ISTITUTO DI ANALISI DEI SISTEMI ED INFORMATICA
CONSIGLIO NAZIONALE DELLE RICERCHE

V. de Angelis, G. Felici, P. Impelluso

SIMULATION BASED DECISION SUPPORT SYSTEM FOR THE OPTIMAL PLANNING OF A TRANSFUSION CENTRE

R. 561 Dicembre 2001

Vanda de Angelis – Università degli Studi di Roma "La Sapienza", Piazzale Aldo Moro 4, 00185 Roma. Email: Vanda.Deangelis@uniroma1.it.

Giovanni Felici – Istituto di Analisi dei Sistemi ed Informatica "Antonio Ruberti" del CNR, Viale Manzoni 30 - 00185 Roma, Italy. Email :felici@iasi.cnr.it.

Paolo Impelluso – Università degli Studi di Roma "La Sapienza", Piazzale Aldo Moro 4, 00185 Roma.

ISSN: 1128–3378

Collana dei Rapporti dell'Istituto di Analisi dei Sistemi ed Informatica, CNR

viale Manzoni 30, 00185 ROMA, Italy

tel. ++39-06-77161
fax ++39-06-7716461
email: iasi@iasi.rm.cnr.it
URL: http://www.iasi.rm.cnr.it

Abstract

In this paper we address a design and management problem in health care systems, related with multi-server services required by one or many categories of users in different but predetermined sequences. The number of servers of each type assigned to the various services reflects on the overall efficiency of the system and on the total time spent in the system by the various categories of users. We describe a methodology to identify and validate optimal values of configuration parameters. Such methodology interactively uses system simulation, estimation of target function, and optimisation and constitute the core of an effective Decision Support System (DSS) for health care managers.

Key words: Decision Support System, Simulation, Optimization, Tranfusion Centre.

1. Introduction

In this paper we address a design and management problem in health care systems. Many studies have been carried on by ORAHS members in order to assist health system managers in evaluating the results produced by alternative decisions on how to assign a budget to various resources. We like to mention here the works by M. J. Ferreira De Oliveira [1] on size and composition of facilities in a new hospital and by J. Vissers [7] for acute care related to hospitals. M. Lagrergren [5] stresses the importance of giving insight into a system by what-if-models, that enable to evaluate alternative choices.

The problem considered here is related with multi-server services required by one or many categories of users in different but predetermined sequences. Users of the same category go from one service to another in predetermined paths but different types of users may have to follow different paths through the same set of services; the different paths interact in a complex way to determine the final performance of the system. Each service is performed by one or more type of servers (doctors, beds, instruments, etc., according to the service), each one with its cost, and the number of servers of each type that can be assigned to a service can be chosen in different way. Users may have to queue for every service and the total time spent in the system is the sum of the various service times and the waiting times spent in the various queues. Obviously, the number of servers of each type assigned to the various services, i.e. the capacity made available for each service, reflects on the overall efficiency of the health system and on the total time spent in the system by the various categories of users.

How to choose such numbers in order to keep the average total time spent in the system (shortly TiS) as low as possible within the constraints imposed by the budget is the problem tackled in this paper. The tools we interactively use to solve it are on-field survey, system simulation, estimation of target function, and optimisation. Although such techniques are quite sophisticated, we produce an output that can be read and understood quite easily and can be the base of a DSS that enables managers to take good decisions. Starting from the description of the paths followed by the various categories of users, the weight of each category, the costs of the various types of servers and the budget alternatives, the goal to be pursued in the management of such systems in order to optimise their efficiency can be expressed in two ways:

[1] Determine the configuration of servers (number of servers of each type in each service) that minimises the TiS of the users within a given budget constraint;

[2] Determine the configuration of servers of minimum cost that satisfies an upper bound on the TiS of the users.

2. The Proposed Method

In this setting the analytical tools proposed by queuing theory (see [8]) may fail to capture the complex behaviour of the system. Moreover, the solution approach adopted must result in a flexible tool that can be used by the managers to evaluate different configurations and different models of the behaviour of the system. For these reasons, we propose an integrated approach that combines the following four steps:

[1] *On-field survey*. Many service centres have started to regularly collect statistics on users

arrivals and service time. A well-designed survey may integrate the available data and identify the main parameters that are needed to model the system, that are:

[1] Arrival rates of users in the different time slots of the day and for different days of the week;
[2] Distribution of service times.

The information needed to evaluate these parameter is very limited, and can be obtained by small questionnaires that ask each user the arrival time and the starting and ending time of each service. The questionnaires may be filled by staff members or by the users themselves, and can also be used to obtain additional information on service satisfaction to assess a qualitative evaluation of the system.

2) System Simulation. Once the arrival and service time distributions have been obtained, a simulation system is set up that reproduces, with a good degree of reliability, the behaviour of the system. There are many user-friendly tools presently available to implement a discrete event simulation; to be preferred are those that provide a user friendly visual interface (see for example [1] and [6]). The design of a good simulation system is a difficult task by itself, and many such applications have been described in the literature [4]. In this setting the purpose of the deployment of a simulation tool is twofold. While, on the one hand, the simulation is used within a what-if model by the management to reproduce the behaviour of the system for specific configurations, on the other hand we intend to use it as a tool to generate a "training set" from which we can estimate some relevant analytical relation between the parameters of the system and some measure of its performance. Such relations can then be used in an optimisation model to identify specific configurations that minimize a given cost function obeying to a set of constraints. This is described in the next two steps.

3) Estimate of the Target Function. The objective of this phase is to identify the relation that links the configuration parameters of the system and some measure of its performances. For the sake of simplicity, let us consider the number of servers in each service as configuration parameters and theTiS of the users as measure of the performance. In the discussion below we omit for brevity the parameters linked to the arrival rate of users, assuming this rate is constant in the time frame considered. Assume that n services are present in the system, and that $x_1, x_2, .., x_n$ represent the number of servers for each of them; then, for each realistic configuration $x = (x_1, x_2, .., x_n)$ we obtain, using the simulation system, a value of the average time in system, say y. Running an adequate number of simulation sessions on samples of the parameter space, we derive a set of observations (x^k, y^k) , $k = 1,..,n$, from which the functional relation $y = f(x)$ may be estimated. Any valid estimation procedure may be considered to solve this subproblem; due to the nature of the underlying system, it is likely that the relation linking the parameters and the performance indicator is non linear and proper estimation methods must be considered. In this paper we have successfully solved this step using a Radial Basis Function neural network [2], that has proved to be sufficiently stable and accurate for the specific purposes.

4) Optimisation. Once a reliable estimate of the target function is available, we combine this function with additional constraints to define an optimisation model that identifies the desired solutions. One may consider two types of optimisation problems:

Problem 1: Minimise TiS
The constraints considered involve the total cost of the system, expressed by a function of the configuration parameters, integrality conditions, and upper and lower bounds on the parameters:

Min $y = f(x_1, x_2, .., x_n)$
s.t.
$c(x_1, x_2, .., x_n) \leq B$
$l_i \leq x_i \leq u_i$ *for* $i = 1, n$
x_i *integer for* $i = 1, n$

The function $y = f(x_1, x_2, .., x_n)$ is the estimated relation between the TiS and the configuration, $c(x_1, x_2, .., x_n)$ represents the cost of operating configuration $x_1, x_2, .., x_n$, B is the upper bound on the budget available to run the system, and l_i , u_i are the lower and upper bounds on the number of servers in service i, respectively.

Problem 2: Minimise budget
The constraints considered involve the TiS, expressed by a function of the configuration parameters, integrality conditions and upper and lower bounds on the parameters:

Min $c(x_1, x_2, .., x_n)$
s.t.
$f(x_1, x_2, .., x_n) \leq S$
$l_i \leq x_i \leq u_i$ *for* $I = 1, n$
x_i *integer for* $i = 1, n$

The parameter S is an upper bound on the TiS that has to be guaranteed to the users.

Problem 1 and Problem 2 are, in principle, difficult optimisation problem, as they both contain non linear objective function and constraints, and integrality constraints. Standard optimisation tools are able to solve these problems only when the number of variables involved is contained; otherwise, one should resort to local search techniques or ad hoc heuristics. Nevertheless, real applications rarely present a number of services large enough to make optimal solution of the two described problems difficult.

An important issue raises when one considers the validity of the results obtained by solving the optimisation problems described. These solutions may be affected by the reliability of the simulation system and of the estimation process (step 2 and step 3). It is important to point out that the solutions obtained by the optimisation step can be validated using the simulation system. Once an optimal solution has been determined in step 4), it can in fact be simulated and analysed in a number of sessions, and its simulated performance indicators can be computed; if the distance between the solution value of the optimisation model and the simulated one is small, we obtain a confirmation of the validity of the process; otherwise, we can enrich the training set with the new values and produce a new, more accurate, estimate. In this framework, the combination of the estimation and the optimisation (steps 3 and 4) may be viewed as a tool to explore more effectively the parameter space in order to identify good configurations; the role played by the simulation is indeed central in the whole process to create the samples for the estimation and to validate the proposed solutions.

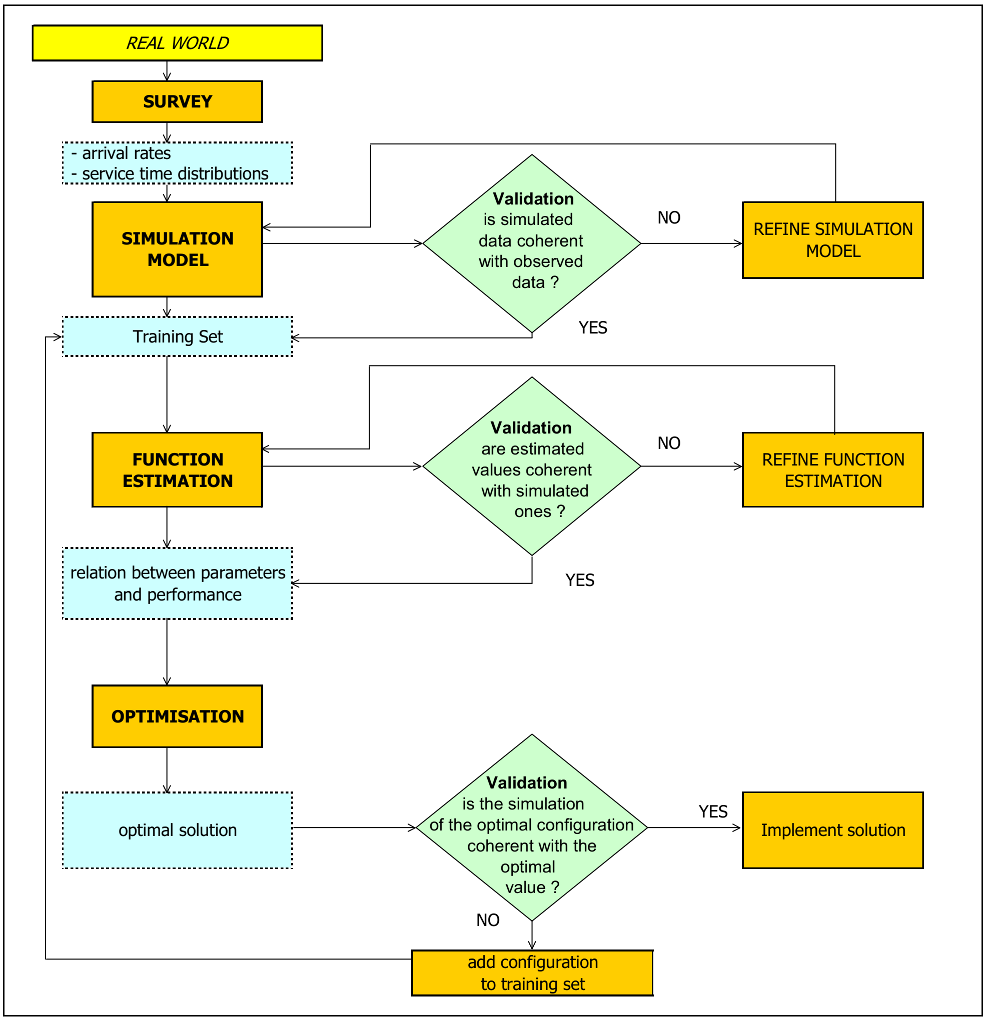

Figure 1: The proposed methodology

In Figure 1 we synthesise the complete process, highlighting the relations between the four steps and the validation loops that have to be implemented to obtain a reliable DSS.

The Transfusion Centre

The discussed methodology has been applied to a transfusion centre run in Rome by the University of Rome. The Ad Spem centre is open to donors 7-days a week from 8:15 to 12:15 and receives an average of sixty potential donors daily. The service is articulated as follows:

[1] reception (donors must fill in a form with general information and questions regarding events occurred in the last six months in the donors' life that might affect the donation);

[2] haemoglobin test (if the level of haemoglobin is low, the donor is invited to delay the donation for a few days);
[3] medical examination (a doctor checks whether the donation can take place);
[4] venipuncture;
[5] rest.

At any of the first three stages the donor may be rejected or accepted. At the rest stage, donors sit in the coffee room for a little snack, to be sure that they get sufficient rest and comfort after the donation, and then they go away. Being this the scheme of how the centre works (see also Figure 1), what determines the total time spent in the system is the quadruple of the number of servers in the first four stages. At the moment the centre works with two receptionists, one doctor or nurse for the haemoglobin test, one or two doctors for the medical examination, eight beds. The operating costs per server are 80, 100, 130, and 100 thousand liras respectively.

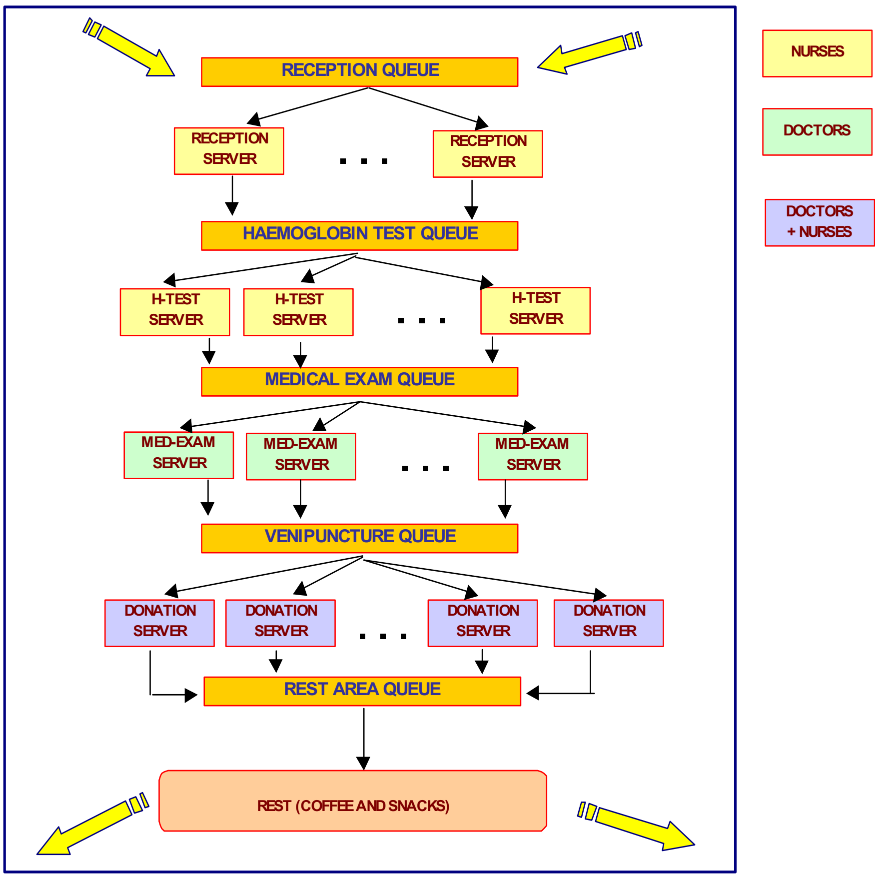

Figure 2: The Donation Centre

3. The survey on donors

In order to investigate the quantitative aspects of the centre, a survey has been carried out to collect data on the interarrival time, and on the time donors spend on the single queues and phases and totally in the centre. The survey took place in February 2001; 717

donors have filled a questionnaire, that included also questions on the qualitative perception of the service efficiency, and on demographic issues. 57% of donors declared that they were satisfied of how the centre works, 15% that they were quite satisfied. The data collected with the survey lead to accept that the interarrival time (referred to three time slots: before 8:00; between 8:00 and 10:00; after 10:00) and the time spent for reception, haemoglobin test and medical examination, follow exponential distributions, with means shown in Table 1, while the time spent for venipuncture and rest follow lognormal distributions with means and variances shown in Table 2.

random variable	*mean*
interarrival time before 8:00	5.52 minutes
interarrival time between 8:00 and 10:00	3.36 minutes
interarrival time after 10:00	10.45 minutes
time for reception	2.54 minutes
time for haemoglobin test	1.80 minutes
time for medical examination	7.97 minutes

Table 1. Parameter of the exponential probability distributions

random variable	*mean*	*variance*
time for venipuncture	19.00 minutes	57.00
time for rest	7.39 minutes	16.10

Table 2. Parameters of the lognormal probability distributions

In Figure 3 and Figure 4 we present, as an example, two charts, showing the agreement between the chosen probability distribution and the data, for medical examination time (exponential distribution) and venipuncture time (lognormal distribution). The same level of fitness has been obtained for the other service times.

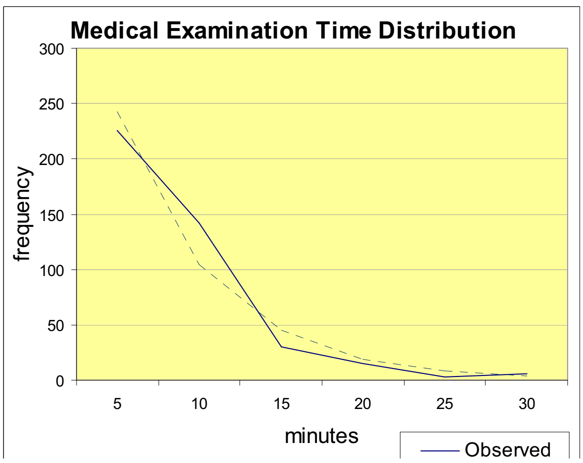

Figure 3: Medical Examination Time Distribution: observed and fitted

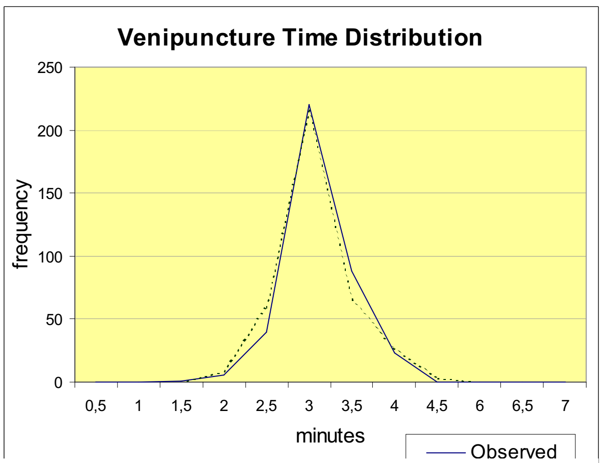

Figure 4: Venipuncture Time Distribution: observed and fitted

The simulation of the centre

The functioning of the donation centre has been modelled by a discrete event simulation system. The system is described by:

[1] the number of servers in each stage, that define the configuration of the system;
[2] the type of inter-arrival time distributions and the value of their parameters, that have been estimated by means of the survey described in the previous section;
[3] the type of service time distribution and the value of its parameters in each service, that have been estimated by means of the survey described in the previous section;
[4] the events and the activities that tale place in the system.

A simulation program has been developed *ad hoc* for this application, using MS Visual Basic. It has been designed as a network of timers, each one associated to the different events in the system (arrivals, start of service, end of service), and queues, that collect the donors waiting to be served between the stages of the donation process. The final version of the system is embedded in a visual simulation tool that shows the progress of donors through the centre, the length of the different queues and the value of several parameters related with the behaviour of the system. Figure 5 depicts the main window of the system.

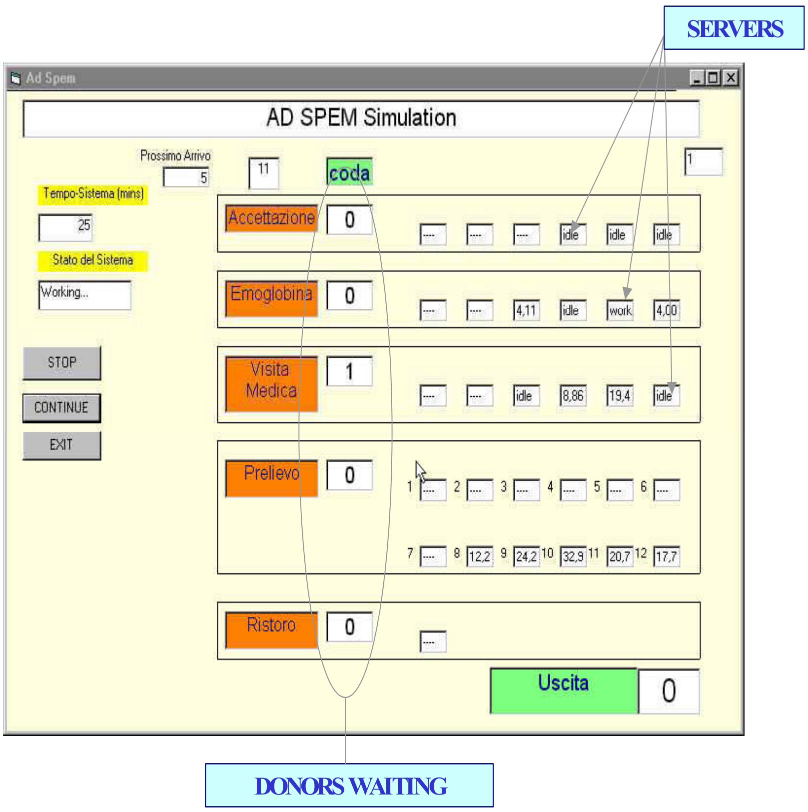

Figure 5: The Main Window of the Simulation System

4. The estimate of the TiS function

The simulation system described in the previous section has been used to generate a number of observations by means of which we could reasonably estimate the relation between the TiS and the configuration parameters. In order to obtain a significant sample of the parameter space, the observations have been generated considering only feasible configurations of the servers; moreover, we have excluded those configuration that would result, *a priori*, inefficient.

The parameters that have been varied are the arrival rate per day and the number of servers in each stage of the process. To produce only observations in interesting regions of the parameter space, we have proceeded as follows. Starting from a minimum number of arrival per day (30), fifteen simulation sessions with the minimum number of servers were run. Then, the number of servers was increased, and for each new configuration fifteen additional simulations were completed. When the TiS of the simulation reached an ideal value (approx. 30 minutes), the arrival rates were increased and the whole process repeated. This has been iterated until the value 120 for the daily arrival rate was reached, counting up to 330 different simulation sessions. Each session was represented by:

[1] y: value of the average time in system:
[2] λ: daily arrival rate;
[3] x_1: number of servers in reception;
[4] x_2: number of servers in haemoglobin test;
[5] x_3: number of servers in medical visit;
[6] x_4: number of servers in venipuncture;

The 330 observations were used to estimated the function $y = f(\lambda, x_1, x_2, x_3, x_4)$. As expected, the relation appeared to be strongly non linear and standard regression techniques were not able to produce an estimate of sufficient quality. We resorted to adopt a Radial Basis Function (RBF), a particular type of neural network that is particularly flexible and effective in estimating non linear functions [2]. We omit for brevity a detailed description of such technique; interested readers may consult [3] for more information. A RBF network is defined by a number of input neurons, a single hidden layer represented by m baricenters, weights γ_i, and one or more output neurons. The baricenters can be chosen at random amongst the input vectors or by more sophisticated techniques. The estimating function is:

$$y(x) = \sum_{i=1}^{m} \gamma_i \sqrt{(\|x - c_i\|^2 + \sigma^2)} \tag{1}$$

where x is the input vector, c_i is the i-th baricenter and σ is a constant. When the baricenters are determined in advance, the training problem amounts to the minimisation of a quadratic function in the weights γ_i. For the available data, we determined a satisfactory approximating function using m = 30 baricenters and σ = 10. This problem was solved exactly with optimisation routines from the NAG library. We compare, in Figure 6, the plotting of the observed values of TiS with the ones of the TiS estimated by the neural network; the agreement between the values is good and it shows how the estimated function is able to capture the essential behaviour of the relation that we want to approximate.

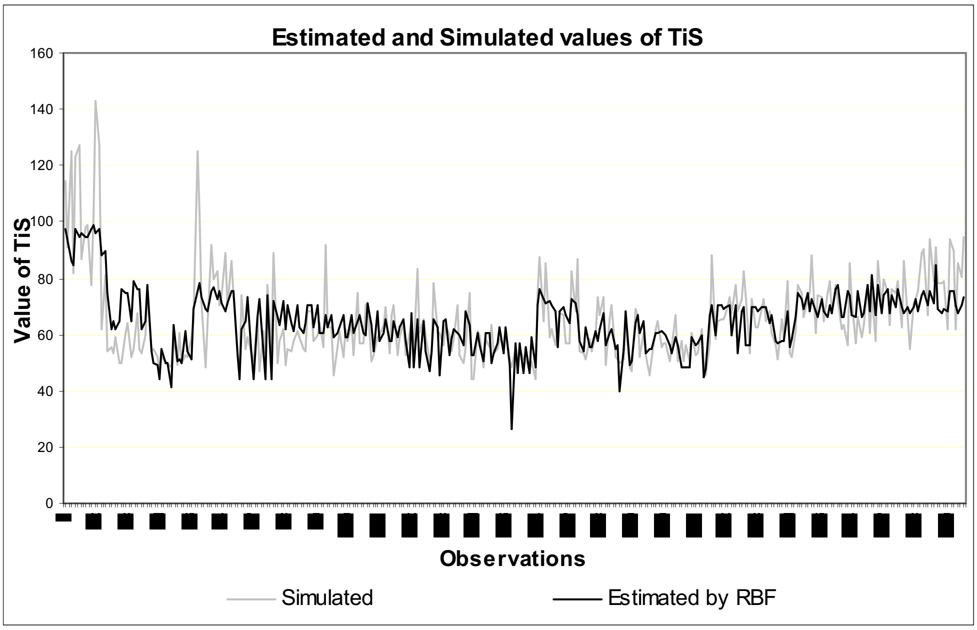

Figure 6: Estimated and Simulated values for Time in System

5. The optimisation models

Two optimisation problems are considered, as already sketched at the beginning of the paper; while the first problem aims at minimizing the TiS given the budget available, the second represents the situation where the management aims to operate the donation centre spending the minimum amount of money that guarantees a desired level of service (as expressed by the TiS). The elements that must be taken into account in both problems are:

$f(\lambda, x_1, x_2, x_3, x_4)$	Estimate of Average Time in System
λ	Arrival Rate
x_1, x_2, x_3, x_4	Number of servers (integer variables)
$\sum_{1=1}^{4} c_i x_i$	Daily Cost
$l_i \le x_i \le u_i,\ i = 1,4$	Bounds on servers

The two optimisation problems are then:

Problem 1) *Minimize Time in System*	Problem 2) *Minimise Cost*
$\min f(\lambda, x_1, x_2, x_3, x_4)$	$\min \sum_{i=1}^{4} c_i x_i$
subject to :	subject to :
$\sum_{i=1}^{4} c_i x_i \le B$	$f(\lambda, x_1, x_2, x_3, x_4) \le S$
$\lambda_{\min} \le \lambda \le \lambda_{\max}$	$\lambda_{\min} \le \lambda \le \lambda_{\max}$
$l_i \le x_i \le u_i,\ i = 1,4$	$l_i \le x_i \le u_i,\ i = 1,4$
x_i integer, $i = 1,4$	x_i integer $i = 1,4$

The values of B, S, λ_{min} and λ_{max} are parameters that need to be fixed before solving the two problems. For example, if one wants to determine the optimal configuration to serve between 40 and 50 donors per day, spending no more than 1000 thousand liras, then Problem 1 should be solved with B = 1000, λ_{min} = 40 and λ_{max} = 50. On the other hand, the solution of problem 2 with the same values of λ_{min} and λ_{max} and a value of S = 35, would determine the configuration that serves 40 to 50 donors per day, with a maximum TiS of 35 minutes, at the minimum cost. Both problems are theoretically difficult, but their small size makes it possible to solve them with commercial solvers in little time (quick results were obtained, in this application, with the MS Excel 2000 solver with Newton descent method).

6. The results

We have conducted a number of experiments varying the parameters of Problem 1 and Problem 2. We restrict for brevity to the results of the optimisation problem 1, presented in Table 3. For different values of B in the budget constraint (750, 1000, 1250, 1500) and different ranges of the arrival rate (from 30-40 to 90-100) we describe the results of the optimisation with respect with the optimal value of TiS and the related cost (the budget constraints is not always tight at the optimal solution). To each pair (B, range of the arrival rate λ) is associated an optimal configuration, that is here omitted for brevity.

	Value of B in the Budget Constraint							
Arrival Rate λ (per day)	*750*		*1000*		*1250*		*1500*	
	Budget	*TiS*	*Budget*	*TiS*	*Budget*	*TiS*	*Budget*	*TiS*
30-40	750,0	39,7	963,3	21,4	1063,3	20,1	1293,3	20,4
40-50	750,0	54,8	996,7	31,1	1223,3	21,4	1223,3	21,4
50-60	740,0	63,4	1000,0	47,6	1243,3	35,8	1456,7	30,9
60-70	743,3	68,3	1000,0	59,8	1230,0	48,6	1490,0	40,4
70-80	743,3	68,3	1000,0	64,8	1230,0	53,7	1490,0	48,4
80-90	750,0	80,1	996,7	70,3	1230,0	65,5	1490,0	59,0
90-100	743,3	79,4	1000,0	73,8	1230,0	67,5	1490,0	62,7

Table 3: Optimisation solution for problem 1 (costs are expressed in thousand of liras)

The results behave in a very reasonable way, and several interesting hints can be extracted from the figures of the table. As an example, consider the following situation. A standard configuration that is currently operated in the centre is (1,1,3,7), that has a operating cost of approximately 803 thousand liras and determines, with arrival rate of 60, a TiS of 83 minutes. We have run optimisation Problem 1 setting $B = 803$, $\lambda_{min} = \lambda_{max} = 60$ and have obtained an optimal value of TiS of 66 minutes with the new configuration (1,1,2,11). One may question the applicability of such configuration. Thus, we have run again a simulation session with the same arrival rate and the proposed optimal configuration. The simulation produced a TiS of 69 minutes, confirming the indication of the optimisation model and validating the new configuration before its application.

References

[1] Ferreira de Oliveira M.J., 3D Visual Simulation Platform for the project of a new hospital facility, in V. De Angelis, N. Ricciardi, G., Storchi (Eds), *Monitoring, Evaluating, Planning Health Services. Proceedings ot the 24th meeting of the ORAHS EURO-WG*, World Scientific, Singapore, 1999, pp.82-95.

[2] Funahashi K., On the approximate realization of continous mappings by neural networks, *Neural Networks*, 2 (1989) 183-192.

[3] Haykin S., *Neural Networks*, Macmillan Publishing Company, Englewood Cliffs, 1994.

[4] Jones, L.M., Hirst A.J., Visual Simulation in Hospital: A managerial or a political Tool?, *European Journal of Operational Research* 29 (1987) 167-177.

[5] Lagergren M., What is the Role and Contribution of Models to Solving Health Care Management Problems?, in A. Kastelein, J. Vissers, G.G. Van Merode, L.Delesie (Eds), *Managing Health Care under Resource Constraints. Proceedings of the 21st annual meeting of the ORAHS EURO-WG*, Eindhoven University Press, Eindhoven, NL, 1966

[6] Law M. , Kelton W. D., *Simulation Modeling & Analysis*, McGraw-Hill Inc., 1991.

[7] Vissers J., Van Der Bij H., Kusters R., Towards Decision Support for Waiting Lists: An Operations Management View, *Health Care Management Science* 4 (2) (2001) 133-142.

[8] Tanner M., *Practical Queueing Analysis*, IBM McGraw-Hill series,1995.