OPTIMISATION OF FOUR QUEUE NETWORK VIA NESTED PARTITIONS METHOD

EBERT BREA & RUSSELL C. H. CHENG
University of Southampton, Department of Mathematics
Southampton SO17 1BJ, UK
E.Brea@maths.soton.ac.uk
R.C.H.Cheng@maths.soton.ac.uk

Abstract: A practical example of searching for the optimal operating conditions of a four queue network via Nested Partitions (NP) method is shown in this paper. NP Method systematically partitions the feasible region and concentrates the search for an optimal solution in regions that are most promising. The example is shown step by step with the aim of giving a didactic explanation of the method. Numerical results demonstrate the efficacy of the method, although actually, there has only been limited evaluation of the effectiveness of this method.

Keywords: Nested partitions method, optimisation, queue network.

1 INTRODUCTION

Searching for the optimal operating conditions of a Discrete Event Dynamic System (DEDS) represented by a simulation model is often difficult, because a large amount of experimentation is required to compute the object function even in the situation where there is just a finite set of points of the feasible region. Moreover, estimation of the object function must be made through a number of simulation replication in order to obtain some statistical measure of the variability of the performance measure of the studied simulation model.

Decision variables can be divided into two categories: continuous decision variables and discrete decision variables. According to these two categories and the information that could be gathered from the simulation model, there has developed two groups of methods for estimating the optimal solution in a simulation model. Firstly these are those which extract information from a single simulation run, as for example: perturbation analysis methods, score function, and frequency domain experimentation [Fu, 1994]. Secondly there exists another group of methods which require performing a number of simulation runs.

We call these "multiple simulation run methods". Of these the most recently developed is the Nested Partitions (NP) Method [Shi and Chen, 2000; Shi and Ólafsson, 2000a; Shi and Ólafsson, 2000b]. To date there has only been limited evaluation of the effectiveness of this method. In this paper we shall show the application of NP method through an example of a four queue network with the aim of giving a didactic explanation of the NP method, showing its potential for more complex situations.

The remainder of the paper is organized as follows. In Section 2 we describe the NP method. In Section 3 we propose a small example of a four queue network with the aim of identifying the optimal solution. The use of the NP method for our problem will be shown through three iterations of the algorithm, in Section 4. Finally, in Section 5 we present our conclusions and recommendations.

2 THE NESTED PARTITIONS METHOD

The NP method proposed by Shi and Ólafsson [Shi and Ólafsson, 2000b] was …rstly developed for solving global optimisation problems in deterministic models. The method systematically partitions the feasible region and concentrates the search for an optimal solution in regions that are most promising. The estimation of which region is most promising is computed through random sampling of the considered feasible region. Is it important to point out that the NP method o¤ers convergence with probability one in …nite time, without having to verify all the feasible regions. We consider the problem of getting the optimal solution in the case that we have a …nite feasible region, in this case, we can get the optimal solution via comparison of all points in order to choose the best feasible solution. This problem is mathematically simple to solve in principal. However, when there are a huge number of alternatives, the problem can become large and tedious. The NP method assumes that in each iteration of the algorithm there is a subset σ of the feasible region Θ considered as the most promising. This subset σ must be partitioned into M+1 subregions if σ is not a singleton subset followed by checking each of these M+1 subregions using some random sampling scheme. The stopping rule of the algorithm is given in [Shi and Ólafsson, 2000a].

3 FOUR QUEUE NETWORK PROBLEM

We have a queue network in which there are four queues. Each queue (Q_1, Q_2, Q_3 and Q_4) is associated with its respective server (S_1, S_2, S_3 and S_4). The set formed by each queue and server will be called a subsystem. Each subsystem, i=1, 2, 3 and 4, requires of a number of resources called R_1, R_2, R_3 and R_4, respectively (See Figure 1).

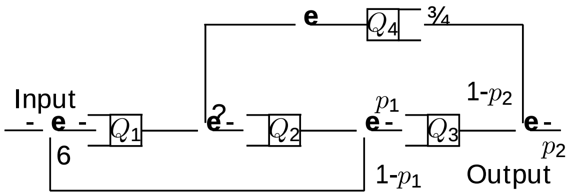


Figure 1. Four Queue Network

Let (1-p_1) and (1-p_2) be the probabilities of that a customer returns to Q_1 and Q_2, after service by S_2 and S_3, respectively. It is important to point out that returned customers from Q_3 are sent to Q_4 before that they can be sent to Q_2. Additionally, customers that are not returned from Q_2 are sent to Q_3 with probability p_1, and customers that are not returned from Q_3 leave with probability p_2. The e¢ciency of each server is dependent on the amount of resource R_i that is allocated to it. We shall assume that the service time has a mean inversely proportional to the amount resource allocated to it. We shall assume that the total resource available for allocation to servers is limited. Our object is to …nd the best con…guration or allocation of resources to the network of queues, to minimize the average time customers spend in the system ($\hat{T}$) under the condition that the sum of all the resources is equal to N. A mathematical formulation of the problem takes the following form:

Let $\hat{T}(\vec{R})$ be the average time customers spend in the system and Let $\vec{R}^T$ be the resource vector given by (R_1, R_2, R_3, R_4).

$$\vec{R}_{Opt} = arg\ \ min\ \ \hat{T}(\vec{R})$$

Subject to

$$\sum_{i=1} R_i = N$$

where

$$R_i \in \mathbb{N}^+ \ \forall\ i = 1, \ldots, 4 \qquad N \geq 4$$

Here $\mathbb{N}^+$ is the set of positive integers.

The inter-arrival times were assumed to be uniformly distributed with minimum 0 and maximum given by the parameter Max Inter-Arrival. A way of representing the e¤ect that has each resource (R_i, i = 1,. . . ,4) on the system is through the activity time associated to each server (S_i, i = 1,. . . ,4). In this sense, each server time was modelled by a uniform random variable with minimum given by a inversely proportional function depending on the quantity of assigned resource and maximum given by twice the last one function. The factor of proportion of these functions was called "time factor" (tf).

Probabilities p_1 and p_2 were 0.90 and 0.85 respectively, the total number of resources was 8, Max Inter-Arrival was equal to 20 time unit (tu) and tf was equal to 5.

4 RESULTS OF THREE ITERATIONS

With the aim of giving a didactic explanation of NP method, it is shown step by step each phase of the algorithm.

Step 0, k=0, De…nition of the feasible region

Let σ(0) be the set of all possible combinations of R_i allocated to the four queues subject to the total being 8.

Mathematically we have:

σ(0)=£=f(R_1, R_2, R_3, R_4)/ R_1=1,. . .,5 ^ R_2=1,. . .,(6-R_1) ^ R_3=1,. . .,(7-R_1-R_2) ^ R_4=8-R_1-R_2-R_3g

Each design experiment point was sequentially denoted by E_i, in our case, i = 1,. . .,35. Each $\vec{R}_i$ is explicitly de…ned by the following algorithm:

```
Start N = 8; i = 1
    Do: (R1=1 to (N-3))
        Do: (R2=1 to (N-2-R1))
            Do: (R3=1 to (N-1-R1-R2)
                R4=N-R1-R2-R3
                Ri=(R1, R2, R3, R4)
                i = i + 1
            End Do
        End Do
    End Do, End
```

Table 1 shows the set of design experiment points which was denoted by £. Each row of the table shows its design experiment point E_i and its respective components of $\vec{R}_i$

Table 1. Set of design experiment points £

Exp	R_1	R_2	R_3	R_4
E_1	1	1	1	5
E_2	1	1	2	4
E_3	1	1	3	3
E_4	1	1	4	2
E_5	1	1	5	1
E_6	1	2	1	4
E_7	1	2	2	3
E_8	1	2	3	2
E_9	1	2	4	1
E_{10}	1	3	1	3
E_{11}	1	3	2	2
E_{12}	1	3	3	1
E_{13}	1	4	1	2
E_{14}	1	4	2	1
E_{15}	1	5	1	1
E_{16}	2	1	1	4
E_{17}	2	1	2	3
E_{18}	2	1	3	2
E_{19}	2	1	4	1
E_{20}	2	2	1	3
E_{21}	2	2	2	2
E_{22}	2	2	3	1
E_{23}	2	3	1	2
E_{24}	2	3	2	1
E_{25}	2	4	1	1
E_{26}	3	1	1	3
E_{27}	3	1	2	2
E_{28}	3	1	3	1
E_{29}	3	2	1	2
E_{30}	3	2	2	1
E_{31}	3	3	1	1
E_{32}	4	1	1	2
E_{33}	4	1	2	1
E_{34}	4	2	1	1
E_{35}	5	1	1	1

Step 1, k=0, Partitioning $\sigma(0)$

Here we partition $\sigma(0)$ into ...ve subsets as follows:

$\sigma_1(0)$= f(1, R_2, R_3, R_4)/ R_2=1,...,(6-R_1) ^ R_3=1,...,(7-R_1-R_2) ^ R_4=8-R_1-R_2-R_3g
$\sigma_1(0)$= f(2, R_2, R_3, R_4)/ R_2=1,...,(6-R_1) ^ R_3=1,...,(7-R_1-R_2) ^ R_4=8-R_1-R_2-R_3g
$\sigma_3(0)$= f(3, R_2, R_3, R_4)/ R_2=1,...,(6-R_1) ^ R_3=1,..., (7-R_1-R_2) ^ R_4=8-R_1-R_2-R_3g
$\sigma_4(0)$= f(4, R_2, R_3, R_4)/ R_2=1,...,(6-R_1) ^ R_3=1,..., (7-R_1-R_2) ^ R_4=8-R_1-R_2-R_3g
$\sigma_5(0)$= f(5, 1, 1, 1)g

Step 2, k=0, Sampling $\sigma_i(0)$, i=1,...,5

We now sample randomly and uniformly from each subset. The randomly sampled subsets are denoted by:

D$\sigma_1(0)$= f(1, R_2, R_3, R_4)/ E_5, E_{15}, E_{13}, E_8, E_3g
D$\sigma_2(0)$= f(2, R_2, R_3, R_4)/ E_{24}, E_{21}, E_{17}g
D$\sigma_3(0)$= f(3, R_2, R_3, R_4)/ E_{31}, E_{27}g
D$\sigma_4(0)$= f(4, R_2, R_3,R_4)/ E_{32}g
D$\sigma_5(0)$= f(5, 1, 1, 1)/ E_{35}g

Step 3, k=0, Ranking and selection of the best design D$\sigma_i(0)$, i=1,...,5

We now carry out simulation runs at each design experiment point of each subset. Ranking was done using 50 simulation replication per design experiment point, and the simulation time per replication was of 100000 tu.

The results summarized in Table 2 to Table 14 give us the average of the average time customers spend in the system ($\hat{T}$) of the simulation replication which is denoted by $\hat{T}_{E_i}$. The minimum value of $\hat{T}$ is given by $\hat{T}_{min}$ and the maximum value if denoted by $\hat{T}_{Max}$. S_{E_i} means the sample standard deviation of the set of sample $\hat{T}$ and S_{error} is the standard error of $\hat{T}$ which is computed as the sample standard deviation divided by the square root of the number of simulation replication.

Table 2. Statistical summary of D$\sigma_1(0)$

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_5	174.01	79.93	11.3	85.17	413.68
E_{15}	45.16	2.29	0.32	38.56	50.29
E_{13}	45.16	2.13	0.3	39.33	49.87
E_8	28.95	1.02	0.11	26.59	31.06
E_3	174.61	80.42	11.37	86.38	418.91

Table 3. Statistical summary of D$\sigma_2(0)$

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_{24}	14.74	0.01	0.01	14.53	15
E_{21}	16.03	0.01	0.01	15.85	16.22
E_{17}	171.07	83.3	11.78	87.25	435.44

Table 4. Statistical summary of D$\sigma_3(0)$

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_{31}	33.53	1.93	0.27	29.95	38.89
E_{27}	169.8	84.02	11.88	86.83	441.76

Table 5. Statistical summary of D$\sigma_4(0)$

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_{32}	169.62	72.68	10.28	92.73	417.64

Table 6. Statistical summary of D$\sigma_5(0)$

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_{35}	168.99	81.89	11.58	93.73	437.58

From a statistical comparative analysis of all $\hat{T}_{E_i}$ de...ned, we concluded that the best subset was $\sigma_2(0)$, because it had the smallest $\hat{T}$. The following step is k = k + 1 and go to step 1, with $\sigma(1)$ = $\sigma_2(0)$, due to the fact that $\sigma_2(0)$ is not a singleton set. We additionally require to check some stopping rule, for instance see [Shi and Ólafsson, 2000a].

Step 1, k=1, Partitioning $\sigma(1)$ = $\sigma_2(0)$

We now partition the best subset identi...ed as $\sigma_2(0)$ in the previous stage.

$\sigma_1(1)$= f(2, 1, R_3, R_4)/ R_3=1,...,(7-R_1-R_2) ^ R_4=8-R_1-R_2-R_3g
$\sigma_2(1)$= f(2, 2, R_3, R_4)/ R_3=1,...,(7-R_1-R_2) ^ R_4=8-R_1-R_2-R_3g

σ_3(1)=f(2, 3, R_3, R_4)/ R_3=1,...,(7-R_1-R_2) ^
R_4=8-R_1-R_2-R_3g
σ_4(1)=f(2, 4, 1, 1)g
σ_5(1)= £ n σ(1)=f(R_1, R_2, R_3, R_4)/ $R_1 \neq 2$ ^
R_2=1,...,(6-R_1) ^ R_3=1,...,(7-R_1-R_2) ^
R_4=8-R_1-R_2-R_3g

Step 2, k=1, Sampling σ_i(1), i=1,...,5
Dσ_1(1)=f(2, 1, R_3, R_4)/ E_{19}, E_{17}g
Dσ_2(1)=f(2, 2, R_3, R_4)/ E_{20}g
Dσ_3(1)=f(2, 3, R_3, R_4)/ E_{24}g
Dσ_4(1)=f(2, 4, R_3, R_4)/ E_{25}g
Dσ_5(1)=ff(R_1, R_2, R_3, R_4)/ E_{11},E_6, E_8, E_{12}, E_2, E_{28}, E_{32}, E_{26}g

Step 3, k=1, Ranking and selection of the best design Dσ_i(1), i=1,...,5
In this step, we also did 50 simulation replication per design experiment point with a simulation time per replication of 100000 tu. It is important point out that the results of before iteration were taken into account.

Table 7. Statistical summary of Dσ_1(1)

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_{19}	169.39	83.27	11.77	86.99	433.16
E_{17}	171.07	83.3	11.78	87.25	435.44

Table 8. Statistical summary of Dσ_2(1)

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_{20}	175.79	79.92	11.3	87.06	416.56

Table 9. Statistical summary of Dσ_3(1)

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_{24}	14.74	0.1	0.01	14.53	15

Table 10. Statistical summary of Dσ_4(1)

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_{25}	34.39	1.84	0.26	30.19	39.4

Table 11. Statistical summary of Dσ_5(1)

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_{11}	28.73	0.99	0.14	26.13	30.85
E_6	47.45	2.16	0.31	41.02	52.63
E_8	28.95	1.02	0.14	26.59	31.06
E_{12}	27.76	0.85	0.12	25.14	29.1
E_2	175.79	79.92	11.3	87.06	416.56
E_{28}	169.94	83.95	11.87	86.06	441.07
E_{32}	169.62	72.68	10.28	92.73	417.64
E_{26}	175.91	74.24	10.50	78.49	369.56

In this iteration σ_3(1)was the most promising region. Because σ_3(1)is not a singleton set, we have to do k = k + 1 and go to step 1, with σ(2) = σ_3(1).

Step 1, k=2, Partitioning σ(2) = σ_3(1)
σ_1(2)=f(2, 3, 1, 2)g
σ_2(2)=f(2, 3, 2, 1)g
σ_3(2)= £ n σ(2)=f(R_1, R_2, R_3, R_4)/ $R_1 \neq 2$ ^
$R_2 \neq 3$ ^ R_3=1,...,(7-R_1-R_2) ^
R_4=8-R_1-R_2-R_3g

Step 2, k=2, Sampling σ_i(2), i=1,...,3
Dσ_1(2)=f(2, 3, 1, 2)/ E_{23}g
Dσ_2(2)=f(2, 3, 2, 1)/ E_{24}g
Dσ_3(2)=ff(R_1, R_2, R_3, R_4)/ E_7,E_{15}, E_5, E_{22}, E_9, E_{14}, E_{19}, E_{30}, E_{25}, E_{32}g

Step 3, k=2, Ranking and selection of the best design Dσ_i(1), i=1,...,3
Of the same manner that we did the before iteration, we did 50 simulation replication per design experiment point with a simulation time per replication of 100000 tu. Results of before iterations were considered in this step.

Table 12. Statistical summary of Dσ_1(2)

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_{23}	34.67	1.9	0.27	30.14	39.89

Table 13. Statistical summary of Dσ_2(2)

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_{24}	14.74	0.1	0.01	14.53	15

Table 14. Statistical summary of Dσ_3(2)

Exp	$\hat{T}_{E_i}$	S_{E_i}	S_{error}	$\hat{T}_{min}$	$\hat{T}_{Max}$
E_7	30.29	0.92	0.13	28.19	32.33
E_{15}	45.16	2.29	0.32	38.56	50.29
E_5	174.01	79.93	11.30	85.17	413.68
E_{22}	15.05	0.11	0.02	14.82	15.34
E_9	28.83	0.87	0.12	26.77	31.20
E_{14}	28.64	0.96	0.14	26.34	30.61
E_{19}	169.39	83.28	11.77	86.99	433.16
E_{30}	15.20	0.11	0.02	14.98	15.48
E_{25}	34.39	1.84	0.26	30.19	39.4
E_{32}	169.62	72.68	10.28	92.73	417.64

From a statistical approach, we can say the best design was σ_2(2) because it has the smallest $\hat{T}$. Note that in this iteration σ_2(2) is a singleton set. Therefore, we must review if σ_2(2) is a optimal solution according to some stopping rule. An example of such a rule is given in [Shi and Ólafsson, 2000a].

We assume that σ_2(2) is a optimal solution which corresponds with $\vec{R}^T_{Opt}$ = (2, 3, 2, 1). In this example we needed to estimate the performance measure of 28 design experiment points out of a total of 35.

Figure 2 depicts the range of $\hat{T}$ per design experiment point E_i for 50 simulation replication with a simulation time per replication of 100000 tu. The vertical axis displays the design experiment points and the horizontal axis the values of $\hat{T}$ from the minimal $\hat{T}$ to the maximum $\hat{T}$.

As can be seen from Figure 2, there exists a great di¤erence between the group of design experiment points given by (R_1, 1, R_3, R_4) and the group of design experiment points given by (R_1, R_2, R_3, R_4) for R_2 = 2, 3, 4. If we analyse the queue network we can see the following: Firstly the second subsystem is the most in‡uential, because all returned customers are serviced by S_2. Secondly the …rst subsystem is the second most in‡uential, because it receives a portion of all returned customers, and …nally, the subsystem 1 is less in‡uential than the subsystem 3, however, is more important than subsystem 4. Hence the best allocation of resources must maximize R_2 taking into consideration the described aspects.

Another explanation we could give about the behaviour of the system when R_2 = 1 and R_2 = 2, 3, 4, is based on the transient period of the system. A question which needs proo…ng is if there exists some relationship between transient period and optimal operating of the DEDS. Because we can see that when the system operates at R_2 = 1 the transient period is bigger than when it operates at R_2 = 2, 3, 4. Hence, the performance measure has an important noise when R_2 = 1. This last reason needs further study in order to …nd a relationship between the transient period and optimal operation of the DEDS.

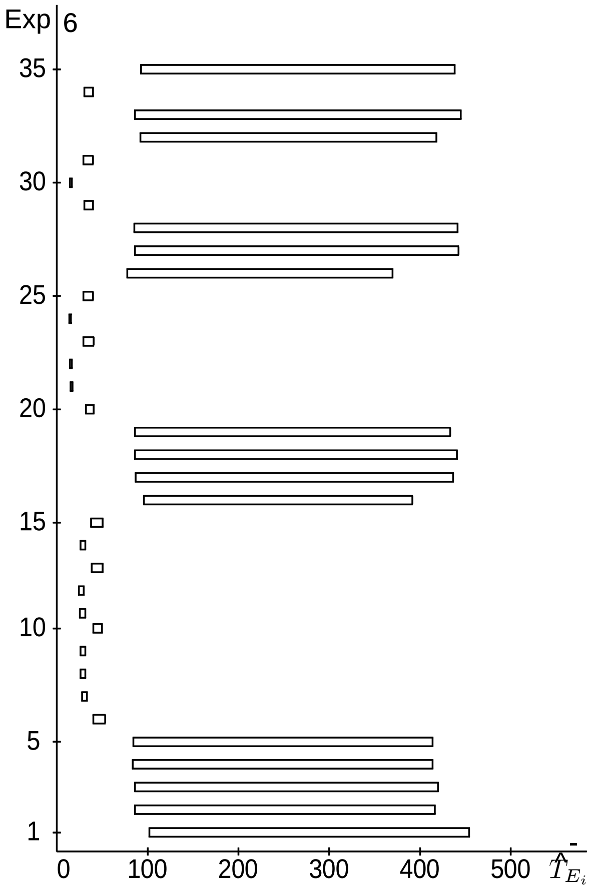


Figure 2. Minimal and maximum of $\hat{T}_{E_i}$

5 CONCLUSION

This paper shows an approach for the identi…cation of the best solution in a simulation model, without having to evaluate all feasible points. There was only limited bene…t in using the NP technique in this example, because of the small feasible region. The NP method might be expected to problem more e¤ectively when the feasible region is large but we are only interested in a relatively course identi…cation of the optimal operating point. We hope to investigate this further. Additionally, a studying of the relationship between transient period and optimal operating of DEDS is suggested in order to improve the estimation of the performance measure of DEDS. This kind of example can be implemented in simulation course with the aim of showing the use of simulation models in optimisation.

REFERENCES

Fu, M. C. 1994. "Optimization via simulation: A review". Annals of Operations Research: Simulation and Modeling. Vol. 53. Pp199-247.

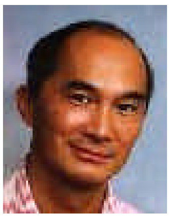
Shi, L., C-H Chen. 2000. "A New Algorithm for Stochastic Discrete Resource Allocation Optimization". Discrete Event Dynamic Systems: Theory and Applications. Vol. 10 (3) Pp271-294.

Shi, L., S. Ólafsson. 2000a. "Stopping Rules for the Stochastic Nested Partitions Method". Methodology and Computing in Applied Probability. Vol. 2 (1) Pp37-58.

Shi, L., S. Ólafsson. 2000b. "Nested Partitions Method for Global Optimization". Operations Research. Vol. 48 (3) Pp390-407.

Author Biographies

Ebert Brea is a research student at the Department of Mathematics, the University of Southampton. He received an M.Sc degree in Operational Research and the degree of Electrical Engineer from Central University of Venezuela. He has worked as a consultant at Corpoven, S.A., a subsidiary company of Petróleos de Venezuela S.A. (PDVSA) for developing of simulation models of gas network and forecast crude oil berth operations. He has received awards from Central University of Venezuela in 1999 (PEI-1999) and in 1997 (PEI-1997) for his research activities, and from the National Commission for the Development of the Education in Venezuela, in 1998 (FRA-CONADE-1998). He has been coauthor of seven articles in journals and he has participated in twenty one presentation, most of them in the area of simulation model. He is a member of INFORMS and the College of Simulation of INFORMS.



Russell C. H. Cheng is Professor of Operational Research at the University of Southampton. He has an M.A. and the Diploma in Mathematical Statistics from Cambridge University, England. He obtained his Ph.D. from Bath University. He is former Chairman of the U.K. Simulation Society, a Fellow of the Royal Statistical Society, Member of the Operational Research Society. His research interests include: variance reduction methods and parametric estimation methods. He is Joint Editor of the IMA Journal on Mathematic Applied to Business and Industry.
Home page:
http://www.maths.soton.ac.uk/sta¤/cheng