Proceedings of the 2012 Winter Simulation Conference
C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, eds.

SIMULATION WITH DATA SCARCITY: DEVELOPING A SIMULATION MODEL OF A HOSPITAL EMERGENCY DEPARTMENT

Yong-Hong Kuo,
Janny M.Y. Leung

Department of Systems Engineering
and Engineering Management
The Chinese University of Hong Kong
Shatin, New Territories, HONG KONG

Colin A. Graham

Accident and Emergency Medicine Academic Unit

The Chinese University of Hong Kong
Shatin, New Territories, HONG KONG

ABSTRACT

Our research was motivated by the resource allocations problem in the Emergency Department at the Prince of Wales Hospital in Hong Kong. We adopted a simulation approach to analysis how the allocation decisions impact patient's experience in the department. The development of the model is complicated by the fact that there are different categories of patients (with different time-varying arrival rates, treatments and procedures), and the data records were incomplete to allow direct estimation of many of the key operational parameters (e.g. the duration of doctor's consultation). To tackle the first issue, patients' arrivals are modelled as Poisson processes with category and time-dependent arrival rates. The second issue is resolved by positing a general distribution (Weibull) for some key processes, and developing meta-heuristic approaches to *jointly* estimate the distribution parameters. Our computational results show that accurate estimates of the distribution parameters are found using our proposed search procedure, in that the simulated results and the actual data were consistent.

1 INTRODUCTION

Effective management of Emergency Departments is an important problem in health-care. Patients, many with life-threatening conditions, are directed to hospital emergency rooms every day. Inefficiencies in their operations can lead to serious consequences, perhaps even unnecessary deaths. With reduced government financial support, it is increasingly difficult for hospital administrators to ensure that sufficient resources and manpower are available to maintain service quality. As a result, decision makers have to make sure that valuable resources (such as doctors' time) in the department are best-utilized. Moreover, in many hospitals around the world, there are often quite a lot of patients visiting emergency rooms for routine medical consultation. This increase in non-urgent cases leads to long waiting times and degrades the overall quality of service provided.

The Prince of Wales Hospital (PWH) in Hong Kong is a general hospital and the teaching hospital of the Medical Faculty of the Chinese University of Hong Kong. It is equipped with 1,200 beds, has a total workforce of around 3,500 and serves around 1.5 million people in the Eastern New Territories of Hong Kong. The Emergency Department (ED) at PWH provides 24-hour Accident and Emergency (A&E) services, handling around 400 cases per day. Similar to situations faced by many emergency units in hospitals around the world, it is not easy for the ED at PWH to strike the proper balance between providing a good quality of services (paying required attention to each patient) and at the same time improving the patients' experience (e.g., shortening the waiting time), particularly when resources are limited. Since one of the most valuable and limited resources is doctors' and nurses' time, matching the availability of doctors to the (uncertain and varying) demands of emergency patients is a critical issue. In order to evaluate the impact of staffing and scheduling decisions made by the ED, we are developing a simulation model to analyze the patient flows and medical services provided by the department. However, development of the simulation model is complicated by the following factors:

1. There are different categories of patients, which have different time-varying arrival rates, treatments and procedures.
2. The data records kept by the department were incomplete for many key operational processes. For example, the duration of key services (e.g., doctor's consultations) were not recorded directly.

To tackle the first issue, the patients' arrivals are modelled as Poisson processes with time and category-dependent arrival rates. To address the second issue of data scarcity, we assume a Weibull distribution for the key activities (such as consultations, triage, etc.), since the Weibull distribution fits many continuous probability density functions over the domain of positive real numbers. Next we develop two meta-heuristic search procedures to tune the distribution parameters to obtain a good estimate of the probability distribution of these key activities. Our results indicate that our search procedure enabled an accurate model to be built. This simulation model would allow various staff rosters and schedules to be evaluated without jeopardizing actual operations, and provide some insights into ED staffing policies.

We first give a brief literature review in Section 2. In Section 3, we describe the ED operations under study. We then present our simulation model in Section 4, and discuss our approach to resolving the problems due to data scarcity in Section 5. Our search procedure for parameter estimation and computational results are then presented in Section 6. Finally, in Section 7, we conclude by summarizing our contribution and discuss future directions.

2 LITERATURE REVIEW

Modelling Emergency Departments is challenging due to the unpredictable demand (patients do not make appointments) and the huge array of operations involved. Immediate care has to be given to critical patients, other patients with different levels of urgency require different treatments and procedures. Moreover, EDs usually lack sufficient physicians and resources. The many complexities of the system makes it difficult to develop analytical models with closed-form solutions. Therefore, simulation models play a very important role in the investigation of A&E operations.

The use of simulation in health-care research has been growing in the past decades. We refer the reader to Jun et al. (1999) and Standridge (1999) for an overview. In this brief literature review, we focus on papers related to staffing and operations. Rohleder et al. (2011) used simulation to determine the staffing levels and patient schedules at an outpatient orthopedic clinic where, unlike in an ED, patients normally have pre-scheduled appointments and only a small proportion of patients "walk-in". Angelis et al. (2003) used simulation to estimate function parameters of the average time a patient spent in the system (ATiS). Research on aspects of ED operations include, for example, improving scheduling rules for physicians (Carter and Lapierre 2001, Sinreich and Jabali 2007) and setting queuing priorities to reduce waiting times (Siddharthan et al. 1996). Researchers have also investigated staff scheduling and assignment (Kumar and Kapur 1989, Draeger 1992 and Evans et al. 1996), policy /decision making (Komashie and Mousavi 2005) and capacity (Baesler et al. 2003) of an ED. Some researchers have used simulation to evaluate potential system improvements, e.g. introducing a "fast track" in an ED (García et al. 1995). There is also some research integrating simulation with other methodologies. Testi et al. (2007) proposed a three-phase approach that combines simulation and optimization modules into a decision-support system for the weekly scheduling of operating rooms; where the first phase used bin-packing to determine the number of sessions for each ward, the second phase considers surgeon preference in generating a master surgical schedule, and the final stage analyzes different patient assignments using simulation. Centeno et al. (2003) used simulation to determine the parameters in an integer linear program to help scheduling staff in an ED. Yeh and Lin (2007) used simulation and genetic algorithms for nurse rostering in a hospital ED.

In all the papers mentioned above, it appears that researchers could obtain the data and information which were necessary for simulation or made assumptions if they were not available. To the best of our knowledge, the problem when some necessary data and information are missing or incomplete has not been explicitly addressed. In this paper, we propose a way to obtain a good estimate of the required parameters for simulation, when records are missing and such parameters cannot be directly estimated.

3 CURRENT DAILY OPERATIONS AND DATA COLLECTION

In the ED at PWH, patients are assessed by a triage nurse and categorized into 5 groups according to levels of urgency — Category 1 (critical), 2 (emergency), 3 (urgent), 4 (standard) and 5 (non-urgent). According to their categories, patients follow specific procedures and are directed to different treatment areas.

Since Category 1 and 2 patients are in a critical situation, whenever they are received into the ED, they are directed instantly to trauma and resuscitation rooms and given immediate attention by a doctor. Category 3, 4 and 5 patients are separated into 2 groups: walking and non-walking. After registering at the main entrance of the ED, walking patients are assessed by a nurse at the triage station. Afterwards, they have to wait in the waiting area until called to see a doctor in a consultation room. Patients from Category 3 have a higher priority over patients from Category 4 and 5, and occupy the front of the queue for consultation. Finally, after being seen by a doctor, they may have extra treatments such as X-rays and blood tests followed by a second doctor's consultation, or they can be discharged. Non-walking patients, in wheelchairs or stretchers, also follow a similar procedure for treatment, but they are seen by a different group of doctors and in a different area of the ED.

To support efficient data retrieval for prompt treatment, patient records are electronically stored on a central computer system. These records include personal information and case histories of the patients. The data are input into the system by the staff (e.g., nurses and doctors) manually at different stations at the department at different stages to keep track of the patients. Whenever a staff starts to input the information, most likely at the time a patient arrives at that particular station for service, a time-stamp is recorded to the system. These time-stamps are the only temporal data we have to extract useful information to develop our simulation model.

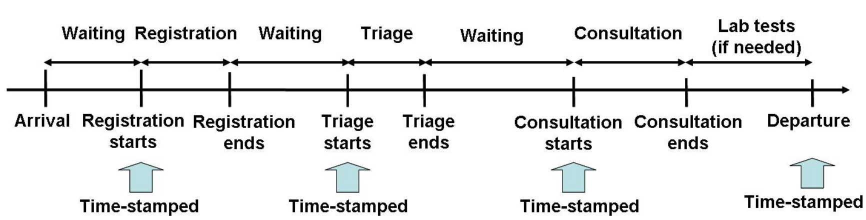

Figure 1: Time-stamps recorded when a category 3, 4 or 5 walking patient arrives.

To illustrate the temporal data collected, we describe the time-stamp record for a non-critical walking patient, representing the majority of the patients received by the department. When the patient arrives at the Emergency Department, he/she has to queue, on a first-come-first-served basis, for registration. When the patient reaches the front of the queue, the employee at the registration counter starts the system, which records a time-stamp for the start of the patient's registration, and inputs the personal information of the patient. Then the patient has to wait in the waiting area to be called for triage. When the patient is seen by the triage nurse, the nurse will check the patient's information on the computer system; hence at this time, the time-stamp for the start of triage is recorded. After triage, the patient will be assigned a category according to his/her level of urgency, and then has to wait for consultation. Next, at the time the doctor starts consultation, she/he will check the patient's information and extract his/her case history from the computer system, whence the time-stamp for the beginning of doctor's consultation is recorded. After consultation, the patient may need to receive extra treatments, such as X-Ray and blood tests. Finally, the patient can get his/her medicine and at that time a time-stamp is recorded indicating the patient is discharged. Figure 1 illustrates the time-stamps we obtained from the computer system.

4 SIMULATION MODEL

4.1 Arrival Events

From the data provided by the ED at PWH, the interarrival time of each patient category is observed to follow an exponential distribution. We also observe that the arrival rates of different categories are heterogeneous. From the ED at PWH, the numbers of Categories 1, 2 and 5 patients are relatively low whereas the vast majority of the patients are from Categories 3 and 4. Another finding from the data is that the arrival rates vary over time. These two findings are consistent with the literature (e.g., Kumar and Kapur 1989 and Rossetti et al. 1999).

The arrivals of patients with time-dependent arrival rates can be regarded as non-homogeneous Poisson processes (NHPP), which has been well studied (see e.g., Cinlar 1975, Leemis 1991 and Eick et al. 1993), although analytical results are known for only specialized cases. In our simulation, we model the arrival events by category and by time of day. To simulate an arrival of a patient, we generate an interarrival time, which follows an exponential distribution with arrival rate λ, corresponding to the arrival rate in the time period of the previous arrival for the particular patient category. Specifically, we let $\lambda_k(t)$ be the arrival rate of patient category k at time t and $A_{k,n}$ be the arrival time of the n-th patient of category k. If $A_{k,n} = t$, we let $I_{k,n}(t)$ be the interarrival time between the n-th patient and the $(n+1)$-th patient from category k. The time of the next arrival from the same category is $A_{k,n+1} = A_{k,n} + I_{k,n}(t)$, where $A_{k,1} \sim Exp(\lambda_k(0))$ and $I_{k,n}(t) \sim Exp(\lambda_k(t))$. Whenever we generate an interarrival time, we use the arrival rate that was in effect when the previous arrival occurred. This allows us to incorporate the effects of the non-stationary time-varying arrival rates of patients.

4.2 Service Activities

Service activities refer to the operations such as registration, triage, consultations and resuscitation for walking and non-walking patients. We would like to model their durations since these service times have a great impact on the queueing time for patients. The duration of each service activity has its own probability distribution. To accommodate the variety of distributions, we assume that the service times follow Weibull distributions, which can fit many continuous functions on the positive real line. Moreover, some literature also reported that some service durations follow Weilbull distributions. (e.g. Rohleder et al. 2011)

Let $S_{i,n}$ be the service time for the n-th execution of service operation i. We assume that $S_{i,n} \sim Weibull(\alpha_i, \beta_i)$. (Note that a Weibull distributions reduces to an Exponential distribution if $\beta = 1$.) For each service operation, we choose the appropriate distribution parameters α (the scale parameter) and β (the shape parameter). We assume the service rates, hence α and β, are constant for all time periods for a given service operation. Implicitly, this assumes that the doctors and nurses maintain the same level of effectiveness throughout their shifts.

5 PARAMETER ESTIMATION

5.1 Data Scarcity

To build an accurate simulation of the ED, we need the probability distributions of its activities. Ideally, these distribution parameters can be obtained from historical data. However, data from the ED at PWH on patient movements were incomplete. The available data included only the time-stamps when patients started service for the activities (triage, consultation, etc.) in the ED. Unfortunately, the time-stamps when patients completed the service were not recorded, so the service times cannot be measured directly. In this section, we present some search methods to estimate the parameters for the distributions required for our simulation model, and discuss the challenges of such parameter estimations.

5.2 Estimation of Service Time Distributions

The available data we have are the times in between the start of two different services for each patient (such as from triage to consultation). These "time differences" consist of the service time in triage and the waiting time before consultation. Thus, the actual service durations are not measured, and hence their distribution parameters cannot be directly estimated. Suppose we simulate with a "guesstimate" for the parameters of the service durations (e.g., for triage and for consultation), the resultant "time differences" can then be measured from the simulation. If the patterns of the time differences are consistent with the ones of actual data, it is likely that we used good estimates of the parameters of service times for our simulation. Since the parameters for different service distributions interact with each other to influence these "time differences", we need to consider the whole set of parameters simultaneously to determine if the parameters for all the service durations are estimated accurately. By trying different sets of values of α_i and β_i for ALL the different distributions needed for the simulation model, we choose the set for which the simulated result is consistent with the actual data.

Specifically, we let α and β be the vectors of parameters α_i and β_i; we denote by $\alpha\beta$ the appended vector $\alpha\beta = (\alpha^T \quad \beta^T)^T$. Let $\bar{x}_{k,n}$ and $x^{\alpha\beta}_{k,n}$ be the actual time and the simulated time with parameters $\alpha\beta$ from triage to consultation for the n-th patient from Category k respectively, where $k \neq 1,2$. (Since patients from Categories 1 and 2 can receive immediate treatment, with no triage and no waiting time, these categories are not considered in our consistency measure.) Let $\overline{\mu}_k$ ($\mu^{\alpha\beta}_k$ respectively) and $\overline{\sigma}_k$ ($\sigma^{\alpha\beta}_k$ respectively) be the average and the standard deviation of these $\bar{x}_{k,n}$ ($x^{\alpha\beta}_{k,n}$ respectively).

Instead of simply comparing means, we use a more detailed consistency measure to compare the distribution profiles. We divide the domains of the probability distributions into intervals. Let $\overline{p}_{k,j}$ ($p^{\alpha\beta}_{k,j}$ respectively) be the proportion of patients with $\bar{x}_{k,n}$ ($x^{\alpha\beta}_{k,n}$ respectively) in the j-th interval $[l_j, l_{j+1})$, i.e., $\overline{p}_{k,j} = \frac{\overline{N}_{k,j}}{\overline{N}_k}$, where $\overline{N}_{k,j} = |\{n : l_j \leq \bar{x}_{k,n} < l_{j+1}\}|$ and $\overline{N}_k$ is the total number of patients in the actual data. Similarly, $p^{\alpha\beta}_{k,j} = \frac{N^{\alpha\beta}_{k,j}}{N^{\alpha\beta}_k}$, where $N^{\alpha\beta}_{k,j} = |\{n : l_j \leq x^{\alpha\beta}_{k,n} < l_{j+1}\}|$ and $N^{\alpha\beta}_k$ is the total number of patients in the simulation with parameters $\alpha\beta$. We use the following function to measure the consistency between the actual data and the simulated result.

$$c(x^{\alpha\beta}, \bar{x}) = \gamma_1 \sum_k w_{1,k} \frac{|\mu^{\alpha\beta}_k - \overline{\mu}_k|}{\overline{\mu_k}} + \gamma_2 \sum_k w_{2,k} \frac{|\sigma^{\alpha\beta}_k - \overline{\sigma}_k|}{\overline{\sigma_k}} + (1 - \gamma_1 - \gamma_2) \sum_k w_{3,k} \sum_j a_j \frac{|p^{\alpha\beta}_{k,j} - \overline{p}_{k,j}|}{\overline{p}_{k,j}} \tag{1}$$

where $0 \leq \gamma_i, w_{m,k}, a_j \leq 1$, and $\sum_k w_{m,k} = \sum_j a_j = 1$. The function $c(x^{\alpha\beta}, \bar{x})$ is a weighted average of the absolute values of the relative errors of the mean, standard deviation and the proportions of classes for each category of patients. The lower the value, the higher is the consistency between the data and the simulated result.

For a given set of parameter values, we use a single simulation replication with a time horizon of a month (31 days) to evaluate the consistency function. By its very nature, two different simulation runs using the same set of parameters may produce different outcomes (e.g. using a different seed) and hence different values of the consistency function. A more statistically robust approach would be to make several simulation runs for each set of parameter values, and use the average values of $x^{\alpha\beta}_{k,n}$ to compute the consistency function. In our search procedure for parameter estimation (to be described in Section 5), the consistency function is evaluated at each search step; using several simulation runs to compute the consistency function would greatly increase the computational time for the search.

5.3 Search Procedure for Parameter Estimation

In the literature on simulation models in health-care, we did not find much discussion on the problem of unobtainable service times. One related work is Angelis et al. (2003), who estimated the parameters of the approximated ATiS function using simulation and optimization routines. In our case, the estimation problem is more complicated since we are estimating parameters of probability distributions and not just a deterministic function; the parameters of the Weibull distributions to be estimated do not explicitly appear in our consistency function $c(x^{\alpha\beta},\bar{x})$.

We estimate the set of parameters *jointly* using a search procedure. Starting with an initial guess of the distribution parameters, we compute the consistency function to evaluate "goodness of fit". We explore iteratively a search neighborhood by adding/subtracting an increment to/from the current set of parameters, until the stopping criterion is satisfied, and retain the set of parameters with the smallest value of the consistency function. We tested two widely-used search methods — the Descent Method and Simulated Annealing — and report on their effectiveness in identifying parameters for our simulation model.

5.3.1 Parameter Estimation by Descent Method

In the descent method, one moves along a search direction until no improvement can be made. At that point, the algorithm selects another search direction. The procedure stops when there is no further improvement in all directions examined. In our implementation, a search direction corresponds to increasing or decreasing the value of a single parameter. If an increase (a decrease) in the parameter value improves the consistency objective $c(x^{\alpha\beta},\bar{x})$, then we keep increasing (decreasing) the value until there is no improvement. When this occurs, we select another parameter to adjust and repeat the above search procedure. The increments of the search step should be small enough to be able to identify optima in the search neighborhood, but at the same time not too small to make our search procedure inefficient. We stop when we achieve no further improvement by changing any one of the parameter values. The procedure is summarized in Figure 2. By the descent method, we can obtain a "good" solution quickly. However, the solution can become trapped at a local optimal point and global optimality is in general not guaranteed.

Descent Method	
Step 1	Initialization: $\alpha\beta = \left(\alpha^0,\beta^0\right)^T > \mathbf{0}$ and $iter = 1$. $\Delta(\alpha\beta_i)$ = increment or decrement of the i-th component of vector $\alpha\beta$ at each iteration. Simulate a scenario with parameters $\alpha\beta$ and evaluate the consistency function $c(x^{\alpha\beta},\bar{x})$.
Step 2	Choosing a Direction: Randomly choose d. If $\alpha\beta_d - \Delta(\alpha\beta_d) < 0$, then *direction* = *up*. Else *direction* = *down*.
Step 3	Neighbour Generation: If $direction = up$, then $\tilde{\alpha\beta}_d = \alpha\beta_d + \Delta(\alpha\beta_d)$. Else $\tilde{\alpha\beta}_d = \alpha\beta_d - \Delta(\alpha\beta_d)$. Other parameters remain unchanged, $\tilde{\alpha\beta}_i = \alpha\beta_i, i \neq d$. $iter = iter + 1$.
Step 4	Comparison: If $iter = max.iter$, then terminate the procedure. Else simulate scenarios with parameters $\tilde{\alpha\beta}$ to obtain $x_{\tilde{\alpha\beta}}$. If $c(x^{\alpha\beta},\bar{x}) - c(x^{\tilde{\alpha\beta}},\bar{x}) > 0$, then keep d and *direction*. $\alpha\beta = \tilde{\alpha\beta}. \Rightarrow Step3$. Else if $direction = down$, keep d and $direction = up$. $\Rightarrow Step3$. Else $\Rightarrow Step2$.

Figure 2: The Descent Method for Parameter Estimation.

Using a more general search direction (instead of co-ordinate direction only) may improve convergence and find better local optima, but determining the search direction may not be straightforward. It would certainly be interesting to considering more general descent directions in our future research.

5.4 Parameter Estimation by Simulated Annealing

Simulated annealing (SA), introduced by Kirkpatrick et al. (1983), is a probabilistic meta-heuristic widely-adopted for global optimization problems, especially in combinatorial optimization. SA is designed to avoid the search process being trapped at a local optimum. To apply the algorithm, a neighborhood structure is defined. If there is an improvement when moving from the current solution to a neighboring point, we always make the move. However, if the move would worsen the objective value, the move is still accepted with some probability, which depends on the change in the objective function and the current number of iterations. A temperature, T, is used to associate the acceptance chance of a worse move with the number of iterations. For comparison with the descent method, we define the neighborhood as differing by one parameter value only. We summarize the Simulated Annealing search procedure in Figure 3.

Simulated Annealing	
Step 1	Initialization:
	$\alpha\beta = \begin{pmatrix} \alpha^0 \\ \beta^0 \end{pmatrix} > \mathbf{0}$, $T = T_0 > 0$, $r = r_0 (0 < r_0 < 1)$ and $iter = 1$.
	$\Delta(\alpha\beta_i)$ = increment or decrement of the i-th component of vector $\alpha\beta$ at each iteration.
	Simulate a scenario with parameters α and β and evaluate the consistency function $c(x^{\alpha\beta}, \bar{x})$.
Step 2	Choosing Direction:
	Randomly choose d.
	If $\alpha\beta_d - \Delta(\alpha\beta_d) < 0$, then $direction = up$. Else $direction = down$.
Step 3	Neighbor Generation:
	If $direction = up$, then $\widetilde{\alpha\beta}_d = \alpha\beta_d + \Delta(\alpha\beta_d)$. Else $\widetilde{\alpha\beta}_d = \alpha\beta_d - \Delta(\alpha\beta_d)$.
	The other parameters remain unchanged $i.e., \widetilde{\alpha\beta}_i = \alpha\beta_i, i \neq d$.
	$T = rT$ and $iter = iter + 1$.
Step 4	Comparison:
	If $iter = max.iter$, then terminate the procedure.
	Else simulate scenarios with parameters $\widetilde{\alpha\beta}$ to obtain $x_{\widetilde{\alpha\beta}}$.
	$\delta = \frac{c(x^{\alpha\beta}, \bar{x}) - c(x^{\widetilde{\alpha\beta}}, \bar{x})}{c(x^{\alpha\beta}, \bar{x})}$.
	If $R < e^{\frac{\delta}{T}}$, where $R \sim Uniform(0,1)$, then $\alpha\beta = \widetilde{\alpha\beta}$. $\Rightarrow Step3$.
	Else if $direction = down$, keep d and $direction = up$. $\Rightarrow Step3$.
	Else go to Step 2.

Figure 3: Simulated Annealing for Parameter Estimation.

Note that the acceptance criterion ensures that all improved moves are accepted since $e^{\frac{\delta}{T}} > 1$ as $\delta > 0$. The temperature T acts to control the search process. As the number of iterations increases, T becomes smaller and smaller, so non-improving moves have a lower chance of being accepted and the algorithm tends to resemble the descent method. When the algorithm terminates, we can also save the final solution as an initial solution to restart the procedure again. A good initial guess of the parameters $\alpha\beta$ can lead to a good solution using the above two methods. In our computations, the initial values of the Weibull parameters were estimated based on discussions with the ED operations manager at PWH.

6 COMPUTATIONAL RESULTS

To validate the simulation model and the search methodology proposed, we compare the simulated results and the actual data to see whether they are consistent. Data for a month were provided by the ED at PWH. The data consisted of the triage category, arrival date and time, and time-stamp records of triage start time, consultation start time and departure time of each patient. Moreover, the ED also provided us the actual staffing levels and schedule. We developed a program, written in EXCEL VBA, to simulate the patient flows in the ED at PWH. For each set of parameters, we simulated a whole month of patient flows.

From the data provided by the ED, there were 12882 patient visits. They were mainly from Categories 3 and 4 (95.26 % of the total). In our simulation results, there were 12548 patients and most of them

(95.24 %) were also from Categories 3 and 4. Table 1 shows that the proportions of patients observed in the simulation in each category were quite close to the actual data.

Table 1: Proportions of patients in each category: Actual data v.s. Simulated results.

	Cat 1	Cat 2	Cat 3	Cat 4	Cat 5
Actual data	1.11%	2.27%	26.57 %	68.69 %	1.35 %
Simulated results	1.08%	2.57%	26.03 %	69.21 %	1.11 %

From the actual data, we observed that the interarrival time in each category followed an Exponential distribution, where the arrival rates varied from time period to time period. In our simulation model, we assumed the arrivals are Poisson with category and time-dependent arrival rates. Figures 4, 5, 6 and 7 depict that, in terms of statistical measures, our simulated arrival process was similar in profile to the actual situation. The number of Category 1 and 2 patients were too few to be statistically significant, but results in Table 1 indicate that the Poisson assumption is valid. Similar arrival patterns were also observed by Sinreich and Jabali (2007), Draeger (1992) and Komashie and Mousavi (2005).

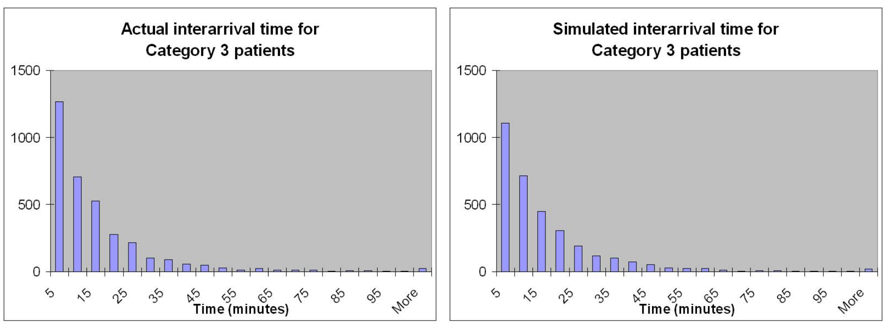

Figure 4: Interarrival time of patients from category 3: Actual data v.s. Simulated results.

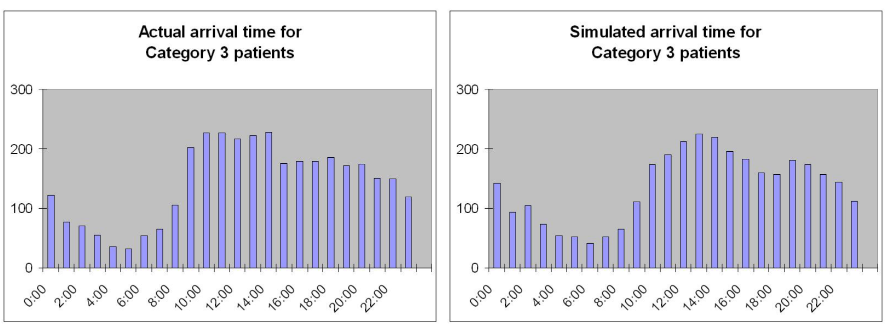

Figure 5: Arrival time of patients from category 3: Actual data v.s. Simulated results.

To determine the probability distribution of the service activities, we implemented the search methods of Section 5. For our computation of the consistency function, we regard the simulated averages, standard deviations and patterns of the net time distributions as equally important so that $\gamma_1 = \gamma_2 = \frac{1}{3}$ in equation (1). We also set $w_{1,k} = w_{2,k} = w_{3,k} = \frac{1}{2}$ for $k = 3,4$ and equal to 0 for $k = 1,2$. (Note that we put patients from Category 5 into Category 4 since the patients in these two groups are basically the same in all the procedures in the department.) The weight was taken to be a geometric sequence, $a_j = \frac{1}{2^j}$, so that more weight was placed on time intervals with less duration. [There were some *outliers* in the data where patients waited for an exceptionally long time, for example, when a nurse forgot to put a patient into the queue for

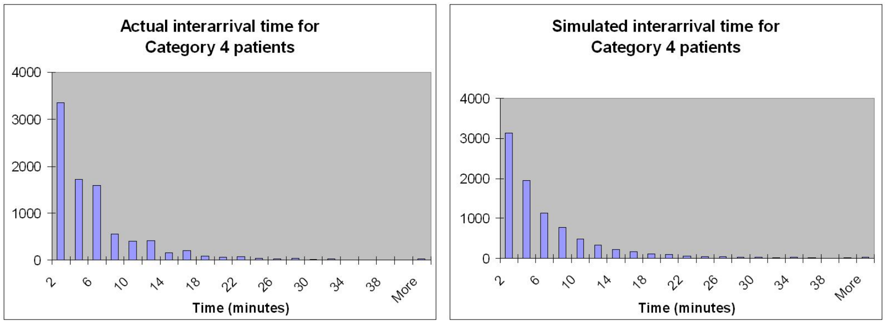

Figure 6: Interarrival time of patients from category 4: Actual data v.s. Simulated results.

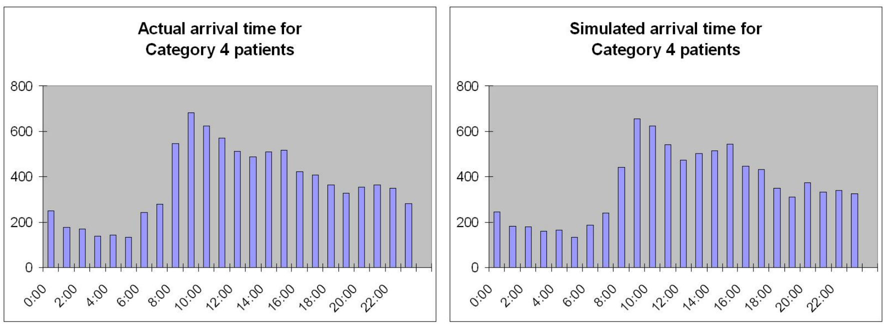

Figure 7: Arrival time of patients from category 4: Actual data v.s. Simulated results.

consultation. Therefore, we put less weight on intervals with longer duration so as to reduce the outlier effects.] In general, other weightings, such as an uniform or an arithmetic weighting, can also be used.

In defining our search neighborhood, we discretized the domains of the parameters. For each parameter $\alpha\beta_i$, we increased or decreased it by a fixed increment when adjusting its value. These values were chosen to be small enough (within 10% and 20% of the initial guesses for α and β respectively) to explore fully the terrain of the search space, and not too small to make the search procedure time consuming.

In comparing the descent method and simulated annealing for parameter estimation, we observed that the descent method became trapped at local optima, as other researchers have reported. In our computational study, we observed that this solution was quite far from the solution that we obtained by simulated annealing, in terms of the objective value and also the individual values. However, one advantage of using a descent method is that the algorithm terminates in a relatively short time since local optima can always be found easily by just moving in a beneficial direction. In terms of solution quality, simulated annealing almost always provides a better solution but the solution time may be long. Another way to reduce the solution time is to choose a smaller value of the initial temperature, T_0, and cooling factor, r_0. This can reduce the time to reach the final solution, but with the disadvantage of having a higher chance of being trapped at a local optimum. In our computational tests, T_0 and r were set to 0.1 and 0.999 respectively in order to search for solutions with better quality with acceptable search time.

We applied both the descent method and simulated annealing to search for the best parameters to fit the real data. The descent method made no improvement after 2000 iterations, which took around 3 hours. We let the simulated annealing run 5000 iterations and then used the final solution to restart the process again. We repeated this process 3 times, which took about a day in total. The consistency value of the best solutions reported by the descent method and simulated annealing are $c(x^{\alpha\beta}, \bar{x}) = 0.1736$ and 0.0738 respectively. Although the descent method was faster, simulated annealing performed much better in terms of the solution quality.

To verify that our simulated results are consistent with the actual data, we plotted the histograms of time spent from triage to consultation for Categories 3 and 4 patients using the best estimate reported by simulated annealing. Figures 8 and 9 illustrate the comparisons between the actual data and simulated results of the time spent from triage to consultation for Categories 3 and 4 patients. The figures suggest we obtained reasonably good estimates of the parameters.

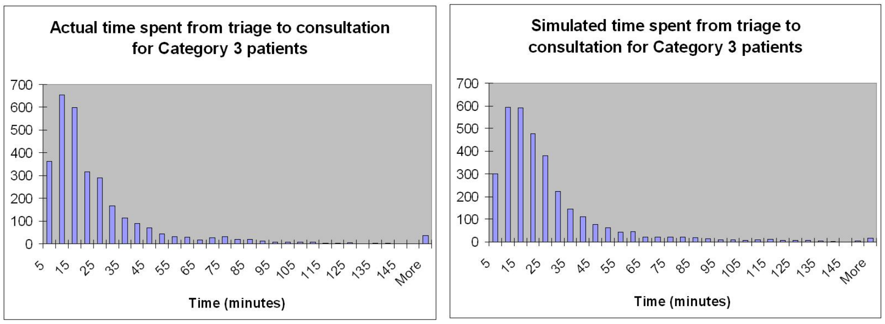

Figure 8: Time spent from triage to consultation for Category 3 patients: Actual data v.s. Simulated results.

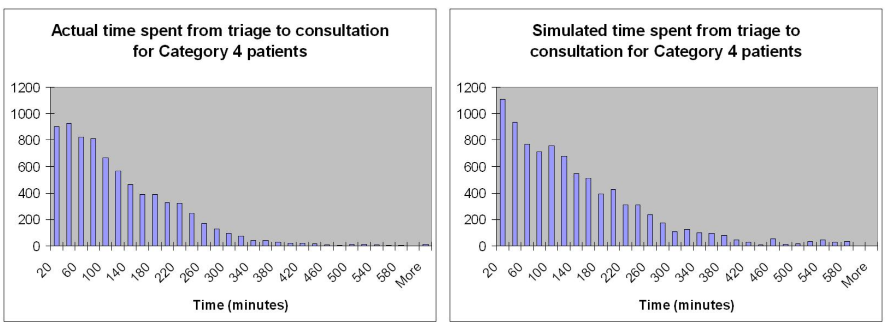

Figure 9: Time spent from triage to consultation for Category 4 patients: Actual data v.s. Simulated results.

7 CONCLUSIONS AND FUTURE WORK

In developing a simulation model for the operations of an Emergency Department of a hospital in Hong Kong, two challenges we face are the highly time-varying nature of the arrivals and data scarcity. We incorporate the complex arrival processes of patients in our simulation by modelling arrivals as Poisson processes with time and category-dependent arrival rates. Computational results in Section 6 show that the arrival patterns of our simulated results are consistent with the actual arrival data. Secondly, we address the issue of parameter estimation with paucity of data. We applied two search procedures, descent method and simulated annealing, to estimate the set of parameters for the service-time distributions when service-time data are not directly obtainable. Simulated annealing performed much better than the descent method in our experimental tests and a good set of parameters was obtained. Computational results in Section 6 show that our simulated results are consistent with the actual data. The parameter estimation procedures described in this paper has been incorporated into a simulation model being built for the PWH to measure the efficiency and performance of the daily operations of the ED. Some preliminary findings from that simulation study are:

1. the utilization of some doctors are over 100 % (requiring their scheduled breaks to be shortened);

2. adding an extra doctor to the ED and adjusting his/her shifts hours, we can reduce around 10% of average waiting time for consultation of patients; and
3. a 10% increase in patient numbers leads to about 20% increase in average total waiting time.

Our simulation model can help the operations manager at the department analyze different "what-if" scenarios before making any important decisions. This helps to enhance the quality of service that the ED is providing (e.g. reduce waiting times) since the manpower and resources can be well-allocated, and more patients are expected to be treated as a consequence.

ACKNOWLEDGMENT

The authors would like to thank Mr. Stones Wong, the operations manager at the emergency department of the Prince of Wales hospital, for his assistance in data collection.

REFERENCES

Angelis, V. D., G. Felici, P. Impelluso, V. De Angelis, G. Felici, and P. Impelluso. 2003. "Integrating simulation and optimisation in health care centre management". *European Journal of Operational Research* 150 (1): 101–114.

Baesler, F., H. Jahnsen, and M. DaCosta. 2003, December. "The use of simulation and design of experiments for estimating maximum capacity in an emergency room". In *Proceedings of the 2003 Winter Simulation Conference*, edited by S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, Volume 2, 1903 – 1906 vol.2. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Carter, M. W., and S. D. Lapierre. 2001. "Scheduling Emergency Room Physicians". *Health Care Management Science* 4:347–360.

Centeno, M., R. Giachetti, R. Linn, and A. Ismail. 2003, December. "A simulation-ilp based tool for scheduling ER staff". In *Proceedings of the 2003 Winter Simulation Conference*, edited by S. Chick, P. J. Sánchez, D. Ferrin, and D. J. Morrice, 1930–1938. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Cinlar, E. 1975. *Introduction to stochastic processes*. Prentice-Hall.

Draeger, M. A. 1992, December. "An emergency department simulation model used to evaluate alternative nurse staffing and patient population scenarios". In *Proceedings of the 1992 Winter Simulation Conference*, edited by J. J. Swain, D. Goldsman, R. C. Crain, and J. R. Wilson, 1057–1064. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Eick, S., W. Massey, and W. Whitt. 1993. "The physics of the Mt/G/∞ queue". *Operations Research* 41:731–742.

Evans, G. W., T. B. Gor, and E. Unger. 1996, December. "A simulation model for evaluating personnel schedules in a hospital emergency department". In *Proceedings of the 1996 Winter Simulation Conference*, edited by J. M. Charnes, D. J. Morrice, D. T. Brunner, and J. J. Swain, 1205–1209. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

García, M. L., M. A. Centeno, C. Rivera, and N. DeCario. 1995, December. "Reducing time in an emergency room via a fast-track". In *Proceedings of the 1995 Winter Simulation Conference*, edited by C. Alexopoulos, K. Kang, W. R. Lilegdon, and D. Goldsman, 1048–1053. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Jun, J. B., S. H. Jacobson, and J. R. Swisher. 1999. "Application of Discrete-Event Simulation in Health Care Clinics: A Survey". *The Journal of the Operational Research Society* 50 (2): pp. 109–123.

Kirkpatrick, S., C. Gelatt Jr, and M. Vecchi. 1983. "Optimization by simulated annealing". *Science* 220 (4598): 671–680.

Komashie, A., and A. Mousavi. 2005, December. "Modeling emergency departments using discrete event simulation techniques". In *Proceedings of the 2005 Winter Simulation Conference*, edited by M. E.

Kuhl, N. M. Steiger, F. B. Armstrong, and J. A. Joines, 2681–2685. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Kumar, A. P., and R. Kapur. 1989, December. "Discrete simulation application-scheduling staff for the emergency room". In *Proceedings of the 1989 Winter Simulation Conference*, edited by E. A. MacNair, K. J. Musselman, and P. Heidelberger, 1112–1120. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Leemis, L. 1991. "Nonparametric estimation of the cumulative intensity function for a nonhomogeneous Poisson process". *Management Science* 13 (7): 886–900.

Rohleder, T., P. Lewkonia, D. Bischak, P. Duffy, and R. Hendijani. 2011. "Using simulation modeling to improve patient flow at an outpatient orthopedic clinic". *Health Care Management Science* 14 (2): 135–145.

Rossetti, M. D., G. F. Trzcinski, and S. A. Syverud. 1999, December. "Emergency department simulation and determination of optimal attending physician staffing schedules". In *Proceedings of the 1999 Winter Simulation Conference*, edited by P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. Evans, 1532–1540. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Siddharthan, K., W. Jones, and J. Johnson. 1996. "A priority queuing model to reduce waiting times in emergency care". *International Journal of Health Care Quality Assurance* 9 (5): 10–16.

Sinreich, D., and O. Jabali. 2007. "Staggered work shifts: a way to downsize and restructure an emergency department workforce yet maintain current operational performance". *Health care management science* 10 (3): 293–308.

Standridge, C. R. 1999, December. "A tutorial on simulation in health care: applications issues". In *Proceedings of the 1999 Winter Simulation Conference*, edited by P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. Evans, 49–55. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Testi, A., E. Tanfani, and G. Torre. 2007. "A three-phase approach for operating theatre schedules". *Health Care Management Science* 10 (2): 163–172.

Yeh, J.-Y., and W.-S. Lin. 2007. "Using simulation technique and genetic algorithm to improve the quality care of a hospital emergency department". *Expert Systems with Applications* 32 (4): 1073–1083.

AUTHOR BIOGRAPHIES

YONG-HONG KUO is a Ph.D. candidate and obtained his M.Phil. degree in the Department of Systems Engineering and Engineering Management at the Chinese University of Hong Kong. He received his Bachelor degree from the Department of Mathematics at the Chinese University of Hong Kong. In the summer of 2007, he worked at the Oak Ridge National Laboratory as an intern. His research focuses are Optimization, Manpower Planning and Simulation. His email address is yhkuo@se.cuhk.edu.hk.

JANNY M.Y. LEUNG is a Professor in the Systems Engineering and Engineering Management Department of the Chinese University of Hong Kong. Her major research interests are combinatorial optimisation and transportation logistics. Her research has been supported by the Hong Kong Research Grants Council and the (US) National Science Foundation. Her email address is janny@se.cuhk.edu.hk.

COLIN ALEXANDER GRAHAM has been a Professor in the Accident and Emergency Academic Unit at CUHK since 2007. He qualified in 1994 and undertook higher specialist training in emergency medicine in Glasgow, Scotland before moving to Hong Kong in 2004. He combines clinical practice in emergency medicine with research into trauma and emergency medicine and is an active clinical teacher. He is Editor-in-Chief of the European Journal of Emergency Medicine and has contributed to several medical books. His email address is cagraham@cuhk.edu.hk.