Proceedings of the 1998 Winter Simulation Conference
D.J. Medeiros, E.F. Watson, J.S. Carson and M.S. Manivannan, eds.

STATISTICAL SCREENING, SELECTION, AND MULTIPLE COMPARISON PROCEDURES IN COMPUTER SIMULATION

David Goldsman

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

Barry L. Nelson

Dept. of Industrial Engineering & Management Sciences
Northwestern University
Evanston, IL 60208-3119, U.S.A.

ABSTRACT

We present a state-of-the-art review of screening, selection, and multiple-comparison procedures that are used to compare system designs via computer simulation. We describe methods for three broad classes of problems: screening a large number of system designs, selecting the best system, and comparing all systems to a standard (either known or unknown). We concentrate primarily on recent methods that we would be likely to use in practice. Where possible, we unify the screening, selection, and multiple-comparison perspectives.

1 INTRODUCTION

Simulation experiments are often performed to compare, in some fashion, two or more system designs. The statistical methods of *screening*, *selection*, and *multiple comparisons* are applicable when we are interested in making comparisons among a finite and typically small number of systems (say 2 to 30). The particular method that is appropriate depends on the type of comparison desired and properties of the simulation output data. In this state-of-the-art review we describe methods for three broad classes of problems: screening a large number of system designs, selecting the best system, and comparing all systems to a standard (either known or unknown). We focus herein on the methods that we would be likely to use on real problems. And where possible, we unify the screening, selection, and multiple-comparison perspectives.

Screening and selection procedures (SSPs) are statistical methods designed to find the "best" (or "nearly the best") system from among a collection of competing alternatives. For example, such procedures could be efficacious in any of the following practical situations:

- A manufacturer would like to know which of three potential plant layouts will maximize expected revenues.
- A polling service wishes to determine the most popular candidate before a certain election.
- A medical research team conducts a clinical study comparing the success rates of five different drug regimens for a particular disease.

Informally speaking, SSPs are used to

- *screen* the competitors in order to find a small subset of those systems that contains the best system (or at least a "good" one).
- *select* outright the best system.

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, SSPs usually guarantee a user-specified probability of advertised performance—i.e., with high probability, a screening procedure will choose a subset containing the best (or a good) alternative, and a selection procedure will pick the best.

Multiple-comparison procedures (MCPs) treat the comparison problem as an inference problem on the performance parameters of interest. MCPs account for the error that arises when making simultaneous inferences about differences in performance among the systems. Usually, MCPs report to the user simultaneous confidence intervals for the differences. Recent research has shown that MCPs can be combined with SSPs for a variety of problems—including the manufacturing, polling, and medical examples outlined above. In fact, a number of significant advances have appeared in the literature since our last WSC tutorial, Goldsman and Nelson (1994).

SSPs and MCPs are relevant and statistically valid in the context of computer simulation because the assumptions behind the procedures can frequently be satisfied: the assumption of normally distributed data can

often be secured by batching large numbers of (cheaply generated) outputs. Independence can be obtained by controlling random-number assignments. And multiple-stage sampling—which is required by some methods—is feasible in computer simulation because a subsequent stage can be initialized simply by retaining the final random-number seeds from the preceding stage. As a bonus, it is possible to enhance (in a theoretically rigorous way) the performance of some of the procedures through the use of common random numbers, a popular variance reduction technique sometimes used in simulation.

The remainder of this article is organized as follows. In the next section, we will establish relevant notation and ground rules. Sections 3–5 discuss screening methods, procedures for selecting the best alternative, and methods for comparing alternatives with a standard, respectively. Some final thoughts end the discussion in Section 6.

2 PRELIMINARIES

To facilitate what follows we define some notation: Let Y_{ij} represent the jth simulation output from system design i, for $i = 1, 2, \ldots, k$ alternatives and $j = 1, 2, \ldots$. For fixed i, we will always assume that the outputs from system 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 $\mu_i = \mathrm{E}[Y_{ij}]$ denote the expected value of an output from the ith system, and let $\sigma_i^2 = \mathrm{Var}[Y_{ij}]$ denote its variance. Further, let

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

be the probability that Y_{ij} is the largest of the jth outputs across all systems *when* $Y_{1j}, Y_{2j}, \ldots, Y_{kj}$ *are mutually independent*.

The methods we describe make comparisons based on either μ_i or p_i. Although not a restriction on either SSPs or MCPs, we will only consider situations in which there is no known functional relationship among the μ_i or p_i (other than $\sum_{i=1}^{k} p_i = 1$). Therefore, there is no potential information to be gained about one system from simulating the others—such as might occur if the μ_i were a function of some explanatory variables—and no potential efficiency to be gained from fractional-factorial experiment designs, group screening designs, etc.

3 SCREENING PROBLEMS

Example 1 *A brain-storming session produces* 25 *potential designs for the architecture of a new computer system. Expected response time is the performance measure of interest, but there are so many designs that a careful simulation study will be deferred until a pilot simulation study determines which designs are worth further scrutiny. Smaller response time is preferred.*

The expected response time is the performance measure of interest, and the goal of the pilot study is to determine which designs are the better performers, which have similar performance, and which can be eliminated as clearly inferior.

3.1 Multiple Comparison Approach

Let μ_i denote the expected response time for architecture i. MCPs attack the screening problem by forming simultaneous confidence intervals on the parameters $\mu_i - \mu_j$ for all $i \neq j$. These $k(k-1)/2$ confidence intervals indicate the magnitude and direction of the difference between each pair of alternatives. The most widely used method for forming the intervals is Tukey's procedure, which is implemented in many statistical software packages and which is known to be more efficient than other methods such as those involving Bonferroni's inequality. We review Tukey's procedure here and cite some recent advances. General references include Hochberg and Tamhane (1987) and Miller (1981); most standard statistics textbooks also contain some version of the procedure.

Suppose that the systems are simulated independently, and we obtain i.i.d. outputs $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ from system i, $i = 1, 2, \ldots, k$. Let $\bar{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$ be the sample mean from system i, and let

$$S^2 \;=\; \frac{1}{k}\sum_{i=1}^{k} \frac{1}{n_i - 1}\sum_{j=1}^{n_i}(Y_{ij} - \bar{Y}_i)^2$$

be the pooled sample variance. Tukey's simultaneous confidence intervals are

$$\mu_i - \mu_j \;\in\; \bar{Y}_i - \bar{Y}_j \pm \frac{Q_{k,\nu}^{(\alpha)}}{\sqrt{2}} S \sqrt{\frac{1}{n_i} + \frac{1}{n_j}}$$

for all $i \neq j$, where $Q_{k,\nu}^{(\alpha)}$ is the $1-\alpha$ quantile of the Studentized range distribution with parameter k and $\nu = \sum_{i=1}^{k}(n_i - 1)$ degrees of freedom (see for instance Hochberg and Tamhane 1987, Appendix 3, Table 8, or Goldsman and Nelson 1998, Table 8.1).

When the Y_{ij} are normally distributed with *common* (unknown) variance, and $n_1 = n_2 = \cdots = n_k$, these intervals achieve simultaneous coverage probability $1 - \alpha$. Hayter (1984) made the important finding that the coverage probability is strictly greater than $1 - \alpha$ when the sample sizes are not equal.

3.2 Subset Selection Approach

The subset selection approach is a screening device that attempts to select a (random-size) *subset* of the $k = 25$ competing designs of Example 1 that contains the design with the smallest expected response time. 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., $n_1 = \cdots = n_k = n$) and are normal with common (unknown) variance σ^2. Nelson, et al. (1998) handle more general cases—in particular, that in which the unknown variances σ_i^2, $i = 1, 2, \ldots, k$, are not necessarily equal. Their procedure is illustrated below.

(Random) Subset Procedure

1. Specify the common sample size n, as well as the desired probability $1-\alpha$ of actually including the best design 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 systems simulated independently.

3. Calculate the k sample means $\bar{Y}_i = \sum_{j=1}^{n} Y_{ij}/n$, and sample variances

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

for $i = 1, 2, \ldots, k$. In addition, calculate the quantity

$$W_{ij} = t\left(\frac{S_i^2 + S_j^2}{n}\right)^{1/2}$$

for all $i \neq j$.

4. Include the ith design in the selected subset if

$$\bar{Y}_i \leq \bar{Y}_j + W_{ij} \text{ for all } j \neq i.$$

If we had been interested in selecting responses with the *largest* expected values, then the final step above would instead be

4′. Include the ith design in the selected subset if

$$\bar{Y}_i \geq \bar{Y}_j - W_{ij} \text{ for all } j \neq i.$$

Notice that as the common sample size n increases, the W_{ij} tend to decrease, and so the random subset size also tends to decrease.

Sullivan and Wilson (1989) proposed *restricted* subset procedures, in which the experimenter can specify an upper bound on the subset size.

4 SELECTING THE BEST

Example 2 (Goldsman, Nelson, and Schmeiser 1991) *For the purpose of evaluation prior to purchase, simulation models of four different airline-reservation systems have been developed. The single measure of system performance is the time to failure (TTF), so that larger TTF is better. A reservation system works if either of two computers works. The four systems arise from variations in parameters affecting the TTF and time-to-repair distributions. Differences of less than about two days are considered practically equivalent.*

4.1 Indifference-Zone Selection Approach

If expected TTF is taken as the performance measure of interest, then the goal in this example is to select the system with the largest expected TTF. In a stochastic simulation such a "correct selection" can never be guaranteed with certainty. A compromise solution offered by *indifference-zone selection* is to guarantee to select the best system with high probability whenever it is at least a user-specified amount better than the others; this "practically-significant" difference is called the indifference zone. In the example the indifference zone is $\delta = 2$ days. Law and Kelton (1991) describe a number of indifference-zone procedures that have proven useful in simulation, while Bechhofer, Santner, and Goldsman (BSG) (1995) provide a comprehensive review of SSPs.

MCPs approach the problem of determining the best system by forming simultaneous confidence intervals on the parameters $\mu_i - \max_{j \neq i} \mu_j$ for $i = 1, 2, \ldots, k$, where μ_i denotes the expected TTF for the ith reservation system. These confidence intervals are known as *multiple comparisons with the best (MCB)*, and they bound the difference between the expected performance of each system 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).

Matejcik and Nelson (1995) and Nelson and Matejcik (1995) established a fundamental connection between indifference-zone selection and MCB by showing that *most indifference-zone procedures can simultaneously provide MCB confidence intervals with the width of the intervals corresponding to the indifference zone*. The procedures we display below are combined indifference-zone selection and MCB procedures. The advantage of a combined procedure is that we not only select a system as best, we also gain information about how close each of the inferior systems is to being the best. This information is useful if secondary criteria that are not reflected in the performance measure (such as ease of installation, cost to maintain, etc.) may tempt us to choose an inferior system if it is not deficient by much.

4.1.1 Independent Sampling

The next two combined procedures use sampling strategies in which the normal observations between 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 indifference-zone 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.

Rinott + MCB

1. Specify the indifference-zone parameter δ, the desired probability of correct selection $1-\alpha$, and the common first-stage sample size $n_0 \geq 2$. Let h_α solve Rinott's integral for n_0, k, and α (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 systems 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

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

for $i = 1, 2, \ldots, k$.

4. Compute the final sample sizes

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

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

5. Take $N_i - n_0$ additional i.i.d. observations from system i, independently of the first-stage sample and the other systems, for $i = 1, 2, \ldots, k$.

6. Compute the overall sample means $\bar{\bar{Y}}_i = \sum_{j=1}^{N_i} Y_{ij}/N_i$ for $i = 1, 2, \ldots, k$.

7. Select the system with the largest $\bar{\bar{Y}}_i$ 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 $i = 1, 2, \ldots, k$, where $(a)^+ = \max\{0, a\}$ and $-(b)^- = \min\{0, b\}$.

If we had been interested in selecting the system with the *smallest* expected value, then the final steps above would instead be

7′. Select the system with the smallest $\bar{\bar{Y}}_i$ as best.

8′. Simultaneously form the MCB confidence intervals

$$\mu_i - \min_{j \neq i} \mu_j \;\in$$

$$\left[-\left(\bar{\bar{Y}}_i - \min_{j \neq i} \bar{\bar{Y}}_j - \delta\right)^-, \left(\bar{\bar{Y}}_i - \min_{j \neq i} \bar{\bar{Y}}_j + \delta\right)^+ \right]$$

for $i = 1, 2, \ldots, k$.

Nelson, et al. (1998) show how to combine Section 3.2's subset procedure with the Rinott+MCB procedure. This combined 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 systems, after which Rinott selects the best from the survivors.

Subset + Rinott + MCB

1. Specify the overall desired probability of correct selection $1-\alpha$, the indifference zone δ, the common initial sample size $n_0 \geq 2$, and the initial number of competing systems 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. Take an i.i.d. sample $Y_{i1}, Y_{i2}, \ldots, Y_{in_0}$ from each of the k systems simulated independently.

3. As in Step 3. of Rinott+MCB.

4. Calculate the quantity

$$W_{ij} \;=\; t\left(\frac{S_i^2 + S_j^2}{n_0}\right)^{1/2}$$

for all $i \neq j$. Form the screening subset I, containing every alternative i such that $1 \leq i \leq k$ and

$$\bar{Y}_i^{(1)} \;\geq\; \bar{Y}_j^{(1)} - (W_{ij} - \delta)^+ \text{ for all } j \neq i.$$

5. If I contains a single index, then stop and return that system as the best. Otherwise, for all $i \in I$, compute the second-stage sample sizes

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

6. Take $N_i - n_0$ additional i.i.d. observations from all systems $i \in I$, independently of the first-stage sample and the other systems.

7. Compute the overall sample means $\bar{\bar{Y}}_i = \sum_{j=1}^{N_i} Y_{ij}/N_i$ for $i \in I$.

8. Select the system with the largest $\bar{\bar{Y}}_i$ as best.

9. 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 systems 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 $i \in I$,

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

(Thus, these confidence intervals bound the difference between each alternative and the best of the others in I.)

4.1.2 Correlated Sampling

A fundamental assumption of the Rinott+MCB and the Subset+Rinott+MCB procedures is that the k systems are simulated independently (see Step 2 in both of the above procedures). In practice this means that different streams of (pseudo)random numbers are assigned to the simulation of each system. However, under fairly general conditions, assigning common random numbers (CRN) to the simulation of each system decreases the variances of estimates of the pairwise differences in performance. Unfortunately, CRN also complicates the statistical analysis when $k > 2$ systems are involved. The following procedure from Nelson and Matejcik (1995) provides the same guarantees as Rinott+MCB under a more complex set of conditions, but has been shown to be quite robust to departures from those conditions. And unlike Rinott+MCB, it is designed to exploit the use of CRN to reduce the total number of observations required to make a correct selection.

NM + MCB

1. Specify the constants δ, α, and n_0. Let $g = T^{(\alpha)}_{k-1,(k-1)(n_0-1),0.5}$, an equicoordinate critical point of the equicorrelated multivariate central t-distribution; this constant can be found in Hochberg and Tamhane (1987), Appendix 3, Table 4; BSG (1995); or by using the FORTRAN program AS251 of Dunnett (1989).

2. Take an i.i.d. sample $Y_{i1}, Y_{i2}, \ldots, Y_{in_0}$ from each of the k systems *using CRN across systems*.

3. Compute the approximate sample variance of the difference of the sample means

$$S^2 = \frac{2\sum_{i=1}^{k}\sum_{j=1}^{n_0}\left(Y_{ij} - \bar{Y}_{i\cdot} - \bar{Y}_{\cdot j} + \bar{Y}_{\cdot\cdot}\right)^2}{(k-1)(n_0-1)},$$

where $\bar{Y}_{i\cdot} = \sum_{j=1}^{n_0} Y_{ij}/n_0$, $\bar{Y}_{\cdot j} = \sum_{i=1}^{k} Y_{ij}/k$, and $\bar{Y}_{\cdot\cdot} = \sum_{i=1}^{k}\sum_{j=1}^{n_0} Y_{ij}/kn_0$.

4. Compute the final sample size

$$N = \max\left\{n_0, \left\lceil (gS/\delta)^2 \right\rceil\right\}.$$

5. Take $N - n_0$ additional i.i.d. observations from each system, using CRN across systems.

6. Compute the overall sample means $\bar{\bar{Y}}_i = \sum_{j=1}^{N} Y_{ij}/N$ for $i = 1, 2, \ldots, k$.

7. Select the system with the largest $\bar{\bar{Y}}_i$ as best.

8. Simultaneously form the MCB confidence intervals as in Rinott+MCB.

4.2 Multinomial Selection Approach

Another approach to the airline-reservation problem is to select the system that is most likely to have the largest *actual* TTF (instead of the largest *expected* TTF). To this end, one can define p_i as the probability that design i will produce the largest TTF from a given vector-observation $\boldsymbol{Y}_j = (Y_{1j}, Y_{2j}, \ldots, Y_{kj})$. 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 θ 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.

BEM

1. For the given k, and (α, θ) 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 alternative i, $i = 1, 2, \ldots, k$. Turn these into n independent multinomial observations, $\boldsymbol{X}_j = (X_{1j}, X_{2j}, \ldots, X_{kj})$, $j = 1, 2, \ldots, n$, 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 assume (for notational convenience) that there are never ties for the maximum observation within a particular vector $\boldsymbol{Y}_j$.

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

A more efficient procedure, due to Bechhofer and Goldsman (1986), uses *closed, sequential* sampling; that is, the procedure stops when one design is "sufficiently ahead" of the others.

BG

1. For the given k, and (α, θ) specified prior to the start of sampling, find the *truncation number* (i.e., an upper bound on the number of vector-observations) n_0 from the tables in Bechhofer and Goldsman (1986) or BSG (1995).

2. At the mth stage of experimentation ($m \geq 1$), take the random multinomial observation $\boldsymbol{X}_m = (X_{1m}, X_{2m}, \ldots, X_{km})$ (defined above) and calculate the *ordered* category totals $W_{[1]m} \leq W_{[2]m} \leq \cdots \leq W_{[k]m}$; also calculate

$$Z_m = \sum_{i=1}^{k-1} (1/\theta)^{(W_{[k]m} - W_{[i]m})}.$$

3. Stop sampling at the first stage when *either*

$$Z_m \leq \alpha/(1-\alpha) \quad or \quad m = n_0$$

$$or \quad W_{[k]m} - W_{[k-1]m} \geq n_0 - m,$$

whichever occurs first.

4. Let N (a random variable) denote the stage at which the procedure terminates. Select the design that yielded the largest W_{iN} as the one associated with the largest p_i (randomize in the case of ties).

Miller, Nelson, and Reilly (1998) present a remarkably efficient procedure that directly uses the original Y_{ij} observations (instead of the 0-1 X_{ij}, which lose information). Their procedure AVC, based on all possible vector comparisons of the observations, always results in an increased probability of correct selection when compared to the analogous implementation of the BEM procedure.

AVC

1. For the given k, and (α, θ) specified prior to the start of sampling, use the same n as in BEM.

2. Take a random sample of n observations $Y_{i1}, Y_{i2}, \ldots, Y_{in}$ from each alternative i, $i = 1, 2, \ldots, k$. Consider all n^k vectors of the form $\boldsymbol{Y}_j^{'} = (Y_{1j}^{'}, Y_{2j}^{'}, \ldots, Y_{kj}^{'})$, $j = 1, 2, \ldots, n^k$, where $Y_{ij}^{'}$ is one of the n observations from alternative i. Turn these into n^k (non-independent) multinomial observations, $\boldsymbol{X}_j^{'} = (X_{1j}^{'}, X_{2j}^{'}, \ldots, X_{kj}^{'})$, $j = 1, 2, \ldots, n^k$, 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 again assume that there are never ties for the maximum observation within a particular vector $\boldsymbol{Y}_j^{'}$.

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

5 COMPARISONS WITH A STANDARD

Example 3 *Several different investment strategies will be simulated to evaluate their expected rates of return. The strategy ultimately chosen may not be the one with the largest expected return—since factors such as risk could be considered—but none of the strategies will be chosen unless its expected return is larger than that of a conservative U.S. Government bond fund.*

Here the goal is to select the best investment strategy *only if it is better than the standard* (bond fund); if no strategy is better than the standard, we continue with the standard. More precisely, we have the following probability requirement: Denote the (known or unknown) expected value of the standard by μ_0 and the ordered means of the other investment strategies by $\mu_{[1]} \leq \mu_{[2]} \leq \cdots \leq \mu_{[k]}$. For specified constants α and δ with $\delta > 0$, we require

$$P\{\text{Select the standard}\} \geq 1 - \alpha \quad \text{whenever} \quad \mu_{[k]} \leq \mu_0$$

and

$$P\{\text{Select best strategy}\} \geq 1 - \alpha \quad \text{whenever}$$

$$\mu_{[k]} \geq \max\{\mu_0, \mu_{[k-1]}\} + \delta.$$

We now present a generic procedure for the problem of comparison with a standard due to Nelson and Goldsman (1997); this generic procedure can be used for a variety of situations (e.g., μ_0 known or unknown, unequal variances across systems, etc.), thus generalizing earlier results due to Bechhofer and Turnbull (1978).

Comparison with a Standard

1. Given k alternative systems and a standard (call it system 0), specify an initial first-stage sample size n_0, an indifference zone δ, and a confidence level $1 - \alpha$. Determine appropriate constants g and h, and let $c = \delta h / g$.

2. Generate a random sample $Y_{i1}, Y_{i2}, \ldots, Y_{in_0}$ from system i, for $i = 0, 1, 2, \ldots, k$.

3. Compute an appropriate variance estimator S_i^2 to be associated with system i.

4. Determine the required total sample size from system i as

$$N_i \;=\; \max\{n_0, \lceil (gS_i/\delta)^2 \rceil\}.$$

5. Take $N_i - n_0$ additional i.i.d. observations from system i if needed, and compute the overall sample mean $\bar{\bar{Y}}_i = \sum_{j=1}^{N_i} Y_{ij}/N_i$ for $i = 0, 1, 2, \ldots, k$.

6. With confidence level of at least $1 - \alpha$, apply the following rule:

* If $\max_{1 \le i \le k} \bar{\bar{Y}}_i \le \bar{\bar{Y}}_0 + c$, then choose the standard and form the one-sided confidence intervals

$$\mu_0 - \mu_i \;\le\; \bar{\bar{Y}}_0 - \bar{\bar{Y}}_i + c$$

for $i = 1, 2, \ldots, k$.

* Otherwise, choose the alternative associated with the largest sample mean $\max_{1 \le i \le k} \bar{\bar{Y}}_i$, and form the MCB intervals of the form given in Step 8 of Rinott+MCB for $i = 0, 1, 2, \ldots, k$.

6 FINAL THOUGHTS

Space limitations preclude detailed discussion, but we also mention the interesting technical results to be found in Damerdji, et al. (1997ab), Damerdji and Nakayama (1996), and Nakayama (1997), in which the authors rigorously show that certain selection-of-the-best procedures satisfy probability requirements similar to those in Section 4. Chen, et al. (1997) propose a completely different approach in their discussion on optimal budget strategies. Chick (1997) takes a decision-theoretic view on the selection-of-the best problem. And Boesel and Nelson (1998) use the techniques discussed in the current paper to present a methodology for optimization of stochastic systems.

ACKNOWLEDGMENTS

Portions of this article were published previously in Goldsman and Nelson (1994). Our work was supported by National Science Foundation Grants DMI-9622065 and DMI-9622269.

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. *Ann. Math. Stat.* 30:102–119.

Bechhofer, R. E., and D. M. Goldsman. 1986. Truncation of the Bechhofer-Kiefer-Sobel sequential procedure for selecting the multinomial event which has the largest probability (II): Extended tables and an improved procedure. *Comm. Stat.—Simul. and Comp.* B15:829–851.

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 and Sons.

Bechhofer, R. E., and B. W. Turnbull. 1978. Two $(k+1)$-decision selection procedures for comparing k normal means with a specified standard. *J. Amer. Stat. Assoc.* 73:385–392.

Boesel, J., and B. L. Nelson. 1998. Accounting for randomness in heuristic simulation optimization. Technical Report. Dept. of Industrial Engineering and Management Sciences, Northwestern Univ., Evanston, Illinois.

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

Chick, S. E. 1997. Selecting the best system: A decision-theoretic approach. In *Proc. 1997 Winter Simulation Conference*, ed. S. Andradóttir, K. J. Healy, D. H. Withers, and B. L. Nelson, 326–333. Piscataway, New Jersey: IEEE.

Damerdji, H., P. W. Glynn, M. K. Nakayama, and J. R. Wilson. 1997a. Selection of the best system in steady-state simulations. Technical Report 97–5, Dept. of Industrial Engineering, North Carolina State Univ., Raleigh, North Carolina.

Damerdji, H., P. W. Glynn, M. K. Nakayama, and J. R. Wilson. 1997b. Selection of the best system in transient simulations. Technical Report 97–6, Dept. of Industrial Engineering, North Carolina State Univ., Raleigh, North Carolina.

Damerdji, H. and M. K. Nakayama. 1996. Two-stage multiple-comparison procedures for steady-state simulations. Technical Report, Dept. of Industrial Engineering, North Carolina State Univ., Raleigh, North Carolina.

Dunnett, C. W. 1989. Multivariate normal probability integrals with product correlation structure. *Applied Statistics* 38:564–579. Correction: 42:709.

Gibbons, J. D., I. Olkin, and M. Sobel. 1977. *Selecting and Ordering Populations: A New Statistical Methodology*. New York: John Wiley and Sons.

Goldsman, D., and B. L. Nelson. 1994. Ranking, selection, and multiple comparisons in computer simulation. In *Proc. 1994 Winter Simulation Conference*, ed. J. Tew, M. S. Manivannan, D. A. Sadowski, and A. F. Seila, 192–199. Piscataway, New Jersey: IEEE.

Goldsman, D., and B. L. Nelson. 1998. Comparing systems via simulation. To appear in *Handbook of Simulation*, ed. J. Banks, Chapter 8. New York: John Wiley and Sons.

Goldsman, D., B. L. Nelson, and B. Schmeiser. 1991. Methods for selecting the best system. In *Proc. 1991 Winter Simulation Conference*, ed. B. L. Nelson, W. D. Kelton, and G. M. Clark, 177–186. Piscataway, New Jersey: IEEE.

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.

Hayter, A. J. 1984. A proof of the conjecture that the Tukey-Kramer multiple comparisons procedure is conservative. *Annals of Statistics* 12:61–75.

Hochberg, Y., and A. C. Tamhane. 1987. *Multiple Comparison Procedures*. New York: John Wiley and Sons.

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.

Miller, J. O., B. L. Nelson, and C. H. Reilly. Efficient multinomial selection in simulation. 1998. To appear in *Naval Research Logistics*.

Miller, R. G. 1981. *Simultaneous Statistical Inference*. 2nd ed. New York: Springer-Verlag.

Nakayama, M. K. 1997. Multiple-comparison procedures for steady-state simulations. *Annals of Statistics* 25:2433–2450.

Nelson, B. L., and D. Goldsman. 1997. Comparisons with a standard in simulation experiments. Technical Report. Dept. of Industrial Engineering and Management Sciences, Northwestern Univ., Evanston, Illinois.

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 Industrial Engineering and Management Sciences, Northwestern Univ., Evanston, Illinois.

Rinott, Y. 1978. On two-stage selection procedures and related probability-inequalities. *Comm. Stat.—Thy. and Meth.* A7:799–811.

Sullivan, D. W., and J. R. Wilson. 1989. Restricted subset selection procedures for simulation. *Operations Research* 37:52–71.

Wilcox, R. R. 1984. A table for Rinott's selection procedure. *J. Quality Technology* 16:97–100.

AUTHOR BIOGRAPHIES

DAVID GOLDSMAN is an Associate Professor in the School of Industrial and Systems Engineering at the Georgia Institute of Technology. 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 is an active participant in the Winter Simulation Conference, having been Program Chair in 1995.

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. He is interested in the design and analysis of computer simulation experiments, particularly statistical efficiency, multivariate output analysis, and input modeling. He was Program Chair of the 1997 Winter Simulation Conference.