Proceedings of the 2006 Winter Simulation Conference
L. F. Perrone, F. P. Wieland, J. Liu, B. G. Lawson, D. M. Nicol, and R. M. Fujimoto, eds.

RANKING AND SELECTION PROCEDURES FOR SIMULATION

Kirk C. Benson

Center for Army Analysis
6001 Goethals Street
Fort Belvoir, VA 22060-5230, U.S.A.

David Goldsman
Amy R. Pritchett

School of Industrial and Systems Engineering
Georgia Institute of Technology
Atlanta, GA 30332-0205, U.S.A.

ABSTRACT

We present sequential ranking and selection statistical procedures that determine the best simulated model configuration among competing alternatives. The best in this context denotes the largest expected value of a given performance metric. In order to run the procedures efficiently, we give algorithms using batched observations, which under certain conditions, exhibit the characteristics necessary for the appropriate application of ranking and selection procedures. We present empirical results that indicate that the sequential procedures are quite parsimonious, in terms of the number of required observations.

1 INTRODUCTION

The goal of any selection, screening, and multiple comparison problem is to determine the "best" of several competing configurations. In this context, a configuration implies that we have two or more competing systems that are compared by the mean value of some metric describing performance, where simulation is required to assess the value of this metric. Bechhofer et al. (1995) highlight several problem formulations appropriate to various experimental designs. Here, our focus is on the indifference-zone formulation where the objective is to select the configuration with the highest/lowest (interpreted "best") expected value. In this realm, an expectation offers insight on long-term system performance.

We describe various Ranking and Selection (RS) procedures, culminating in the fully sequential KN+ procedure (see Kim and Nelson 2001 and Goldsman et al. 2002). To demonstrate the efficacy of such procedures., we present an empirical performance analysis using independent, identically distributed (iid) normal processes and autoregressive processes. The RS procedures presented here outperform other procedures in terms of reducing the number of required observations necessary to discriminate between competing simulated system configurations. We also show how embedded statistical estimators enable easy application of the RS procedures.

This paper is organized as follows. Sections 2 and 3 present a brief background discussion on RS procedures and embedded statistical analysis, respectively. Section 4 discusses RS procedure performance analysis. Section 5 summarizes this effort.

2 RANKING AND SELECTION PROCEDURES

RS procedures may be single or multistage. In this context, a stage denotes the execution of a simulated configuration for a number of observations. A single-stage method determines the number of required observations from parameters specified by the experimenter. Adaptive control of an experiment is not possible with a single-stage RS procedure. However, a multistage RS procedure updates the required number of observations from the simulated configuration output, thereby enabling adaptive control of the comparison problem.

For the indifference-zone formulation, the practitioner provides the constants (δ^*, P^*). The quantity δ^* is the indifference-zone parameter, where the indifference zone indicates some comparative region in which the practitioner would not discriminate between competing configurations; here the indifference zone is defined as the set of mean vectors $\Omega \equiv \{(\mu_1, ..., \mu_k) : \mu_{[k]} - \mu_{[k-1]} < \delta^*\}$, where $\mu_{[1]} \le \mu_{[2]} \le \cdots \le \mu_{[k]}$ are the ordered means. The quantity P^* denotes the desired probability of correctly selecting (CS) the best of the competing configurations, when the means of those configurations fall outside of the indifference zone. So in other words, we want to satisfy the probability requirement $P(CS) \ge P^*$ whenever it is the case that $\mu_{[k]} - \mu_{[k-1]} \ge \delta^*$.

To highlight a single-stage RS procedure, suppose we are interested in determining which of *k* competing normal populations has the largest mean. Further suppose for now

that we make the additional strong assumption that the normal distributions all have the same *known* variance, σ^2. Then the classical method of Bechhofer (1954) is appropriate. In particular, this method determines the number of required observations, *n*, from each of the competing populations via the following formula:

$$n = \left\lceil 2\left(\frac{\sigma Z_{k-1,0.5}^{(1-P^*)}}{\delta^*} \right)^2 \right\rceil,$$

where the constant $Z_{k-1,\rho}^{(1-P^*)}$ is the $(1-P^*)$ equicoordinate point of the $k-1$ dimensional multivariate standard normal distribution with off-diagonal correlation ρ. Values for $Z_{k-1,\rho}^{(1-P^*)}$ may be obtained from table lookup from Bechhofer et al. (1995). For example, if the variance, σ^2, is known to be 2.25 and the experimenter sets $(\delta^*,P^*)=(0.306,0.95)$ with $k=6$ configurations, then 262 observations are needed for statistical comparison. Examples of two-stage and multistage methods follow.

If the variance of a predetermined metric is unknown, then Rinott's (1978) procedure provides a well-known two-stage technique for comparing configurations. This method relies on the assumptions that obtained data are independent, identically distributed, and from a normal distribution. Goldsman et al. (2002) present an extended version of this two-stage method (*R+*) and the extended version of the multistage Kim and Nelson (*KN+*) (2001) method. The following subsections give details.

2.1 Extended Rinott's Procedure (R+)

Rinott's procedure is conducted in two-stages. In the first stage, we take an initial sample of observations with the intent of estimating the variance parameters v_i^2, $i=1,2,\ldots,k$, of the systems (see Section 3); and these estimates determine the number of observations from each competing system that need to be taken in the second stage in order to satisfy the probability-of-correct-selection requirement. In this extended version of Rinott's procedure, the observations will eventually be divided into *batch means* (i.e., the sample averages from, say, *m* consecutive simulation observations from a particular system), which are assumed to be approximately normally distributed. See Section 3 for details on batching and variance estimation.

Setup: Select the confidence level $1-\alpha = P^*$, indifference-zone parameter $\delta^* > 0$, first-stage sample size $n_0 \geq 2$, and batch size $m < n_0$.

Initialization: Obtain Rinott's constant (from, e.g., Bechhofer et al. 1995) $h = h(d,k,1-\alpha)$, where *d* is the degrees of freedom for the necessary variance estimators (see Section 3), and *k* is the number of systems.

Obtain n_0 observations X_{ij}, $j=1,2,\ldots,n_0$, from each system $i=1,2,\ldots,k$. For $i=1,2,\ldots,k$, compute mV_i^2, the sample asymptotic variance of the data from system *i* using one of the variance estimators discussed in this paper (Section 3).

Calculate the total number of observations from system $i=1,2,\ldots,k$,

$$N_i = \max\left\{ n_0, \left\lceil \frac{h^2 m V_i^2}{(\delta^*)^2} \right\rceil \right\}.$$

Stopping Rule: If $n_0 \geq \max_i N_i$ then stop and select the system with the largest sample mean, $\overline{X}_i(n_0)$, as the best. Otherwise, take $N_i - n_0$ additional observations $X_{i,n_0+1}, X_{i,n_0+2}, \ldots, X_{i,N_i}$ from each system *i,* where $N_i > n_0$. Select the configuration with the largest overall sample mean, $\overline{X}_i(N_i)$, as the best.

2.2 Extended Kim and Nelson's Procedure (KN+)

The Kim and Nelson procedure KN+ is conducted in multiple stages. As in *R+,* the purpose of the initial stage is to obtain estimates for the variance parameters v_i^2, $i=1,2,\ldots,k$, of the competing systems. At the end of this and subsequent stages, any systems that are deemed as being inferior can be eliminated from future sampling and consideration. These variance estimates also help to determine bounds on the numbers of observations that can be taken from each system. As before, for purposes of variance estimation, the observations will eventually be divided into batches.

For two systems *i* and *l*, the asymptotic variance of the difference, $v_i^2 + v_l^2$, is estimated by forming the differenced series $D_{ilj} = X_{ij} - X_{lj}$, $j=1,2\ldots,$ and then applying a variance estimator (see Section 3).

Setup: Select the confidence level $1-\alpha$, indifference-zone parameter $\delta^* > 0$, first-stage sample size $n_0 \geq 2$, and batch size $m < n_0$. Calculate the constant as follows:

$$\eta = \frac{1}{2}\{[2(1-(1-\alpha)^{1/(k-1)}]^{-2/d} - 1\}.$$

Initialization: Let $I = \{1,2,\ldots,k\}$ be the set of systems still in contention, and let $h^2 = 2\eta d$, where d is the degrees of freedom for the necessary variance estimators (see Section 3).

Obtain n_0 observations $X_{ij}, j = 1,2,\ldots,n_0$, from each system $i = 1,2,\ldots,k$. For all $i \neq l$, compute mV_{il}^2, the sample asymptotic variance of the difference of systems i and l. Let

$$N_{il} = \left\lfloor \frac{h^2 m V_{il}^2}{(\delta^*)^2} \right\rfloor$$

and

$$N_i = \max_{l \neq i} N_{il}.$$

Here $N_i + 1$ is the maximum number of observations that can be taken from system i. If $n_0 \geq \max_i (N_i + 1)$, then stop and select the system with the largest sample mean $\overline{X}_i(n_0)$ as the best. Otherwise, set the observation counter $r = n_0$ and go to Screening.

Screening: Set $I^{old} = I$. Let

$$I = \{i : i \in I^{old} \text{ and } \overline{X}_i(r) \geq \overline{X}_l(r) - W_{il}(r), \text{ for all } l \in I^{old}, l \neq i\}$$

where

$$W_{il}(r) = \max\left\{0, \frac{\delta}{2r}\left(\frac{h^2 m V_{il}^2}{(\delta^*)^2} - r\right)\right\}.$$

Stopping Rule: If the cardinality $|I| = 1$, then stop and select the system whose index is in I as the best. Otherwise, take one additional observation $X_{i,r+1}$ from each system $i \in I$ and set $r = r + 1$. If $r = \max_i N_i + 1$, then stop and select the system whose index is in I and has the largest $\overline{X}_i(r)$ as the best. Otherwise, repeat the Screening process. Note that variance estimation only depends on data collected in the initialization stage of this procedure.

3 EMBEDDED STATISTICAL ANALYSIS

RS procedures often require calculations on both individual and batched observation data. Note that in the presence of autocorrelated data, all of the popular variance estimators from the literature are typically biased. The remainder of this section details methods for obtaining "good" estimators for the variance parameter of a stationary stochastic process.

Suppose that $X_{i1}, X_{i2}, \ldots$ is the simulation output stream from a single replication of the i^{th} alternative, e.g., X_{ij} could denote the j^{th} customer's waiting time in a simulation of the i^{th} system. Then after appropriate initialization (to eliminate simulation start-up bias), we might feel comfortable making the following assumptions:

Stationarity: $X_{i1}, X_{i2}, \ldots$ form a stationary stochastic process.

(Strong) Consistency: $\overline{X}_i(r) \to \mu_i$ as $r \to \infty$ with probability one, where μ_i is the steady-state mean from system i.

Functional Central Limit Theorem (FCLT): There exist constants μ_i and $v_i^2 > 0$ such that

$$\frac{\sum_{j=1}^{\lfloor rt \rfloor}(X_{ij} - \mu_i)}{\sqrt{r}} \Rightarrow v_i W(t).$$

for $0 \leq t \leq 1$, where $\Rightarrow$ denotes weak convergence as $r \to \infty$ and $W(t)$ is a standard Brownian motion (Weiner) process (cf. Glynn and Iglehart 1990).

In the current paper, we make comparisons based on the steady-state means $\mu_1, \mu_2, \ldots, \mu_k$, which is reasonable due to the consistency assumption. The variance parameter, $v_i^2 \equiv \lim_{r \to \infty} rVar\left(\overline{X}_i(r)\right)$, can be estimated by the well-known methods of batch means (BM) or overlapping batch means (OBM), among others (see Law 2006). We proceed with a description of these methods.

3.1 Batch Means

If n observations from system i, $X_{i1}, X_{i2}, \ldots, X_{in}$, are divided into b batches of length m, then the j^{th} batch mean from system i is:

$$\overline{X}_{i,j,m} \equiv \frac{1}{m}\sum_{p=1}^{m} X_{i,(j-1)m+p}.$$

The observations $X_{i,(j-1)m+1}, X_{i,(j-1)m+2}, \ldots, X_{i,jm}$ comprise the j^{th} batch, $j = 1,2,\ldots,b$, for system i. For $b > 1$, the batch means estimator for the variance parameter v_i^2 from system i is:

$$mV_B^2 \equiv \frac{m}{b-1}\sum_{j=1}^{b}\left(\overline{X}_{i,j,m} - \overline{X}_i(n)\right)^2 \xrightarrow{D} \frac{v_i^2\chi_{b-1}^2}{b-1},$$

where χ_d^2 is a chi-squared random variable with $d = b-1$ degrees of freedom and $\xrightarrow{D}$ indicates convergence in distribution as m becomes large (see, e.g., Glynn and Whitt 1990).

3.2 Overlapping Batch Means

The method of overlapping batch means re-uses observations in multiple batches, resulting in an estimator for the variance parameter that has lower variance than that of the regular batch means method. To this end, consider all batch means of the form:

$$\overline{X}_i(j,m) \equiv \frac{1}{m}\sum_{p=0}^{m-1} X_{i,j+p} .$$

The observations $X_{i,j+1}, X_{i,j+2}, \ldots, X_{i,j+m-1}$ from system i comprise the j^{th} overlapping batch, for $i = 1,2,\ldots,k$ and $j = 1,2,\ldots,n-m+1$. The OBM variance estimator is:

$$mV_O^2 \equiv \frac{nm}{(n-m+1)(n-m)}\sum_{j=1}^{n-m+1}\left(\overline{X}_i(j,m) - \overline{X}_i(n)\right)^2 .$$

Note that $mV_O^2 \approx \frac{v_i^2\chi_d^2}{d}$, where $d = \lfloor 3(b-1)/2 \rfloor$ degrees of freedom (see Meketon and Schmeiser 1984).

4 COMPARING RANKING AND SELECTION PROCEDURES

This section focuses on the testing and comparison of RS procedures.

4.1 Assumptions and Goals

The RS procedures in this paper obtain observations, $X_{ij}, j = 1,2,\ldots,$ for competing system configurations $i = 1,2,\ldots,k$. We assume that, for a specific system *i,* the observations can be batched, so that we can obtain approximately independent and normally distributed batch means, at least for a large enough batch size. We place no restrictions on the variance parameters of the systems.

4.2 Assessing Procedure Performance

The relative difference between competing system configurations directly impacts RS procedure performance. When evaluating the performance of a particular procedure, it is a good idea to see how well the procedure does under various special configurations of the means. Of interest are the so-called slippage configuration (SC),

$$\mu_{[1]} = \mu_{[k-1]} = \mu_{[k]} - \delta^*$$

(often also referred to as the least favorable (LF) configuration), the equal means (EM) configuration,

$$\mu_{[1]} = \mu_{[k]},$$

and the equally spaced (ES) configuration,

$$\mu_{[i]} = \mu_{[1]} + (i-1)\delta^*$$

(when comparing the performance of procedures, we can assume, with little loss of generality, that the smallest mean $\mu_{[1]} = 0$). The SC is of interest because, in terms of the ability to decide between systems, it can be thought of as the "most difficult" configuration of means that satisfy the conditions of the probability requirement. The EM configuration is difficult in the sense that it might require a sequential procedure a great deal of time to make a decision as to which system is the best, even though there is no "wrong answer". And the ES configuration is interesting in that it is a "favorable" configuration that ought to allow for easy determination of the best system.

The user-specified choice of the indifference-zone parameter δ^* also impacts RS procedure performance. Recall that δ^* is a comparative region where the experimenter would not discriminate between competing system configurations. If δ^* is "too small", then the number of required observations can be prohibitively high; if δ^* is "too large", then it is difficult to make a "useful" differentiation between systems. The desired probability of correct selection, P^*, also affects procedure performance in an obvious way — the higher the desired *P(CS)*, the more observations will likely be required.

4.3 Examples: Comparing R+ and KN+ Procedure Performance

We illustrate the performance of the R+ and KN+ procedures with two simple examples.

4.3.1 Independent normal observations

In our first experiment, we varied the initial number of observations, n_0. Parameterization for this experiment includes $k = 6$ underlying independent normal processes in the slippage configuration with $\mu_1 = \mu_{k-1} = 0$, $\mu_k = 0.306$, all with variance 1, and each producing independent and identically distributed observations. We took batch size $m = 1$, indifference-zone parameter $\delta^* = 0.306$, and a desired probability of correct selection of $P^* = 0.95$. Here, the "best" system configuration is $\mu_6 = 0.306$. Performance metrics are estimators of (1) the probability of correct selection, $\hat{P}(CS)$, and (2) the average number of required raw (unbatched) observations, $\hat{\bar{T}}$, obtained from 1000 independent experiment replications. Since the KN+ method is multistage, the upper bound on the number of required unbatched observations (i.e., the largest *possible* number of observations), determined at the end of the first stage, is also reported.

As shown in Table 1, the KN+ method requires fewer total raw observations, $\hat{\bar{T}}$, than the R+ method except when n_0 is large (where both methods require the same number). The achieved $\hat{P}(CS)$ is statistically equivalent to or exceeds the desired probability, P^*, in all cases. Also, a large number of initial observations creates computational inefficiency, i.e., a large total observation requirement, in both methods. While the upper bound for required observations for the KN+ method is always larger than that for the R+ method, the screening/elimination process within the KN+ method allows for increased computational efficiency.

4.3.2 Using Batched Data

What if the underlying data are not iid normal, i.e., non-normal and/or serially correlated? Then batching methods such as BM or OBM (among others) might be appropriate to the situation at hand. In this example, we applied batching to a first-order autoregressive [AR(1)] process — a process that produces autocorrelated observations, with the correlations decaying exponentially with the time lag.

Table 1: R+ and KN+ Comparison Varying Initial Number of Observations

$\delta^* = 0.306$, $P^* = 0.95$, SC, iid normal case, $m = 1$			
n_0	Method	$\hat{P}(CS)$	$\hat{\bar{T}}$ / Upper Bound
8	R+	0.982	420
	KN+	0.990	277/928
20	R+	0.952	292
	KN+	0.968	162/533
40	R+	0.967	274
	KN+	0.992	149/468
60	R+	0.976	258
	KN+	0.957	137/420
80	R+	0.948	262
	KN+	0.967	141/428
100	R+	0.951	262
	KN+	0.968	143/419
150	R+	0.955	261
	KN+	0.969	168/405
250	R+	0.963	262
	KN+	0.969	252/395
400	R+	1.000	400
	KN+	1.000	400

If we denote the mean for system *i* by μ_i, an AR(1) process generates each observation X_{ij}, $j = 1,2,\ldots$ for competing system configurations, $i = 1,2,\ldots,k$, from the relationship:

$$X_{i,j} = \mu_i + \phi(X_{i,j-1} - \mu_i) + Z_{i,j},$$

where, to preserve stationarity, the error terms, $Z_{i,j}$, are distributed iid $N(0,1-\phi^2)$, the $X_{i,1}$'s, $i = 1,2,\ldots,k$, are initialized as $N(\mu_i,1)$, and we must have $-1 < \phi < 1$. In that case, the autocovariance function is $R_k \equiv Cov(X_{i,j}, X_{i,j+k}) = \phi^k$, for $k = 0,1,2,\ldots$. It can easily be shown that the variance parameter for this system turns out to be

$$v_i^2 = \frac{1+\phi}{1-\phi},$$

which of course would be unknown in a real-world, practical application.

The BM method obtains batched observations of size *m*. The number of initial batches may be obtained from the

relationship $b_0 = \lceil n_0 / m \rceil$ where n_0 is the number of initial unbatched or raw observations.

Table 2 presents the experimental results for the R+ and KN+ methods applied to an AR(1) process while obtaining observations with the BM method. Experiment parameterization involved setting $\phi = 0.22$, $n_0 = 4200$, $P^* = 0.95$, $k = 6$ competing system configurations, and $\delta^* = 0.0193 = \sqrt{v_i^2 / 4200}$ (a reasonable test value for purposes of this example), while varying both the batch size, *m*, and the initial number of batches b_0. The required raw observation upper bound for the KN+ method is also reported. Intuitively, as b_0 decreases, the number of required unbatched or raw observations, $\hat{\bar{T}}$, increases. This follows since the variance estimator is based on a χ^2 distribution with $b_0 - 1$ degrees of freedom, and thus the variance of that distribution is high for a low b_0. This is consistent with results found in the iid case. Of special interest is the relatively poor performance, in terms of achieving the desired probability of correct selection, of the R+ method when the batch size is small. This can be attributed to fact that the autocorrelated process has caused the distribution of the BM estimator to be something other than the limiting χ^2 distribution. Note that, in this experiment, the KN+ method is not as susceptible to poor $\hat{P}(CS)$ performance as the R+ method. In addition, KN+ requires far fewer raw observations due to its screening process.

Similarly, Table 3 presents the experimental results for the R+ and KN+ methods applied to an AR(1) process while obtaining batched observations using the OBM method. Parameterization for this experiment includes $\phi = 0.22$, $n_0 = 8400$, $P^* = 0.95$, $k = 6$ competing system configurations, and $\delta^* = 0.0136 = \sqrt{v_i^2 / 8400}$. The table shows the effects of varying batch size, *m*; note that in the context of OBM, the ratio quantity $b_0 = n_0 / m$, is still meaningful, but can no longer be interpreted as "the number of batches".

This experiment highlights the necessity of asymptotic variance convergence when using the R+ and KN+ RS procedures. Such convergence for the OBM variance estimator in the case of an AR(1) process is obtained once we have both sufficiently large *m* and b_0. In fact, the performance of the RS methods is poor when these quantities are not sufficiently large. Observe that the estimated $\hat{P}(CS)$ is nominally achieved with a large n_0 / m ratio, implying the necessity for an increase in the number of initial unbatched observations.

Table 2: R+ and KN+ Comparison Using Batch Means while Varying Batch Size

$\delta^* = 0.0193$, $P^* = 0.95$, $n_0 = 4200$, SC, AR(1), $\phi = 0.22$				
	R+		**KN+**	
(m, b_0)	$\hat{P}(CS)$	$\hat{\bar{T}}$	$\hat{P}(CS)$	$\hat{\bar{T}}$ / Upper Bound
(10,420)	0.887	43244	0.968	20912 / 64676
(25,168)	0.940	44861	0.964	22834 / 69939
(50,84)	0.936	45917	0.972	23005 / 73263
(100,42)	0.952	46825	0.976	25219 / 82739
(150,28)	0.976	51321	0.955	30191 / 94992
(200,21)	0.956	51826	0.974	29919 / 99284
(300,14)	0.941	53732	0.962	33371/110908

Table 3: R+ and KN+ Comparison Using Overlapping Batch Means while Varying Batch Size, $n_0 = 8400$

$\delta^* = 0.0136$, $P^* = 0.95$, $n_0 = 8400$, SC, AR(1), $\phi = 0.22$				
	R+		**KN+**	
(m, b_0)	$\hat{P}(CS)$	$\hat{\bar{T}}$	$\hat{P}(CS)$	$\hat{\bar{T}}$ /Upper Bound
(10, 840)	0.951	86819	0.953	41872/127280
(25, 336)	0.935	89075	0.952	43514/132585
(50, 168)	0.944	89861	0.933	42938/135185
(100, 84)	0.933	88517	0.956	43259/135788
(150, 56)	0.944	87038	0.942	42403/134688
(200, 42)	0.914	85283	0.939	41699/135308
(300, 28)	0.916	85254	0.913	42456/137350
(400, 21)	0.938	86436	0.925	41390/138420
(600, 14)	0.928	82064	0.924	40433/136469

5 SUMMARY

Ranking and selection procedures enable differentiation between competing simulated system configurations. They are simple to use, especially considering the fact that batching methods allow transformation of correlated data into normal observations under certain conditions. For more mathematical details as well as numerous extensions and augmentations of the elementary procedures discussed herein, see for example, Kim and Nelson (2006).

ACKNOWLEDGMENTS

This work was partly funded by the NASA Ames Research Center, under Grant NAG-2-1291, with Irv Statler and Mary Conners administering and serving as technical points of contact. Partial support was also provided by National Science Foundation grant DMI-0400260.

REFERENCES

Bechhofer, R. E. 1954. A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances. *Ann. Math. Stat.,* 25:16-39.

Bechhofer, R. E., T. J. Santner, and D. M. Goldsman. 1995. *Design and Analysis of Experiments for Statistical Selection, Screening, and Multiple Comparisons*. New York: John Wiley & Sons, Inc.

Glynn P. W., and D. L. Iglehart. 1990. Simulation Output Analysis Using Standardized Time Series. *Mathematics of Operations Research,* 15:1-16.

Glynn P. W., and W. Whitt. 1990. Estimating the Asymptotic Variance with Batch Means. *Operations Research Letters,* 10:431-435.

Goldsman, D., Kim, S., Marshall, W. S., and B. L. Nelson. 2002. Ranking and Selection for Steady-State Simulation: Procedures and Perspectives. *INFORMS Journal on Computing,* 14(1):2-19.

Kim, S., and B. Nelson. 2001. A Fully Sequential Procedure for Indifference-Zone Selection in Simulation. *ACM TOMACS*, 11:251-273.

Kim, S., and B. Nelson. 2006. Selecting the Best System, *Handbooks in Operations Research and Management Science: Simulation* (ed. S. G. Henderson and B. L. Nelson), Chapter 17, Oxford: Elsevier Science.

Law, A. M. 2006. *Simulation Modeling and Analysis*. 4th Edition. Boston, MA: McGraw-Hill.

Meketon, M. S., and B. W. Schmeiser. 1984. Overlapping Batch Means: Something for Nothing? In *Proceedings of the 1984 Winter Simulation Conference*, ed. S. Sheppard, U. W. Pooch, and C. D. Pegden, 227-230, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers.

Rinott, Y. 1978. On Two-Stage Selection Procedures and Related Probability-Inequalities. *Commun. Stat. – Theory and Methods,* A8:799-811.

AUTHOR BIOGRAPHIES

KIRK C. BENSON is a lieutenant-colonel in the U.S. Army. Currently, he is forward deployed from the Center for Army Analysis to Afghanistan. He received a B.S. from the United States Military Academy, an M.S. degree from the Naval Postgraduate School, and a Ph.D. from the School of Industrial and Systems Engineering at the Georgia Institute of Technology.. His research interests include simulation, statistical selection techniques, and distributed computing methods. His e-mail address is <kirk@bensco.com>.

DAVID GOLDSMAN is a 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 also studies applications arising in the healthcare field. Dave has been an active participant in the Winter Simulation Conference; he is currently on the Board of Directors, and was the 1995 Program Chair and 1992 Associate *Proceedings* Editor. His e-mail and web addresses are <sman@gatech.edu> and <http://www.isye.gatech.edu/~sman>.

AMY R. PRITCHETT is an Associate Professor in the Schools of Industrial and Systems Engineering and Aerospace Engineering at the Georgia Institute of Technology. She is an area editor of *SIMULATION* for the areas of air traffic and aviation. She received her S.B., S.M., and Sci.D. degrees from the Department of Aeronautics and Astronautics at the Massachusetts Institute of Technology. Her research specialties include cockpit design, air traffic control, flight simulation, and large-scale agent-based simulation of hybrid systems. Her e-mail and web addresses are <amy.pritchett@isye.gatech.edu> and <http://www.isye.gatech.edu/people/faculty/Amy_Pritchett/>.