## A RANKING AND SELECTION PROJECT: EXPERIENCES FROM A UNIVERSITY-INDUSTRY COLLABORATION

David Goldsman

School of Industrial and Systems Engineering Georgia Institute of Technology Atlanta, GA 30332, U.S.A.

Tracy Opicka

Schneider National Inc. 3101 S. Packerland Dr. Green Bay, WI 54313, U.S.A. Barry L. Nelson

Department of Industrial Engineering & Management Sciences Northwestern University Evanston, IL 60208, U.S.A.

A. Alan B. Pritsker

Symix Systems, Inc. 9032 E. Cedar Waxwing Drive Sun Lakes, AZ 85248, U.S.A.

# ABSTRACT

We describe the experiences and results from a long-term collaboration between two universities and Pritsker Corporation on a grant funded by the National Science Foundation. The goal of the joint work is to make state-of-the-art research in the area of ranking and selection available to practicing engineers and management scientists.

## **1 INTRODUCTION**

This paper describes a common problem encountered in university research settings and in software companies: How does one develop software that converts new algorithmic developments into a product that is easily usable by practitioners? The key problem lies in determining how the customer and application characteristics interact with algorithm limitations. The customer and application characteristics also define the requirements for the user interfaces. In this paper the specific problem is the delivery of ranking and selection (R&S) software for use by simulation modelers and analysts. The goal of R&S is to determine the best simulated scenario, or a subset containing the best, with precise statistical guarantees about the validity of the choice.

The make-up of the project team—containing both industry and academic participants—led to a more-precise definition of the problem statement. This refined problem statement is to develop a software module in a specific environment, which allows for the use of several specific R&S algorithms for comparing multiple scenarios, where each algorithm and each scenario may have its own model and statistical input and output characteristics. The algorithms that were implemented were limited to those studied and created by Goldsman and Nelson (G&N). The simulation system selected was AweSim, and the simulation language was Visual SLAM, based on the participation of Pritsker and Opicka.

This paper describes the module, Scenario Selector, and examples of its use to meet the needs of practitioners as presented in the problem statement. In Section 2 we discuss how the team was formed. We give the algorithms included in Scenario Selector in Section 3. Sections 4 and 5 describe the Scenario Selector interface and our experiences with its use, including examples. We conclude by describing the interactions of the team while performing the project and potential future collaboration.

## 2 VISION

In October of 1995, G&N submitted a proposal to the National Science Foundation (NSF) to undertake research on five classes of problems encountered when trying to compare scenarios via computer simulation: screening a large number of scenarios to eliminate the clearly inferior ones; finding the best scenario; comparing alternative scenarios to a fixed standard of performance; comparing alternatives to a default; and comparing scenarios that are functionally related. In the proposal, they pointed out that research on basic methods was needed to address the special problems and opportunities that arise in the analysis of simulation experiments, problems and opportunities that do not arise in, for instance, biological or medical experiments (e.g., dealing with dependent data and the use of common random numbers in experiment

design). However, the key premise of the proposal, which led to the collaboration described in this paper, was the following: "... advanced analysis tools will not be applied routinely unless they are included in the software packages that simulation analysts use. Further, these tools will not appear in simulation software until (a) their use can be made transparent to the analyst; (b) they are robust enough to be generally applicable, or they are self-diagnosing when they are not applicable; and (c) they provide results that are easily interpreted for making decisions."

In other words, G&N wanted to develop analysis tools that would be easy to include in general-purpose simulation packages. Further, instead of producing special-purpose research software or unsupported public-domain programs to facilitate the technology transfer that NSF desired, G&N sought to insure that the research results would be included in commercial software packages. This approach was inspired by the many successful collaborations that exist between academic statisticians and the companies that produce general-purpose statistical analysis packages.

Consistent with the premise, G&N obtained the support of two leading software vendors. Both vendors develop commercial simulation software and provide consulting in manufacturing and service applications. Both agreed to support the proposed research by (i) donating commercial simulation software, (ii) participating in an annual review of the research progress, and (iii) providing guidance and feedback during the course of the research, specifically guidance related to software development and feedback regarding the needs of practitioners. Items (ii) and (iii) were accomplished via formal presentations and extended discussions at the vendors' offices during the summers of 1996-1998 and at each year's Winter Simulation Conference. In turn, G&N agreed to help with internal development of the commercial software, but not develop the software themselves, so that each company would have clear ownership of, and the expertise to support, the software it produced. This paper is the result of collaboration with one of the companies, Pritsker Corporation, currently a division of SYMIX Systems, Inc.

## **3 PROCEDURES**

#### 3.1 Introduction and Notation

Informally speaking, ranking and selection procedures are used to *screen* the alternative scenarios in order to find a small subset of those scenarios that contains the best (or at least a "good" one), or *select* outright the best scenario.

In practice, we could invoke a screening procedure to pare down a large number of alternatives into a palatable number; at that point, we might use a selection procedure to make the more fine-tuned choice of the best. Provided that certain assumptions are met, a screening procedure will choose a subset containing the best (or a good) scenario, and a selection procedure will pick the best, with a user-specified confidence level.

R&S procedures allow for the specification of a "practical-significant" difference, often denoted by  $\delta$ . Any scenario whose performance is within  $\delta$  of the best can be considered as a candidate for the best.

To facilitate what follows we define some notation: Let  $Y_{ii}$  represent the *j*th simulation output from scenario *i*, for  $i = 1, 2, \ldots, k$  scenarios and  $j = 1, 2, \ldots$  For fixed i, we will always assume that the outputs from scenario *i*,  $Y_{i1}, Y_{i2}, \ldots$ , are independent and identically distributed (i.i.d.). These assumptions are plausible if  $Y_{i1}, Y_{i2}, \ldots$ are outputs across independent replications, or if they are appropriately defined batch means from a single replication after accounting for initialization effects. Let  $\mu_i = E[Y_{ii}]$ denote the expected value of an output from the *i*th scenario, and let  $\sigma_i^2 = \text{Var}[Y_{ij}]$  denote its variance. The methods we describe in Sections 3.2-3.3 make comparisons based on estimates of  $\mu_i$ . We will only consider situations in which there is no known functional relationship among the  $\mu_i$ . Therefore, there is no potential information to be gained about one scenario from simulating the otherssuch as might occur if the  $\mu_i$  were a function of some explanatory variables-and no potential efficiency to be gained from fractional-factorial experiment designs, group screening designs, etc.

We now describe the three mean-based procedures that will be used by Scenario Selector; in Section 3.4 we describe an alternative approach that was also implemented.

## 3.2 Subset Selection

The subset selection approach is a screening device that attempts to select a (random-size) *subset* of the k competing designs that contains the design with the largest (or smallest) expected response  $\mu_i$ . Gupta (1956, 1965) proposed a single-stage procedure for this problem that is applicable in cases when the data from the competing designs are balanced (i.e., use equal sample sizes) and are normal with common (unknown) variance  $\sigma^2$ . Nelson, et al. (1998) handle more general cases—in particular, those in which the unknown variances  $\sigma_i^2$ , i = 1, 2, ..., k, are not necessarily equal. A variation of their procedure as implemented in Scenario Selector is given below:

#### 3.2.1 Subset

1. Specify the common sample size *n*, the practicalsignificant difference  $\delta$ , and the desired probability  $1 - \alpha$  of actually including some design within  $\delta$  of the best in the selected subset. Further, calculate the following quantile from the *t*-distribution with n - 1 degrees of freedom:

$$t = t_{1-(1-\alpha)^{\frac{1}{k-1}}, n-1}.$$

- 2. Take an i.i.d. sample  $Y_{i1}, Y_{i2}, \ldots, Y_{in}$  from each of the *k* scenarios simulated independently.
- 3. Calculate the k sample means  $\bar{Y}_i = \sum_{j=1}^n Y_{ij}/n$ . For all  $i \neq \ell$ , calculate the sample variance of the difference

$$S_{i\ell}^{2} = \frac{\sum_{j=1}^{n} \left( Y_{ij} - Y_{\ell j} - (\bar{Y}_{i} - \bar{Y}_{\ell}) \right)^{2}}{n-1}$$

In addition, calculate the quantity

$$W_{i\ell} = t S_{i\ell} / \sqrt{n}$$

for all  $i \neq \ell$ .

4. Include the *i*th design in the selected subset if

$$\bar{Y}_i \geq \bar{Y}_\ell - (W_{i\ell} - \delta)^+$$
 for all  $\ell \neq i$ ,

where  $(a)^+ = \max\{0, a\}$ .

If we had been interested in selecting responses with the *smallest* expected values, then the final step above would instead be to include the *i*th design if  $\bar{Y}_i \leq \bar{Y}_\ell + (W_{i\ell} - \delta)^+$ .

#### 3.3 Select the Best

Suppose we want to select the single scenario with the largest (or smallest) expected value. In a stochastic simulation such a "correct selection" can never be guaranteed with certainty. The selection procedures considered here guarantee to select the best scenario with high probability whenever it is at least a user-specified amount better than the others; this "practical-significant" difference  $\delta$  is called the indifference zone in the statistical literature. Law and Kelton (1991) describe a number of selection procedures that have proven useful in simulation, while Bechhofer, Santner, and Goldsman (BSG) (1995) provide a comprehensive review of such procedures.

Multiple comparison procedures (MCPs) approach the problem of determining the best scenario by forming simultaneous confidence intervals on the means  $\mu_i - \max_{j \neq i} \mu_j$  for i = 1, 2, ..., k. These confidence intervals are known specifically as *multiple comparisons with the best (MCB)*, and they bound the difference between the expected performance of each scenario and the best of the others. The first MCB procedures were developed by Hsu (1984); a thorough review is found in Hochberg and Tamhane (1987). MCPs are often used in conjunction with selection procedures (Matejcik and Nelson 1995 and Nelson and Matejcik 1995).

The next two procedures use sampling strategies in which the (normal) observations among scenarios are independent, i.e.,  $Y_{ij}$  is independent of  $Y_{i',j}$  for all  $i \neq i'$  and all j. The first combines Rinott's (1978) two-stage selection procedure with accompanying MCB intervals, simultaneously guaranteeing a probability of correct selection and confidence-interval coverage probability of at least  $1 - \alpha$  under the stated assumptions.

#### 3.3.1 Rinott

- 1. Specify the practical-significant difference  $\delta$ , the desired probability of correct selection  $1 \alpha$ , and the common first-stage sample size  $n_0 \ge 2$ . Let  $h_{\alpha}$  solve Rinott's integral for  $n_0$ , k, and  $\alpha$  (see the tables in Wilcox 1984 or BSG 1995).
- 2. Take an i.i.d. sample  $Y_{i1}, Y_{i2}, \ldots, Y_{in_0}$  from each of the *k* scenarios simulated independently.
- 3. Calculate the first-stage sample means  $\bar{Y}_i^{(1)} = \sum_{j=1}^{n_0} Y_{ij}/n_0$ , and marginal sample variances

$$S_i^2 = \frac{\sum_{j=1}^{n_0} \left( Y_{ij} - \bar{Y}_i^{(1)} \right)^2}{n_0 - 1},$$

for i = 1, 2, ..., k.

4. Compute the final sample sizes

$$N_i = \max\left\{n_0, \left\lceil (h_{\alpha}S_i/\delta)^2 \right\rceil\right\}$$

for i = 1, 2, ..., k, where  $\lceil \cdot \rceil$  is the integer "round-up" function.

- 5. Take  $N_i n_0$  additional i.i.d. observations from scenario *i*, independently of the first-stage sample and the other scenarios, for i = 1, 2, ..., k.
- 6. Compute the overall sample means  $\overline{\overline{Y}}_i = \sum_{j=1}^{N_i} Y_{ij}/N_i$  for i = 1, 2, ..., k.
- 7. Select the scenario with the largest  $\bar{Y}_i$  as best.
- 8. Simultaneously form the MCB confidence intervals

$$\mu_{i} - \max_{j \neq i} \mu_{j} \in \left[ -\left(\bar{\bar{Y}}_{i} - \max_{j \neq i} \bar{\bar{Y}}_{j} - \delta\right)^{-}, \left(\bar{\bar{Y}}_{i} - \max_{j \neq i} \bar{\bar{Y}}_{j} + \delta\right)^{+} \right]$$

for i = 1, 2, ..., k, where  $(a)^+ = \max\{0, a\}$  and  $-(b)^- = \min\{0, b\}$ .

If we had been interested in selecting the scenario with the *smallest* expected value, then the final steps above would instead be to select the scenario with the smallest  $\overline{Y}_i$  as best, and then form the MCB confidence intervals by substituting min's for max's.

Nelson, et al. (1998) show how to combine Section 3.2's subset procedure with the Rinott procedure. This *two-phase* procedure is of great utility when the experimenter is initially faced with a large number of alternatives—the idea is for the subset procedure to pare out non-contending scenarios, after which Rinott selects the best from the survivors.

#### 3.3.2 Two-Phase: Subset + Rinott

1. Specify the overall desired probability of correct selection  $1 - \alpha$ , the practical-significant difference  $\delta$ , the common initial sample size  $n_0 \ge 2$ , and the initial number of competing scenarios k. Further, set

$$t = t_{1-(1-\alpha/2)^{\frac{1}{k-1}}, n_0-1}$$

and let  $h_{\alpha/2}$  solve Rinott's integral for  $n_0$ , k, and  $\alpha/2$  (see Wilcox 1984 or BSG 1995).

2. Carry out algorithm Subset with  $n = n_0$ . Call the retained subset *I*. If *I* contains a single index, then stop and return that scenario as the best. Otherwise, for all  $i \in I$ , compute the second-stage sample sizes

$$N_i = \max\left\{n_0, \left\lceil (h_{\alpha/2}S_i/\delta)^2 \right\rceil\right\}$$

and take  $N_i - n_0$  additional i.i.d. observations from all scenarios  $i \in I$ .

- 3. Compute the overall sample means  $\overline{\bar{Y}}_i = \sum_{j=1}^{N_i} Y_{ij}/N_i$  for  $i \in I$ , and select the scenario with the largest  $\overline{\bar{Y}}_i$  as best.
- 4. With probability at least  $1 \alpha$ , we can claim that
  - \* for all  $i \in I^c$ , we have  $\mu_i < \max_{j \neq i} \mu_j$  (i.e., the scenarios excluded by the screening are not the best), and
  - \* if we define  $J_i = \{j : j \in I \text{ and } j \neq i\}$ , then for all  $i \in I$ ,

$$\begin{split} \mu_i &- \max_{j \in J_i} \mu_j \in \\ \left[ - \left( \bar{\bar{Y}}_i - \max_{j \in J_i} \bar{\bar{Y}}_j - \delta \right)^-, \left( \bar{\bar{Y}}_i - \max_{j \in J_i} \bar{\bar{Y}}_j + \delta \right)^+ \right] \end{split}$$

## 3.4 Multinomial Approach

Another approach is to select the scenario that is most likely to have the largest *actual* realization (instead of the largest *expected* realization). To this end, one can define  $p_i$  as the probability that design *i* will produce the largest realization from a given vector-observation  $(Y_{1j}, Y_{2j}, \ldots, Y_{kj})$ ; i.e.,

$$p_i = \Pr\left\{Y_{ij} > \max_{\ell \neq i} Y_{\ell j}\right\}$$

when  $Y_{1j}$ ,  $Y_{2j}$ , ...,  $Y_{kj}$  are mutually independent. The goal now is to select the design associated with the largest  $p_i$ value. This goal is equivalent to that of finding the multinomial category having the largest probability of occurrence, and there is a rich body of literature concerning such problems.

More specifically, suppose that we want to select the correct category with probability  $1 - \alpha$  whenever the ratio of the largest to second largest  $p_i$  is greater than some user-specified constant, say  $\theta > 1$ . The indifference constant  $\theta$  can be regarded as the smallest ratio "worth detecting."

The following *single-stage* procedure was proposed by Bechhofer, Elmaghraby, and Morse (BEM) (1959) to guarantee the above probability requirement.

#### 3.4.1 BEM

- 1. For the given k, and  $(\alpha, \theta)$  specified prior to the start of sampling, find *n* from the tables in BEM (1959), Gibbons, Olkin, and Sobel (1977) or BSG (1995).
- 2. Take a random sample of *n* observations  $Y_{i1}$ ,  $Y_{i2}, \ldots, Y_{in}$  from each scenario *i*, *i* = 1, 2, ..., *k*. Turn these into *n* independent multinomial observations,  $(X_{1j}, X_{2j}, \ldots, X_{kj})$ , *j* = 1, 2, ..., *n*, by setting

$$X_{ij} = \begin{cases} 1, & \text{if } Y_{ij} > \max_{\ell \neq i} \{Y_{\ell j}\} \\ 0, & \text{otherwise,} \end{cases}$$

where we randomize in the case of ties.

3. Let  $Q_i = \sum_{j=1}^n X_{ij}$  for i = 1, 2, ..., k. Select the design that yielded the largest  $Q_i$  as the one associated with the largest  $p_i$  (randomize in the case of ties).

#### 4 SCENARIO SELECTOR

Scenario Selector (Opicka and Pritsker 1998) is a computer program which implements the procedures described in the previous section for statistically comparing scenarios defined by one or more AweSim models. Scenario Selector provides dialog boxes to define the conditions under which screening and selection are to be performed, the scenarios to be considered as candidates for selection or the Visual SLAM model parameters that are to be used for defining new scenarios. Given the candidate scenarios and the statistical confidence for detecting differences among scenarios, Scenario Selector performs screening, selection, or screening and then selection, based on a single performance measure. Most of the procedures assume that the simulation output data are normally distributed and comparisons are based on mean performance, but Scenario Selector can also compare scenarios based on the probability that a scenario

will have the best performance in a one-shot comparison (multinomial selection).

To use the screening and selection algorithms, it is necessary to supply a practical-significant difference level  $\delta$ , and a confidence level with which the practical-significant difference is to be detected. For the multinomial procedure, a practical-significant difference ratio must be specified. The remaining specifications are the initial number of simulation runs to be made and the objective function, e.g., cost. The objective function is defined in the Visual SLAM model as the internally defined Visual SLAM array variable XX[objective] where objective is an index.

A more difficult interface design issue is defining the scenarios to be evaluated. There are two ways to define the set of scenarios to be compared: (1) a list of distinct models from a specified project; or (2) a single model with scenarios defined by different parameter settings, in particular, different combinations of Visual SLAM XX and LL variables and RESOURCE capacities. Scenarios may also be defined using a combination of (1) and (2). An illustration of Scenario Selector's user interface to define scenarios to be compared is shown in Figure 1.

The pull-down list on the left provides the means for selecting a project. AweSim maintains this list as the directory of all AweSim projects available on the computer being used. After a project is selected, the AweSim scenarios that have been developed under that project are available in the scenario name pull-down list. For comparing existing AweSim scenarios, an AweSim scenario is selected and added to the list of selected scenarios. The procedure for defining scenarios in terms of variable values requires selecting an AweSim scenario and specifying the number of XX and LL variables and RESOURCE capacities that need to be defined for each alternative. The user can then set the values by invoking the SET VARIABLES or SET RESOURCES button, which causes screens to be displayed for these inputs. If it is desired to set the variables at evenly spaced intervals rather than by value, then the FILL VARIABLES button is invoked.

Once the selection conditions and scenario definition information have been entered, the simulations are executed by invoking the RUN button shown on the bottom of Figure 1. When Scenario Selector has made a selection, the selected scenario is presented to the user. The user may then terminate the Scenario Selector application and proceed to retrieve the results.

All of the parameters that have been entered into the Scenario Selector interface, including Report Header data, can be saved to a file using File-Save As from the menu bar and providing a directory path and file name where the input parameters should be saved. For example, one could select C:\Projects\Selector\ as the directory path and enter EX51 as the file name. The file is saved with an ".ss" extension. When this file is opened, the Scenario Selector input screens are populated with the saved data.

The user may also provide a file name for storing the output by selecting Report-File Name for Output from the menu bar and providing a directory path and file name where the output file(s) are to be saved. A Word document is created in this location. If the user checks the Report-Detailed Report menu item, then an Excel file containing results from each run is also produced.

This brief description of Scenario Selector illustrates the need for classifying the inputs, procedures and outputs when employing R&S procedures. Having researchers, industrial engineers, software developers and simulation practitioners on the project team provided the information and interaction necessary to develop usable interfaces for Scenario Selector.



Figure 1: User Interface for Defining Alternative Scenarios

### 5 EXPERIENCES WITH SCENARIO SELECTOR APPLICATIONS

This section presents illustrations of the use of Scenario Selector. The three model types used in the examples are inventory; queueing; and project management. Due to space limitations, it is not possible to provide detailed problem statements or the Visual SLAM models. The focus of the examples is on the use of Scenario Selector and the presentation of outputs.

**Example 1.** The problem is the setting of a stock control level, reorder point and time between reviews for a periodic review inventory control system for a single product, a radio (Pritsker and O'Reilly 1999, pp. 180–186, 392–395). The performance measure is average weekly profit, where profit is equal to revenues minus costs; revenue is from the sale of radios and costs are inventory carrying costs, lost sales imputed cost, cost of reviewing inventory, and ordering and purchasing costs. The calculation of average weekly profit is included in the Visual SLAM model.

Subset selection is used in this example which requires the input to Scenario Selector of the practicalsignificant difference level ( $\delta = \$1000$ ), confidence level ( $1 - \alpha = 0.90$ ) and initial number of runs ( $n_0 = 20$ ). The objective is to select a subset of the scenarios that contains the alternative that yields the highest expected weekly profit. The number of scenarios to be considered is four, consisting of combinations of three model variables: reorder point, stock control level and time between reviews. Scenarios 1 through 4 are associated with the following triples of these variables, respectively: (18, 72, 4); (18, 72, 2); (10, 72, 2); and (10, 72, 4).

Figure 2 shows the output report from Scenario Selector, which includes the problem and the inputs that were specified. The scenarios that are retained in the subset of possible best alternatives are also identified. Scenario 2 is the only retained alternative and hence it is the best scenario. Finally, the scenarios are listed along with their decision variables and the average weekly profit from each of the 20 runs. Scenario 2 had an average profit of \$85,069 per week for the simulated time period.

**Example 2.** The model for this example involves a banking situation in which there are two tellers each with their own queue. Arriving customers to the bank select the teller subsystem that has the fewest customers in it. If there is a difference of two between the numbers of customers in the two subsystems, then the last customer in the longer queue jockeys to the queue of the other teller. The number of customers who can be in the system is limited to twenty and, when the system is full, an arriving customer balks from the bank. The system is simulated for a period of 1,000 minutes.

Scenario Selector Detailed Report

Simulation Project: INVENTORY SYSTEM Modeler: PRITSKER Date: 2/05/99 Number of Scenarios: 4 Procedure:Avg (SubsetSelection) Scenario Directory: C:\PROJECTS\EXAMPLE\ Larger Better? Yes Practical Significant Difference: 1000.0 Confidence Level: 0.90

Selection Objective: Maximize profit

 Retained Scenarios: Scenario2								
Scenario Name S	cenariol	Scenario2	Scenario3	Scenario4				
Input Values:								
Reorder Point	18	18	10	10				
Stock Control level	72	72	72	72				
Time Between Review	в 4	2	2	4				
Results:								
Mean	82050	85069	77171	74777				
Variance	5559653	4643530	5157343	3101496				
Standard Deviation	2358	2155	2271	1761				
Total Sample Size	20	20	20	20				
Run 1	78825	82865	74810	74981				
Run 2	83485	88445	79532	77197				
Run 19	84080	85979	78345	76780				
Run 20	80290	85874	77475	73041				

Figure 2: Scenario Selector Output for Inventory Analysis

The problem is to determine the best way to estimate the steady-state expected number of customers in the banking system based on a restricted-time simulation. Two decision variables in making this estimate are the initial number of customers in the banking system and the time to clear initial results, since they are not representative of the steady-state system behavior. A measure that has been proposed for evaluating initial conditions and clear (truncation) times is the mean square error (MSE) (Wilson and Pritsker 1978ab). To compute the MSE, it is necessary to know the theoretical mean or to have an estimate of it. For the bank teller model, an analysis has been performed that computes the theoretical mean.

For Example 2 we use the algorithm Two-Phase. It is desired to detect differences in the MSE of 0.04 with a confidence of 0.9. The Phase 1 number of runs is set at 20. The scenario for this example is stored in the AweSim project whose name is EXAMPLE. Eight variable sets are defined by two decision variables: XX[1], the clear time, and LL[1], the initial number in the system. The eight values are combinations of four initial conditions representing empty and idle (LL[1]=0); half-full (LL[1]=10); the closest integer to the steady-state average (LL[1]=4); and the steady-state median (LL[1]=3); and two clear times of 0 and 100. Figure 3 shows a portion of the SET VARIABLES screen which specifies the eight scenarios defined by combinations of XX[1] and LL[1].

The output report from Scenario Selector is shown in Figure 4. At the top of the page the header information input to Scenario Selector is provided. Scenario Selector statistically determined that Scenario 3 was the best of the scenarios; it has a clear time of 0 and an initial number in the system of 4. The detailed results show the estimated MSE for all runs performed for a given

	Scenario1	Scenario2	Scenario3	Scenario4	Scenario5	Scenario6	Sc
×	0	0	0	0	100	100	
LL	0	3	4	10	0	3	
•							Þ
•							Þ
<u>.</u>							Þ

Figure 3: Decision Variables for the Scenarios of the Bank Teller Model

scenario. After the first phase of sampling, Scenario Selector retained Scenarios 1–4 in the subset as being statistically indistinguishable within the practical-significant MSE difference of 0.04. For each scenario, the estimated MSE, as well as its variance and standard deviation based on the total sample size, are also shown in Figure 4. The confidence interval for multiple comparisons with the best (MCB) and the differences from the best mean are also outputs from the Two-Phase selection algorithm.

**Example 3.** Scenario Selector is used in this example to compare alternative designs for an Armored Vehicle Launched Bridge (AVLB). The original risk analysis was performed at the Army Mobility Research Development Center which provided the basis for the Visual SLAM models (Pritsker and O'Reilly 1999, pp. 301–306). Four different scenarios are available to develop an AVLB and it is desired to select the scenario which has the highest probability of having the least total cost. To accomplish this,

Scenario Selector Detailed Report

we use a multinomial selection algorithm. The multinomial procedure requires a confidence level which is the probability of making a correct selection (0.9), a difference ratio (1.6) and the number of scenarios (4). Based on these input values, the number of runs to be made is determined to be 98 by Scenario Selector, which includes the appropriate tables to determine the smallest number of runs required. Since the scenarios were previously built, the user need only identify them from the AweSim project name. Activating the list of existing scenarios of the project from the Scenario Selector interface allows the user to pick the set of scenarios to be compared.

Figure 5 shows the detailed report from Scenario Selector for comparing the four AVLB design programs using the multinomial algorithm. After presenting the input values, Scenario Selector identifies the best scenario as AVLB1. The output provides the cost for each scenario for the 98 runs of which only the first three runs are shown in Figure 5. The number of times each scenario had the

Scenario	Selector	Detailed	Report
Scenario	Selector	Decarred	report

Simulation Project: CO Modeler: PRITSKER Number of Scenarios: - Scenario Directory: C Indifference Ratio: 1 Selection Objective: 	CTS Date: 2 Procedu Larger Confide hat has t of times	/08/99 re: Multir Better? No nce Level: he least c	omial 0.90		
Best Scenario: AVLB1					
Scenarios Results:	AVLB1	AVLB2	AVLB3	AVLB4	
Mean	1769	3086	2461	1672	
Variance	196016	265715	45044	77613	
Std.Dev.	443	515	212	279	
Sample Size	98	98	98	98	
Run 1	1443	3292	2538	2709	
Run 2	2614	3277	2607	1497	
Run 3	2555	2448	2548	1562	
Total Wins out of 98	51	2	1	44	



beenario beree	JUDI DUCUI	ICG REPOIL									
Simulation Project: Bank Tellers			Modeler	Modeler: Pritsker							
Date: 10/29/98											
Number of Scenarios: 8			Procedu	Procedure: Avg(Two Phase)							
Scenario Directory	y: C:\PROJ	ECTS\EXAMPLE\	Larger	Larger Better? No							
Significant Difference: 0.0400			Confide	Confidence Level: 0.90							
Initial Number of	Runs: 20										
Selection Objectiv	ve:	Minimize Mea	an Square Ern	ror							
Best Scenario:	Scenario	3									
Retained Scenarios	s: Scen	ariol Scer	nario2 So	cenario3	Scenario4						
Scenario Name Sc	cenario1	Scenario2	Scenario3	Scenario4	Scenario5 Scenari	o6 Scenario7	Scenario8				
Input Values:											
XX[1]	0	0	0	0	100	100	100	100			
LL[1]	0	3	4	10	0	3	4	10			
Results:											
Mean	0.1810	0.1764	0.1289	0.2087	0.1680	0.1741	0.1481	0.1830			
Variance	0.0636	0.0594	0.0253	0.1058	0.0285	0.0389	0.0145	0.0699			
Standard Dev	0.2522	0.2436	0.1592	0.3252	0.1689	0.1971	0.1205	0.2644			
Interval (MCB)	0.0922	0.0875	0	0.1198	-	-	-	-			
Diff from Best	0.0522	0.0475	0	0.0798	-	-	-	-			
Total Sample Size	281	343	113	350	20	20	20	20			

Figure 4: Scenario Selector Output for Bank Teller Model

lowest cost is displayed in the last row which indicates that Scenario AVLB1 has the least cost in 51 of the 98 runs while the other 3 scenarios had the least cost 2, 1 and 44 times, respectively.

# 6 EPILOG

All of the participants in this joint effort felt that it was successful, and the creation of Scenario Selector a major enhancement to the output analysis capability of AweSim—is tangible proof that it was. Why was this industry-university collaboration able to succeed when so many others fail? In this section we list some of the barriers to such collaboration and how we believe we were able to overcome them.

• The difference between what is publishable and what is useful: The desire to publish research findings pushes academics toward topics and results that are different, in a substantial way, from what has been published previously. The need to prove that results are better than the status quo often narrows the class of problems or assumptions that are considered. This approach to research is not always compatible with what is useful, meaning robust, widely applicable, usable and easily understood.

Our collaboration started with a National Science Foundation proposal by G&N to produce results that are useful in applications, and to work with the sponsoring software vendors to see that technology transfer actually occurred. In fact, two of the four procedures implemented in Scenario Selector were not results of G&N's work, but were recommended by them because they satisfied Pritsker Corporation's needs. This approach was possible because neither academic needed to be concerned with achieving tenure, so they were willing to spend time on matters such as implementation and testing that did not lead directly to publications. NSF's requirement that all engineering proposals contain a technology transfer component also encouraged the academics to focus on applications.

• The language barrier: The languages of research and practice can be remarkably different, and this presents an impediment to collaboration and technology transfer. Research communication emphasizes precision and formality because of the need to rigorously prove correctness and to state results unambiguously. Communication for practice emphasizes interpretation and how a result will be applied, because for the end user the software is just a tool and not important in and of itself.

The language barrier existed in our collaboration, but because Pritsker had been an academic researcher he was able to understand G&N's approach to problems and insist on a translation into practical terms. His experience in software product implementation provided the joint effort with basic information on procedure invocation, input/output specifications, and testing and evaluation. He also felt that the struggle to communicate was worth the effort and did not hesitate to restate and refine until all parties understood.

• The difference between academic and industry time scales: Academics often work on a much longer time scale than industry, with minimum time units of quarters (3 months) or semesters (4 months). It is not unusual for a university researcher to put aside a research project for months or even years when progress is slow or other work interferes. Obviously this kind of schedule is incompatible with industry's need to set product release dates, for instance.

In this project, G&N were willing to work to Pritsker Corporation's schedule, particularly with respect to answering questions in a timely manner when Opicka was implementing the procedures. And as Scenario Selector neared completion, G&N scheduled additional meetings with Pritsker and Opicka and helped in coding pieces of the algorithms.

Bureaucracy: On any research project universities strive to protect the publishability of results in the open literature and to control the patent rights to any new ideas. On the other hand, industry wants to own the results it pays for in order to maintain market edge. Although Pritsker was clearly aware of competitive market issues, he understood the academics' need to publish. More importantly, he felt that competitive advantage comes less from the theoretical results themselves and more from how effectively the results are implemented in a product, as measured for instance by ease of use, understandability, accuracy, etc. By having the procedures implemented by Opicka as an employee of Pritsker Corporation, Pritsker Corporation maintained ownership of the implementation, allowing the ideas to be freely published.

The members of the team all felt that the collaboration garnered a number of positive experiences. In particular, the academic-practitioner interaction spawned by the project was invaluable-the academics were exposed to real-world problems, while the practitioners implemented in Scenario Selector provably rigorous statistical algorithms. The academics even helped do some production-level computer coding, and they assisted in the design and development of Scenario Selector's user-interface screens. The practitioners also brought a number of interesting issues to the fore. They suggested new, ongoing research avenues, e.g., the development of multi-stage algorithms that explicitly use sample-mean differences between alternatives to cut down on subsequent sampling effort. Further, the practitioners came up with other performance criteria under which comparisons between alternatives can be made, e.g., an indifference zone for means based on ratios rather than differences. A final testament to the success of the joint effort is that it has formed the basis for Opicka's M.S. thesis (Opicka 1999) and resulted in continuing research activities among the four collaborators.

## ACKNOWLEDGMENTS

This research was partially supported by National Science Foundation grants DMI-9622065 and DMI-9622269, and by SYMIX Systems/Pritsker Division.

## REFERENCES

- Bechhofer, R. E., S. Elmaghraby and N. Morse. 1959. A single-sample multiple-decision procedure for selecting the multinomial event which has the highest probability. *Annals of Math. Stat.* 30:102–119.
- Bechhofer, R. E., T. J. Santner and D. Goldsman. 1995. Design and Analysis of Experiments for Statistical Selection, Screening and Multiple Comparisons. New York: John Wiley.
- Gibbons, J. D., I. Olkin and M. Sobel. 1977. Selecting and Ordering Populations: A New Statistical Methodology. New York: John Wiley.
- Gupta, S. S. 1956. On a Decision Rule for a Problem in Ranking Means. Ph.D. Dissertation (Mimeo. Ser. No. 150). Inst. of Statist., Univ. of North Carolina, Chapel Hill, North Carolina.
- Gupta, S. S. 1965. On some multiple decision (selection and ranking) rules. *Technometrics* 7:225–245.
- Hochberg, Y., and A. C. Tamhane. 1987. *Multiple Com*parison Procedures. New York: John Wiley.
- Hsu, J. C. 1984. Constrained simultaneous confidence intervals for multiple comparisons with the best. *Annals of Statistics* 12:1136–1144.
- Law, A. M., and W. D. Kelton. 1991. *Simulation Modeling & Analysis*. 2nd ed. New York: McGraw-Hill, Inc.

- Matejcik, F. J., and B. L. Nelson. 1995. Two-stage multiple comparisons with the best for computer simulation. *Operations Research* 43:633–640.
- Nelson, B. L., and F. J. Matejcik. 1995. Using common random numbers for indifference-zone selection and multiple comparisons in simulation. *Management Science* 41:1935–1945.
- Nelson, B. L., J. Swann, D. Goldsman and W.-M. Song. 1998. Simple procedures for selecting the best system when the number of alternatives is large. Technical Report. Dept. of IEMS, Northwestern Univ., Evanston, Illinois.
- Opicka, T. L. 1999. Evaluating Ranking and Selection Procedures in Computer Simulation. M.S. Thesis, School of Industrial Engineering, Purdue University, West Lafayette, Indiana.
- Opicka, T. L., and A. A. B. Pritsker. 1998. *Scenario Selector User's Guide*. SYMIX Systems/Pritsker Division, West Lafayette, Indiana.
- Pritsker, A. A. B., and J. J. O'Reilly. 1999. Simulation with Visual SLAM and AweSim. 2nd ed. New York: John Wiley; West Lafayette, Indiana: Systems Publishing Corp.
- Rinott, Y. 1978. On two-stage selection procedures and related probability-inequalities. *Comm. Stat.*—*Thy. and Meth.* A7:799–811.
- Wilcox, R. R. 1984. A table for Rinott's selection procedure. J. Quality Technology 16:97–100.
- Wilson, J. R., and A. A. B. Pritsker. 1978a. A survey of research on the simulation startup problem. *Simulation* 31:55–58.
- Wilson, J. R., and A. A. B. Pritsker. 1978b. A procedure for evaluating startup policies in simulation experiments *Simulation* 31:79–89.

### AUTHOR BIOGRAPHIES

**DAVID GOLDSMAN** is an Associate Professor in the School of Industrial and Systems Engineering at Georgia Tech. His research interests include simulation output analysis and ranking and selection. He is the Simulation Department Editor for *IIE Transactions*, and an Associate Editor for *Operations Research Letters*. He was the Program Chair for the 1995 Winter Simulation Conference.

**BARRY L. NELSON** is a Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University, and is Director of the Master of Engineering Management Program there. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems. He has published numerous papers and two books. Nelson has served the profession as the Simulation Area Editor of *Operations Research* and President of the INFORMS (then TIMS) College on Simulation. He has held many positions for the Winter Simulation Conference, including Program Chair in 1997.

**TRACY L. OPICKA** is a logistics engineer with Schneider National, Inc. She is involved primarily with the Business Development team, which provides solutions to supply chain problems for various customer groups. She received her M.S. in Industrial Engineering from Purdue University in 1999 and her B.S. in Mathematics and Computer Science from St. Norbert College, De Pere, Wisconsin in 1995. She was also a computer consultant with Andersen Consulting from 1995 to 1997 and an intern at SYMIX Systems, Inc. in 1998.

**A. ALAN B. PRITSKER** was Chairman and CEO of Pritsker Corporation. He was employed by Battelle Memorial Institute, Arizona State University, Virginia Tech, and Purdue University. He has published more than 150 technical papers and 12 books, and is the developer or codeveloper of the following techniques and languages: GERT, Q-GERT, GASP IV and Visual SLAM. In 1985, Pritsker was elected to the National Academy of Engineering for development of the basic concepts underlying combined discrete/continuous simulation languages. He is a Fellow of IIE and, in 1991, he was selected the Frank and Lillian Gilbreth Award winner, IIE's highest award. He received a Doctor of Science, *honoris causa*, from Arizona State University in May 1992 and a Doctor of Engineering, *honoris causa*, from Purdue University in December 1998.