Proceedings of the 1990 Winter Simulation Conference
Osman Balci, Randall P. Sadowski, Richard E. Nance (eds.)

BATCH-SIZE EFFECTS ON SIMULATION OPTIMIZATION USING MULTIPLE COMPARISONS WITH THE BEST

Lynne Goldsman
Barry L. Nelson

Department of Industrial and Systems Engineering
The Ohio State University
Columbus, Ohio 43210

ABSTRACT

Multiple comparisons with the best (MCB) is one approach for selecting the best of a finite number of system designs via simulation. However, there is currently no MCB procedure for steady-state simulation experiments. This paper proposes a procedure based on batching and derives the basic theory needed to design a batching algorithm.

1. INTRODUCTION

In practice, stochastic simulation is invariably used to compare alternative system designs and select a good, or ideally the best, design. Because stochastic simulations are sampling experiments, selection or *optimization* of system design must be done in the presence of sampling error, meaning the best design can never be determined with certainty. When the number of alternative designs is large and the simulation model is complex, the simulation experiment may be computationally expensive. However, the most significant cost is often the cost of unknowingly selecting an inferior system design. Thus, there is a need for statistically efficient methods for optimization via simulation.

Optimization via simulation is currently an area of intense research interest. Much of this research concentrates on systems where the design or decision variables are continuous, and system performance measures are (at least assumed to be) continuous functions of the decision variables. Developments in this area include perturbation analysis, likelihood-ratio methods, importance-sampling methods, frequency-domain methods, stochastic approximation and response-surface methods. Some of these methods may be adapted to discrete-decision-variable problems, but they are most naturally applied when the variables are continuous.

In many practical problems, however, there are a finite number of alternative system designs; these designs may arise from combinations of equipment, schedules or facilities, subject to constraints on the available budget or technology (several examples are listed below). Research on optimizing this class of problems has concentrated on the methodology of ranking and selection. Hsu and Nelson [1988] demonstrated that *multiple comparisons with the best (MCB)* is a viable alternative to ranking and selection. MCB is one of a family of multiple-comparison procedures that are applicable in single-stage experiments and provide simultaneous inference about relationships among all systems. However, there are unsolved problems limiting the usefulness of multiple-comparison procedures in general simulation experiments. One of these problems is extending multiple-comparison procedures to the class of steady-state simulation experiments. We propose a solution to this problem based on batching.

2. BACKGROUND

In his survey of optimization via simulation, Glynn [1986] classified optimization problems based on their decision space. Within his classification scheme, this paper addresses problems with a finite-dimensional, discrete decision space, in which the number of possible decisions (system designs) is also finite. Many practical problems are of this type. Examples include, but are not restricted to:

1. A combination of robots and material-handling equipment is to be chosen to form a flexible manufacturing system that maximizes the expected production rate, subject to a budget constraint on the equipment purchased.
2. Similar to the previous example, a computer system is to be designed by choosing components from different manufacturers to minimize expected response time, subject to a budget constraint.
3. A limited amount of buffer space is to be allocated between work stations in an assembly system in order to maximize expected throughput. If buffer space is measured in units of number of items stored, then there are a finite number of possible system designs.
4. One of a number of possible schedules is to be chosen for a job shop to minimize the expected total tardiness of all jobs due that week. The possible schedules may result from applying a number of scheduling heuristics.
5. A reorder point, s, and a stock level, S, are to be chosen to minimize expected cost per period for an inventory system containing discrete items. If there is a physical upper limit on the stock level, then there are a finite number of (s, S) combinations.

Unless some relationship among the system designs is known *a priori*, any optimization procedure for this class of problems must simulate all candidate system designs to guarantee a chance of selecting the best design. We propose a method for the most general case where no such relationships are known in advance.

For exposition, assume that k competing systems are to be compared in terms of their expected performance, and denote the expected performance of the ith system by θ_i, $i = 1, 2, \ldots, k$. The goal is to determine which design has the maximum expected performance. The statistical methods of ranking and selection have often been suggested to solve this problem.

Ranking and selection procedures treat the optimization problem as a decision problem, typically either deciding which system design is the best (indifference-zone ranking) or on a subset of system designs that contains the best design (subset selection). The decisions are guaranteed to be correct with a prespecified probability. Achieving this goal often requires two-stage sampling, which means restarting simulation experiments after initial runs

of all systems. In addition, ranking and selection inference is typically limited to the system or subset of systems selected as best. A summary of the many ranking and selection procedures is given by Gupta and Panchapakesan [1979]. Extensions of ranking and selection procedures specifically to stochastic simulation include Clark and Yang [1986], Goldsman [1985], Iglehart [1977], Koenig and Law [1985] and Sullivan and Wilson [1989]. Some of these papers address ranking and selection in steady-state simulation experiments; Clark and Yang combine variance reduction with ranking and selection.

Multiple-comparison procedures, on the other hand, treat the optimization problem as an inference problem on the parameters of interest. Suppose that larger expected performance implies a better system. For system i, the quantity $\theta_i - \max_{j\neq i}\theta_j$ can be termed *system i performance minus the best of the other systems' performance*. In optimization problems, the parameters $\theta_i - \max_{j\neq i}\theta_j$, for $i = 1,\ldots,k$, are often the quantities of primary interest. This can be seen as follows: If $0 < \theta_i - \max_{j\neq i}\theta_j$, then system i is the best, for it is better than the best of the other systems. If $\theta_i - \max_{j\neq i}\theta_j < 0$, then system i is not the best, since there is another better system. Even if $\theta_i - \max_{j\neq i}\theta_j < 0$, if $-\Delta < \theta_i - \max_{j\neq i}\theta_j$, where Δ is a positive number, then system i is within Δ of the best. Simultaneous statistical inference on $\theta_i - \max_{j\neq i}\theta_j$, for $i = 1,\ldots,k$, is termed multiple comparisons with the best (MCB).

An important property of multiple-comparison procedures is that inference about the relative performance of all systems is provided. In fact, Hsu and Nelson [1988] show that MCB inference implies both the indifference zone and subset selection inference of ranking and selection. Such inference may be critical if the performance measure of interest does not account for differences in the cost or convenience of alternative designs; e.g., expected throughput of a manufacturing system does not account for the cost of maintaining the system. In such cases an inferior system design, in terms of expected performance, may still be the one to choose if the difference in performance relative to the optimal design is small. A second property is that multiple-comparison procedures can be implemented in a single-stage of sampling, which is important if restarting the simulation experiment is cumbersome or expensive.

Let the output of the simulation of system design i be denoted by $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$, $i = 1,2,\ldots,k$. In simulation research and practice there are two important cases:

1. For system i, $Y_{i\ell}$, $\ell = 1,2,\ldots,n_i$ are independent and identically distributed (i.i.d.) random variables with expectation $\theta_i = \mathrm{E}[Y_{i\ell}]$. This case would typically arise in a *terminating* simulation experiment. The natural experiment design is to generate i.i.d. replications of the process, yielding i.i.d. observations of system performance $Y_{i\ell}$, $\ell = 1,2,\ldots,n_i$.

2. For system i, $Y_{i\ell} \Longrightarrow \mathcal{Y}_i$ as $\ell \longrightarrow \infty$, and $\theta_i = \mathrm{E}[\mathcal{Y}_i]$. This case would typically arise in a *steady-state* simulation experiment. The natural experiment design is to generate a single (long) replication of the process. However, the outputs $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$ from a single replication may be neither independent, due to the nature of the stochastic process, nor identically distributed, due to the choice of the initial state.

This paper derives and analyzes an MCB procedure for the steady-state simulation case. MCB, in the form it might be used in terminating simulations, is described next.

Suppose that we can represent the simulation output as

$$Y_{i\ell} = \theta_i + \epsilon_{i\ell}, \tag{1}$$

for $\ell = 1,\ldots,n_i$ and $i = 1,\ldots,k$, where $\epsilon_{11},\ldots,\epsilon_{kn_k}$ are i.i.d. normal random variables with mean 0 and unknown variance σ^2. This model might be tenable in terminating simulations when $Y_{i\ell}$ is a summary statistic from the ℓth replication of system design i. Let

$$\begin{aligned}\bar{Y}_i &= n_i^{-1}\sum_{\ell=1}^{n_i} Y_{i\ell}\\ S^2 &= k^{-1}\sum_{i=1}^{k}(n_i-1)^{-1}\sum_{\ell=1}^{n_i}(Y_{i\ell}-\bar{Y}_i)^2\end{aligned}$$

be the sample mean for the ith system and the pooled sample variance, respectively. Let $x^- = \min\{x,0\}$ and $x^+ = \max\{x,0\}$. Hsu [1984] showed that, under model (1), the closed intervals

$$\left[D_i^-, D_i^+\right] \tag{2}$$

where

$$\begin{aligned}D_i^+ &= \left(\min_{j\neq i}\left\{\bar{Y}_i - \bar{Y}_j + d_i^\alpha S\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\right\}\right)^+\\ D_i^- &= \left(\min_{\substack{j\in\mathcal{G}\\ j\neq i}}\left\{\bar{Y}_i - \bar{Y}_j - d_i^\alpha S\sqrt{\frac{1}{n_i}+\frac{1}{n_j}}\right\}\right)^-\\ \mathcal{G} &= \{j : D_j^+ > 0\}\end{aligned}$$

for $i = 1,\ldots,k$, form a set of $(1-\alpha)100\%$ simultaneous confidence intervals for $\theta_i - \max_{j\neq i}\theta_j$ given appropriate constants d_i^α (if the set $\mathcal{G} = \{i\}$ then D_i^- is defined to be 0). Result (2) is the foundation for our research. Examples in Hsu and Nelson [1988] and Yang and Nelson [1989a] show that the MCB intervals are easy to interpret for decision making.

3. MULTIPLE COMPARISONS FOR STEADY-STATE SIMULATION

When system designs are compared on the basis of long-run performance, problems related to the design and analysis of steady-state simulation experiments arise. Standard multiple-comparison procedures, which are based on observing i.i.d. replications of system performance, can be directly adapted to steady-state simulation by employing an experiment design specifying replications. However, because of the still unsolved initial-condition bias problem, single-replication designs are often the only practical designs (initial-condition bias is typically minimized by allocating the entire computation budget to one replication). In the subsections below we derive and analyze an MCB procedure for steady-state simulation experiments employing a single replication for each system design.

Assuming that initial-condition bias has been mitigated by making a single replication and taking appropriate remedial measures (such as deletion), the fundamental output analysis problem is estimating the $\mathrm{Var}[\bar{Y}_i]$. This is a long-standing problem in simulation research which arises because the outputs from a single replication are typically dependent with unknown dependence structure.

The problem is more difficult in the context of multiple com-

parisons because variance estimators are required for all k systems and they must have suitable properties for deriving multiple-comparison inference. Specifically, if the point estimators are the sample means from the simulation of each system—$\bar{Y}_1, \ldots, \bar{Y}_k$ as defined above—then exact MCB inference can be derived if the joint distribution of the estimators is multivariate normal

$$\begin{pmatrix} \bar{Y}_1 \\ \vdots \\ \bar{Y}_k \end{pmatrix} \sim N\left[\begin{pmatrix} \theta_1 \\ \vdots \\ \theta_k \end{pmatrix}, \sigma^2 \begin{pmatrix} \delta_1 & 0 & \cdots & 0 \\ 0 & \delta_2 & \cdots & 0 \\ & & \ddots & \\ 0 & \cdots & 0 & \delta_k \end{pmatrix} \right] \quad (3)$$

with $\delta_1, \delta_2, \ldots, \delta_k$ known, and an estimator of the unknown parameter σ^2 is available that is independent of $\bar{Y}_1, \ldots, \bar{Y}_k$ and (appropriately scaled) has a chi-squared distribution.

For model (1), σ^2 is the common variance of the observations $Y_{i\ell}$, $\delta_i = 1/n_i$, and S^2 is an independent estimator of σ^2 that is distributed as $\sigma^2\chi^2_{df}/df$, where $df = \sum_{i=1}^{k}(n_i - 1)$. In steady-state simulation, where the $Y_{i\ell}, \ell = 1, 2, \ldots, n_i$, are dependent for fixed i, there is no unique interpretation for the parameters σ^2 and δ_i. The batch means method described below provides an interpretation for σ^2 and δ_i and a way to estimate σ^2.

3.1 Batching

Batching or "batch means" is a well-known method for estimating the variance of the sample mean of a stationary output process (e.g., Schmeiser [1982]). In numerous studies this method has been shown to be robust and dependable (e.g., Sargent, et al. [1987]).

The idea behind batching is to transform the dependent output process from system design i, $Y_{i\ell}, \ell = 1, 2, \ldots, n_i$, into a (nearly) independent and normally distributed batch-means process. Let the jth *batch mean* from system i be defined as

$$\bar{Y}_{ij}(m_i) = b_i^{-1} \sum_{\ell=(j-1)b_i+1}^{jb_i} Y_{i\ell}, \quad (4)$$

for $j = 1, 2, \ldots, m_i$, where $n_i = m_i b_i$; the integers b_i and m_i are called the *batch size* and the *number of batches*, respectively. For convenience, we assume from here on that the number of observations from each system is the same, n, but the number of batches, m_i (and the size of each batch, b_i), may be different for different systems. If m_i is small enough, then it is hoped that the following model approximately describes the batch means (4) from the simulation of system i:

$$\bar{Y}_{ij}(m_i) = \theta_i + \bar{\epsilon}_{ij}(m_i), \quad (5)$$

for $j = 1, 2, \ldots, m_i$, where the $\bar{\epsilon}_{ij}(m_i)$, are (nearly) independent and identically normally distributed. Let $\sigma_i^2(m_i) = \mathrm{Var}[\bar{\epsilon}_{ij}(m_i)]$, the variance of the batch means from system i when the number of batches is m_i.

Selecting a number of batches m_i to achieve approximate independence and normality has been a longstanding research problem. In the multiple-comparison context, we propose to select the number of batches for each system, $m_i, i = 1, 2, \ldots, k$, to satisfy the additional condition that $\sigma_1^2(m_1) \approx \sigma_2^2(m_2) \approx \cdots \approx \sigma_k^2(m_k)$; that is, we chose the m_i so that the variances of the batch means from all systems are nearly equal. This transforms a problem with (possibly) unequal variances into a problem with equal variances but unequal sample sizes. Thus, $\delta_i = 1/m_i$ and $\sigma^2 = \sigma_i^2(m_i)$ in (3). The MCB intervals follow directly from (2).

In summary, we propose to use batching to obtain approximately independent and normally distributed batch means that have *approximately equal variances across system designs*, and then to apply the standard MCB procedure to model (5).

The batching approach has one major advantage over other methods: Estimators of $\mathrm{Var}[\bar{Y}_i]$ from many standard methods do not lend themselves to deriving a pooled variance estimator with well-defined degrees of freedom. The proposed batching approach facilitates using a standard pooled variance estimator (based on the batch means) with $\sum_{i=1}^{k}(m_i - 1)$ degrees of freedom.

The next subsection presents the basic research necessary to design a batching algorithm to implement this MCB procedure. In particular, batch-size effects on the probability of correct and useful inference (defined below) are derived so that a batching algorithm can search for an appropriate number of batches within a range that balances the trade-off between degrees of freedom (and thus sharp inference) and the need to have nearly independent batch means. This analysis is analogous to the batch-size effects examined by Schmeiser [1982], Yang and Nelson [1988, 1989b] and Nelson [1989] in other contexts, but is somewhat more difficult because of the performance criterion we consider.

3.2 Batch-Size Effects

Multiple-comparison procedures construct simultaneous confidence intervals for selected differences in expected system performance. If $\mathcal{C}$ is the event that the intervals simultaneously contain all of the selected differences, then $\Pr\{\mathcal{C}\} = 1 - \alpha$ when the assumptions underlying the procedure are satisfied. The event $\mathcal{C}$ could be called correct inference, since the procedure correctly identifies a region that contains all of the parameters. Let $\mathcal{U}$ be the event that the intervals exclude zero when the true difference is not zero. The event $\mathcal{U}$ could be called useful inference, since the procedure distinguishes differences in expected performance. Given two multiple-comparison procedures for the same estimation problem that both have $\Pr\{\mathcal{C}\} = 1 - \alpha$, the superior procedure is the one with larger $\Pr\{\mathcal{C} \cap \mathcal{U}\}$; that is, the one that has a larger probability of *correct and useful inference*. Notice that $\Pr\{\mathcal{C} \cap \mathcal{U}\} \leq 1 - \alpha$.

The event $\mathcal{C} \cap \mathcal{U}$ implies that differences, and the direction of the differences, are correctly identified. For MCB, this means that the true best system design is identified as the best. In this section we examine the effect of the number of batches (equivalently batch size) on $\Pr\{\mathcal{C} \cap \mathcal{U}\}$. We are looking for a range in which the performance of MCB *is not sensitive to the number of batches*, m_i. Thus, nothing is lost if we assume a common batch size $m = m_1 = \cdots = m_k$ in the analysis, which implies that the critical values d_i^α in (2) will be a common d^α, which is $\sqrt{2}$ times the $1 - \alpha$ quantile of the maximum of a $(k-1)$-dimensional multivariate-t random variable with correlation $1/2$ and $k(m-1)$ degrees of freedom.

Since $\Pr\{\mathcal{C} \cap \mathcal{U}\}$ is problem dependent—depending on the actual differences among the parameters $\theta_1, \theta_2, \ldots, \theta_k$—we consider a related measure

$$\Pr\{\mathcal{E} \cap \mathcal{N}\} = \Pr\{Y_{(k)} - \theta_{(k)} \geq Y_i - \theta_i - d^\alpha S(m)/\sqrt{m}, \forall i \neq (k) \text{ and } d^\alpha S(m)/\sqrt{m} \leq \Delta\} \quad (6)$$

where (k) is the index of the unknown true best (maximum θ) system design, $S^2(m)$ is the pooled variance estimator based on the batch means, and Δ is a fixed value. The expression (6) is the probability that the sample best system design (largest sample

mean) will be correctly inferred to be within Δ of the true best system design [Hsu 1988]. Stated differently, it is the probability that intervals are correct ($\mathcal{E}$) and sufficiently short ($\mathcal{N}$) so that the system design selected as best is no further than Δ from the true best system design. The $\Pr\{\mathcal{E} \cap \mathcal{N}\}$ is independent of $\theta_1, \theta_2, \ldots, \theta_k$.

To investigate the effect of number of batches on (6) we did a worst-case analysis:

- We assumed that the output data $Y_{i\ell}, \ell = 1, 2, \ldots, n$, from each system design i is actually i.i.d. normal with common variance, meaning that batching is not needed at all. Thus, batching only *penalizes* us in terms of $\Pr\{\mathcal{E} \cap \mathcal{N}\}$ because of the loss of degrees of freedom.
- We set Δ to be the *smallest attainable difference* that can be distinguished with a fixed total sample size. Recall that the total sample size, n, is fixed and does not increase with the number of batches, m; some Δ values are too small to distinguish without increasing n. The minimum value of Δ, in units of $\sqrt{\mathrm{Var}[\bar{Y}]}$, is equal to the critical value d^α at an infinite degrees of freedom. By taking the smallest possible Δ we make degrees of freedom (number of batches) the most important.

Goldsman [1990], using results in Hsu [1988], derived an expression for (6) that can be evaluated numerically. Figure 1 shows a plot of $\Pr\{\mathcal{E} \cap \mathcal{N}\}$ versus number of batches m for $k = 3, 5$ and 10 system designs. Notice that $\Pr\{\mathcal{E} \cap \mathcal{N}\}$ initially increases rapidly as m increases, but then slows so that the marginal increase from additional batches is negligible. This is comforting, because it is at small numbers of batches that the approximations of independence and normality in batching are most likely to be valid. In addition, performance is relatively insensitive to the number of system designs for $3 \leq k \leq 10$; however, larger k does cause $\Pr\{\mathcal{E} \cap \mathcal{N}\}$ to level off more quickly.

Another view of the same results is given in Table 1, which shows the increase in $\Pr\{\mathcal{E} \cap \mathcal{N}\}$ obtained by going from $m = m_0$ to $m = \infty$ batches; $m = \infty$ gives the theoretical upper bound on $\Pr\{\mathcal{E} \cap \mathcal{N}\}$. The increase is dramatic for less than 20 batches, but negligible for greater than 40 batches.

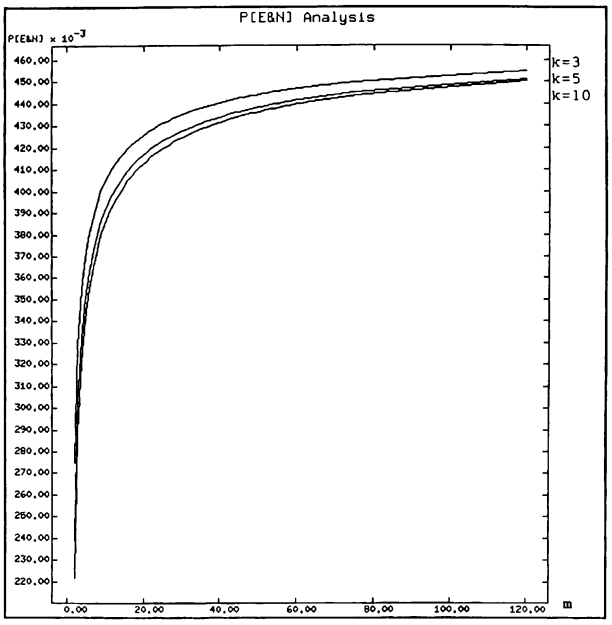

Figure 1. $\Pr\{\mathcal{E} \cap \mathcal{N}\}$ as a Function of m for $k = 3, 5$ and 10

Table 1. Percentage Increase in $\Pr\{\mathcal{E} \cap \mathcal{N}\}$ from $m = m_0$ to $m = \infty$

	k		
m_0	3	5	10
5	0.39	0.33	0.26
10	0.23	0.19	0.15
15	0.18	0.14	0.11
20	0.15	0.10	0.09
30	0.12	0.09	0.07
40	0.10	0.07	0.05
60	0.08	0.05	0.04
120	0.06	0.03	0.02

4. BATCHING ALGORITHM

The results above have been used to design an algorithm to compute MCB intervals for steady-state simulation experiments. This section gives a brief overview of the algorithm, which is still a topic of ongoing research; details are given in Goldsman [1990].

For $3 \leq k \leq 10$ systems designs—which is all that we consider here—the algorithm batches the output from a single replication of each system into $20 \leq m \leq 40$ batch means. A premise of the algorithm is that the replication lengths, $n_1, n_2, \ldots, n_k$, are long enough that, within the range of $20 \leq m \leq 40$ batches, the dependence and nonnormality of the batch means are negligible. However, a statistical test of independence is applied to the batch means to check for a serious departure from the independence assumption.

In addition to obtaining approximately independent and normal batch means, the algorithm attempts to obtain batch means with approximately equal variances across system designs. More precisely it searches for a vector $\vec{m} = (m_1, m_2, \ldots, m_k)$ such that $\sigma_1^2(m_1) \approx \sigma_2^2(m_2) \approx \cdots \approx \sigma_k^2(m_k)$, with $m_i \in [20, 40]$.

Let $\sigma_i^2(m_i)$ be estimated by the sample variance

$$S_i^2(m_i) = (m_i - 1)^{-1} \sum_{j=1}^{m_i} (\bar{Y}_{ij}(m_i) - \bar{Y}_i)^2.$$

The problem formulation we use is

$$\begin{aligned} \text{minimize } Z(\vec{m}) &= \sum_{i=1}^{k} (S^2(\vec{m}) - S_i^2(m_i))^2 \qquad (7) \\ \text{subject to} \quad & S^2(\vec{m}) = k^{-1} \sum_{i=1}^{k} S_i^2(m_i) \\ & m_i \in [20, 40] \end{aligned}$$

which attempts to minimize the sum-of-squared deviations between the sample variance from each system design and the resulting pooled sample variance. We use a branch and bound technique on (7) to search for $\vec{m}^* = (m_1^*, m_2^*, \ldots, m_k^*)$ that minimizes $Z(\vec{m})$. A test for variance homogeneity is applied to $S_1^2(m_1^*), S_2^2(m_2^*), \ldots, S_k^2(m_k^*)$ to check for a serious departure from the equal variance assumption.

The algorithm follows the outline below:
Algorithm BMEV

1. Collect a single replication, $Y_{i1}, Y_{i2}, \ldots, Y_{in_i}$, from each system $i = 1, 2, \ldots, k$ (the algorithm assumes that remedial measures for initial-condition bias, such as deletion, have already been applied).
2. Batch the data from each system into (about) 40 batch means. Apply the test of independence. If it fails, reduce the maximum number of batches further.
3. Solve (7) for $\vec{m}^*$. Apply the variance homogeneity test. If it fails, note this in the final results.
4. Form the MCB intervals as in (2) with $S^2(\vec{m}^*)$ substituted for S^2 and m_i^* substituted for n_i.

5. EXPERIMENT RESULTS

To gain some insight into the performance of BMEV we applied it to the output of an AR(1) process

$$Y_{i\ell} = \theta_i + \phi_i(Y_{i,\ell-1} - \theta_i) + \varepsilon_{i\ell},$$

for $i = 1, 2, \ldots, k$ and $\ell = 1, 2, \ldots, n$, where $\varepsilon_{i\ell} \sim$ i.i.d $N(0, \sigma_i^2)$, and $-1 < \phi_i < 1$. The AR(1) process was selected to control the properties of the output process.

For each system the AR(1) process was initialized in steady-state and a single run of $n = 1000$ observations was generated. MCB 95% confidence intervals were formed from the data, and a success was recorded if the intervals provided correct and useful inference. We ran 1000 replications of the entire experiment to obtain a standard error of approximately 0.01 for the estimate of $\Pr\{\mathcal{C} \cap \mathcal{U}\}$.

The results from BMEV were compared with results from an algorithm that ignored the problem of unequal variances (BM). Algorithm BM batches the observations from each system into $m = 20$ batch means and forms MCB intervals.

Results were obtained for $k = 3$ and 5 systems. For $k = 3$ systems the variances σ_i^2 were fixed so that $\vec{m}^* = (20, 25, 40)$ gives equal variances; for $k = 5$ systems the variances σ_i^2 were fixed so that $\vec{m}^* = (20, 20, 25, 25, 40)$ gives equal variances.

In the experiments reported here, $\phi_1 = \phi_2 = \cdots = \phi_k = \phi$. Hence, for a given ϕ, the systems have identical dependence structure.

To set values for the process means, $\theta_1, \theta_2, \ldots, \theta_k$, the least favorable configuration (LFC) from ranking and selection was used:

$$\theta_{(k)} - \theta_i = \begin{cases} 0 & \text{if } i = (k) \\ \gamma & \text{otherwise} \end{cases}$$

where $\theta_{(k)}$ is the largest mean. Based on the analysis of $\Pr\{\mathcal{E} \cap \mathcal{N}\}$, we chose a worst-case initial value of γ to be $2\Delta = 2d^{0.05} \times \max_i \left\{\sqrt{\mathrm{Var}[\bar{Y}_i]}\right\}$, where $d^{0.05}$ is the critical value at an infinite degrees of freedom. We also ran experiments with values of $\gamma < 2\Delta$, values of $\gamma > 2\Delta$, and an additional configuration where the θ_i's were equally spaced 2Δ apart.

Results were obtained for the following two cases:

1. $\theta_{(r)}$ is associated with the system that has the minimum variance, and
2. $\theta_{(r)}$ is associated with the system that has the maximum variance.

Tables 2 and 3 summarize the results of our experiments for both cases for $\phi = 0.9$ with $r = 3$ and 5 systems, respectively (Goldsman [1990] presents results for a range of ϕ values).

When $\theta_{(r)}$ is associated with the system having the minimum variance (case 1, the top half of each table), BMEV shows an improvement over BM. This improvement is more dramatic for small differences in expected performance; that is, when γ is much smaller than 2Δ. As γ approaches 2Δ the improvement from BMEV over BM decreases until it is negligible. When $\theta_{(r)}$ is associated with the system having the maximum variance (bottom half of each table), the difference between BMEV and BM is negligible. These results are preliminary, and a more thorough empirical evaluation is required.

Table 2. Estimated $\Pr\{\mathcal{C} \cap \mathcal{U}\}$ for $r = 3$ Systems

			$\Pr\{\mathcal{C} \cap \mathcal{U}\}$		
θ_1	θ_2	θ_3	BMEV	BM	% Increase
0.0	0.0	Δ	0.23	0.19	0.21
0.0	0.0	2Δ	0.94	0.92	0.02
0.0	0.0	3Δ	1.0	1.0	0.0
0	2Δ	4Δ	0.96	0.95	0.01
Δ	0.0	0.0	0.22	0.19	0.16
2Δ	0.0	0.0	0.93	0.92	0.01
3Δ	0.0	0.0	1.0	1.0	0.0
4Δ	2Δ	0.0	0.95	0.95	0.0

Table 3. Estimated $\Pr\{\mathcal{C} \cap \mathcal{U}\}$ for $r = 5$ Systems

					$\Pr\{\mathcal{C} \cap \mathcal{U}\}$		
θ_1	θ_2	θ_3	θ_4	θ_5	BMEV	BM	% Increase
0.0	0.0	0.0	0.0	Δ	0.10	0.07	0.43
0.0	0.0	0.0	0.0	2Δ	0.91	0.88	0.03
0.0	0.0	0.0	0.0	3Δ	1.0	1.0	0.0
0	2Δ	4Δ	6Δ	8Δ	0.98	0.96	0.02
Δ	0.0	0.0	0.0	0.0	0.10	0.09	0.11
2Δ	0.0	0.0	0.0	0.0	0.88	0.87	0.01
3Δ	0.0	0.0	0.0	0.0	1.0	1.0	0.0
8Δ	6Δ	4Δ	2Δ	0.0	0.95	0.96	-0.01

6. DISCUSSION

The batch-size analysis presented in Section 3.2 provides the framework for an algorithm to form MCB confidence intervals for steady-state simulations. Deriving a comparable theory for batching outputs from different system designs to obtain batch means with nearly equal variances, and an analysis of the effect of such batching on the resulting inference, remains to be done. However, the empirical studies completed to date indicate that even our heuristic approach to the batching problem is effective.

ACKNOWLEDGMENTS

The authors acknowledge the help of Prof. Jason Hsu. This material is based upon work supported by National Science Foundation Grant No: DDM 8922721.

REFERENCES

Clark, G.M. and W. Yang (1986), "A Bonferroni Selection Procedure when using Common Random Numbers with Unknown Variances," In *Proceedings of the 1986 Winter Simulation Conference*, J.R. Wilson, J.O. Henriksen, and S.D. Roberts, Eds. IEEE, Piscataway, NJ, 313-315.

Glynn, P.W. (1986), "Optimization of Stochastic Systems," *Proceedings of the 1986 Winter Simulation Conference*, J.R. Wilson, J.O. Henriksen, and S.D. Roberts, Eds. IEEE, Piscataway, NJ, 52-59.

Goldsman, D. (1985), "Ranking and Selection Procedures Using Standardized Time Series," In *Proceedings of the 1985 Winter Simulation Conference*, D.T. Gantz, S.L. Solomon, and G.C. Blais, Eds. IEEE, Piscataway, NJ, 120-124.

Goldsman, L. (1990), "Batching Strategies for Multiple Comparisons in Steady-State Simulation," M.S. Thesis, Department of Industrial and Systems Engineering, The Ohio State University, Columbus, OH.

Gupta, S.S. and S. Panchapakesan (1979), *Multiple Decision Procedures - Theory and Methodology of Selecting and Ranking Populations*, Wiley, New York, NY.

Hsu, J.C. (1984), "Constrained Two-Sided Simultaneous Confidence Intervals for Multiple Comparisons with the Best," *Annals of Statistics 12*, 1136-1144.

Hsu, J.C. (1988), "Sample Size Computation for Designing Multiple Comparison Experiments," *Computational Statistics & Data Analysis 7*, 79-91.

Hsu, J.C. and B.L. Nelson (1988), "Optimization over a Finite Number of System Designs with One-Stage Sampling and Multiple Comparisons with the Best," In *Proceedings of the 1988 Winter Simulation Conference*, M.A. Abrams, P.L. Haigh, and J.C. Comfort, Eds. IEEE, Piscataway, NJ, 451-457.

Iglehart, D.L. (1977), "Simulating Stable Stochastic Systems, VII: Selecting the Best System," In *TIMS Studies in the Management Sciences 7*, North Holland, Amsterdam, The Netherlands, 37-49.

Koenig, L.W. and A.M. Law (1985), "A Procedure for Selecting a Subset of Size m Containing the l Best of k Independent Normal Populations, with Applications to Simulation," *Communications in Statistics - Simulation and Computation 14*, 719-734.

Nelson, B.L. (1989), "Batch Size Effects on the Efficiency of Control Variates in Simulation," *European Journal of Operational Research 43*, 184-196.

Sargent, R.G., K. Kang and D. Goldsman (1987), "An Investigation of Small Sample Behavior of Confidence Interval Estimation Procedures," Working Paper #87-005, Department of Industrial Engineering and Operations Research, Syracuse University, Syracuse, NY.

Schmeiser, B. (1982), "Batch Size Effects in the Analysis of Simulation Output," *Operations Research 30*, 556-568.

Sullivan, D.W. and J.R. Wilson (1989), "Restricted Subset Selection Procedures for Simulation," *Operations Research 37*, 52-71.

Yang, W. and B.L. Nelson (1988), "Multivariate Estimation and Variance Reduction in Terminating and Steady-State Simulation," In *Proceedings of the 1988 Winter Simulation Conference*, M.A. Abrams, P.L. Haigh, and J.C. Comfort, Eds. IEEE, Piscataway, NJ, 466-472.

Yang, W. and B.L. Nelson (1989a), "Optimization using Common Random Numbers, Control Variates and Multiple Comparisons with the Best," In *Proceedings of the 1989 Winter Simulation Conference*, E.A. MacNair, K.J. Musselman, and P. Heidelberger, Eds. IEEE, Piscataway, NJ, 444-449.

Yang, W. and B.L. Nelson (1989b), "Multivariate Batch Means and Control Variates," Working Paper Series No. 1989-005, Department of Industrial and Systems Engineering, The Ohio State University. Columbus, OH.