Proceedings of the 2001 Winter Simulation Conference
B. A. Peters, J. S. Smith, D. J. Medeiros, and M. W. Rohrer, eds.

SELECTION-OF-THE-BEST PROCEDURES FOR OPTIMIZATION VIA SIMULATION

Juta Pichitlamken
Barry L. Nelson

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

ABSTRACT

We propose fully sequential indifference-zone selection procedures that are specifically for use within an optimization-via-simulation algorithm when simulation is costly and partial or complete information on solutions previously visited is maintained. *Sequential Selection with Memory* guarantees to select the best or near-best alternative with a user-specified probability when some solutions have already been sampled and their previous samples are retained. For the case when only summary information is retained, we derive a modified procedure. We illustrate how our procedure can be applied to optimization-via-simulation problems and compare its performance with other methods by numerical examples.

1 INTRODUCTION

Simulation is merely a tool for problem solving; by itself, it cannot provide an answer. In addition to a good model, one also needs a sound technique to utilize the information from a simulation to make a decision. One such technique is optimization via simulation.

As discussed in Azadivar (1999), optimization-via-simulation problems typically lack a closed-form expression for the objective function, any known regularity conditions on the response surface (e.g., convexity), or other special structure. The response and constraints are subject to uncertainties and can only be estimated by simulation; therefore, linear or non-linear programming techniques cannot be directly applied. Although the number of feasible solutions of optimization-via-simulation problems may not be as large as the number of solutions encountered in deterministic combinatorial optimization, simulation of a single solution can take so long that complete enumeration of all of the alternatives is impossible. Thus, an inefficient optimization algorithm may be able to visit only a small portion of the feasible space and be forced to terminate far from the optimum.

Several general-purpose optimization algorithms for deterministic problems, including tabu search (Glover 1989) and simulated annealing (see van Laarhoven and Aarts 1987), are based on a *neighborhood search*. At each iteration, the search may move from a current solution to a better solution that is chosen from among the neighbors of the current solution. In the deterministic setting, selection of the "best" neighbor is accomplished by a straightforward evaluation of the objective function.

If we want to use neighborhood search strategies in optimization via simulation, then uncertainty dramatically complicates the selection of the best neighbor. Because of randomness, multiple replications (or lengthy runs) may be required to obtain a useful estimate of the objective function. Ideally, we want to obtain enough observations to be confident of our choice of the best neighbor. However, if too much computational effort is spent on the neighborhood selection problem, then the overall search may not be able to make much progress in the time available. Thus, the efficiency of the local search—in terms of correct selections and computational effort required—is critical to the overall performance of an optimization algorithm applied to stochastic simulation.

One way to save some simulation effort is to maintain partial or complete information on solutions previously encountered; for instance, all observations of the current solution, or of the best solution seen so far, or of all solutions that have been simulated, might be retained. The selection of the best can then be considered a comparison with a standard, where the solutions on which we retain data are the standards. Nelson and Goldsman (2001) address such comparisons when there is a single standard. However, there does not exist a selection procedure for the case when we maintain "memory" of more than one solution. Most statistical selection procedures assume that none of the alternatives, or at most one (a standard), have already been sampled.

We have designed a new indifference-zone selection procedure specifically to supplement optimization-via-

simulation algorithms. *Sequential Selection with Memory* (SSM) guarantees to select the best or near-best alternative with a user-specified probability when some or all of the alternatives have previously been sampled. SSM is intended to aid optimization algorithms in making correct local search steps, not to provide any overall correct selection guarantee.

The procedure is fully sequential: it takes one observation at a time from every solution that is still in play and eliminates clearly inferior ones as soon as their inferiority is apparent. Bechofer et al.'s (1990) numerical assessment shows that fully sequential procedures usually perform well. Nevertheless, their use has been limited by the high overhead in switching among solutions. As a result, the number of fully sequential procedures is small, and there was a large time gap between the first work on this topic and subsequent work. However, optimization-via-simulation software must be able to switch among solutions, which makes fully sequential procedures viable.

Paulson (1964) proposed fully sequential procedures for the selection-of-the-best problem when all solutions have equal variances. He assumed normality and independence within and across solutions. His procedures are applicable for both known and unknown variance cases. Hartmann (1988, 1991) improved Paulson's results by using tighter probability bounds; Boole's inequality, which Paulson used, was replaced by a geometric inequality, and a large deviation bound was replaced by a Brownian motion bound. The most recent work by Kim and Nelson (2001a) further extends Hartmann (1988, 1991) to the problem of unequal and unknown variances, with special emphasis on use within the simulation context. The key difference between their work and ours is that, in their case, no solutions have been previously sampled.

This paper is organized as follows: We define the selection-of-the-best problem in Section 2 and then present our procedures in Sections 3–4. We illustrate how our procedures can be applied in the optimization-via-simulation context with numerical examples in Section 5, followed by future research directions in Section 6.

2 FRAMEWORK

We consider the problem of selecting the "best" from a finite number of solutions. Let π_i denote solution i, $i = 1, 2, \ldots, k$. The observations taken from π_i, X_{ip}, are i.i.d. normally distributed with mean μ_i and variance σ_i^2, where μ_i and σ_i^2 are unknown and not necessarily equal for different solutions. Without loss of generality, we assume that the true means of the solutions are indexed such that

$$\mu_1 \leq \mu_2 \leq \cdots \leq \mu_k.$$

As is traditional, the best solution is defined as the one with the largest mean, which is μ_k in the current problem. Our procedure guarantees to select π_k with probability at least $1 - \alpha$ whenever the difference between the best and the next best solution is worth detecting:

$$\Pr\{\text{select } \pi_k\} \geq 1 - \alpha \text{ whenever } \mu_k - \mu_{k-1} \geq \delta. \qquad (1)$$

The indifference-zone parameter is denoted by $\delta > 0$. Even when the indifference-zone condition is not satisfied ($\mu_k - \mu_{k-1} < \delta$), our procedures still select a "good" solution whose mean is within δ of μ_k with probability at least $1-\alpha$.

In the context of optimization via simulation, $\pi_1, \pi_2, \ldots, \pi_k$ represent solutions to be compared in a neighborhood search step of the optimization algorithm. At the start of the procedure, we may have already sampled some of the solutions. Let n_i be the number of observations obtained on solution i and let n_0 be the *minimum* number of observations that we take on any solution. If $n_i = 0$ for *all* $i = 1, 2, \ldots, k$, our procedure becomes Kim and Nelson (2001a). If some solution i, $n_i = 0$, we obtain n_0 observations on all such solutions before the elimination step begins.

3 PROCEDURE SSM

In SSM, we sequentially take one observation at a time from *surviving* solutions, immediately followed by screening. Depending on their sample sizes at the start of the procedure, we may or may not take more observations from the solutions we have already sampled. Screening eliminates the solutions whose cumulative sums fall short of the best of the rest minus some positive tolerance level. As more observations are taken, this tolerance level decreases.

The *continuation region* (see Figure 1) illustrates the elimination step. Suppose we have only two solutions, π_i and π_j. Our procedure continues as long as the sum of the difference between solution i and solution j, $\sum_{p=1}^{r}(X_{ip} - X_{jp})$, stays within the triangular region. The sum can leave this region in three ways: First, if the sum drifts below the lower boundary, we eliminate π_i (recall that bigger is better). Second, if the sum goes above the upper boundary, we eliminate π_j. Lastly, if the sum exits the continuation region to the right of the triangular area ($r > N_{ij}$), we select the solution with the maximum average as the better one. The procedure is finite; at most, we take one step beyond the continuation region.

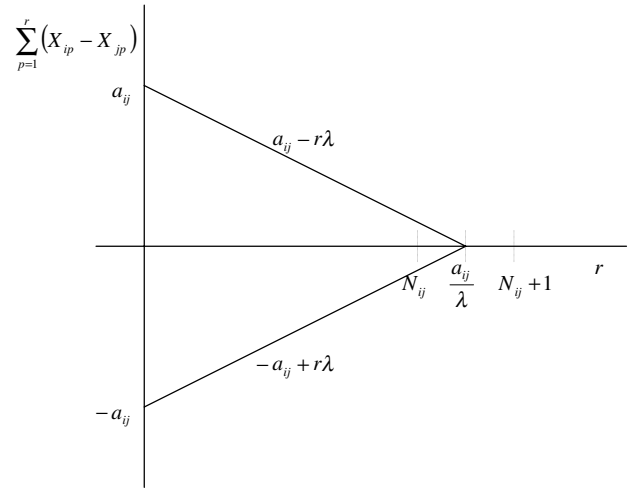

Figure 1: Continuation Region

To make this precise, we define some notation. Let $i \in \{1, 2, \ldots, k\}$.

$$
\begin{aligned}
V &= \text{set of solutions we have "visited" before;} \\
&\quad n_i > 0 \text{ for } i \in V \\
V^c &= \text{set of solutions we see for the first time;} \\
&\quad n_i = 0 \text{ for } i \in V^c \\
&= \{1, 2, \ldots, k\} \setminus V \\
N_{ij} &= \left\lfloor \frac{a_{ij}}{\lambda} \right\rfloor \\
N_i &= \max_{j \neq i} \{N_{ij}\} \\
N &= \max_i N_i, \qquad (2) \\
\bar{X}_i(r) &= \frac{1}{r} \sum_{p=1}^{r} X_{ip}, \\
\sigma_{ij}^2 &= \text{Var}\,(X_{ip} - X_{jp}) \\
n_0 &= \text{minimum initial number of observations} \\
&\quad \text{from any solution} \\
S_{ij}^2 &= \text{estimator of } \sigma_{ij}^2 = \text{Var}\,(X_{ip} - X_{jp}) \\
&= \frac{1}{n_0 - 1} \sum_{p=1}^{n_0} \left(X_{ip} - X_{jp} - [\bar{X}_i(n_0) - \bar{X}_j(n_0)]\right)^2 \\
&\qquad (3) \\
f &= n_0 - 1
\end{aligned}
$$

Note that $\lfloor x \rfloor$ is the largest integer less than or equal to x, and $N_i + 1$ is the maximum number of observations taken from π_i.

Procedure SSM

1. *Initialization:* Take initial samples to estimate the variance σ_{ij}^2.
   - For each π_i, $i \in V^c$, take $n_0 \geq 2$ observations, and set $n_i = n_0$.
   - For $i \in V$, if $n_i < n_0$, take additional observations so that $n_i = n_0$, and update V and V^c:

$$
\begin{aligned}
V^c &= V^c \cup \{i\} \\
V &= V \setminus \{i\}.
\end{aligned}
$$

   Compute S_{ij}^2 using (3).
2. *Procedure parameters:*
   Let c be a positive integer. We choose λ and a_{ij} as follows:

$$\lambda = \frac{\delta}{2c} \quad \text{and} \quad a_{ij} = \frac{\eta f S_{ij}^2}{4(\delta - \lambda)} \qquad (4)$$

   where η satisfies

$$\sum_{\ell=1}^{c} (-1)^{\ell+1} \left(1 - \frac{1}{2}\mathbb{1}(\ell = c)\right) \cdot \left(1 + \frac{(2c - \ell)\ell\eta}{2c - 1}\right)^{-f/2} = \frac{\alpha}{k - 1}. \qquad (5)$$

   The indicator function is represented as $\mathbb{1}(\cdot)$. Equation (5) has a closed-form solution for $c = 1$:

$$\lambda = \frac{\delta}{2} \text{ and } a_{ij} = \frac{f S_{ij}^2}{4(\delta - \lambda)} \left[\left(\frac{k - 1}{2\alpha}\right)^{2/f} - 1\right].$$

   If $n_0 > N$ (defined in (2)), stop and select the solution with the largest $\bar{X}_i(n_i)$ as the best. Otherwise, let $I = \{1, 2, \ldots, k\}$ be the set of surviving solutions, set $r = n_0$, and proceed to *Screening*. From here on V represents the set of solutions on which we have obtained more than r observations, while V^c is the set of solutions with exactly r observations.
3. *Screening:* Set $I^{\text{old}} = I$. Let

$$I = \left\{i : i \in I^{\text{old}} \text{ and } R_i \geq \max_{j \in I^{\text{old}}, j \neq i} \left(R_j - a_{ij}\right) + r\lambda\right\}$$

   where

$$R_j = \begin{cases} \sum_{p=1}^{r} X_{jp} & \text{for } j \in V^c \\ r\bar{X}_j(n_j) & \text{for } j \in V. \end{cases}$$

In essence, for π_i with $n_i > r$, we substitute $r\bar{X}(n_i)$ for $\sum_{p=1}^{r} X_{ip}$.

4. *Stopping rule:* If $|I| = 1$, then stop and report the only survivor as the best; otherwise, for each $i \in (I \cap V^c)$, take one additional observation from π_i and set $r = r + 1$. If $r = N + 1$, terminate the procedure and select the solution in I with the largest sample average as the best; otherwise, for each $i \in (I \cap V)$ with $n_i = r$,

$$\begin{aligned} V^c &= V^c \cup \{i\} \\ V &= V \setminus \{i\}, \end{aligned}$$

and go to *Screening*.

We show that our procedure satisfies the probability guarantee (1) in Theorem 1, and that a good solution is selected even if the indifference-zone condition does not hold via Corollary 1. The proofs are given in Pichitlamken (2001). The impact of c is discussed in Kim and Nelson (2001a).

Theorem 1. *Suppose that X_{ip}, $p = 1, 2, \ldots$, are i.i.d. normally distributed, and that X_{ip} and X_{jq} are independent for $i \neq j$. Then SSM guarantees that (1) holds.*

Corollary 1. *Suppose $\mu_k - \mu_{k-1} < \delta$. Then SSM selects a solution whose mean is within δ of μ_k with probability at least $1 - \alpha$.*

4 APPROXIMATE PROCEDURE

The SSM procedure uses the variance estimator S_{ij}^2 that requires retaining individual observations, specifically $(X_{i1}, X_{i2}, \ldots, X_{in_0})$. This requirement may impose a storage problem if the number of solutions is large and we retain data on many of them. Welch's (1947) approximation offers an alternative that avoids saving the raw data by estimating σ_{ij}^2 from the marginal variance estimators for each solution, S_i^2 and S_j^2. Therefore, instead of $(X_{i1}, X_{i2}, \ldots, X_{in_0})$, we are only required to maintain the triplet $(n_i, \sum_{p=1}^{n_i} X_{ip}, \sum_{p=1}^{n_i} X_{ip}^2)$. Notice that the approximate procedure allows us to estimate σ_{ij}^2 using *all* available observations, not just n_0 observations from each solution, π_i and π_j. Therefore, the variance estimator used in Welch's approximation may be more accurate than (3). However, the modified procedure is approximate in the sense that we can no longer prove the probability of correct selection guarantee (although it appears to hold in experiments).

Welch's approximation was originally intended to solve the Behren-Fisher problem (a test for equality of means of two normal populations with unequal variances and unequal sample sizes). In it, the distribution of the standardized difference between sample means is approximated as a Student's t distribution.

We hypothesize that our approximate procedure is asymptotically valid as the indifference zone δ approaches zero using the regime described in Kim and Nelson (2001b). Kim and Nelson (2001b) extends their fully sequential procedure to steady-state simulation (stationary, but dependent and non-normal data), and they are able to prove the statistical validity of their procedure under certain conditions, one of which is that $\delta \to 0$. See Goldsman, Kim, Marshall and Nelson (2001) for an empirical evaluation of this procedure.

Although letting $\delta \to 0$ is only a proof technique in Kim and Nelson (2001b), actually decreasing δ over the course of an optimization run may make sense. Recall that δ is defined as the difference in performance between the best and the next-best solution that is worth detecting. By letting δ get small, the procedure becomes increasingly demanding when selecting the best solution in the neighborhood. As the search gets close to the optimum, or a long exploration time has elapsed, the search becomes more and more certain about the best solution. Thus, it should become more and more difficult to accept a new solution as the optimal solution. Having $\delta \to 0$ is one way to enforce this preference, with δ playing a role similar to a "cooling schedule" in simulated annealing.

The approximate procedure differs from the exact one only through a_{ij}. The modification is explained below:

1. The variance of $X_{ip} - X_{jp}$ is estimated from the marginal variance estimators for each solution as

$$\begin{aligned} S_{ij}^2 &= S_i^2(n_i) + S_j^2(n_j). \qquad (6) \\ S_i^2(n_i) &= \frac{1}{n_i - 1}\left(\sum_{p=1}^{n_i} X_{ip}^2 - n_i \bar{X}_i^2(n_i)\right) \end{aligned}$$

2. The resulting degrees of freedom are approximated by

$$f_{ij} = \left\lfloor \frac{\left(S_i^2(n_i) + S_j^2(n_j)\right)^2 - 2\left(\frac{S_i^4(n_i)}{f_i+2} + \frac{S_j^4(n_j)}{f_j+2}\right)}{\frac{S_i^4(n_i)}{f_i+2} + \frac{S_j^4(n_j)}{f_j+2}} \right\rfloor \qquad (7)$$

$$f_i = n_i - 1$$

3. With the approximate procedure, the degrees of freedom are unequal, so f is replaced by f_{ij}. We substitute S_{ij}^2 from (6) and f_{ij} from (7) in (4) to yield a_{ij}.

5 ILLUSTRATIVE EXAMPLES

We show how our procedures can be applied to select the best alternative in a neighborhood via numerical examples. Consider the following optimization problem:

$$\max_{\mathbf{x}\in\Theta} \mathrm{E}\left[G(\mathbf{x})\right],$$

where the feasible space Θ is two-dimensional, finite and discrete,

$$\begin{aligned}\Theta &= \{(x_1, x_2): \\ &\quad x_i \in \{-N, -N+1, \ldots, N-1, N\}, i = 1, 2\}\end{aligned}$$

where N is a positive integer we set to 10 in the experiments reported here. The observed response $G(\mathbf{x})$ is normally distributed with the random noise being a function of the true response:

$$G(\mathbf{x}) = g(\mathbf{x}) + N\left(0, \sigma^2(g(\mathbf{x}))\right).$$

The response surface $g(\mathbf{x})$ is the paraboloid,

$$g(\mathbf{x}) = -(0.5x_1 - 1)^2 - (1.5x_2 - 1)^2 + x_1x_2 - 4$$

which is unimodal with a maximum $\mathbf{x}^*$ at (6,2), and $g(\mathbf{x}^*) = 0$. The standard deviation of the random errors, $\sigma(g(\mathbf{x}))$, is either directly or inversely proportional to the true response as follows:

$$\begin{aligned}\sigma_1(\mathbf{x}) &= \sqrt{|g(\mathbf{x})|} + 1 \\ \sigma_2(\mathbf{x}) &= \frac{1}{\sqrt{|g(\mathbf{x})|} + 1}.\end{aligned}$$

For simplicity, we use a hill-climbing algorithm as our global search strategy; the search iteratively moves from the current solution to one of its neighbors that is better than itself and any other solution in the neighborhood. In other words, the pool of alternatives considered by a selection procedure consists of the current solution and its neighbors. If the current solution is still deemed best, the search is trapped, and it terminates. The other way that the search may stop is when the total budget is expended. At the end of the search, we compare the sample averages of the current solution and the best seen so far and select the one with the larger sample average as the optimum.

The neighbors of solution (x_1, x_2) are (x_1-1, x_2), (x_1+1, x_2), (x_1, x_2-1) and (x_1, x_2+1). We exclude any neighbor that lies outside the feasible space Θ from the selection process. Only the previous observations of the current solution and the best solution seen so far are maintained; other solutions are re-sampled upon each visit. The past information on the current solution is re-used in the selection step, and further sampling may be done if required by a selection method.

We compare SSM to three other schemes for neighborhood selection of the best: Nelson and Goldsman's (2001) comparison with a standard (NG), a naïve approach (NA) and a sequential t-test (TT). In the naïve approach, we take equal numbers of observations from each alternative and select the one with the maximum average. The sequential t-test is modified from Costa and Silver's (1998) neighborhood selection method which was developed in the context of applying tabu search in the presence of randomness. Each neighbor receives equal numbers of replications and then competes with the best neighbor seen so far on a 1-1 basis using a one-side null hypothesis and a t-test. The differences between our TT and Costa and Silver's (1998) version are the following: They exclude the current solution from the competition, and they perform two rounds of sampling and hypothesis testing while we have one; in their scheme, the second round is needed only if a contestant beats the incumbent in the first round.

For SSM we set $c = 1$, the initial number of replications to $n_0 = 10$, the probability of correct selection to $1-\alpha = 0.9$ and the indifference zone $\delta = 0.2$ (smaller than $g(\mathbf{x}')-g(\mathbf{x}^*)$ where $\mathbf{x}'$ is a neighbor of $\mathbf{x}^*$). We used the same settings for n_0, $1-\alpha$ and δ for NG. For NA and TT, the number of replications per solution was 1000, and we used confidence level 0.9 for TT.

For each optimization problem (directly and inversely proportional standard deviation) and computational budget (B total observations), the search is repeated 1000 times using each local search strategy (SSM, NG, NA and TT). The initial solution in each search is sampled uniformly from the feasible space.

The performance measures we considered were $\bar{n}$, the number of observations used per search averaged over 1000 searches, and CP, the number of convergent paths out of 1000 (a convergent path is one that terminates with the optimal solution).

The simulation results for the directly proportional standard deviation problem ($\sigma_1(\mathbf{x})$) are shown in Table 1, and for the inversely proportional standard deviation problem ($\sigma_2(\mathbf{x})$) in Table 2. As expected, the number of convergent paths increased with the available budget B across all selection methods for both types of standard deviation. SSM outperforms the other schemes for both $\sigma_1(\mathbf{x})$ and $\sigma_2(\mathbf{x})$; given the same budget, SSM is able to achieve a much higher CP and lower $\bar{n}$.

NA and TT are comparable for either type of standard deviation. This finding suggests an assertion that if a statistical test (in this case, a t-test) is naïvely incorporated into a search scheme without taking into consideration *the multiplicity effect*—the number of alternatives in the selection pool—the resulting scheme may not do any better than one without a statistical test at all. In addition, NA and TT

Table 1: Performance of Selection-of-the-best Procedures with Directly Proportional Standard Deviation, $\sigma_1(\mathbf{x})$

B	SSM		NG		NA		TT	
	$\bar{n}$	CP	$\bar{n}$	CP	$\bar{n}$	CP	$\bar{n}$	CP
20,000	12,000	660	20,000	7	19,500	116	19,500	104
30,000	14,000	868	30,000	8	28,000	220	28,000	230
40,000	15,000	935	39,800	27	36,000	320	36,000	317
50,000	15,000	975	50,000	31	43,000	498	43,000	443
100,000	15,000	973	98,000	72	56,000	971	56,000	851
200,000	15,000	972	190,000	125	56,000	993	56,000	867
1,000,000	16,000	978	770,000	412	58,000	993	56,000	882

Table 2: Performance of Selection-of-the-best Procedures with Inversely Proportional Standard Deviation, $\sigma_2(\mathbf{x})$

B	SSM		NG		NA		TT	
	$\bar{n}$	CP	$\bar{n}$	CP	$\bar{n}$	CP	$\bar{n}$	CP
20,000	1,000	978	4,900	974	19,600	94	19,500	110
30,000	1,000	974	4,800	973	28,000	249	29,000	229
40,000	1,000	976	4,800	973	37,000	294	36,000	330
50,000	1,000	980	4,800	981	42,000	481	42,000	485
100,000	1,000	987	4,800	973	56,000	973	57,000	972
200,000	1,000	981	4,800	970	57,000	1000	58,000	1000
1,000,000	1,000	982	4,800	968	57,000	1000	56,000	1000

do not adapt to the level of randomness; even when $\sigma_2(\mathbf{x})$ is generally much lower than $\sigma_1(\mathbf{x})$ for most $\mathbf{x} \in \Theta$, NA and TT obtain approximately equal numbers of convergent paths (for a given B) for both standard deviation functions.

NA and TT are superior to NG in the case of $\sigma_1(\mathbf{x})$, but NG performs better than either of them for inversely proportional standard deviation, $\sigma_2(\mathbf{x})$. When the standard deviation is relatively high, NG does not work well as it requires a large sample and, therefore, uses up its computational budget before approaching the optimum. This is a consequence of how NG is designed: it is a two-stage procedure whose second-stage sample size is based on the first-stage variance estimates, among other factors, but not the relative performance of other solutions under consideration. However, for low variability, such as $\sigma_2(\mathbf{x})$, NG does not fare drastically worse than a fully sequential procedure like SSM which takes into account the relative performance of every surviving candidate in the selection.

6 THE FUTURE

We have designed *Sequential Selection with Memory* specifically for use in a context that is common in optimization via simulation: Due to expensive simulation, an optimization algorithm maintains some information on solutions previously visited. In addition, there is a continuing need to select the best from a number of (neighboring) solutions as part of a local search step. SSM is highly efficient in this problem setting because it is fully sequential; i.e., every solution is allowed to simultaneously eliminate every other solution, and one observation at a time is taken from each surviving solution. Moreover, SSM lets the search re-use the past information it has gained, thereby avoiding re-sampling at every encounter. We expect that SSM will similarly enhance the performance of better search heuristics than the hill-climbing algorithm employed here.

However, SSM may pose a storage problem when the number of solutions is large. With Welch's approximation, we are able to circumvent this issue by maintaining either the triplet $(n_i, \sum_{p=1}^{n_i} X_{ip}, \sum_{p=1}^{n_i} X_{ip}^2)$ for each solution i or keeping $(X_{i1}, X_{i2}, \ldots, X_{in_i})$ only for the best solution seen so far and the current solution. We would like to prove the statistical validity of SSM modified with Welch's approximation and assess of the trade-offs between different record keeping options via numerical studies.

ACKNOWLEDGMENT

We would like to thank General Motors R&D Center for their support of this work.

REFERENCES

Azadivar, F. 1999. Simulation optimization methodologies. In *Proceedings of the 1999 Winter Simulation Conference*, ed. P. A. Farrington, H. B. Nembhard, D. T. Sturrock, and G. W. Evans, 93–100. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers. Available online via <http://www.informs-cs.org/wsc99papers/012.PDF> [accessed June 17, 2001].

Bechhofer R. E., C. W. Dunnett, D. M. Goldsman, and M. Hartmann. 1990. A comparison of the performances of procedures for selecting the normal population having the largest mean when the populations have a common unknown variance. *Communications in Statistics*, B19: 971–1006.

Costa, D., and E. A. Silver. 1998. Tabu search when noise is present: An illustration in the context of cause and effect analysis. *Journal of Heuristics*, 4: 5–23.

Glover, F. 1989. Tabu search—part I. *ORSA Journal on Computing*, 1: 190–206.

Goldsman, D., S.-H. Kim, W. S. Marshall, and B. L. Nelson. 2001. Ranking and selection for steady-state simulation: Procedures and perspectives. Working paper, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Hartmann, M. 1988. An improvement on Paulson's sequential ranking procedure. *Sequential Analysis*, 7: 363-372.

Hartmann, M. 1991. An improvement on Paulson's procedure for selecting the population with the largest mean from k normal populations with a common unknown variance. *Sequential Analysis*, 10: 1-16.

Kim, S.-H. and B. L. Nelson. 2001a. A fully sequential procedure for indifference-zone selection in simulation. *ACM TOMACS*, in press.

Kim, S.-H. and B. L. Nelson. 2001b. On the asymptotic validity of fully sequential selection procedures for steady-state simulation. Working paper, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

Nelson, B. L., and D. Goldsman. 2001. Comparisons with a standard in simulation experiments. *Management Science*, 47: 449-463.

Paulson, E. 1964. A sequential procedure for selecting the population with the largest mean from k normal populations. *Annals of Mathematical Statistics*, 35: 174-180.

Pichitlamken, J. 2001. *Combined procedures for simulation optimization*, Ph.D. Dissertation Proposal, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, Illinois.

van Laarhoven, P. J. M., and E. H. L. Aarts. 1987. *Simulated Annealing: Theory and Applications*. Dordrecht, the Netherlands: D. Reidel Publishing.

Welch, B. L. 1947. The generalization of 'Student's' problem when several different population variances are involved. *Biometrika*, 34: 28-35.

AUTHOR BIOGRAPHIES

JUTA PICHITLAMKEN is a Ph.D. candidate in the Department of Industrial Engineering and Management Sciences at Northwestern University. She received her B.S. in chemical engineering from Cornell University and M.S. in chemical engineering from University of Washington. Her research interests include ranking and selection procedures and optimization via simulation. Her e-mail and web addresses are <juta@northwestern.edu> and <http://users.iems.nwu.edu/~ying/>.

BARRY L. NELSON is a Professor in the Department of Industrial Engineering and Management Sciences at Northwestern University, and is Director of the Master of Engineering Management Program there. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems. He has published numerous papers and two books. Nelson has served the profession as the Simulation Area Editor of *Operations Research* and President of the INFORMS (then TIMS) College on Simulation. He has held many positions for the Winter Simulation Conference, including Program Chair in 1997 and current membership on the Board of Directors. His e-mail and web addresses are <nelsonb@northwestern.edu> and <www.iems.northwestern.edu/ñelsonb/>.