highway and conventional roadway crashes.

Over the recent decades, researchers have integrated simulation and optimization methods to simultaneously evaluate results based on performance and optimality (McCormack & Coates, 2015; Pinto et al., 2015; Bélanger et al., 2020; Boutilier et al., 2020; Amorim et al., 2019; Kamran et al., 2016). Simulation models must be constructed appropriately to allow the actual system performance to be determined with high precision. Several studies have focused on estimating system parameters. Matteson et al. (2011) proposed integer-valued time-series models to estimate emergency call arrival rates. Weinberg et al. (2007) estimated call arrival rates to model call centers efficiently. They proposed a multiplicative Gaussian model to measure and estimate the arrival rates of inhomogeneous Poisson processes. For parameter estimation, Olava-Rojas and Nickel (2021) suggested machine learning methods to simulate EMS systems in North Germany. Yang et al. (2020) developed a simulation optimization model for the Shanghai EMS system using a Gaussian mixture model to predict uncertain spatial demands. They showed that allocation costs can be reduced by approximately 41% through spatial distribution. Zhu et al. (1992) created a simulation model that matched the operating characteristics of an actual system by partitioning a day into 24-hour periods.

Our literature review indicates that parameter estimation methods have recently gained importance in simulation modeling. Classical methods that estimate arrival rates at any call center can be classified into time series, machine learning, and statistical methods. Generally, the arrival rates of EMSs are estimated based on single-parameter distributions using time-series methods (Channouf et al., 2007, Matteson 2011). However, the main contribution of this study is to estimate the arrival parameter using a Poisson mixture distribution on an hourly basis. A Poisson mixture distribution enables the use of two or more parameters for the arrival rate, thus increasing the accuracy and precision of the estimation.

EMS SYSTEM

The sequence of events after receiving an emergency call is shown in Fig. 1. When a call arrives at the call center, if necessary, the operator dispatches an available ambulance based on its proximity to the scene. Once the ambulance is ready, it is dispatched to the scene. After on-scene treatment, the ambulance typically transports and drops off the patient at the hospital. In some cases, the patient does not need to be transported to a hospital after initial care is administered. Finally, an ambulance is considered available for the next call once it departs from the hospital or returns to the base station.

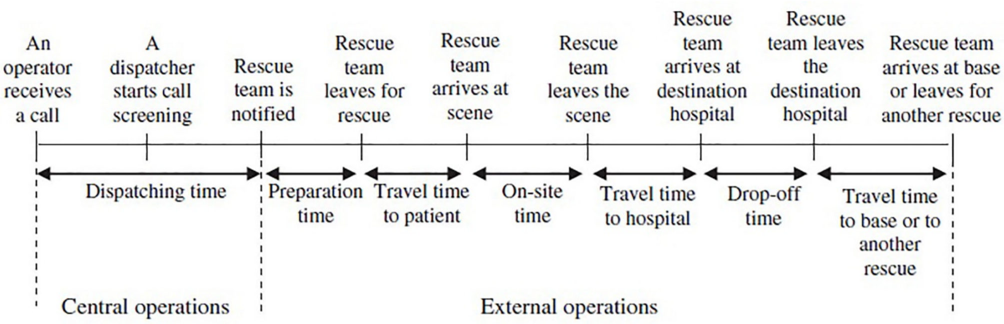


Figure 1. EMS system process.

The performance of an EMS is typically assessed by its response time. The EMS performance criterion is the percentage of cases that respond within the target response time limit. The EMS performance standards of Turkey are to respond to 95% of the cases in 10 min and 96% of the cases in 30 min in urban and rural regions, respectively. The Adana City zone is segmented into 65 urban and 11 rural sites. The EMS system includes 49 active ambulances and 40 stations. A total of 23 hospitals in the city are equipped with EMS units. In total, 33 and 16 ambulances are provided in the urban and rural regions, respectively (Fig. 2).

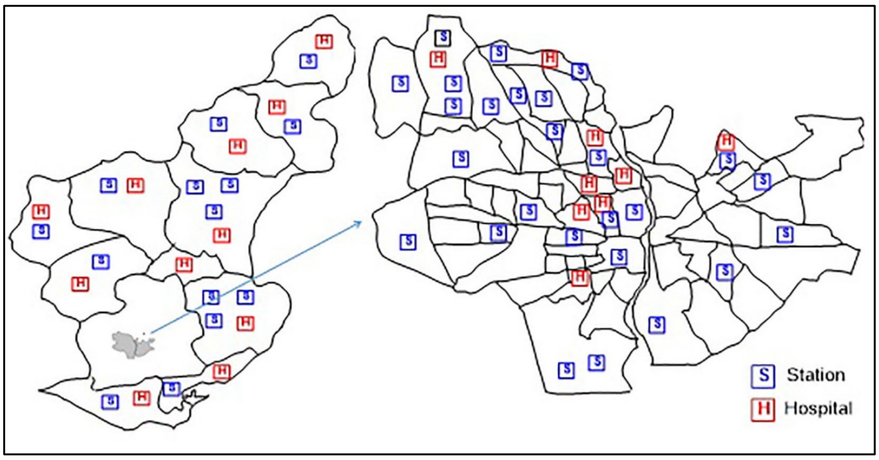


Figure 2. Location of ambulance stations and hospitals in rural (left) and urban (right) regions.

Based on our data analysis, ambulances were dispatched for 90% of calls, among which 70% involved dispatches to hospitals, and 30% involved on-scene treatment. Although rare in daily practice, an ambulance can be dispatched for a new case before the patient returns to the station. However, because such cases are rare, we assume that the ambulance routing policy is static.

EMS SYSTEM SIMULATION OPTIMIZATION MODEL

The EMS simulation optimization framework is shown in Fig. 3. The simulation and optimization modules are indicated separately using dashed lines. The simulation model was created using the Arena software. The system performance outputs were obtained via simulation and then evaluated and updated using the OptQuest optimization tool embedded in Arena.

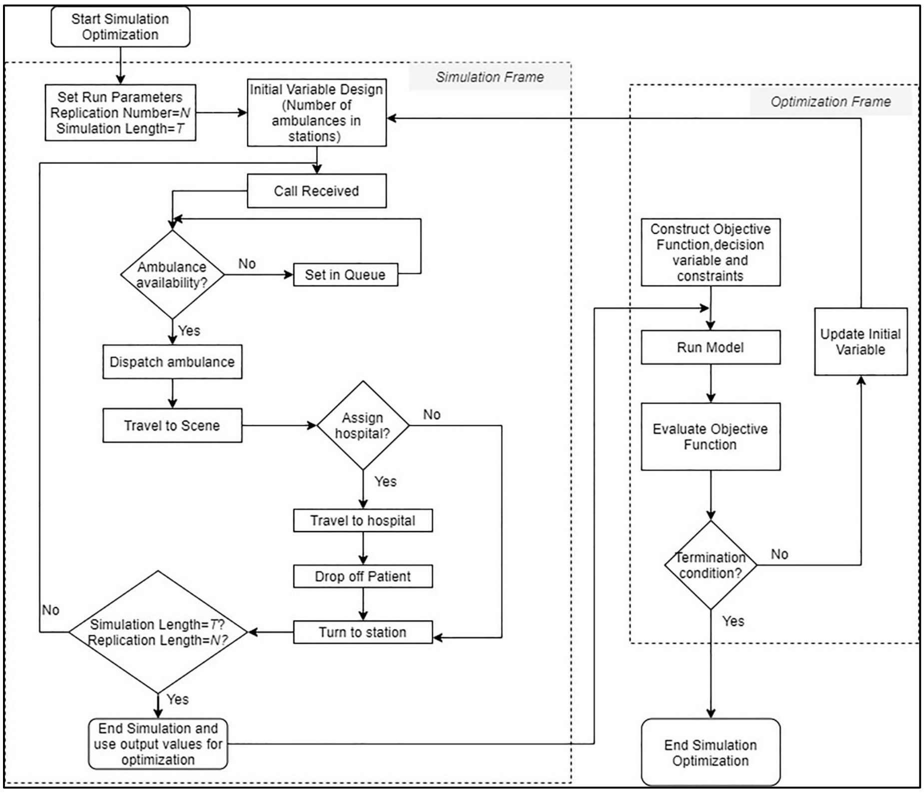


Figure 3. Simulation optimization framework.

Decision variable

x_n: number of ambulances assigned to station *i, i* = 1...*n*

Input variables

c: unit cost of an ambulance

Auxiliary variables

$\vec{\boldsymbol{X}}$: vector for any candidate solution

Simulation Outputs

$J\,(\vec{\boldsymbol{X}})$: expected cost function of candidate solution vectors

$RT^U(\vec{\boldsymbol{X}})$: average response time achieved for urban regions with input set $\vec{\boldsymbol{X}}$

$RT^R(\vec{\boldsymbol{X}})$: average response time achieved for rural regions with input set $\vec{\boldsymbol{X}}$

$SR^U(\vec{\boldsymbol{X}})$: average success rate for urban regions with input set $\vec{\boldsymbol{X}}$

$SR^R(\vec{\boldsymbol{X}})$: average success rate for rural regions with input set $\vec{\boldsymbol{X}}$

The aim of the objective function is to minimize the cost of ambulances allocated to EMS stations, as expressed in Eq. (1).

$$Min\ J(\vec{\boldsymbol{X}}) = \sum_{n=1}^{N} c\, x_n \tag{1}$$

$J(\vec{\boldsymbol{X}})$ is a function that represents the cost of allocating ambulances to n stations and is calculated via simulation. $\vec{\boldsymbol{X}}$ is a vector representing a candidate solution. Each $\vec{\boldsymbol{X}}$ includes n one-dimensional solution sets of x_n, which is randomly assigned at the beginning of the simulation optimization. The constraints expressed in Eqs. (2) and (3) provide an average response time shorter than the target response time for urban and rural regions, respectively.

$$E[RT^U(\vec{\boldsymbol{X}},)] \leq RT^U_{target} \tag{2}$$

$$E[RT^R(\vec{\boldsymbol{X}})] \leq RT^R_{target} \tag{3}$$

The percentage of cases that respond to the target response time is determined using Eqs. (4) and (5) for the urban and rural regions, respectively:

$$E[SR^U_n(\vec{\boldsymbol{X}})] \geq SR^u_{target} \tag{4}$$

$$E[SR^R_n(\vec{\boldsymbol{X}})\,] \geq SR^R_{target} \tag{5}$$

Poisson Mixture Distribution

The general function of the mixture distribution model is expressed as shown in Eq. (6).

$$f(x, \Psi) = \sum_{i=1}^{k} \pi_i\, f(x; \theta_i) \tag{6}$$

where $\Psi = (\theta, \pi)$ denotes the vector of all parameters; for a mixed distribution of k components, $\theta = (\theta_1, \theta_2\ldots, \theta_k)$ represents the parameters of k components and $\pi = (\pi_1, \pi_2, \ldots, \pi_k)$ represents the proportion of each distribution.

Suppose that the sample data $x_1, x_2\ldots, x_m$ are selected from a population comprising k components. The probability mass function of the mixture model with k Poisson distribution components is expressed as shown in Eq. (7), and Eq. (8) below should be satisfied:

$$P(x;\pi,\lambda) = \sum_{i=1}^{k} \pi_i \frac{\lambda_i^x e^{-\lambda_i}}{x!} \qquad x=0,1\ldots.m\ ,\ \lambda_i > 0 \tag{7}$$

$$\sum_{i=1}^{k} \pi_i = 1 \qquad\qquad 0 \leq \pi_i \leq 1 \tag{8}$$

Here, λ_i is the parameter of the i^{th} Poisson distribution for $i = 1\ldots k$. Each component k is determined using model selection criteria such as the Akaike information criterion (AIC), Bayesian information criterion (BIC), and log-likelihood (LL).

Let L(Ω) for the parameter set Ω comprising k components be the maximum value of the likelihood function. Hence, the likelihood and log-likelihood functions can be expressed as shown in Eqs. (9) and (10), respectively:

$$L(\Psi)=\prod_{j=1}^{m} f(x_j,\Psi) = \prod_{j=1}^{m}\left[\sum_{i=1}^{k} \pi_i f(x_j,\theta_i)\right] \tag{9}$$

$$\ln L(\Psi)=\sum_{j=1}^{m} \ln\ f(x_j,\Psi) = \sum_{j=1}^{m} \ln \sum_{i=1}^{k} \pi_i\, f(x_j,\theta_i) \tag{10}$$

The AIC and BIC values are calculated to be $2k$ −2LL and -ln(n)k-2LL, respectively. Finally, a k value corresponding to the smallest AIC and BIC values and the largest LL value is selected.

Expectation-Maximization (EM) Algorithm

In a Poisson mixture distribution, parameters $\pi_1, \pi_2 \ldots \pi_k$ and $\lambda_1, \lambda_2 \ldots, \lambda_k$ must be estimated. Dempster et al. (1977) developed the EM algorithm to estimate parameters in mixture distribution models. The EM algorithm is implemented based on the complete data likelihood function shown in Eq. (11), where the mixture likelihood $lnL(\Psi)$ is augmented with $z_{i,k}$ binary variables (0,1).

$$lnL(\Psi, Z) = \sum_{i=1}^{N}\sum_{k=1}^{K} z_{i,k}\, ln\ (\pi_k f(x,\theta_k)) \tag{11}$$

Beginning from $\widehat{\Psi}^{(0)}$, the EM algorithm iterates between the expectation step (E) and maximization step (M). In step (E), the expected missing probabilities are calculated. An estimate of $\hat{z}_{i,k}^{(r)}$ can be obtained using Eq. (12).

Step E

$$\hat{z}_{i,k}^{(r)} = \frac{\hat{\pi}_{i,k}^{(r-1)} f(x_i,\theta_k^{(r-1)})}{\sum_{k=1}^{k} \hat{\pi}_k^{r-1} f(x_i,\theta_k)} \tag{12}$$

Subsequently, based on the estimated $\hat{z}_{i,k}^{(r)}$, all unknown parameters $\widehat{\Psi}^{(r)} = (\hat{\theta},\ \hat{\pi})$ are updated using Eq. (13) until they converge to the maximum.

Step M

$$\sum_{i=1}^{N}\sum_{k=1}^{K} \hat{z}_{i,k}^{(r)}\, ln\ (z_k f(x_i,\theta_k)) \tag{13}$$

Modeling Call Arrival Process

We used an hourly dataset of call arrivals recorded in one year. Using the Poisson mixture model, data from components 1 to 10 were tested to determine whether they fit the Poisson mixture distribution. The AIC, BIC, and LL values indicate that two components existed for incoming calls in 17 of the 24-hour periods, and the remaining fitted the pure Poisson distribution shown in Table 1. We applied the EM algorithm and used Flexmix embedded in the R software to obtain the parameters (π, λ) of the Poisson mixture distribution.

Table 1. Hourly calls and Poisson mixture distribution parameters.

Time Interval	k	Proportions		Poisson Distribution Parameters		Selection Criteria		
		π_1	π_2	λ_1	λ_2	AIC	BIC	LL
$00:^{00}$-$01:^{00}$	2	0.62	0.38	13.43	17.78	2136.88	2140.78	-1067.44
$01:^{00}$-$02:^{00}$	2	0.89	0.11	11.71	17.91	2085.25	2089.15	-1041.62
$02:^{00}$-$03:^{00}$	1	1	-	9.82	-	1940.76	1944.66	-969.38
$03:^{00}$-$04:^{00}$	1	1	-	8.28	-	1900.38	1904.28	-949.19
$04:^{00}$-$05:^{00}$	1	1	-	6.95	-	1780.82	1784.72	-889.40
$05:^{00}$-$06:^{00}$	1	1	-	6.53	-	1738.01	1741.91	-868.00
$06:^{00}$-$07:^{00}$	1	1	-	7.23	-	1790.76	1794.66	-894.38
$07:^{00}$-$08:^{00}$	1	1	-	10.10	-	1937.01	1940.41	-967.50
$08:^{00}$-$09:^{00}$	2	0.18	0.82	13.91	9.37	2294.35	2298.25	-1146.17
$09:^{00}$-$10:^{00}$	2	0.31	0.69	24.09	16.94	2369.84	2373.74	-1183.92
$10:^{00}$-$11:^{00}$	2	0.91	0.09	20.67	35.03	2456.09	2459.99	-1227.04
$11:^{00}$-$12:^{00}$	2	0.21	0.79	30.55	21.02	2472.03	2475.93	-1235.01
$12:^{00}$-$13:^{00}$	2	0.16	0.84	31.72	21.68	2433.48	2437.38	-1215.74
$13:^{00}$-$14:^{00}$	2	0.18	0.82	32.86	22.47	2438.33	2442.23	-1218.16
$14:^{00}$-$15:^{00}$	2	0.70	0.30	21.92	29.12	2416.66	2420.56	-1207.33
$15:^{00}$-$16:^{00}$	2	0.20	0.80	30.18	22.11	2394.20	2398.10	-1196.10
$16:^{00}$-$17:^{00}$	2	0.93	0.07	22.38	37.80	2450.17	2454.07	-1224.08
$17:^{00}$-$18:^{00}$	2	0.85	0.15	21.94	32.82	2483.40	2487.30	-1240.70
$18:^{00}$-$19:^{00}$	2	0.10	0.90	35.45	22.73	2511.43	2513.33	-1254.71
$19:^{00}$-$20:^{00}$	2	0.22	0.88	32.49	22.50	2570.32	2574.22	-1284.16
$20:^{00}$-$21:^{00}$	2	0.54	0.46	21.72	29.23	2489.03	2492.98	-1243.54
$21:^{00}$-$22:^{00}$	2	0.66	0.34	26.47	19.81	2379.39	2383.29	-1188.69
$22:^{00}$-$23:^{00}$	1	1	-	21.42	-	1738.01	1741.91	-868.00
$23:^{00}$-$00:^{00}$	2	0.48	0.52	15.61	21.80	2335.92	2339.82	-1166.96

Ambulance Dispatching and Travel Features

The dispatching time was assumed to be log-normally distributed, with an average of 1.81 min. Moreover, the preparation time was disregarded since it was short. The total travel time was segmented into five intervals. The related time statistics were determined using distance matrices and data analysis. Only the onsite delay time was assumed to be exponentially distributed with an average of 9.68 min.

MODEL VERIFICATION AND VALIDATION

For each replication, the simulation model was performed for one year with a 50-day warm-up period and eight replications. First, using the parameters listed in Table 1, we generated the number of arrivals per hour in the simulation model. Second, we applied the Mann–Whitney U and Kolmogorov–Smirnov tests to determine whether the simulated arrival calls represented arrival calls in the actual system. The test hypotheses are as follows:

H_0: The actual and simulated data have the same distribution for the i^{th} time interval.

H_1: The actual and simulated data do not have the same distribution for the i^{th} time interval.

The MWU and K-S test p-values are listed in Table 2. To accept H_0, the MWU p-value should be between 0.15 and 1 or the K-S p-value should exceed 0.10. Based on the results, most of the MWU p-values were approximately 1, and the K-S p-values were greater than 0.10. Thus, we confirmed that the Poisson mixture distributions accurately represented the actual system.

Table 2. Comparison of results of Mann–Whitney U and Kolmogorov–Smirnov tests.

Time interval	Means		Standard deviations		p values	
	Real data	Simulated data	Real data	Simulated data	MWU	K-S
$00:^{00}$-$01:^{00}$	15.09	14.99	4.43	4.13	0.98	$p>0.10$
$01:^{00}$-$02:^{00}$	12.42	12.37	4.14	3.88	0.88	$p>0.10$
$08:^{00}$-$09:^{00}$	18.65	11.93	5.09	5.06	0.73	$p>0.10$
$09:^{00}$-$10:^{00}$	19.21	18.99	5.67	5.24	0.68	$p>0.10$
$10:^{00}$-$11:^{00}$	22.00	22.07	6.41	6.22	0.77	$p>0.10$
$11:^{00}$-$12:^{00}$	23.05	22.82	6.37	6.13	0.71	$p>0.10$
$12:^{00}$-$13:^{00}$	23.38	23.15	6.30	6.10	0.49	$p>0.10$
$13:^{00}$-$14:^{00}$	24.35	24.18	6.35	6.39	0.59	$p>0.10$
$14:^{00}$-$15:^{00}$	24.04	24.16	6.06	5.52	0.80	$p>0.10$
$15:^{00}$-$16:^{00}$	23.73	23.55	6.05	5.78	0.58	$p>0.10$
$16:^{00}$-$17:^{00}$	23.44	23.43	6.46	5.99	0.81	$p>0.10$
$17:^{00}$-$18:^{00}$	23.50	23.71	6.37	6.12	0.67	$p>0.10$
$18:^{00}$-$19:^{00}$	24.15	23.71	6.63	6.31	0.15	$p>0.10$
$19:^{00}$-$20:^{00}$	24.68	24.28	6.84	6.44	0.23	$p>0.10$
$20:^{00}$-$21:^{00}$	25.15	25.14	6.52	5.99	0.97	$p>0.10$
$21:^{00\text{-}}22:^{00}$	24.20	23.80	5.93	5.96	0.47	$p>0.10$
$22:^{00\text{-}}23:^{00}$	21.41	21.03	5.31	5.17	0.48	$p>0.10$
$23:^{00}$-$00:^{00}$	18.77	18.78	5.43	5.13	0.89	$p>0.10$

RESULTS AND DISCUSSION

The model was tested by reallocating 49 ambulances to stations without increasing their number. However, no significant improvement was observed. Subsequently, we increased the number of ambulances and reallocated them to each station. The changes in success rates after varying numbers of ambulances and reallocation combinations are shown in Fig. 4. The optimal system that satisfied the expected performance criteria was achieved using one of the combinations of 89 ambulances reallocated to existing stations.

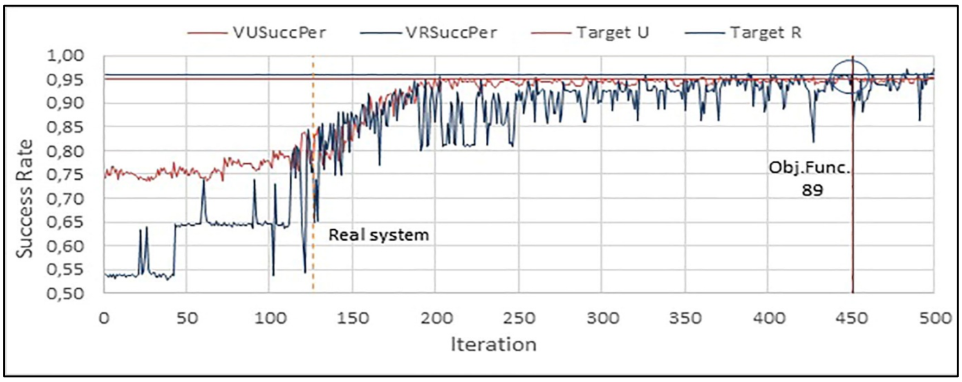


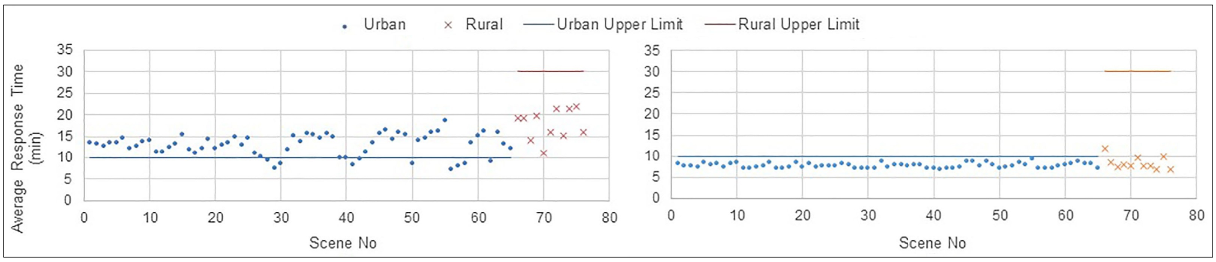
Figure 4. Success rates for different simulation optimization trials.

The time-dependent performance outputs for both the actual and optimized systems are listed in Table 3. The average response times for the urban and rural areas improved by 40% and 45%, respectively. Therefore, emergency cases can be responded within the desired response time, which is a performance target of the Ministry of Health.

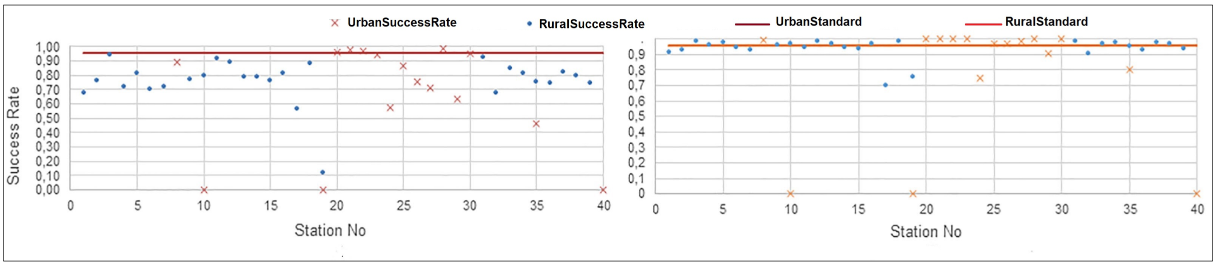
Table 3. Time-dependent system performances.

Time Intervals	Average		Minimum Average		Maximum Average	
	Actual	Optimal	Actual	Optimal	Actual	Optimal
Response Time in Rural Region	15.42	8.44	14.50	8.23	16.45	8.77
Response Time in Urban Region	13.01	7.79	12.73	7.72	13.39	7.85

In Fig. 5, a comparison of the time standard performances of the actual system and optimized model is presented. The red and blue markers represent the stations responsible for cases in the rural and urban regions, respectively. As shown on the left side of Fig. 5(a), most urban cases did not respond within the time limits of the actual system. However, the time standards for both regions were satisfied when the proposed model was used, as shown on the right side of the figure. The success rates of the actual and optimized models are shown in Fig. 5(b) on the left and right sides, respectively. The success rates of the actual system were 84% and 78% in the rural and urban areas, respectively. Using the proposed model, the success rates improved by 96% and 95% for most stations in the rural and urban areas, respectively. Because of their geographical locations, a few stations did not indicate improvements in their success rates. However, station 17 indicated a slight improvement of approximately 14% in terms of its success rate.



(a)



(b)

Figure 5. Comparison of number of ambulances for actual and optimal systems. (**a**) Average response times to scenes. (**b**) Success rates of stations.

A graph illustrating the actual and optimal allocations of ambulances to stations is shown in Fig. 6. The number of ambulances in only 3 among 40 stations remained the same, whereas it decrease in the third station and increased in the other stations.

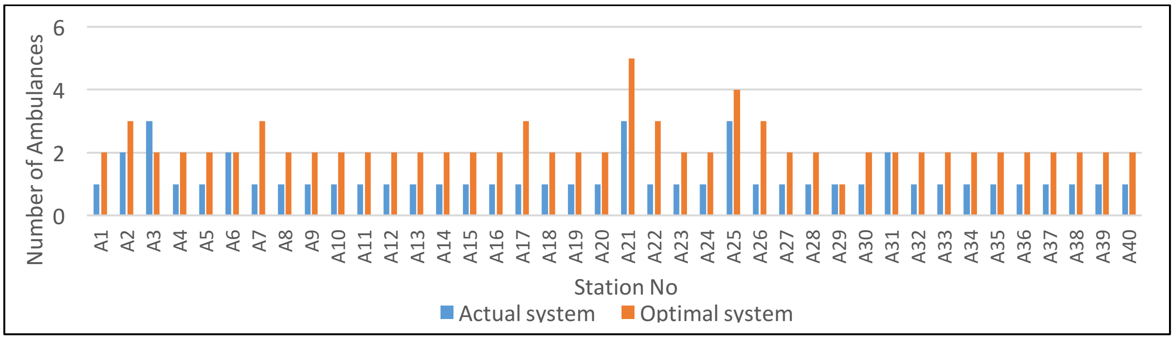


Figure 6. Comparison of number of ambulances for actual and optimal systems.

CONCLUSION

In this study, the performance of an EMS system was analyzed based on time standards and the percentage of cases responded. This study was limited by insufficient recorded data in some cases and mistakenly recorded events. Thus, some approximate values were substituted for the missing data. However, we focused on modeling the system as accurately as possible to represent the actual situation using the Poisson mixture model and EM algorithms. An optimal EMS system was proposed after performing simulation optimization. Using the proposed allocation model, the percentage of cases responded by the EMS system within a specified time period was high.

In future studies, additional mixture distributions can be considered for random data at different time intervals on weekdays or weekends. In addition, the relocation of stations and dynamic reallocation policies should be considered for further improvement.

ACKNOWLEDGMENTS

We are grateful for the support provided by the Ministry of Health of the Adana Division.

REFERENCES

Amorim, M., Antunes, F., Ferreira, S., & Couto, A. 2019. An integrated approach for strategic and tactical decisions for the emergency medical service: Exploring optimization and metamodel-based simulation for vehicle location. Computers & Industrial Engineering, 137, 106057.

Bélanger, V., Lanzarone, E., Nicoletta, V., Ruiz, A., & Soriano, P. 2020. A recursive simulation-optimization framework for the ambulance location and dispatching problem. European Journal of Operational Research, 286(2): 713-725.

Boutilier, J.J. & Chan, T.C. 2020. Ambulance emergency response optimization in developing countries. Operations Research, 68(5): 1315-1334.

Dempster, A. P., Laird, N. M., & Rubin, D. B. 1977. Maximum likelihood from incomplete data via the EM algorithm. Journal of the Royal Statistical Society: Series B (Methodological), 39(1): 1-22.

Kamran, M. A., Karimi, B., Bakhtiari, H., & Masoumzadeh, S. 2016. A resource allocation model in a healthcare emergency center using goal programming. Journal of Engineering Research, 4(4).

Matteson, D. S., McLean, M. W., Woodard, D. B., & Henderson, S. G. 2011. Forecasting emergency medical service call arrival rates. The Annals of Applied Statistics, 5: 1379-1406.

McCormack, R., & Coates, G. 2015. A simulation model to enable the optimization of ambulance fleet allocation and base station location for increased patient survival. European Journal of Operational Research, 247(1): 294-309.

Olave-Rojas, D., & Nickel, S. 2021. Modelling a pre-hospital emergency medical service using hybrid simulation and a machine learning approach. Simulation Modelling Practice and Theory, 109: 102302.

Pinto, L. R., Silva, P. M. S., & Young, T. P. 2015. A generic method to develop simulation models for ambulance systems. Simulation Modelling Practice and Theory, 51: 170-183.

Sánchez-Mangas, R., García-Ferrrer, A., De Juan, A., & Arroyo, A. M. 2010. The probability of death in road traffic accidents. How important is a quick medical response? Accident Analysis & Prevention, 42(4): 1048-1056.

Weinberg, J., Brown, L. D., & Stroud, J. R. 2007. Bayesian forecasting of an inhomogeneous Poisson process with applications to call centre data. Journal of the American Statistical Association, 102(480): 1185-1198.

Yang, W., Su, Q., Huang, S. H., Wang, Q., Zhu, Y., & Zhou, M. 2019. Simulation modeling and optimization for ambulance allocation considering spatiotemporal stochastic demand. Journal of Management Science and Engineering, 4(4): 252-265.

Zhu, Z., McKnew, M. A., & Lee, J. 1992. Effects of time-varied arrival rates: an investigation in emergency ambulance service systems. In Proceedings of the 24th Conference on Winter simulation.