Proceedings of the 2007 Winter Simulation Conference
S. G. Henderson, B. Biller, M.-H. Hsieh, J. Shortle, J. D. Tew, and R. R. Barton, eds.

A TOURNAMENT FRAMEWORK FOR THE RANKING AND SELECTION PROBLEM

Enver Yücesan

Technology and Operations Management Area
INSEAD
Boulevard de Constance
77305 Fontainebleau Cedex, FRANCE

ABSTRACT

A tournament can be broadly defined as a procedure that ranks agents, where they exhibit their performance in a noisy environment. By observing the agents' performance, the organizer computes their ranking and rewards them according to the revealed ranking. The organizer's challenge is therefore to determine the optimal tournament format that identifies the best agent in the most effective fashion. Tournaments thus provide a natural framework for ranking and selection (R&S) via simulation, which represents a set of approaches developed to complement the modeling flexibility of simulation with the efficiency of statistical techniques for effective decision making. In this paper, following the introduction of a general framework to represent various tournament formats and to assess their predictive power, we will report preliminary experimental results on the effectiveness of tournaments in identifying the best simulated system with the desired probability of correct selection in the presence of costs.

1 INTRODUCTION

Ranking and selection (R&S) techniques are statistical methods developed to select the best system, or a subset of systems from among a set of alternative system designs. R&S via simulation is particularly appealing as it combines modeling flexibility of simulation with the efficiency of statistical techniques for effective decision making. Furthermore, it is relatively straightforward to satisfy the underlying technical assumptions of these techniques (such as normally distributed data and independence) in simulation experiments, which also allow for multi-stage sampling as required by some R&S methods.

Goldsman and Nelson (1998) classify the comparison of alternative system designs into five categories:

- *Screening problems*, where the objective is to compare a large number of competing designs and eliminate the inferior performers.
- *Selecting the best*, where the objective is to identify the system with the largest or smallest performance measure.
- *Comparisons with a standard*, where the objective is to find the best system, provided that its performance exceeds a known, fixed threshold.
- *Comparisons with a default*, where the objective is to compare alternative systems to the current system in place.
- *Estimating functional relationships*, where the objective is to represent the difference between competing designs in terms of the parameters of a linear model.

A tournament can be broadly defined as a procedure that ranks agents. During a tournament, agents exhibit their performance, typically in a noisy environment. By observing the agents' performance, the organizer calculates their ranking and rewards them according to the revealed ranking. The choice of a tournament format may indeed affect the outcome. The organizer's challenge is therefore to choose the optimal tournament format that identifies better agents in the most effective fashion. Tournaments also provide a natural framework for ranking and selection, which refers to the area in simulation research that complements the modelling flexibility of simulation with the efficiency of statistical techniques for effective decision making. Within this area, many traditional statistical techniques have been adapted to boost efficiency by exploiting special properties of simulations through, for example, variance reduction, importance sampling, and parallel experimentation. In this paper, we report preliminary results of our investigation into the possibility of using tournaments for ranking and selection among simulated systems. Following the introduction of a general framework to represent various tournament formats and to assess their predictive power, we will report preliminary experimental results on the effectiveness of tournaments in identifying the best simulated system with the desired probability of correct selection in the presence of costs.

1-4244-1306-0/07/$25.00 ©2007 IEEE

The remainder of the paper is organized as follows. Section 2 provides a general framework to define and analyze various tournament formats. This framework is then used to establish the natural fit for R&S via simulation, which is discussed in Section 3. Section 4 reports on some preliminary numerical experiments under selected tournament formats. Section 5 concludes the paper with a discussion on future directions.

2 TOURNAMENTS

A tournament can be defined as a procedure that ranks agents (Ryvkin 2005). Tournaments are widely used in sports, recruiting, promotions in organizations, and research competition. During a tournament, agents (players) produce an output (performance), typically in a noisy environment. By comparing the agents' performance, the organizer (the principal) computes the ranking of the agents and rewards them according to the revealed ranking. This process is executed within a tournament format, which is a scheme or a set of rules by which agents are matched and their ranking is calculated.

For any number of players $N \geq 2$, there exist infinitely many tournament formats (Appleton 1995). A complex tournament format consists of smaller sub-tournaments or building blocks, which can be depicted as a black box $B(N,M)$ with N contenders at the input and M "winners" ($M < N$) at the output. For example, the black box could be a *contest*, where all M players perform together once and the ranking is generated directly by their relative performance (Szymanski 2003). Alternatively, the black box could represent a *round robin* format, which is a complete pairwise matching format where the winner is determined by point counting (Harary and Moser 1966). Finally, the black box may represent a *knock out* (or *binary elimination*) format, where the number of participants decreases by half as the winners of every match advance to the next stage while the losers are eliminated (Rosen 1986). Such building blocks can be joined in parallel or concatenated in series to constitute the entire tournament format.

Following Ryvkin (2005), a noisy tournament with heterogeneous agents can then be characterized as a principal-agent game with the following components: (i) players; (ii) costs; (iii) noise; and (iv) the organizer's decision problem.

There are N players characterized by abilities $\mathbf{x} = (x_1, \ldots, x_N) \in \Re^N$. The distribution of players' abilities, $f(\cdot)$, is assumed to be known.

We consider two types of costs: time and measurement. Since it is generally more costly to run a tournament that lasts longer, it is natural to assume that time costs are proportional with the number of stages, S. In particular, we can take

$$E_t = c_t S,$$

where c_t is the unit time cost. Measurement costs reflect the effort necessary in determining M best players out of N contenders at every stage. The best players are determined by comparing their performance levels. E_m therefore depends on the number of comparisons that must be carried out under a particular tournament format. For instance, in a contest or in a binary elimination tournament with N players, (N-1) comparisons would be necessary to identify the best player. On the other hand, in a round robin tournament with N players, where the winner of each match gets one point and the player with the highest total score is the winner, $N(N-1)/2$ comparisons would be required.

It is further assumed that there is some noise in observing tournament outcomes. One way to represent this uncertainty is to add a noise term to each player's ability:

$$y_{is} = x_i + \varepsilon_{is},\ i=1, \ldots, N;\ s=1, \ldots, S,$$

where y_{is} depicts the observed performance of player i at stage s with noise terms that are IID across players and across stages.

The organizer's challenge is to maximize the probability that the best alternative is indeed chose by selecting an appropriate tournament format. Ryvkin (2005) refers to this probability as the predictive power, ρ, of a tournament. Assuming that the organizer will experience a gain of V by identifying the best agent, her expected payoff is given by

$$\pi = V\rho - E_t - E_m. \quad (1)$$

3 RANKING AND SELECTION IN SIMULATION

Simulation plays a central role in designing and efficiently managing complex man-made systems such as communication networks, traffic systems, and manufacturing facilities since closed-form analytical solutions are scarce for such problems. Ranking-and selection procedures are statistical methods specifically developed to select the best design or a subset that contains the best design from a set of N competing design alternatives (Goldsman and Nelson 1994). Unfortunately, simulation can be both expensive and time consuming. Suppose we conduct k simulation replications for each of the N designs. Therefore, we need kN simulation replications. Simulation results become more precise as k increases. If the precision requirement is high (k is not small), and if the total number of designs in a decision problem is large (N is large), then kN can be very large, which may easily render the total simulation cost prohibitively high and preclude the feasibility of using simulation for R&S problems. The effective reduction of computation costs in the course of obtaining a good decision is therefore crucial.

There exists a rich literature on R&S. Dudewicz and Dalal (1975) developed a two-stage procedure for selecting the best design or a design that is very close to the best system. In the first stage, all systems are simulated through k_0 replications. Based on the results of the first stage, the number of additional simulation replications to be conducted for each design in the second stage is determined in order to reach the desired probability of correct selection. Note that *correct selection* is the event that the best system is indeed identified as such. Rinott (1978) presents an alternative way to compute the required number of simulation replications in the second stage. Many researchers have extended this idea to more general R&S settings in conjunction with new developments. Chiu (1974), Gupta and Panchapakesan (1979), Matejcik and Nelson (1993, 1995), Bechhofer et al. (1995), and Hsu (1996) present methods based on the classical statistical model adopting a frequentist view. Berger (1980), Berger and Deely (1994), Bernardo and Smith (1994), Gupta and Berger (1988), Chick (1997), and Chick and Inoue (2001), on the other hand, use a Bayesian framework for constructing ranking-and-selection procedures.

Chen (1995) formulates the process of selecting the best design as an optimization problem. Gupta and Miescke (1996) give an effective approach for a class of problems where the simulation budget is allocated to minimize an expected linear loss assuming a common known variance. The main difficulty in dealing with this optimization problem is the lack of a closed-form expression for the probability of correctly selecting the best design. Chen et al. (1996) approximate this probability using Chernoff bounds. The steepest-descent method is then applied to solve the approximated optimization problem. Alternatively, Chen et al. (1997) obtain the gradient of the objective function through finite differencing, and then apply the steepest-descent method to solve the budget allocation problem.

4 TOURNAMENTS FOR R&S

4.1 Common Threads

The objective of the current paper is to investigate the adequacy of a tournament framework for R&S. This investigation is motivated by a number of common threads between the two frameworks, as summarized in Table 1. In an R&S problem, *N* competing designs with their performance can indeed be viewed as *N* contestants with their abilities in a tournament. Analogous to the costs in organizing and running a tournament, there are inherent costs in running computer experiments for R&S. While these costs might be much smaller in a digital environment, simulation costs (as a function of the number of runs conducted) and evaluation costs (as a function of the R&S scheme deployed) can be viewed as time and comparison costs, respectively. In tournaments, there are typically two sources of uncertainty. First, each player's ability is a random variable. Second, a player may strategically choose not to perform at the height of her abilities at all stages of the tournament. Therefore, in modeling a player's observed performance in a tournament, a noise terms is typically added to the random variable depicting the player's ability. Since such strategic gaming is not an issue in computer simulations (assuming that the simulation model is adequately validated and verified), an additive noise term is not necessary. However, computer simulations are typically stochastic in nature. One must therefore deal with sampling error, which may be significant in large and heterogeneous variance conditions.

In both settings, the organizer has an expected payoff. What Ryvkin (2005) labels as the 'predictive power' of a tournament format is referred to as the probability of correct selection, *P{CS}*, in the R&S literature. Alternatively, one can replace *P{CS}* with some loss function, as illustrated by Chick (1997), that depicts the loss resulting from erroneously identifying a lower-performance system as the best design.

Table 1: Common threads between tournaments and R&S

Parameter	Tournament	R&S
Players	*N* contestants	*N* competing designs
Costs	E_t and E_m	Simulation cost, Evaluation cost
Variability	Noise	Inherent stochasticity
Organizer's Decision Problem	Max π	Max π

4.2 Tournament Formats under Consideration

The simplest tournament format is a contest, where all *N* players perform together once and the ranking is generated directly by their relative performance with the identification of a "winner." While a natural benchmark, this is a notoriously inefficient setting in R&S that necessitates a large number of simulation replications to reach the desired correct selection probability. We will therefore not consider it. Instead, we will investigate round robin tournaments, knock out (binary elimination) tournaments, and a composite tournament that adopts a round robin format in the first round followed by a knock out format in the second round.

In our preliminary experiments, we have considered design alternatives (or competing players) with performance measures that are normally distributed with mean μ and standard deviation σ. We considered cases where competing alternatives have a constant and common variance (i.e., $N(i,1)$), cases where better alternatives (i.e., those with a larger mean performance) have lower variance (i.e., $N(i,N-i+1)$), and cases where better alternatives (i.e., those with a larger mean performance) have higher variance (i.e., $N(i,i)$).

We experimented with round robin tournaments with 16 and 20 competing alternatives as well as knock out tournaments with 4 and 16 competing alternatives. We also experimented with a composite tournament format where 20 competing alternatives first went through a round robin tournament in two groups; the winner of each group then met in a knock out tournament to determine the "champion."

4.3 Experimental Results

In our experiments, we have conducted 10,000 independent replications and recorded the proportion of the time where the system with the highest mean performance was actually identified as such. We use this ratio as the estimate of *P{CS}* for a particular tournament format.

For any variance structure, round robin tournaments provide perfect correct selection probability, i.e., they are consistently able to identify the alternative with the highest mean performance as the best design. This is because, in a round robin tournament, players are matched a relatively large number of times (with many opportunities to upset one another) without any elimination. As a result, more frequent upsets do not necessarily lead to more frequent overall losses for the best design, particularly when N is large. However, while it is the most robust tournament format, it is also the most expensive one, requiring $N(N-1)/2$ comparisons to identify the winner. Hence, round robin tournaments may not be the most efficient R&S scheme when the number of competing designs is very large.

Knock out (binary elimination) tournaments provide such efficiency as they require (N-1) comparisons to identify the best design. In experimenting with knock out tournaments, we considered two scenarios. In the first scenario, better designs (i.e., those with higher mean performance) have been assigned to higher seeds with the hope of making them harder to beat, as it is done in the NBA playoffs. In most applications, seeding may not be possible, as prior information on system performance (e.g., team rankings during the regular season) is not typically available. However, in comparisons with a default, where the objective is to compare alternative system designs with the system in place, one may wish to assign to the current system the highest seed to ensure that the suggested alternatives have a better performance than the existing system before any costly physical modification is undertaken. In the second scenario, competing designs have been assigned different seeds in a random fashion.

P{CS} for the first (seeded) scenario for 4 and 16 competing designs is shown in Figures 1 and 3, respectively. In both cases, the highest correct selection probability is achieved when competing designs have constant (and common) variances. In cases where better designs have higher or lower variances, *P{CS}* drops dramatically. In the former case, a higher volatility in the performance of the best design makes the likelihood of an upset by inferior systems higher. Similarly, in the latter case, a higher volatility in the performance of inferior systems increases the likelihood that they may upset a better system. While in sports such upsets have some 'entertainment value,' it makes the knock out tournament format in its simplest form inefficient for R&S problems.

P{CS} for the second (unseeded) scenario for 4 and 16 competing designs is shown in Figures 2 and 4, respectively. In both cases, the outcome of the tournament is similar to the seeded case; that is, the highest correct selection probability is achieved when competing designs have constant (and common) variances. In cases where better designs have higher or lower variances, *P{CS}* drops significantly. While cost effective, the knock out tournament format in its simplest form has thus low power for identifying the best system in R&S problems.

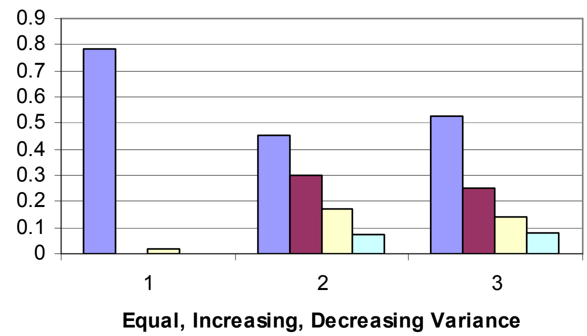

Figure 1: P{CS} in a seeded knock out tournament with four competing designs where better systems have (1) constant, (2) increasing, and (3) decreasing variances

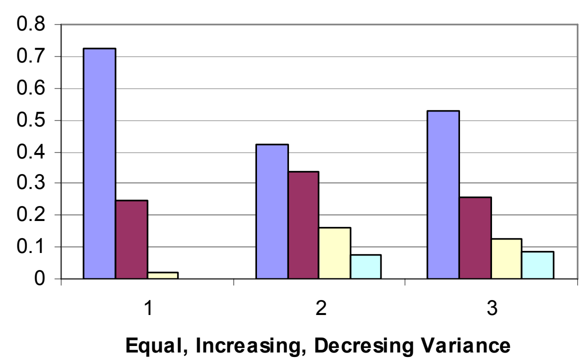

Figure 2: *P{CS}* in an unseeded knock out tournament with four competing designs where better systems have (1) constant, (2) increasing, and (3) decreasing variances

The knock out tournament format could, however, be 'good enough' for screening or subset selection problems, where the objective is to quickly eliminate inferior designs or to identify a subset of the initial set of competing alternatives that contains the best design.

In our final experiment, we have concatenated two round robin tournaments in parallel with a knock out tournament. Twenty competing designs were split into two groups, where the "winner" in each group is determined by a round robin tournament. The winners of the two groups then participated in a knock out tournament to identify the best design –or the "champion." Figure 5 depicts the correct selection probability in a setting where the first ten designs were included in the first group and the remaining ten designs were included in the second group. Recall that design 20 is the design with the highest mean performance.

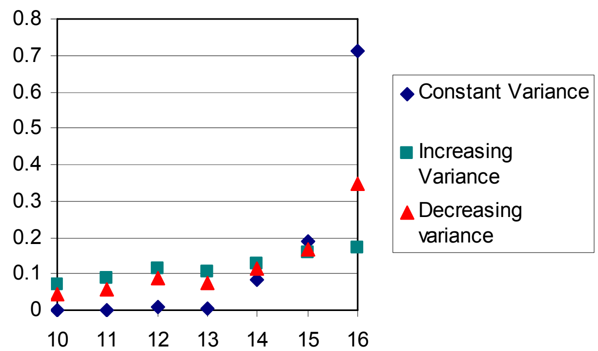

Figure 3: *P{CS}* in a seeded knock out tournament with sixteen competing designs where better systems have constant, increasing, and decreasing variances

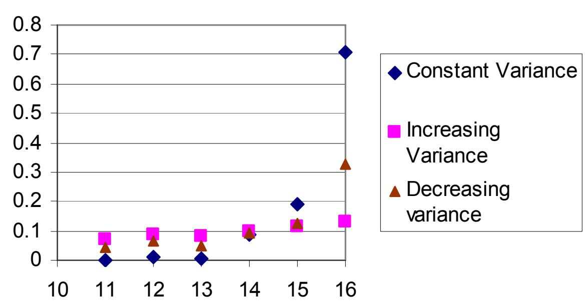

Figure 4: *P{CS}* in an unseeded knock out tournament with sixteen competing designs where better systems have constant, increasing, and decreasing variances

In this composite tournament, the settings with constant variance and with decreasing variance turn out to be quite robust in identifying the best system. In the setting with increasing design, however, there is increasing likelihood that the best system is upset by an inferior design. This, in turn, reduces *P{CS}* –although not as drastically as in the knock out tournament format.

5 CONCLUSIONS

In this paper, we report on some preliminary experimental results on the adequacy of using a tournament format for ranking and selection problems. While it is the most robust tournament format, the round robin tournament is also the most expensive one, requiring $N(N-1)/2$ comparisons to identify the best system. Hence, round robin tournaments may not be the most efficient R&S scheme when the number of competing designs is very large. The knock out tournament format, on the other hand, is parsimonious, as it requires $(N-1)$ comparisons to identify the best design. This reduction in cost, however, comes with a reduced correct selection probability.

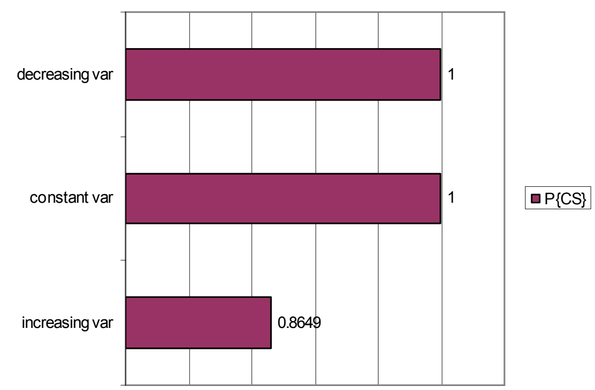

Figure 5: *P{CS}* in a composite tournament with twenty competing designs where better systems have constant, increasing, and decreasing variances

Such a trade-off may be mitigated through a composite tournament where simpler building blocks are concatenated to construct a more efficient tournament design. Our experiment combing a round robin first round with a knock out second round shows that this is indeed possible. However, given that there exist infinitely many tournament formats, a more structured approach is needed to quantify and mitigate this trade-off. To this end, our current work focuses on identifying efficient tournament formats by maximizing π in (1).

REFERENCES

Appleton, D. R. 1995. May the best man win? *The Statistician* 44:529-538.

Bechofer R. E., T. J. Santner, and D. M. Goldsman. 1995. *Design and analysis of experiments for statistical selection, screening, and multiple comparisons.* New York, NY: John Wiley & Sons, Inc.

Berger, J. O. 1980. *Statistical decision theory: foundations, concepts, and methods.* New York, NY: Springer Verlag.

Berger, J. O., and J. Deely. 1994. A Bayesian approach to ranking and selection of related means with alternative to AOV methodology. *Journal of American Statistics Association* 83:364-373.

Bernardo, J. M., and A. F .M. Smith. 1994. *Bayesian theory.* New York, NY: Wiley.

Chen, C. H. 1995. An effective approach to smartly allocate computing budget for discrete event simulation. In *Proceedings of the 34th IEEE Conference on Decision and Control*, 2598-2605. Piscataway, NJ: IEEE.

Chen, C. H. 1996. A lower bound for the correct subset-selection probability and its application to discrete event system simulations. *IEEE Transactions on Automatic Control* 41:1227-1231.

Chen, C. H., H. C. Chen, and L. Dai. 1996. A gradient approach for smartly allocating computing budget for discrete event simulation. In *Proceedings of the 1996 Winter Simulation Conference*, ed. J. M. Charnes, D. M. Morrice, D. T. Brunner, and J. J. Swain, 398-405. Piscataway, NJ: IEEE.

Chen, H. C., C. H. Chen, L. Dai, and E. Yücesan. 1997. New development of optimal computing budget allocation for discrete event simulation. In *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradottir, K. J. Healy, D. H. Withers, and B. L. Nelson, 334-341. Piscataway, NJ: IEEE.

Chick, S. E. 1997. Selecting the best system: a decision-theoretic approach. In *Proceedings of the 1997 Winter Simulation Conference*, ed. S. Andradottir, K. J. Healy, D. H. Withers, and B. L. Nelson, 326-333. Piscataway, NJ: IEEE.

Chick, S. E. and K. Inoue. 2001. New procedures to select the best simulated system using common random numbers. *Management Science* 47: 1133-1149.

Chiu, W. K., 1974. The ranking of means of normal populations for a generalized selection goal. *Biometrika* 61: 579-584.

Dudewicz, E. J. and S.R. Dalal. 1975. Allocation of observations in ranking and selection with unequal variances. *Sankhya* B37:28-78.

Goldsman, D., and B. L. Nelson. 1994. Ranking, selection, and multiple comparison in computer simulation. In *Proceedings of the 1994 Winter Simulation Conference*, ed. J. D. Tew, S. Manivannan, D. A. Sadowski, and A. F. Seila, 192-199. Piscataway, NJ: IEEE.

Goldsman, D. and B. L. Nelson. 1998. Comparing systems via simulation. In *Handbook of Simulation*, ed. J. Banks, 273-306. New York, NY: John Wiley and Sons.

Gupta, S. S. and J. O. Berger. 1988. *Statistical decision theory and related topics IV.* New York, NY: Springer Verlag.

Gupta, S. S. and K. J. Miescke. 1996. Bayesian look ahead one-stage sampling allocations for selection of the best population. *Journal of Statistical Planning and Inference* 54: 229-244.

Gupta, S. S. and S. Panchapakesan. 1979. *Multiple decision procedures: theory and methodology of selecting and ranking populations.* New York, NY: John Wiley.

Harary F. and L. Moser. 1966. The theory of round robin tournaments. *American Mathematical Monthly* 73: 231-246.

Hsu, J. C. 1996. *Multiple comparisons theory and methods*. London: Chapman & Hall.

Matejcik, F. J. and B. L. Nelson. 1993. Simultaneous ranking, selection and multiple comparisons for simulation. In *Proceeding of the 1993 Winter Simulation Conference*, ed. G.W. Evans, M. Mollaghasemi, E.C. Russell, and W.E. Biles, 386-391. Piscataway, NJ: IEEE.

Matejcik, F. J. and B. L. Nelson. 1995. Two-stage multiple comparisons with the best for computer simulation. *Operations Research* 43: 633-640.

Rinott, Y. 1978. On Two-stage selection procedures and related probability inequalities. *Communications in Statistics* A7: 799-811.

Rosen, S. 1986. Prizes and incentives in elimination tournaments. *The American Economic Review* 76: 701-715.

Ryvkin, D. 2005. The predictive power of composite noisy tournaments. Technical Report, Center for Economic Research and Graduate Education, Charles University and Academy of Sciences, Prague, Czech Republic.

Szymanski, S. 2003. The economic design of sporting contests: a review. *Journal of Economic Literature* 41: 1137-1187.

AUTHOR BIOGRAPHY

ENVER YÜCESAN is a professor of Operations Management at and coordinator of the Technology and Operations Management Area of INSEAD. His research interests are in the area of computer simulation experiments for stochastic systems and, more specifically, in the simulation methodology for ranking and selection as well as for optimization via simulation. He has served as a Proceedings co-editor for WSC'02. His web page can be accessed via <www.insead.edu/facultyresearch/faculty/profiles/eyucesan>