# ESTIMATION OF MAXIMUM QUEUE LENGTHS USING SIMULATION-BASED REGRESSION

# JAVIER OTAMENDI

# JOSÉ MANUEL PASTOR

Universidad Rey Juan Carlos, Campus Vicálvaro Facultad de Ciencias Jurídicas y Sociales Departamento Economía Aplicada I Paseo Astilleros s/n 28032 Madrid, Spain E-mail: franciscojavier.otamendi@urjc.es Universidad Castilla La Mancha Escuela Universitaria Politécnica de Cuenca Departamento de Sistemas Informáticos Campus Universitario s/n 16071 Cuenca, Spain E-mail: josemanuel.pastor@uclm.es

**Abstract:** A real application of simulation in the design of public transportation facilities is presented. The flow of passengers is modelled using discrete-event simulation, which facilitates the abstraction of the emptying and filling of the passengers into the buses, rather than the more usual continuous simulation for crowd movement. The decision variables include not only the physical layout of the queues but also some tactical variables that will help in future negotiations with labour unions and owners of the lines. A regression analysis has been performed using the simulation results to study the relationship between the maximum queue length and the input/output ratio in the gates.

Keywords: Dynamics of flow, queuing, simulation, regression analysis

# 1. INTRODUCTION

Among the design activities for one of the future underground stations in a large city of Spain, it is necessary to study the capacity of the waiting area so that the formation of waiting queues does not interfere with the typical movement of the users of the facilities. While the design of the building and the individual platforms is being performed from a static, architectural point of view, there is also a need to validate the design from a dynamic point of view.

Within the Technical Specifications Document, the development of mathematical models to perform the validation stage is required. Specifically these model should estimate the maximum length of the waiting queues at the gates. The input values come from historical data about demand, frequency, bus capacities and relevant decisions on the management of the line.

The system that is the focus of this article is a bus station. The station belongs to the Administration but the bus lines are privately owned. There exists the need that both parties negotiate along with the drivers of the buses to reach an agreement on how the station is to be run. Buses interarrival time, vehicle sizes, gates opening times are key decision variables, which have an important impact on the size of the waiting queues, and therefore on the size of the station building.

The article starts with a description of the real system as designed by the architects, including the

objectives and the data that has been collected from the public administration. A theoretical study of the maximum queue length follows. The model is then described, which includes an interface developed in MsExcel. There is a separate section for the definition and analysis of the simulation experiment. The analysis of the negotiation process and the development of a quantitative negotiation tool is then detailed. Finally, an allocation of bus lines to gates is proposed.

# 2. THE SYSTEM

#### 2.1 Description

Let us start with a brief description of the facilities (figure 1) and the flow of its users.

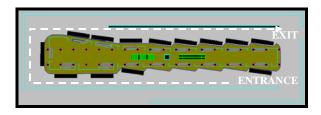


Figure 1: Station layout

The bus station consists of a platform including 16 gates (depicted as a rectangle) with the possibility of parking 12-metre or 15-metre buses. The direction of movement of the vehicles is clockwise with the entrance and exit on the right hand side of the premises.

In front of each of the gates, queues of users form in a zigzag shape (figure 2). The layout of the queues has this shape for three main reasons. First, the movement along the platform is facilitated since the queue should never reach the centre of the platform, thus blocking the movement.

Second, the psychological factor/perception is improved: the rate of movement of the queue appears quicker than the actual rate to the people in the queue because users are moving in opposite directions.

Third, this disposition also helps the flow of passengers moving in and out of the buses, since room is specifically dedicated to the output flows.

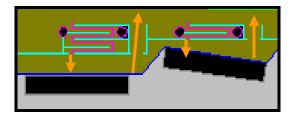


Figure 2: Queues at the gates

To access the platform, the passengers have to use either a mechanical stair or a normal one, which are located in the centre of the platform to reduce the distance between the stairs and the gates. The same set of stairs is used to leave the premises.

#### 2.2 Output Variables and Thresholds

The main objective is to measure the length of the queues. The aim is to prevent the queue from growing so long as to affect the flow in the centre of the platform. That is, once a passenger gets to the bottom of the stairways, it should have a free path towards his/her departing gate.

The maximum number of waiting passengers is set to 56, which is the capacity of a zigzag queue with four rows of 14 passengers each. This number is also similar to the capacity of the small-size buses that are often used. This queue size also agrees with the area in square meters of the designed platform.

The degree to which this objective is met is going to be monitored by measuring the length of the queue every thirty seconds and plotting the corresponding values in a timeseries. There will be a separate plot for each of the gates.

Another objective is to monitor the time that the passengers spend in the queue. However, this

variable depends on the buses interarrival times, which are not design variables but, rather, input data. As long as a passenger is able to catch the next arriving bus, there should be no complaints.

#### 2.3 Available Data

# 2.3.1 Bus capacity

The 12-meter buses have a capacity of 55 and the 15-meter ones might hold up to 71 passengers, if they all are seated. Some 15-meter buses, the so-called mixed buses, might contain 134 users (46 seated and 88 standing).

#### **2.3.2 Demand**

The prediction for the next two decades is that thirteen bus lines will have a stop at this underground station, even if there is room for 17 gates.

Figure 3 shows the data included in the Technical Specifications Document. Each row or record corresponds to one of the thirteen bus lines. An explanation of the different columns follows:

- PHI: PeakHourInterval. It is the time in minutes between consecutive departures
- PeakHour: hour of the day at which the peak time occurs
- Buses: number of departures at peak time
- DailyDemand (arrivals + departures): total volume of users of the facilities
- DailyDepartures
- DailyArrivals
- HourlyDemand: daily data obtained from on-site sampling:
  - o Arrivals: peak rate
  - Departures: peak rate
  - Capacity: maximum cumulative number of arrivals and departures.

From these hourly and daily rates, the variable PeakPercentage is calculated. This variable is the percentage of DailyDepartures at the PeakHour, and it is used to calculate the HourlyRates for those lines for which only daily rates are available.

As this percentage varies between 6.72% for line IV and 11.4% for line IX, this limit is taken for safe calculations (it is also the maximum likelihood estimator).

	Peak Hour Interval	Peak Hour	Buses	Daily Demand	Daily Departures	Daily Arrivals		Hourly Demand		Peak Percentage
	FIII	Departs			טט	νA	Arrivals	Departures	Capacity	
1	12	10	5	3068	2609	968				
2	6	8	10	6186	2284	1827				
3	20	7	3	1365	1248	1075		131	160	0.1050
4	20	14	3	1849	1324		77	89	113	0.0672
5	15	14	4	2843	1501	1269	149	162	212	0.1079
6	30	15	2	1035	727		102	71	115	0.0977
7	30		2	736	441.6					
8	30		2	934	560.4					
9	12	15	5	6295	2009	1638	245	229	275	0.1140
10	20		3	1888	260	497				
11	12	14	5	7947	2199	1908	264	223	300	0.1014
12	30		2	3519	2111.4					
13	15		4	3529	2117.4					

Figure 3: Input data

# 2.3.3 Operation times

This section presents the additional data which is needed in the simulation model. The corresponding values have been provided by the management of the facilities:

- GateOpening: amount of time prior to departure in which the gates open for the users to catch the bus.
- BusSpeed = 0.004 minutes/meter (15 km/hr)
- PassengerVelocity = 0.020 minutes/meter (3 km/hr)
- FillRate = 20 pax/min
- EmptyRate = 50 pax/min

# 3. THEORETICAL ANALYSIS OF QUEUES AT THE STATION

# 3.1 Behaviour of Queues

The theoretical evolution of the length of the queues at the bus station is shown in a timeseries diagram (figure 4). Whenever the door is opened, the length of the queue diminishes. When the bus departs, the length of the queue rises again. The problem with busy bus lines is that at some point, when the bus departs, there are still people waiting in the queue for the next scheduled bus to arrive.

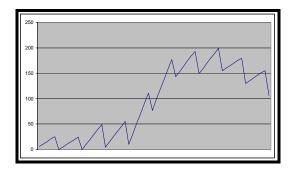


Figure 4: Maximum queue length over time

For this same reason, the difference in the length of the waiting queues when the capacities of the buses are changed is not significant, but it really is crucial when there are passengers who cannot catch the next departing bus. Even the large buses cannot hold then the total amount of waiting users.

# 3.2 Definition of IO Ratio

The previously defined behaviour of queues might be theoretically analyzed via models, which quantify the performance of a service station. They are analytical models that, for specific combinations of input flow, service times, number of servers and queue disciplines, estimate the queue lengths and times in the system [Taha, 1987]. The requisite, however, is that the capacity installed must be enough to handle all the incoming flow, that is, the input flow must be smaller than the output capacity.

The Input/Output Ratio, IO\_Ratio, might be accordingly defined as the ratio between the incoming flow of passengers joining the queuing system (DEMAND) and their rate of service (THROUGHPUT), in total numbers per unit time:

$$IO_Ratio_p = DEMAND_p / THROUGHPUT$$

The demand for the peak period might be defined as:

$$DEMAND_p = TD*PD_p$$
 (1)

where TD stands for the total demand for the time horizon and  $PD_p$  for the percentage of TD that arrives within the p period.

Let us define throughput as the number of customers that a single station can serve per unit time, which might be calculated as follows:

$$THROUGHPUT = min(CAP,SR*min(CT,PT))*Nbuses (2)$$

where CAP stands for the capacity of the station in number of passengers, SR is defined as the service

rate, CT is the length of the period and Nbuses for the number of buses per period.

In other words, during each active period, the number of units that are served is the minimum of the total capacity of the service station and the maximum number of units individually served.

Therefore, using equations (1) and (2), the ratio may be calculated as:

$$IO\_Ratio_p = \frac{TD*PD_p}{min(CAP,R*min(LAP,IAT))*NAP}$$
(3)

If the IO\_Ratio is smaller than 1, all the customers can be served within the period, whereas a value of this index greater than 1 indicates that the queues are going to grow long.

#### 4. THE SIMULATION MODEL

The use of analytical models does not help in all the cases. Specifically, the robustness of the analysis is not appropriate whenever the probability distribution of the different input variables or data does not follow one of the standard ones. Also, its usefulness decreases as the complexity of the relationships between the input and the output variables increases.

Another modelling technique is necessary to correctly account for both uncertainty and logic relationships. Simulation appears to be in the last decades like a very appropriate technique due to its good compromise between efficacy and efficiency.

Design of facilities is one of the complex areas in which simulation fits perfectly as the mathematical model to use, not only due to its very good representation capabilities but also due to its exceptional experimentation possibilities in a design phase [Lin, 2000]. In this particular case, although the model should have been used more in the earlier stages of the design [Kiran, 2000], it is mainly used to validate the model from a dynamic point of view.

Even if continuous modelling looks to be the appropriate choice to model a dynamic system involving the movement of crowds [Chalmet et al 1982; Shen, 2005], it is sometimes possible to use a discrete-event approach [Otamendi, 2005].

If the logic gets too complex, the continuous models have to include events to model the discontinuities. If the modeller feels more comfortable with the discrete-event approach, this technique should also be tried, especially with problems of small size.

# **4.1 Description of the Model**

The buses are generated according to the peak hour intervals, directing them towards a preassigned gate, where the processes of passenger alighting and boarding take place.

The alighting process starts as soon as the bus gets to the gate and the doors open. The number of passengers in the arriving bus is calculated as follows:

When all the passengers have gotten out of the bus, the doors are closed. Several minutes later, at a predefined time prior to departure, *GateOpening*, the doors opened again and the passengers start to board the bus. The rate of passengers is assumed constant and it is calculated as follows:

$$\begin{aligned} & \text{DepartingPax} = \text{ArrivingPax} \\ & \text{DepartingRate} = \frac{60}{\text{DepartingPax}} \end{aligned}$$

The bus then departs at its corresponding time with no delays.

# 4.2 Hypothesis

To facilitate this modelling step, the following restrictions are imposed:

- The peak hour coincide for any bus line.
- The simulation period is 5 hours (300 minutes), with the peak hour being the middle hour, the two contiguous hours with a rate of 50% of the peak rate and the other two with a rate of 25% of the peak (figure 5).

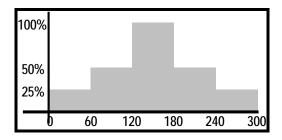


Figure 5: Demand pattern

The hourly demand is randomly assigned.
The total volume of passengers is created
and each passenger is assigned a random
number between 0 and 1. When multiplied
by 60 minutes, the moment of entry in the
system is determined.

- The arriving passengers leave the premises, that is, they do not take another bus.
- The departing passengers access directly to the queue, without waiting in the lounges.
- There are no delays with respect to the arriving or departing schedule.

The manager of the facilities has accepted these hypotheses since the behaviour of the system is still well represented. The model is therefore validated and ready for experimentation.

#### 4.3 Interface

To make use of the model easier, a user-friendly interface has been developed in an MsExcel spreadsheet (figure 6).

The worksheet includes the possibility of defining the length of each simulation run (300 minutes as the initial value) and the number of repetitions of the model to run (the first number is the total number and the second is the run being performed, both being equal at the end of the execution).

The input values to include per bus line are:

- The daily demand
- Peak percentage
- Interarrival time
- Bus capacity
- Gate opening times
- Service rate

With the above variables, the following ones are readily calculated:

- Input
- Number of buses
- Output
- IO\_Ratio

Once the *Reset* button is pressed, the worksheet is cleaned up, and the modelling tool is ready for experimentation. Pressing the *Execute Simulation* button starts the simulation run.

The maximum queue length per run is also presented in the application.

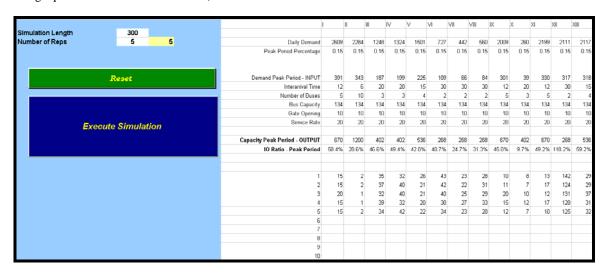


Figure 6: Simulator's interface

# 5. SIMULATION EXPERIMENT

With the model validated and the experimenter ready, it is time to set up an experiment to study the relationships among the variables that affect the negotiation process and the maximum length of the waiting queues. In other words, the model should be executed in order to obtain the data to facilitate the development of a negotiation tool.

# 5.1 Decision Variables

The parameters that a priori might affect the length of the queues are:

- PeakPercentage (2 levels): with an initial estimated value of 11.5%. Foreseeing increments in certain days of the year, the system is analyzed for a peak rate of 15% of the total daily demand.
- BusType (3 levels): the three possibilities have already been mentioned: 12-meters, 15-meters (all seated) or mixed (15-meters with some standing passengers).
- PeakInterval (1 level): with the initial values set by contract and included in the specifications.
- GateOpening (2 levels): with values being 5 or 10 minutes prior to the schedule departure of the bus.

Variable	LINE	N	Mean	StDev	Q1	Median	Q3	Maximum
MAX LQ	1	60	53.52	39.66	20.75	43.50	70.00	142.00
	2	60	5.133	3.223	2.000	5.500	8.000	11.000
	3	60	39.30	10.95	29.75	38.00	45.75	70.00
	4	60	43.52	13.46	33.00	40.50	50.00	79.00
	5	60	28.92	10.40	19.25	29.50	38.50	48.00
	6	60	39.067	6.812	34.000	39.000	43.000	51.000
	7	60	24.550	4.382	22.000	23.500	28.000	34.000
	8	60	29.967	4.438	27.000	29.000	33.750	39.000
	9	60	27.27	14.42	13.00	30.00	36.00	60.00
	10	60	8.867	2.390	7.000	9.000	10.000	14.000
	11	60	33.48	21.97	12.00	35.00	44.75	85.00
	12	60	197.2	80.2	133.3	188.0	256.8	344.0
	13	60	57.18	31.22	32.25	47.00	66.00	138.00

Figure 7: Results for raw data

# **5.2 Planned Experiment**

Five runs of the simulation model are executed for each of the feasible combinations (2\*3\*1\*2=12), accounting for 60 runs in total, each with 13 simulated queue lengths, one per bus line.

#### 5.3 Results with Raw Data

Figure 7 includes a table of descriptive statistics:

- Size of the sample
- Mean and Standard deviation of the sample
- The quartiles (Q1, Median or Q2, and Q3)

The results indicate that the only clearly problematic bus line is Line XII, with a maximum of more than 300 passengers. The boxplot shown in figure 8 shows these findings in graphical format.

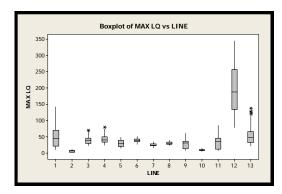


Figure 8: Boxplots for raw data

# 5.4. Raw Data Without Line XII

If trying to develop any type of analytical tool with this set of data, this unusual group, line XII, might bias it. So the decision is to discriminate this group out of the analysis. In fact, the management of the facilities already knew about its behaviour and, as it will be later explained, there is a reserved area in the premises to cater for the length of line XII.

The first attempt is then made to estimate the relationship between IO\_Ratio and Max\_LQ using

all of the data but that for Line XII. Figure 9 shows the estimation with the corresponding 95% prediction bands using linear regression.

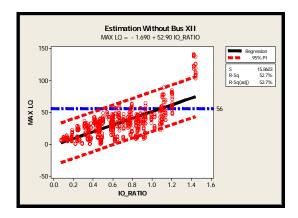


Figure 9: Linear estimation without bus XII

The fit is not yet good, as it explains only half of the variance in the data (*R-Sq*). However, it helps more clearly understand the objective of the study. There is an indication that the IO\_Ratio should be kept below 0.5 if the length is to be always maintained below the maximum queue size of 56. If the IO\_Ratio goes beyond 1.2, the length starts to be outside the prediction area. That behaviour of high input values might be better represented with a quadratic equation, as shown in figure 10.

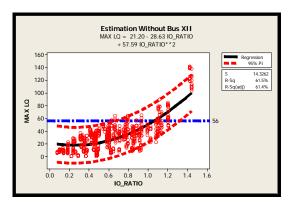


Figure 10: Quadratic estimation without bus XII

# 5.5. Excluding IO\_Ratio < 1.2

Even if the high values of IO\_Ratio are well fitted by the quadratic equation, in order to produce good estimation in the proper domain, it is reasonable to discriminate values with an IO\_Ratio greater than 1.2 (figure 11).

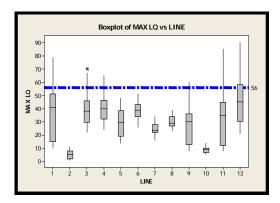


Figure 11: Boxplot excluding IO\_Ratio ≤ 1.2

Once set in the proper domain, both the linear model (figure 12) and the quadratic model (figure 13) can be estimated again.

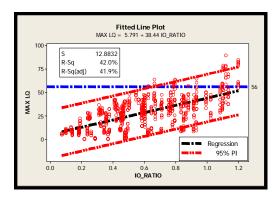


Figure 12: Linear estimation without bus XII and  $IO_Ratio \le 1.2$ 

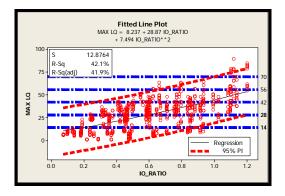


Figure 13: Quadratic estimation without bus XII and  $IO\_Ratio \le 1.2$ 

This last graph also includes one horizontal line per number of queues of 14 passengers.

Figure 14 includes the equation resulting from the last adjustment in analytical form. The regression model is significant at almost any degree of confidence (p-value = 0).

The regression equation is								
MAX LQ =	MAX LQ = 5.791 + 38.44 IO_RATIO							
S = 12.8832	2 R-Sq = 42.0	0%	R-Sq(adj) = 41.9%					
Analysis of	Analysis of Variance							
Source	DF	SS	MS	F	P			
Regression	1	82547	82547.1	497.34	0.000			
Error	688	114191	166.0					
Total	689	196738						

Figure 14: Final model in analytic form

#### 6. NEGOTIATION STRATEGIES

With the proper data already identified, the negotiation tool might be now developed. A thorough study calls for the identification of the variables, and their study, both separately and jointly.

# **6.1** Components in the Negotiation Process

None of the variables are directly controlled at this validation stage by the management of the station. The *PeakPercentage* has been estimated from current data and is clearly influenced by uncontrolled factors, such as weather, holidays... Maximum quantities, like peak values, are usually underestimated.

The owner of the line, which rents the facilities, controls the *BusType*. In that sense, the management wants to know if they should force the owners to use different types of buses. The *PeakInterval* has been negotiated with the owners of the lines. A new negotiation might be initiated.

The drivers control the opening of the gate, with the time being agreed with the labour unions. The management wants to know however how they can negotiate with the drivers in order to improve the behaviour of the system.

The two possible components that could be subject to negotiation are the bus type and the gate opening time, since the demand is clearly beyond the management's capabilities.

# 6.2 One-At-a-Time Analysis

Figure 15 shows that the increase in *Gate Opening* is statistically more important that the increase in capacity of the buses. However, a one-at-a-time study of the queue lengths for each of the two variables might be undertaken.

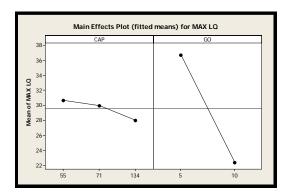


Figure 15. Comparison of Means

Figure 16 shows the distribution of all the values using a boxplot at both levels of *GateOpening*.

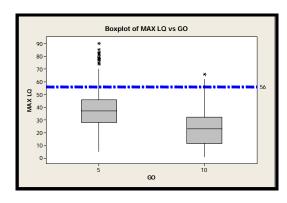


Figure 16. Boxplot for the levels of *GateOpening* 

The probability of the maximum queue length being above 56 is almost non-existent if the gates are opened 10 minutes before the departure of the bus.

Figure 17 shows that the difference between the two levels is significant (p-value = 0) with an estimated difference between 12.13 and 16.70 at 95% confidence.

Two-sample T for MAX LQ								
GO	N	Mean	StDev	SE Mean				
5	345	36.7	17.2	0.93				
10	345	22.3	13.1	0.71				
Difference = mu (5) - mu (10)								
Estimate for	difference:	14.4174						
95% CI for difference:		(12.1311, 16.7037)						
T-Test of difference = 0 (vs not =):								
T-Value = 12	38	P-Value = 0.0	DF = 642					

Figure 17. Test of Hypothesis for Levels of GateOpening

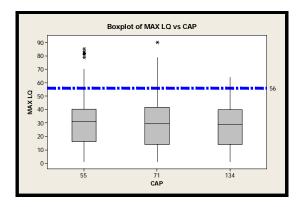


Figure 18. Boxplot for the levels of BusCapacity

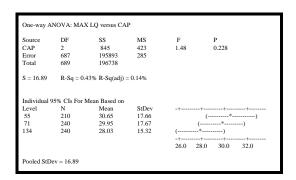


Figure 19: Test of hypothesis for levels of BusCapacity

Figure 20 shows that there is also no significant interaction between the two variables (parallel lines).

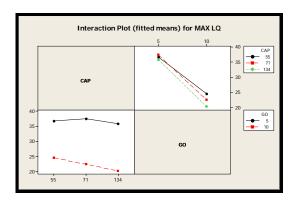


Figure 20: Interaction

# **6.3 Negotiation Tool**

The objective of this research is to develop an analytical tool that graphically shows the relationship between IO\_Ratio and Maximum Length of Queue (MAX\_LQ). The resulting graph (figure 21) is the result of the cleaning process of figure 13. It has two possible uses. The first one is to determine the maximum IO\_Ratio that should be permitted if a maximum queue length is set. For example, for a maximum of 56, an IO\_Ratio should be kept below 0.6.

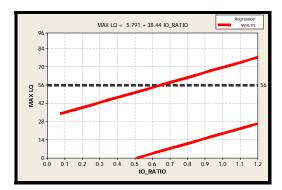
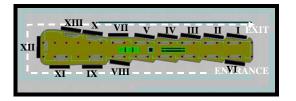


Figure 21: Final negotiation tool

The second use is to provide an estimation for a maximum queue length if an IO\_Ratio is provided. For example, an IO\_Ratio of 0.9 yields a queue of almost 70 passengers.

# 7. ALLOCATION OF LINES TO GATES

The last objective of this research is to develop a proper assignment of line buses to gates. Figure 22 includes the final assignment.



Line XII is located in the wider area of the platform. The rest are located in three areas, one for each of the three different line owners.

#### 8. CONCLUSIONS AND FUTURE WORK

The static design has been dynamically validated with the use of a simulation model that has been executed to calculate the space required in front of the gates.

Experimentation with the model has allowed for understanding not only the actual system but future situations with higher peaks in demand.

Since the maximum number of waiting passengers is set to 54 (four waiting lines of 14), currently, only line XII (queue of de 156 users) presents problems if 15-meter buses are used for these lines. With mixed buses, the maximum queue length will be 33 passengers, except for line XII with 94.

For the future, it has been possible to identify several variables that affect the behaviour of the system. However, their values might only be modified after hard negotiations with the labour unions and with the owners of the bus lines.

#### **REFERENCES:**

Chalmet, L.G.; Francis, R.L. and P.B. Saunders. 1982. "Network Models for Building Evacuation."

Management Science, Vol. 28, No. 1, pp. 86-105

Kiran, A.S. and S. Org. 2000. "Simulation Modeling and Analysis of a New International Terminal," Proceedings of the 2000 Winter Simulation Conference.

Li, J.P. 2000. "Train Station Passenger Flow Study," Proceedings of the 2000 Winter Simulation Conference.

Otamendi, J. 2005. "A Simulation Model for the Design of the Evacuation Policies in the Case of Fire in Public Integrated Stations", Proceedings of the 2005 SIS, Brussels.

Shen, T.S. 2005. "ESM: a building evacuation simulation model." *Building and Environment 40*, No. 5, 671-680

Taha, H.A. 1987. Operations Research. Macmillan, New York.

# **AUTHOR BIOGRAPHIES:**



JAVIER OTAMENDI received the B.S. and M.S. degrees in Industrial Engineering at Oklahoma State University, where he developed his interests in Simulation and Total Quality Management. Back in his

home country of Spain, he received a B.S. in Business Administration and a Ph.D. in Industrial Engineering. He is currently a simulation and statistics consultant and university professor at the Rey Juan Carlos University in Madrid. His e-mail address is: jotamendi\_30@yahoo.com



JOSE M. PASTOR received the Electronics and Automatic Control Engineering degree from the Polytechnic University of Madrid (UPM) in 1991 and a Ph.D. in Robotics and Artificial Intelligence in 1997. In 2005 Dr. Pastor moves to

the Informatics Systems Department of the Castilla La Mancha University where he is currently Associate Professor. His research interests include intelligent manufacturing systems, and systems simulation and optimization. His e-mail address is: josemanuel.pastor@uclm.es