

A SIMULATION MODEL OF A HELICOPTER AMBULANCE SERVICE

Ersan Gunes

Operations Research Department
Naval Postgraduate School
Monterey, CA 93943, U.S.A.

Roberto Szechtman

Operations Research Department
Naval Postgraduate School
Monterey, CA 93943, U.S.A.

ABSTRACT

We study two different operational scenarios for a regional air ambulance service-company which has bases in northern California. Two of these bases serve the land areas encompassed roughly in a circular area of radius 100 miles centered in Gilroy and Salinas, respectively; with a large part of their coverage areas reachable from either base. The base in Salinas currently operates one helicopter only from Thursday to Monday, whereas the base in Gilroy operates one helicopter 24/7. The company is considering extending the operation of one helicopter to 24/7 for its Salinas base. In this paper we analyze the operational impacts of that extension, and develop a framework that can be applied towards the study of the ambulance assignment problem faced by small operators.

1 INTRODUCTION

Currently, there is a growing public debate over the use, cost, and proliferation of helicopter ambulance service. Helicopter ambulances operating in the United States have doubled to 700 in the last ten years, with the price of an airlift ranging from \$5,000 to \$10,000 (The New York Times, Feb. 28th and May 3rd, 2005). Furthermore, the financial crisis faced by many hospitals is making many rural and suburban communities too distant from high level trauma centers, leaving air ambulances as their only option for emergency transportation (SV/SJ Business Journal, October 22nd, 2004 and Santa Cruz Sentinel, Sep. 23rd, 2004).

Emergency response systems have greatly benefited by management science studies; see, for example, the survey paper Green and Kolesar (2004), and references therein. Regarding ground ambulance transport, the standard setting is that of a unique emergency transport provider, and the problem is to find optimal base locations and the minimum number of ambulances per base to meet certain performance targets; see, for example, Iskander (1989), and Henderson and Mason (2004). The air ambulance assignment problem has traditionally been studied in the context of a regional public emergency medical service

(EMS); Parker and Johnson (1970) is an early reference on the subject and more recently, Wears and Winton (1993), and Stundzia and Lumsden (1994).

Helicopter ambulance service is generally provided by individual operators that operate helicopters located in several bases and are under contract with EMS. When an accident that requires air transport occurs, EMS retrieves the first available helicopter from a priority list. For example, current EMS procedures (County of Monterey EMS, 2004) dictate that accidents located in Monterey county are served by a helicopter based either in Salinas (MEDCAL), Gilroy (MEDCAL), Palo Alto (Stanford Hospital), Modesto (Medi Flight), Paso Robles (California Highway Patrol), Fresno (Sky Life), or Santa Maria (MEDCAL), in that order. The MEDCAL helicopter based in Gilroy is the EMS first choice for accidents occurring in southern Santa Clara and southern Santa Cruz counties, and the second choice for missions located in Monterey and most of San Benito counties. On the other hand, the Salinas based MEDCAL helicopter is the EMS first choice for accidents occurring in Monterey and a large portion of San Benito counties, and second choice in southern Santa Clara and southern Santa Cruz counties (County of Monterey EMS, 2004, and County of Santa Cruz EMS, 2004).

This paper addresses the helicopter assignment problem faced by MEDCAL (the sponsoring company wishes to remain anonymous; MEDCAL is a fictitious name), who wishes to assess the operational impact of extending the Salinas based single helicopter operation from Thursday through Monday to 24/7. More specifically, MEDCAL is interested in gaining knowledge about expected changes in helicopter utilization, number of missions completed, and helicopter response times, by expanding its Salinas operation.

In order to solve the problem we gathered flight-log sheets from MEDCAL's Gilroy and Salinas based helicopters for the years 2002, 2003, and 11 months of 2004. Although these log sheets include detailed information about complete missions, data for incomplete missions is mostly missing. Furthermore, EMS's calls that are rejected by a MEDCAL base due to helicopter unavailability are gener-

ally not recorded. Another relevant source of data is EMS itself; they provided monthly aggregate demand for air transport only in Monterey and Santa Cruz counties and yearly totals in Santa Clara county. A first order approximation to our problem would discard missed and rejected missions. However, the log-sheets indicate that EMS calls that result in incomplete missions fall in the range 17% to 43%, and using EMS's data as a measure of overall demand it can be found that about 3% to 7% of EMS's calls are rejected by MEDCAL. This suggests that incomplete and rejected missions need to be accounted for in our model. Thus, we face the problem of figuring out certain parameters from aggregate data; see Leemis (2001) for more details on this issue. Regarding the estimation of a non-homogeneous Poisson process (NHPP) intensity function, which is needed to model mission demand, non-parametric approaches include Law and Kelton (2000, p. 390-393), Arkin and Leemis (2000), Henderson (2003), and Leemis (2004). The papers Kuhl et al. (1997), (2001), and (2004), deal with the estimation of an intensity function subject to periodic effects and trends. The intensity function estimation problem also can be tackled using kernel techniques; see Lewis and Shedler (1976), Diggle and Marron (1988), Brooks and Marron (1991), Hall et al. (1991), and Jones et al. (1996), for more details. Zhu and McKnew (1997) use a piece-wise linear approximation of the Poisson intensity function to model arrivals in an emergency ambulance service. Lastly, Wu et al. (2005) provide a queueing perspective of our problem.

After the historical data is analyzed, we build a discrete event simulation model that uses the fitted input parameters to generate missions and their associated time variables. Simulation output analysis is performed afterwards.

Our paper is organized as follows: In Section 2 we describe our model, explain the input analysis problem, and present the measures of interest. Section 3 deals with model implementation and, more importantly, presents and discusses our results. The paper conclusions are collected in Section 4.

Once again, to goal of this paper is to present a simulation modeling framework for the ambulance allocation problem faced by an individual operator.

2 MODEL

2.1 Model Description

For the purpose of our analysis missions, rather than patients (there may be multiple patients per mission), are divided into three types. First, "complete missions" are unplanned flights triggered by EMS that occur at a random location where the helicopter picks the patient and transports her to a regional hospital; if no helicopter is available

on base upon receiving the EMS call, the call is lost (i.e., there is no waiting buffer in the queue). A motorcycle accident is a frequent instance of this type of mission. Second, we consider "inter-facility missions", in which a patient is picked from a local hospital and transported to a more sophisticated hospital. Because there often is a contractual relationship between the local hospital and the helicopter-ambulance company, EMS is not involved in these missions. Patients can generally wait up to one hour for helicopter pick up; in practice, this means that buffer size is infinite and the transport occurs once a helicopter becomes accessible. An example of this kind of mission is a trauma patient transported by land ambulance to a local hospital, who is later transported by helicopter to a high complexity trauma center. Last, we have the "incomplete missions". These are unplanned randomly located flights dispatched by EMS for which the helicopter returns to base without picking any patient because air transport is cancelled. If no helicopter is available, the call is lost (there is no waiting buffer in the queue). Mechanical failure, adverse weather conditions, or EMS cancellation can lead to incomplete missions.

The steps followed and recorded in the flight-log sheet for complete missions are the following. First a helicopter is dispatched by the base operator upon receipt of an EMS order. After a random amount of time, called the Reaction time, the helicopter departs from base. Next, the helicopter arrives to the scene of the accident and soon after there occurs nurse-patient contact. The time elapsed between departure from base and nurse-patient contact is called the Waiting time. The last steps are departure from the scene, arrival into hospital, and return to base; we call the cumulative time spent in the latter steps the Service time. The steps and time intervals concerned by complete missions are illustrated in Figure 1.

Due to the nature of inter-facility missions, the only relevant time measurement (recorded in the log-sheet) is the total amount of time spent between helicopter take-off and return to base.

The steps recorded in the flight log sheet for incomplete missions only includes dispatch times. Therefore, there are two steps: dispatch time (time known), and helicopter becomes available (time unknown). The latter is modeled under some reasonable assumptions; see Section 2.3.

2.2 Measures of Interest

The window of time that starts at the time of emergency and finishes when the helicopter reaches the hospital is crucial. If this time window, known as the "golden hour", is less than one hour then chances of survival increase significantly if trained professionals start to treat the patient (Tallon et al., 2002). Of this time, only the part

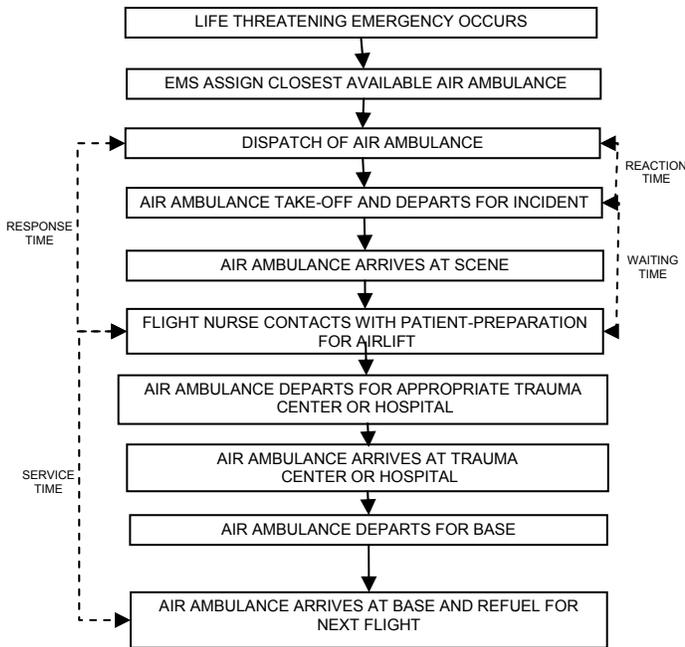


Figure 1: Complete Mission Process Steps

Dispatch → Nurse Patient Contact depends on the helicopter allocation policy; we call that time the Response time, see Figure 1. By configuring Salinas as a 24/7 operation we expect response times to decrease.

The number of complete missions per base per year and the yearly average helicopter utilization are two other measures of interest from the operator’s perspective, because of their impact on revenues. For any given year, the average helicopter utilization is the ratio of the busy times over busy plus idle times, and is found separately for the Gilroy and Salinas based helicopters. Assigning a helicopter 24/7 to the Salinas base will tend to decrease utilizations in Gilroy and Salinas, although this might be compensated by taking additional inter-facility missions from local hospitals (which we don’t include into our model).

2.3 Input Analysis

For the three types of missions, the relevant parameters may depend on the helicopter base and on the emergency location (for which the maximum resolution is the county name). More specifically, we have:

- For complete missions:
 - Reaction time – depends on helicopter base.
 - Waiting time – depends on helicopter base and emergency county name.
 - Service time - depends on helicopter base and emergency county name.

- For inter-facility missions, we capture the total transportation time starting with dispatch time and ending in return to base, this time is base dependent.
- For incomplete missions, because only dispatch times are available, we assume that the time until the helicopter becomes available is uniformly distributed.

For each of the five variables we separately verified independence by using auto-correlation with missing data. The correlation among waiting and service times was checked using Pearson’s correlation test, and we found that for some base-county pairs the waiting times and service times are correlated. In particular, we found that waiting times and service times are conditionally independent given that their waiting times are less (or more) than 20 minutes. All other variables are necessarily independent.

Finally, we used Kernel methods (Wand and Jones, 1995) to fit the data associated with each variable to a density. For example, given n independent and identically distributed Salinas based reaction times R_1, R_2, \dots, R_n , the kernel density estimator is

$$f(r, h) = \frac{1}{nh} \sum_{i=1}^n K\left(\frac{r - R_i}{h}\right),$$

where the kernel $K(\cdot)$ satisfies $\int K(r) dr = 1$, and $h > 0$ is called the bandwidth. The bandwidth is selected to minimize the mean integrated squared error of $f(r, h)$. In particular, we used a Gaussian kernel, and, the optimal bandwidth selection method proposed by Hall et al. (1991).

We believe that our approach, by capturing the randomness of the different stages involved in a mission, is more realistic than other air-ambulance studies for which these parameters are deterministic.

2.4 NHPP intensity function estimation

2.4.1 Goodness-of-Fit Tests

We now describe the procedure followed to analyze the mission arrivals. For each of the four counties that feed the Gilroy and Salinas bases, and for each type of mission, we perform a goodness-of-fit test of a non-homogeneous Poisson process with piecewise constant intensity function. We applied both classical χ^2 test and denominator-free χ^2 as described in Velleman and Hoaglin (1981): Given K independent Poisson random variables with means m_1, \dots, m_K , and observed arrivals n_1, \dots, n_K , the statistic

$$T = \sum_{i=1}^K \left(\sqrt{n_i} + \sqrt{n_i + 1} - \sqrt{4m_i + 1} \right)^2 \stackrel{D}{\sim} \chi_K^2,$$

where $\stackrel{D}{\sim}$ means equality in distribution. We used 2002 and 2003 data to find the rates, and the 2004 data (which consists of only 11 months) to cross-validate our parameters. Table 1 presents a summary of our results, and Figure 2 shows the intensity function for Santa Clara county (the intensity functions for the other three counties only differ slightly; see Gunes, 2005, for a complete scenario).

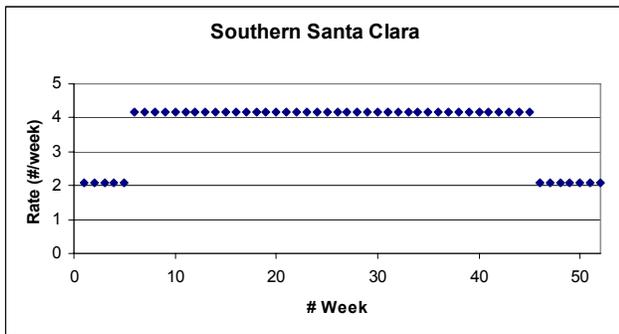


Figure 2: Intensity Function for Southern Santa Clara

Table 1: χ^2 Goodness-of-Fit for Complete-Flights Estimated rates

County	GOF 2002/2003	GOF 2004 (11 months)
SCL	$P\{\chi_{50}^2 \leq T_{SantaClara}\} = 0.83$	$P\{\chi_{48}^2 \leq T_{SantaClara}\} = 0.8$
SCR	$P\{\chi_{50}^2 \leq T_{SantaCruz}\} = 0.593$	$P\{\chi_{48}^2 \leq T_{SantaCruz}\} = 0.435$
MON	$P\{\chi_{50}^2 \leq T_{monterey}\} = 0.499$	$P\{\chi_{48}^2 \leq T_{monterey}\} = 0.305$
SBE	$P\{\chi_{50}^2 \leq T_{SanBenito}\} = 0.42$	$P\{\chi_{48}^2 \leq T_{SanBenito}\} = 0.692$

We conclude that by using 2002 and 2003 data, the rates we find are statistically adequate.

2.4.2 A Kernel Approach to NHPP Intensity Estimation

In the past subsection we found that a piece-wise constant arrival intensity function adequately fits the data at the weekly level. However, after taking a closer look at the data it is clear that the rates are not constant on a daily and intra-day basis. One important reason for this is that most of the arrivals are trauma incidents of which traffic accidents are a large proportion, and these tend to occur on certain days of the week and on certain hours of the day. For this reason, for each of the four counties, and for all kinds of missions, we fit the intra-week intensity function using kernel density estimation procedures (Wand and Jones, 1995) with bandwidths that capture multi-modalities like

morning, evening rush-hour, and Saturday night effects. Figures 3 through 6 illustrate the intra-week intensity function found for each county. Note how the intensities peak on the evenings and on the weekends.

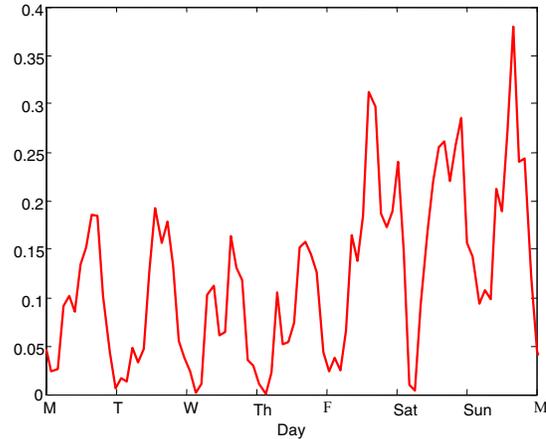


Figure 3: Monterey Intra-Week Intensity

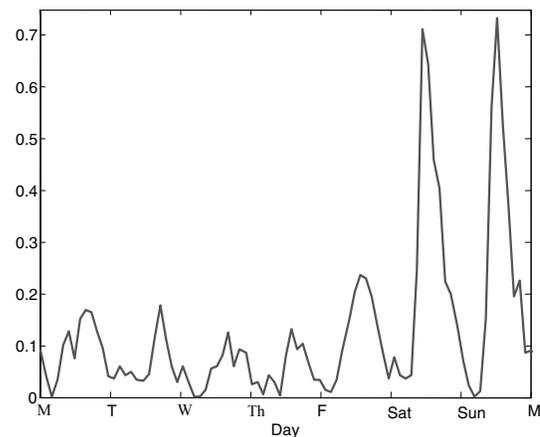


Figure 4: San Benito Intra-Week Intensity

3 IMPLEMENTATION, VALIDATION, AND RESULTS

3.1 Implementation

Our model uses discrete event simulation techniques (Law and Kelton, 2000), and is implemented in Java using the Simkit software package (Buss, 2001). The kernel estimation was done in Matlab, and the simulation samples are drawn employing standard techniques; see Chapter 8 of Law and Kelton (2000).

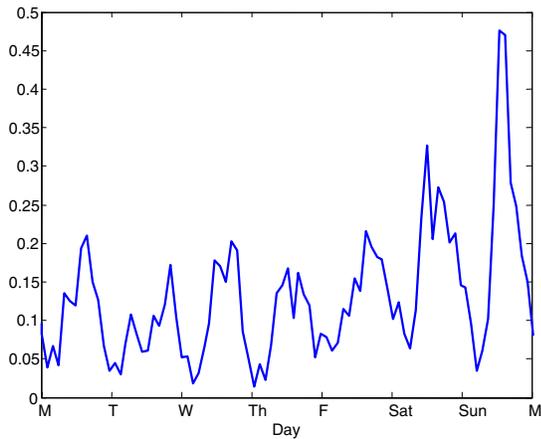


Figure 5: Southern Santa Clara Intra-Week Intensity

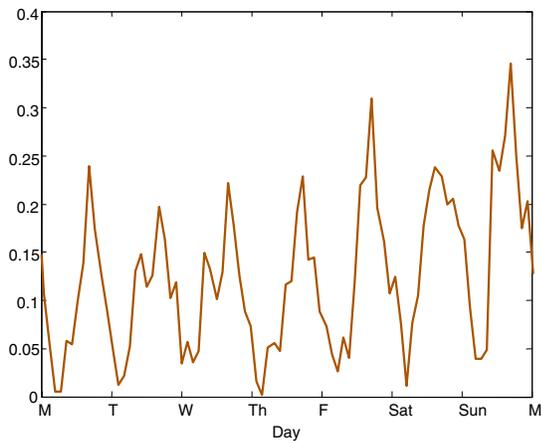


Figure 6: Southern Santa Cruz Intra-Week Intensity

3.2 Model Validation

In order to validate the helicopter assignment policy, we did a trace-driven simulation for the Gilroy and Salinas bases, and compared the number of complete missions with the recorded data, for 11 months of 2004 (which is used because it contains the most complete data set in terms of complete, incomplete, and inter-facility missions). Our results are shown in Table 2. There are about 21 lost flights in our simulation; we conjecture them due to errors in data entry.

In the second phase of the validation, we re-utilized the arrival-epochs, but randomized response times, waiting times, and service times, using their estimated densities.

Table 3 summarizes our results; we concluded that our implementation adequately represents the real system and that the kernel density estimators found for reaction, waiting, and service times are reasonably good.

Table 2: Assignment Model Validation

Base	Actual	Simulated	% coverage
Gilroy	609	575.1	94.4 %
Salinas	202	214.9	105.9 %

Table 3: Estimated Parameter Validation

Base	Actual Data	Estimated Data	% coverage
Gilroy	575.1	571.7	99.4 %
Salinas	214.9	215.5	100.2 %

3.3 Results

In the final stage we randomized mission arrival times and mission dependent variables, with the goal of comparing expected yearly changes in the measures of interest for both the current (Salinas base part-time) and the proposed (Salinas base 24/7) configurations.

We ran 200 sample path iterations, lasting 52 weeks each. Given a random variable X , and a measure of interest $\mu = EX$, we compute the standard sample mean estimator \bar{X} of μ , the sample standard deviation s , and an approximate 95% confidence interval for μ

$$\left(\bar{X} - 1.96s / \sqrt{200}, \bar{X} + 1.96s / \sqrt{200} \right).$$

Our results are summarized in Tables 4 and 5. Under the Salinas 24/7 configuration, the number of complete missions increases by about 4.5%, utilizations decrease in both bases, and response times (denoted as RT in the tables) decrease by about 2 minutes for Monterey county. Accordingly, only a planned increase of inter-facility missions, especially out of Salinas, seems to justify setting Salinas as a full-time base.

Table 4: Simulation with Current Configuration

	\bar{X}	s	95% Conf. Int.
G # Served	521.55	18.7321	(519, 525)
S # Served	210.47	11.6494	(209, 212)
SCL RT	24.7804	0.1395	(24.76, 24.80)
SCR RT	27.4889	0.1721	(27.46, 27.52)
MON RT	29.294	0.1766	(29.27, 29.32)
SBE RT	25.4762	0.1557	(25.45, 25.50)
G Util.	0.1357	0.005	(0.1349, 0.1365)
S Util.	0.1353	0.0076	(0.1341, 0.1365)

Table 5: Simulation with Proposed Configuration

	\bar{X}	s	95% Conf. Int.
G # Served	484.63	18.593	(482, 488)
S # Served	279.495	14.411	(277, 282)
SCL RT	24.872	0.1277	(24.9, 24.9)
SCR RT	27.5226	0.1414	(27.5, 27.6)
MON RT	27.6486	0.1731	(27.62, 27.68)
SBE RT	25.1288	0.1729	(25.10, 25.16)
G Util.	0.125	0.0051	(0.124, 0.126)
S Util.	0.0758	0.0038	(0.0751, 0.0764)

4 CONCLUSION

We developed a model useful to analyze the operations of an air-ambulance service provider. Moreover, our results lent statistical support for MEDCAL's decision to have its Salinas base operate full-time.

ACKNOWLEDGMENTS

We are indebted to Mr. Tom Goff for sharing his knowledge about MEDCAL's operations and for proposing the problem that is the subject of this paper. We would like to thank Professor Patricia Jacobs for suggesting using the goodness-of-fit test and for several valuable discussions. The first author wishes to recognize the support of the Turkish Army. The second author gratefully acknowledges the backing of the Naval Postgraduate School through a RIP grant.

REFERENCES

- Arkin, B. L., and L. Leemis. 2000. Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process from overlapping realizations. *Management Science* 46(7): 989-998.
- Brooks, M. M., and J. S. Marron. 1991. Asymptotic optimality of the least-squares cross-validation bandwidth for kernel estimates of intensity functions. *Stochastic Processes and their Applications* 38:157-165.
- Buss, A. 2001. Discrete event programming with Simkit. *Simulation News Europe* 32/33: 15-25.
- County of Santa Cruz EMS. 2004. Policy no. 7050. Major trauma victim: transport and destination.
- County of Monterey EMS. 2004. Policy no. 500-59. EMS aircraft operations.
- Diggle, P., and J. S. Marron. 1988. Equivalence of smoothing parameter selectors in density and intensity estimation. *Journal of the American Statistical Association* 83(403): 793-800.
- Green, L. V., and P. J. Kolesar. 2004. Improving emergency responsiveness with management science. *Management Science* 50-8: 1001-1014.
- Gunes, E. 2005. A kernel approach to the estimation of performance measures in a helicopter ambulance service with missing data. MS thesis. Naval Postgraduate School.
- Hall, P., Sheather, S. J., Jones, M. C., and J. S. Marron. 1991. On optimal data-based bandwidth selection in kernel density estimation. *Biometrika* 78: 263-269.
- Henderson, S. G. 2003. Estimation for nonhomogeneous Poisson processes from aggregated data. *Operations Research Letters* 31: 375-382.
- Henderson, S. G., and A. J. Mason. 2004. Ambulance service planning: simulation and data visualization. In M. L. Brandeau, F. Sainfort and W. P. Pierskalla, eds, *Operations Research and Health Care: A Handbook of Methods and Applications*, 77-102. Boston: Kluwer Academic.
- Iskander, W. H. 1989. Simulation modeling for emergency medical service systems. In *Proceedings of the 1989 Winter Simulation Conference*, ed. E. A. MacNair, K. J. Musselman and P. Heidelberger, 1107-1111.
- Jones, M. C., Marron, J. S., and S. J. Sheather. 1996. A brief survey of bandwidth selection for density estimation. *Journal of the American Statistical Association* 91-433: 401-407.
- Kuhl, M. E., Wilson, J. R., and M. A. Johnson. 1997. Estimating and simulating Poisson processes having trends or multiple periodicities. *IIE Transactions* 29(3): 201-211.
- Kuhl M. E., and J. R. Wilson. 2001. Modeling and simulating Poisson processes having trends or non-trigonometric cyclic effects. *European Journal of Operational Research* 133: 566-582.
- Kuhl, M. E., Sumant, S. G., and J. R. Wilson. 2004. An automated multiresolution procedure for modeling nonhomogeneous Poisson processes. To appear in *INFORMS Journal on Computing*.
- Law, A. M., and W. D. Kelton. 2000. *Simulation Modeling and Analysis*. 3rd ed. New York: McGraw-Hill.
- Leemis, L. 2001. Input modeling techniques for discrete-event simulations. In *Proceedings of the 2001 Winter Simulation Conference*, ed. B. A. Peters, J. S. Smith, D. J. Mederios, and M. W. Rohrer, eds, 62-73.
- Leemis L. 2004. Nonparametric estimation and variate generation for a nonhomogeneous Poisson process from event count data. *IIE Transactions* 36(12): 1155-1160.
- Lewis, P. A. W., and G. S. Shedler. 1976. Statistical analysis of non-stationary series of events in a data base system. *IBM J. Res. Development* 20-5: 465-482.
- Parker, M. W., and L. R. Johnson. 1970. Simulation of a coordinated accident rescue system. *Proceedings of*

the fourth Annual Conference on Applications of Simulation 59-63.

- Silicon Valley/San Jose Business Journal. October 22nd, 2004. Trauma looms in trauma care. Available on-line via <http://www.bizjournals.com> [accessed May 10,2005].
- Stundzia, A. B., and C. J. Lumsden. 1994. A computer-based simulator for air ambulance system analysis, design and optimization. *Air Medical Journal* 13(3): 96-103.
- Tallon, J. M., Lerner, E. B., and R. M. Moscati. 2002. The "golden hour" paradigm. *Academic Emergency Medicine* 9:760.
- The New York Times. February 28th, 2005. Crashes start debate on safety of sky ambulances. Available on-line via <http://www.nytimes.com> [accessed May 20,2005].
- The New York Times. May 3rd, 2005. As medical airlifts proliferate, the public price tag is rising. Available on-line via <http://www.nytimes.com> [accessed May, 20, 2005].
- Velleman, P. F., and D. C. Hoaglin. 1981. Applications, Basics, and Computing of Exploratory Data Analysis. Duxbury Press.
- Wand, M. P., and M. C. Jones. 1995. *Kernel Smoothing*. Boca Raton: Chapman & Hall CRC
- Wears, R. L., and C. N. Winton. 1993. Simulation modeling of pre-hospital trauma care. In *Proceedings of the 1993 Winter Simulation Conference*, ed. G. W. Evans, M. Mollaghasemi, E. C. Russell, and W. E. Biles, eds, 1216-1224.
- Wu, X., Ingolfsson, A., Akhmetshina, E., Budge, S., and Y. Li. 2005. A survey and experimental comparison of service level approximation methods for non-stationary M(t)/M/s(t) Queuing systems. Available online www.bus.ualberta.ca/aingolfsson/documents/PDF/sl_approx.pdf [accessed May 10 ,2005].
- Zhu Z., and M. A. McKnew. 1997. Effects of time-varied arrival rates: An investigation in emergency ambulance service system. In *Proceedings of the 1997 Winter Simulation Conference*, ed. J. J. Swain, D. Goldman, R. C. Crain, and J. R. Wilson, eds, 1180-1186.

AUTHOR BIOGRAPHIES

ERSAN GUNES is First Lieutenant in the Turkish Army, and obtained a MS degree in Operations Research at the Naval Postgraduate School in 2005. His interests are in helicopter operations modeling.

ROBERTO SZECHTMAN is Assistant Professor in the Operations Research Department at the Naval Postgraduate School since 2003. His research interests include applied probability, simulation theory, and military applications.