ELSEVIER

Decision Support

A sequential procedure for neighborhood selection-of-the-best in optimization via simulation

Juta Pichitlamken [a,*], Barry L. Nelson [b], L. Jeff Hong [c]

[a] *Department of Industrial Engineering, Kasetsart University, Bangkok 10903, Thailand*
[b] *Department of Industrial Engineering & Management Sciences, Northwestern University, Evanston, IL 60208-3119, USA*
[c] *Department of Industrial Engineering and Engineering Management, The Hong Kong University of Science and Technology, Kowloon, Hong Kong*

Abstract

We propose a fully sequential indifference-zone selection procedure that is 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, their previous samples are retained, and simulation outputs are i.i.d. normal. For the case when only summary information on solutions is retained, we derive a modified procedure. We illustrate how our procedures can be applied to optimization-via-simulation problems and compare its performance with other methods by numerical examples.
© 2005 Elsevier B.V. All rights reserved.

Keywords: Simulation; Multivariate statistics; Ranking and selection; Optimization via simulation

1. Introduction

Discrete-event, stochastic simulation is a widely used tool to analyze dynamic systems that are subject to uncertainty. In practical problems, the analyst is often interested in finding a system configuration (i.e., a *solution*) that optimizes the expected value of some measure or measures of system performance. In this paper, we only consider problems with a single performance measure. Determining these solutions is known as *optimization via simulation* (see [3,9,22] for surveys). This problem is difficult because there is no

---

* Corresponding author.
 E-mail addresses: juta.p@ku.ac.th (J. Pichitlamken), nelsonb@northwestern.edu (B.L. Nelson), hongl@ust.ht (L.J. Hong).

closed-form expression for the performance measure; instead expected performance is *estimated* via simulation experiments, and is therefore subject to sampling error. Unfortunately, obtaining precise estimates of the objective function at many solutions can take so long that time may run out before reaching the optimal, or even a good, solution if the trade off between searching for improved solutions and controlling the selection error is managed poorly.

Several general-purpose optimization algorithms that were originally designed for deterministic problems, including tabu search [10], simulated annealing [23] and the nested partitions method [20], have been adapted to stochastic problems [1,7,21]. Provably convergent algorithms for stochastic simulation have also been derived [2,19]. These algorithms are based on a *neighborhood search* and selection-of-the-best neighbor from a finite number of alternatives: On each iteration, the algorithm may move from a current best 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. In the optimization-via-simulation context, however, uncertainty dramatically complicates this 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 of the objective function to be confident in our choice of the best neighbor. However, if too much computational effort is spent in the selection-of-the-best neighbor, then the search may not be able to make much progress in the time available. Thus, the efficiency of the selection—in terms of balancing the tradeoff between making a correct selection and the computational effort required—is critical to the overall performance of an optimization algorithm applied to stochastic simulation. *The goal of our work is to provide an efficient selection-of-the-best scheme to be used during the neighborhood search.*

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 solutions have previously been visited, and individual observations taken on those solutions are maintained. When this data storage is prohibitive, we propose an alternative procedure in which only summary statistics of the simulation output are required. SSM is intended to aid optimization algorithms in making a correct selection-of-the-best neighbor, not to provide any global correct-selection guarantee.

SSM is a highly efficient method for selecting the best because it is *fully sequential with elimination*, which means that it takes one observation at a time from every solution that is still in play and eliminates clearly inferior solutions as soon as their inferiority is apparent. A number of studies have established the superiority of such procedures over well-known two-stage ranking-and-selection procedures [5,14]. The additional feature of SSM is that it utilizes "memory" of solutions that have been previously sampled to alleviate the need to obtain new simulation outputs every time it encounters a solution. This approach not only has obvious computational benefits, it also facilitates establishing convergence properties of optimization algorithms [4].

Neighborhood selection of the best can be considered a comparison with a standard, where the solutions on which we retain data are the standards. Nelson and Goldsman [16] address such comparisons when there is a single standard. However, there is no 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, and that they have all been sampled equally.

Our procedure is the most recent in a line of fully sequential procedures with elimination. Paulson [17] proposed fully sequential procedures for the selection-of-the-best problem when all solutions have equal variances. Under normality and independence within and across solutions, Paulson's procedures are applicable for both known and unknown variance cases. Hartmann [11,12] improved Paulson's results by using tighter probability bounds; Boole's inequality, which Paulson used, is replaced by a geometric inequality, and a large-deviation bound is replaced by a Brownian-motion bound. The most recent work by Kim and Nelson [14] further extends Hartmann [11,12] 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. Although our extension sounds straightforward, it is quite difficult to prove that it is valid (see Appendix A). Further, our procedure not only tolerates previous samples, it exploits them so that selections are made more and more efficiently as the optimization search progresses.

The paper is organized as follows: We define the selection-of-the-best problem in Section 2. In Section 3 we present SSM and prove its properties; we also introduce an alternative version of SSM that is designed for the case where only summary statistics are maintained. We illustrate how our procedures can be applied in the optimization-via-simulation context with numerical examples in Section 4, followed by conclusions in Section 5.

2. Framework

In an iteration of an optimization-via-simulation algorithm, suppose that there are k solutions under consideration, denoted by $\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k\}$. We think of $\mathbf{x}_i$ as being a vector of decision variables, say $\mathbf{x}_i = (x_{i1}, x_{i2}, \ldots, x_{id})$, that define the ith solution. The goal is to select the "best" from among this finite number of solutions on the current iteration; on the next iteration there will be another collection of k solutions, which may include some of the ones from this iteration and some new ones.

We assume that the simulation output data from solution $\mathbf{x}_i$, Y_{ip}, are i.i.d. normally distributed with finite mean μ_i and variance σ_i^2, where μ_i and σ_i^2 are unknown and typically unequal for different solutions. The normality assumption will be reasonable when the simulation output Y_{ip} is actually the average of a large number of more basic outputs, as is frequently the case. For instance, Y_{ip} could be the average delay in queue of all customers on the pth day of service, or the average inventory costs over 12 months of operation in a supply chain system. Independence will pertain when Y_{ip}, $p = 1, 2, \ldots$, correspond to results from different replications.

Independence across systems implies that common random numbers (CRN) is not exploited; however, this issue may be irrelevant in the context of optimization via simulation because using CRN may not be practical. Suppose our optimization-via-simulation algorithm retains the current sample best solution at all times. On a given iteration t of the algorithm, the best solution that it has seen so far, say $\mathbf{x}_{B_t}$, is compared with another candidate solution, $\mathbf{x}_{A_t}$. Suppose further that s simulation outputs have already obtained from $\mathbf{x}_{B_t}$, but $\mathbf{x}_{A_t}$ is visited for the first time. To fully exploit CRN enough information must be retained to synchronize runs the runs of $\mathbf{x}_{A_t}$ with all previous runs of $\mathbf{x}_{B_t}$. Depending on the simulation language, the overhead in ensuring this could be substantial.

Without loss of generality, we assume that the true means of the solutions are (unknown to us) indexed such that $\mu_1 \leqslant \mu_2 \leqslant \cdots \leqslant \mu_k$. The best solution is defined as the one with the largest mean, which is therefore μ_k. Our procedure guarantees to select $\mathbf{x}_k$ with a user-specified probability of at least $1 - \alpha$ whenever the difference between the best and the next-best solution is worth detecting:

$$\Pr\{\text{select } \mathbf{x}_k\} \geqslant 1 - \alpha \quad \text{whenever } \mu_k - \mu_{k-1} \geqslant \delta. \tag{1}$$

The user chooses the indifference-zone parameter $\delta > 0$ to be the smallest important difference between solutions; for instance, if the user is indifferent between two inventory policies whose expected costs are less than \$200 apart, then the indifference-zone parameter is set to \$200. Even when the indifference-zone condition is not satisfied (i.e., $\mu_k - \mu_{k-1} < \delta$), our procedure still selects a "good" solution whose mean is within δ of μ_k with probability at least $1 - \alpha$.

In the context of optimization via simulation, $\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_k$ represent settings of the decision variables to be compared in a neighborhood selection of the best. Each time SSM is invoked, there may already be observations available on some of the solutions that were obtained on previous visits of the search

algorithm. For instance, if our optimization algorithm always retains the sample best solution for comparison with new candidate solutions, then the sample best solution must already have been simulated.

3. Procedure SSM

In SSM, we sequentially take at most one observation at a time from *surviving* solutions, immediately followed by screening. 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 Fig. 1) illustrates the elimination step (Step 3 below shows how the slope λ and the intercept a_{ij} are chosen). Suppose we have only two solutions in contention, $\mathbf{x}_i$ and $\mathbf{x}_j$. Our procedure continues as long as the sum of the difference between solution i and solution j, $\sum_{p=1}^{r}(Y_{ip} - Y_{jp})$, 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 $\mathbf{x}_i$ (recall that bigger is better). Second, if the sum goes above the upper boundary, we eliminate $\mathbf{x}_j$. Lastly, if the sum exits the continuation region to the right of the triangular area, we select the solution with the maximum average as the better one. The procedure is finite; at most, SSM takes one step beyond the end of the continuation region. When there are more than 2 solutions in contention, we do pairwise comparisons of all surviving solutions. If all solutions enter SSM with the same number of observations, then our procedure becomes Kim and Nelson [14].

3.1. Procedure SSM

1. *Initialization*: SSM is given k solutions from which to choose the best. Let n_{i0} be the number of observations already obtained on solution i before SSM starts, and let $n_0 \geqslant 2$ be the *minimum* number of observations that we require on any solution before starting SSM. For any solution i with initial number of observations $n_{i0} < n_0$, we obtain $n_0 - n_{i0}$ more observations before SSM begins. Choose confidence level $1/k < 1 - \alpha < 1$ and indifference-zone parameter $\delta > 0$.
2. *Variance estimation*: Let $\underline{n}_0 = \min_{1 \leqslant i \leqslant k} n_{i0}$ and $\underline{n}_{ij} = \min\{n_{i0}, n_{j0}\}$ for all $i \neq j$. Estimate $\sigma_{ij}^2 = \mathrm{Var}(Y_{ip} - Y_{jp})$ with

$$S_{ij}^2 = \frac{1}{\underline{n}_{ij} - 1} \sum_{p=1}^{\underline{n}_{ij}} \left(Y_{ip} - Y_{jp} - [\overline{Y}_{i\underline{n}_{ij}} - \overline{Y}_{j\underline{n}_{ij}}] \right)^2 \tag{2}$$
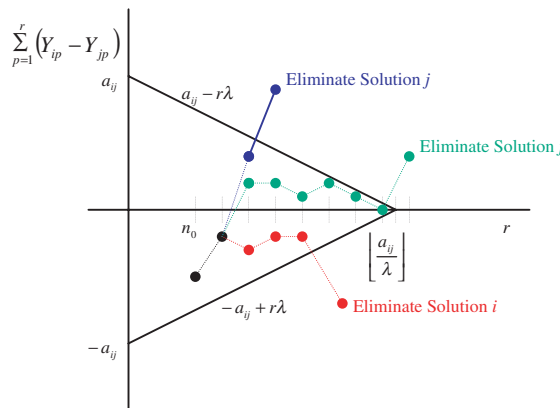

Fig. 1. The continuation region for SSM.

for all $i \neq j$, where $\overline{Y}_{in} = (1/n)\sum_{p=1}^{n} Y_{ip}$. The corresponding degrees of freedom are

$$f_{ij} = \underline{n}_{ij} - 1. \tag{3}$$

3. *Computing procedure parameters*: Let c be a positive integer. We choose λ and a_{ij} as follows:

$$\lambda = \frac{\delta}{2c} \quad \text{and} \quad a_{ij} = \frac{nf_{ij}S_{ij}^2}{4(\delta - \lambda)}, \tag{4}$$

where η satisfies

$$\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1 - \frac{1}{2}\mathscr{I}(\ell = c)\right)\left(1 + \frac{(2c - \ell)\ell\eta}{2c - 1}\right)^{-f_{ij}/2} = \frac{\alpha}{k - 1}. \tag{5}$$

The indicator function $\mathscr{I}(\epsilon)$ is one when ϵ is true and zero otherwise. Eq. (5) has a closed-form solution for $c = 1$, which is the value of c that we recommend: $\eta = ((k - 1)/(2\alpha))^{2/f_{ij}} - 1$.
4. *Define continuation region*: Let $N_{ij} = \lfloor a_{ij}/\lambda \rfloor$, $N_i = \max_{j \neq i}\{N_{ij}\}$, and $N = \max_{1 \leqslant i \leqslant k} N_i$. If $\underline{n}_0 > N$, then stop and select the solution with the largest $\overline{Y}_{in_{i0}}$ as the best. Otherwise, let $I = \{1, 2, \ldots, k\}$ be the set of surviving solutions, set step counter $r = \underline{n}_0$, and set $n_{ir} = n_{i0}$ for $1 \leqslant r \leqslant \underline{n}_0$. Proceed to *Screening*.
5. *Screening*: Set $I^{\text{old}} = I$ and update I as

$$I = \left\{i : i \in I^{\text{old}} \text{ and } r\overline{Y}_{in_{ir}} \geqslant \max_{j \in I^{\text{old}}, j \neq i}(r\overline{Y}_{jn_{jr}} - a_{ij}) + r\lambda\right\}.$$

Notice that $\overline{Y}_{in_{ir}}$ is simply a sample average over *all* available observations, and the set I records the indices of solutions still in contention.
6. *Stopping rule*: If $|I| = 1$, then stop and report the only survivor as the best; otherwise continue as follows:
   (a) For each $i \in I$ such that $n_{ir} < r + 1$, take one additional observation from solution $\mathbf{x}_i$ and set $n_{i,r+1} = n_{ir} + 1$.
   (b) For $i \in I$ such that $n_{ir} \geqslant r + 1$, set $n_{i,r+1} = n_{ir}$.
   (c) 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, go to *Screening*.

We now show that our procedure satisfies the probability guarantee in (1), and it selects a solution whose mean is within δ of the best when $\mu_k - \mu_{k-1} < \delta$. See the Appendix A for all of the proofs.

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

In Corollary 1 below, we extend Theorem 1 to show that SSM also provides a guaranteed good selection when the indifference-zone condition is not satisfied.

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* $1 - \alpha$.

Theorem 1 and Corollary 1 also imply that if a user chooses to terminate SSM with more than one solution on hand, the best or a near-best solution will be in the set of surviving solutions with probability at least $1 - \alpha$. This is because SSM is designed to select the best (or a near-best) solution with some user-specified probability, which requires that the best (or a near-best) is one of the surviving solutions whenever SSM is terminated. This feature is useful when an optimization algorithm only needs to identify a group of good solutions during the neighborhood search.

Consider what happens when the optimization algorithm has visited solution i (perhaps it is the current sample best solution) and then encounters a solution j that it has not visited previously. In order for SSM to compute the sample variance of the difference, S_{ij}^2, based on pairwise differences, we need to have retained all of solution i's individual observations, not just a summary statistic. This could require substantial data storage when the number of feasible solutions is large. Therefore, we next offer an alternative procedure that avoids saving the raw data by estimating σ_{ij}^2 from the marginal variance estimators for each solution, denoted S_i^2 and S_j^2. Instead of $(Y_{i1}, Y_{i2}, \ldots, Y_{in_{i0}})$, we are only required to maintain the summary statistics $(n_{i0}, \sum_{p=1}^{n_{i0}} Y_{ip}, \sum_{p=1}^{n_{i0}} Y_{ip}^2)$. Notice that the alternative procedure allows us to estimate σ_{ij}^2 using *all* available observations, not just $\min(n_{i0}, n_{j0})$ paired observations from each solution, $\mathbf{x}_i$ and $\mathbf{x}_j$ (see (2)). Therefore, the new variance estimator may be more precise than (2).

The alternative procedure differs from the one presented above only through a_{ij}. Procedure SSM is modified at the first step *Variance estimation*: the variance of $Y_{ip} - Y_{jp}$ is now estimated from the marginal variance estimators for each solution as

$$
S_i^2(n_{i0}) = \frac{1}{n_{i0} - 1} \left(\sum_{p=1}^{n_{i0}} Y_{ip}^2 - n_{i0} \overline{Y}_{in_{i0}}^2 \right),
$$

$$
S_{ij}^2 = S_i^2(n_{i0}) + S_j^2(n_{j0}).
$$

(6)

The resulting degrees of freedom is f_{ij} as shown in (3). We substitute S_{ij}^2 and f_{ij} into (4) to yield a_{ij}.

Theorem 2. *Under the same conditions as in Theorem 1, SSM using the variance estimator (6) guarantees to select μ_k or a solution whose mean is within δ of μ_k with probability at least $1 - \alpha$.*

Eq. (3) may underestimate the true degrees of freedom of S_{ij}^2 in (6), resulting in the bound on the probability of incorrect selection being too conservative. This causes SSM to obtain more data than are really necessary to provide the correct-selection guarantee. One alternative is to use the approximate degrees of freedom given by

$$
f_i = n_{i0} - 1,
$$

$$
v_{ij} = \left\lceil \frac{\left(S_i^2(n_{i0}) + S_j^2(n_{j0}) \right)^2 - 2 \left(\frac{S_i^4(n_{i0})}{f_i + 2} + \frac{S_j^4(n_{j0})}{f_j + 2} \right)}{\frac{S_i^4(n_{i0})}{f_i + 2} + \frac{S_j^4(n_{j0})}{f_j + 2}} \right\rceil
$$

(7)

(see Welch [24]). The shortcoming of using (7) to approximate the degrees of freedom of (6) is that we can no longer *guarantee* the probability of correct selection as in Theorem 2.

4. Illustrative examples

In this section we show how SSM can be applied to select the best alternative in a neighborhood via numerical examples, and we also compare the performance of SSM against other selection-of-the-best schemes. For simplicity, we use a random search (RS) algorithm as our optimization strategy (see [19] where SSM is embedded in other more sophisticated global search schemes). We use RS because it allows us to easily examine the impact of (or lack of impact of) SSM. We do not intend to suggest that pure RS is the best, or even a good, choice for practical optimization-via-simulation problems (although some forms of adaptive RS can be quite effective).

We consider the following optimization problem:

$$
\max_{\mathbf{x} \in \Theta} \mu(\mathbf{x}),
$$

where $\mu(\mathbf{x})$ is a scalar, $\mathbf{x}$ is a vector or a scalar decision variable, and the feasible space Θ is finite. The finiteness of Θ allows us (conceptually) to index the solutions $\mathbf{x}$ and the corresponding performance measures as follows: $\Theta = \{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_v\}$, where v denotes the number of feasible solutions in Θ. Let $\mu_i \equiv \mu(\mathbf{x}_i)$, which is unknown (in practice) but can be estimated via a simulation experiment.

After the tth iteration of the RS algorithm, let $B_t \in \{1, 2, \ldots, v\}$ denote the index of the sample best solution on hand, and let S_t be the set of indices of solutions that were under consideration on iteration t. In our experiments we consider the simplest possible neighborhood around $\mathbf{x}_{B_t}$, $\Theta \backslash \{B_t\}$. RS proceeds as follows:

1. *Initialization*: Set $t = 0$, B_t to the index of a user-provided solution (if not given, randomly sample a solution from Θ), and $S_t = \{B_t\}$.
2. *Search*: Repeat Steps 2(a)–(c) until the simulation effort is exhausted:
   (a) Randomly sample ω candidate solutions over $\Theta \backslash \{B_t\}$. Add the indices of these sampled solutions into a set S_t.
   (b) Use a selection-of-the-best method (from the list below) to choose the best solution from S_t.
   (c) Increment t: $t = t + 1$. Let B_t be the index of the solution selected in Step 2(b), and set $S_t = \{B_t\}$.
3. *Estimating the optimal solution*: Upon termination return $\mathbf{x}_{B_t}$.

Note that RS always keeps the best solution that it has seen so far, $\mathbf{x}_{B_t}$, on hand. We consider the following four selection-of-the-best schemes:

1. SSM: Before we run SSM, as described in Section 3, we take Δn observations from each $\mathbf{x}_i$, $i \in S_t$. This prevents the search loop (Step 2 above) from running indefinitely in cases where the number of solutions in Θ is small, and RS is able to visit every solution in Θ. In such scenarios, after RS has seen all solutions in Θ, SSM does not need to take any new observations to select the best because it re-uses past observations. From that point onward, unless we take some observations from each $\mathbf{x}_i$, $i \in S_t$, Step 2 is an infinite loop. Aside from this practical reason, the addition of new observations also makes the estimates of μ_i more precise, and it helps in establishing the global convergence of RS (see [4]).
2. Nelson and Goldsman's [16] comparison with a standard method (NG): We consider NG because $\mathbf{x}_{B_t}$ can be viewed as a standard and we want to determine whether any of the ω sampled solutions beats it. NG is a two-stage indifference-zone ranking-and-selection procedure. Similar to SSM, NG exploits the concept of an indifference zone and using the first-stage sampling to assess the variability of the simulation output. However, SSM is sequential (i.e., having multiple stages), whereas NG has only two stages. NG decides either that no alternative is better than the standard or that one or more of them is better.
3. Naive approach (NA): NA is an intuitive scheme that one may employ instead of a statistical procedure. In the naive approach, we take n observations from each $\mathbf{x}_i$, $i \in S_t$, and select the one with the best sample average (the one with largest average if the objective is to maximize).
4. Sequential t-test (TT): TT is motivated by Costa and Silver's [7] neighborhood selection method which was developed in the context of applying tabu search in the presence of randomness. For each $\mathbf{x}_i$, $i \in S_t$, we take n observations, and then $\mathbf{x}_i$, $i \neq B_t$, competes with $\mathbf{x}_{B_t}$ on a 1–1 basis using a two-sided null hypothesis and a t-test: Let $Z_p = Y_{B_{tp}} - Y_{ip}$, $\overline{Z}(n) = (1/n)\sum_{p=1}^{n} Z_p$, and

$$\widehat{\mathrm{Var}}[\overline{Z}(n)] = \frac{1}{n-1}\left(\sum_{p=1}^{n} Z_p^2 - n\overline{Z}(n)^2\right).$$

If the upper bound of the $100(1 - \alpha)$ percent confidence interval

$$\overline{Z}(n) \pm t_{n-1,1-\alpha/2}\sqrt{\widehat{\mathrm{Var}}[\overline{Z}(n)]}$$

is negative (assuming a maximization context), then TT declares $\mathbf{x}_i$ as better than the current best solution $\mathbf{x}_{B_t}$, and B_t is replaced by i. This comparison continues until every solution $\mathbf{x}_i$, $i \in S_t$, is examined.

Among the selection-of-the-best methods considered above, SSM is the only one that stores and reuses past observations (the others sample anew each time a solution is encountered). Each method is embedded in RS as Step 2(b), and the resulting algorithm is given the same computational budget (number of observations). We then repeat the entire optimization run m^* times. On the mth run, $1 \leqslant m \leqslant m^*$, RS provides an estimate of the optimal solution $\mathbf{x}_{B_t^m}$ which has corresponding *true* performance measure $\mu_{B_t^m}$. The results that we present below are the *averaged* values across m^* optimization runs, specifically,

$$\bar{\mu}_{B_t}(m^*) = \frac{1}{m^*} \sum_{m=1}^{m^*} \mu_{B_t^m}.$$

We compare the performance of the selection-of-the-best selection methods by observing $\bar{\mu}_{B_t}(m^*)$ as the simulation effort increases. We consider an (s, S) inventory problem and a three-stage buffer allocation problem. Table 1 shows parameter settings that we used in the numerical experiment. The variance estimator (2) is employed in SSM (the results are essentially unchanged when we use the variance estimator (6)).

4.1. (s, S) Inventory problem

In this classic problem [15], the level of inventory of some discrete unit is periodically reviewed. Let $s \equiv x_1$ and $S \equiv x_2$ so that a solution is $\mathbf{x} = (x_1, x_2)$. If the inventory position (units in inventory plus units on order minus units backordered) at a review is found to be below x_1 units, then an order is placed to bring the inventory position up to x_2 units; otherwise, no order is placed. Demand per period is Poisson with mean 25. The goal is to select (x_1, x_2) such that the steady-state expected inventory cost per review period is minimized. The constraints on x_1 and x_2 are $x_1 - x_2 \leqslant 0$, $20 \leqslant x_1 \leqslant 80$, $40 \leqslant x_2 \leqslant 100$, and $x_1, x_2 \in \mathcal{Z}^+$, where $\mathcal{Z}^+$ denotes the positive integers. The number of feasible solutions is 2901. The optimal inventory policy is $(20, 53)$ with expected cost/period of 111.1265. To reduce the initial-condition bias, the average cost per period is computed after the first 100 review periods and averaged over the subsequent 30 periods.

Fig. 2 shows the performance comparison of the local selection methods (if confidence intervals were placed around all points, they would not be visible because they are narrower than the size of the series symbols). NA and TT perform slightly better than SSM in the early phase of the search. However, SSM quickly wins as the simulation effort increases. After an initial period of rapid improvement, NA and TT do not seem to get close to the optimal solution even with increasing simulation effort. This suggests that if a statistical test (in this case, a *t*-test) is naively 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. Furthermore, if a selection-of-

Table 1
Parameters used in the numerical experiments

| Parameter	(s, S) inventory problem	Three-stage buffer allocation problem
Number of searches (m^*)	1000	50
Number of candidate solutions (ω)	2	2
NA and TT number of observations per solution(n)	10	10
SSM minimum number of observations taken from a sampled solution (Δn)	2	1
SSM and NG indifference-zone parameter (δ)	1.0	0.5
SSM and NG first-stage number of observations (n_0)	10	4
SSM and NG Confidence level ($1 - \alpha$)	0.9	0.9
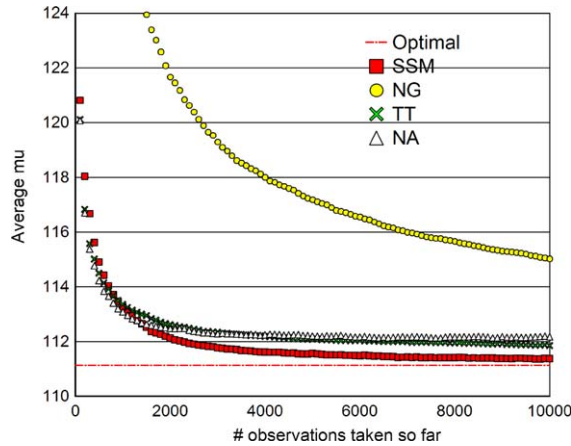
Fig. 2. Evolution of $\bar{\mu}_{B_t}(m^*)$ as the simulation effort increases for the (s, S) inventory problem.

the-best scheme does not adapt to the level of randomness, a global search (i.e., RS, in this experiment) may have great difficulty in finding an optimal solution.

NG is clearly outperformed by the other algorithms because it requires a large number of observations to do a selection, 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 but not the relative performance of the other solutions under consideration.

We can also assess the search performance in terms of the number of searches that find an optimal solution. Fig. 3 shows that SSM distinctly outperforms other selection methods; it is most likely to discover the optimal solution. In contrast, NA appears to find the optimal solution only to lose it later.

4.2. Three-stage buffer allocation problem

We consider a three-stage flow line with finite buffer storage space in front of stations 2 and 3 (denoted by x_4 and x_5, respectively) and an infinite number of jobs in front of station 1. There is a single server at
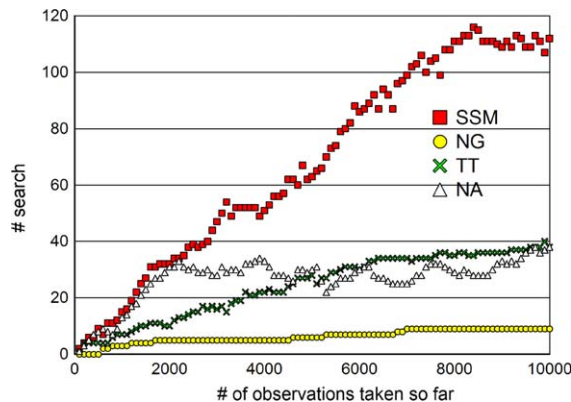

Fig. 3. Plot of the number of searches (out of 1000) that finds the optimal solution as the search progresses for the (s, S) inventory problem.
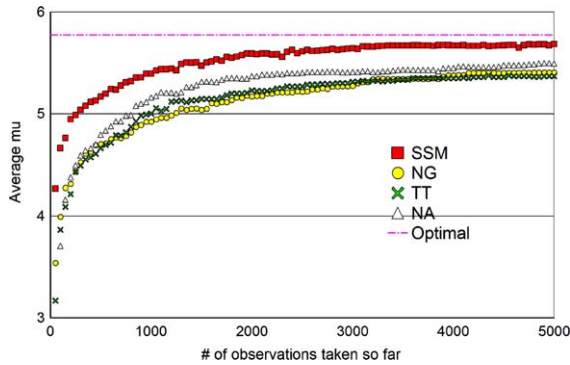
Fig. 4. Evolution of $\bar{\mu}_{B_t}(m^*)$ as the search progresses for the buffer allocation problem.

each station, and the service time at station h is exponentially distributed with service rate x_h, $h = 1, 2, 3$. If the buffer of station h is full, then station $h - 1$ is blocked (production blocking) and a finished job cannot be released from station $h - 1$. The total buffer space and the service rates are limited. The goal is to find a buffer allocation and service rates such that the throughput (average output of the flow line per unit time) is maximized. We obtained the balance equations for the underlying Markov chain from [6]. The constraints are: $x_1 + x_2 + x_3 \leqslant 20$; $x_4 + x_5 = 20$; $1 \leqslant x_h \leqslant 20$, $h = 1, 2, \ldots, 5$; $x_h \in \mathscr{Z}^+$.

The number of feasible solutions is 21,660. The optimal solutions are $(6, 7, 7, 12, 8)$ and $(7, 7, 6, 8, 12)$ with an expected throughput of 5.776. In the simulation, the throughput is estimated after the first 2000 units have been produced, and it is averaged over the subsequent 50 units released.

Fig. 4 shows the evolution of $\bar{\mu}_{B_t}(m^*)$ as the search progresses. Unlike the (s, S) inventory problem, SSM dominates the other selection schemes at all levels of simulation effort considered in this problem. SSM performs well in the buffer allocation problem relative to our other selection schemes because, once SSM has good solutions on hand, it uses them to quickly eliminates inferior solutions. Fig. 5 also supports this find-
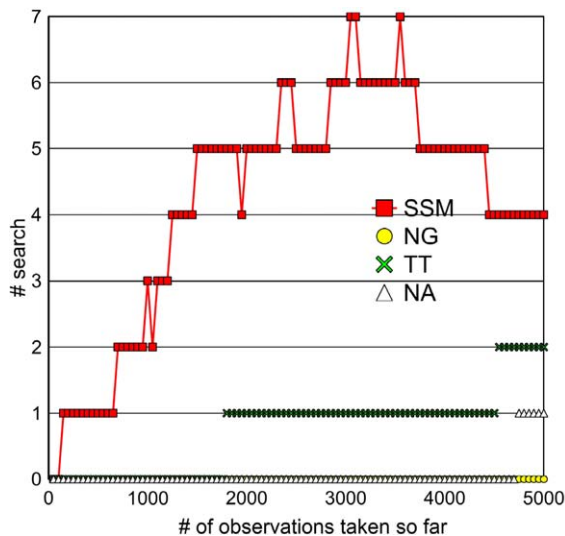

Fig. 5. Plot of the number of searches (out of 50) that finds the optimal solution as the search progresses for the buffer allocation problem.

ing; it shows the number of searches that finds the optimal solution as the search evolves. The RS with SSM is able to find an optimal solution more often than RS which is embedded with other local selection methods.

5. Conclusions

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. SSM is highly efficient in this problem setting because it is fully sequential; 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 algorithms than the RS algorithm employed here, and have shown this to be the case in a version of the nested partitions algorithm described in [18,19].

Acknowledgements

Appendix A

Our proof of Theorem 1 is based on controlling the probability of incorrect selection of a continuous-time Brownian-motion processes, rather than the discrete-time output process SSM actually observes. However, Jennison et al. [13] showed that under very mild conditions—satisfied by SSM—the probability of incorrect selection for a Brownian-motion process observed at discrete points in time is no greater than the corresponding probability for the continuous-time process. Thus, all of our results provide upper bounds on the probability of incorrect selection for SSM. Before providing the specific results, we present an overview of the approach.

We begin by considering a one-to-one elimination between the best solution $\mathbf{x}_k$ and a non-best solution $\mathbf{x}_i$ for $i \neq k$. We use $\mathcal{B}(t; \delta, \sigma^2)$ to denote the Brownian-motion process with $E[\mathcal{B}(t; \delta, \sigma^2)] = \delta t$ and $\mathrm{Var}[\mathcal{B}(t; \delta, \sigma^2)] = \sigma^2 t$. Hartmann [12] shows that the Brownian-motion process $\mathcal{B}(t; \delta, \sigma^2)$ with $\delta > 0$, $m > 0$, $\lambda > 0$, with stopping time defined as

$$T = \inf\{t : |\, \mathcal{B}(t; \delta, \sigma^2)\,| \geqslant \lambda(m - t)\}, \tag{A.1}$$

exits toward the lower boundary of the continuation region with probability

$$\Pr\{\mathcal{B}(T; \delta, \sigma^2) < 0\} = \int_{-\infty}^{\infty} \frac{\mathrm{e}^{-2\lambda\xi/\sigma}}{1 + \mathrm{e}^{-2\lambda\xi/\sigma}} \phi\left(\frac{\xi - (m\delta/\sigma)}{\sqrt{m}}\right) \frac{\mathrm{d}\xi}{\sqrt{m}}, \tag{A.2}$$

where $\phi(x)$ is the probability density function of the standard normal distribution. We use this result to provide a bound on the probability of incorrect selection (PICS) of SSM in one-to-one elimination between $\mathbf{x}_k$ and $\mathbf{x}_i$, $i \neq k$.

Define

$$\sigma_{ki}^2 \equiv \sigma_k^2 + \sigma_i^2, \tag{A.3}$$

$$\delta_{ki} \equiv \mu_k - \mu_i \tag{A.4}$$

and a stochastic process

$$\mathscr{A}(t; \delta, \sigma_{ki}^2, n_k, n_i) \equiv \mathscr{I}\{t \leqslant n_k\} \frac{t \mathscr{B}_k(n_k; \delta, \sigma_k^2)}{n_k} + \mathscr{I}\{t > n_k\} \mathscr{B}_k(t; \delta, \sigma_k^2)$$
$$- \mathscr{I}\{t \leqslant n_i\} \frac{t \mathscr{B}_i(n_i; 0, \sigma_i^2)}{n_i} - \mathscr{I}\{t > n_i\} \mathscr{B}_i(t; 0, \sigma_i^2), \tag{A.5}$$

where $\mathscr{B}_k(t; \delta, \sigma_k^2)$ and $\mathscr{B}_i(t; 0, \sigma_i^2)$ are two independent Brownian-motion processes, and $\mathscr{I}(\cdot)$ is the indicator function. The process $\mathscr{A}(t; \delta, \sigma_{ki}^2, n_k, n_i)$ corresponds to having already observed $\mathscr{B}_k(t; \delta, \sigma_k^2)$ and $\mathscr{B}_i(t; 0, \sigma_i^2)$ up to times n_k and n_i, respectively. The process $\mathscr{A}(\cdot)$ can be thought of as a continuous-time counterpart of the process $\{r \overline{Y}_{kn_{kr}} - r \overline{Y}_{in_{ir}}; r \geqslant 1\}$ that SSM considers during the screening step. Similar to Eq. (A.1), we also define the stopping time of $\mathscr{A}(t; \delta, \sigma_{ki}^2, n_k, n_i)$ as

$$T^* = \inf\{t : | \mathscr{A}(t; \delta, \sigma_{ki}^2, n_k, n_i) | \geqslant \lambda(m - t)\}. \tag{A.6}$$

In the context of SSM, "incorrect selection" corresponds to the event $\{\mathscr{A}(T^*; \delta, \sigma_{ki}^2, n_k, n_i) < 0\}$, and we are interested in three cases of n_i, when $n_k = 0$: (a) $n_i = m$, (b) $n_i < m$, and (c) $n_i > m$. We will derive the PICS of one-to-one elimination for the first case in detail (Lemmas A.1 and A.2), and also consider the scenario where the roles of i and k are reversed (Lemma A.2). We extend these results for the remaining cases of n_k and n_i (Lemma A.3). In the last step, we use the PICS of one-to-one elimination to get the overall PCS guarantee (1).

The proofs of Theorems 1 and 2 depend on a sequence of lemmas whose proofs are available from the authors on request.

Lemma A.1. *When $n_k = 0$ and $n_i = m$, we have that*

$$\Pr\{\mathscr{A}(T^*; \delta, \sigma_{ki}^2, 0, m) < 0\} = \int_{-\infty}^{\infty} \int_{-\infty}^{\infty} \frac{e^{-2\lambda\xi/\sigma_k}}{1 + e^{-2\lambda\xi/\sigma_k}} \phi\left(\frac{\xi - m(\delta - y)/\sigma_k}{\sqrt{m}}\right) d\xi \, \phi\left(\frac{\sqrt{m}y}{\sigma_i}\right) \frac{dy}{\sigma_i}. \tag{A.7}$$

Lemma A.2. *For the Brownian-motion process $\mathscr{B}(t; \delta, \sigma_{ki}^2)$ with $\delta > 0$, $m > 0$, $\lambda > 0$, and the stopping times T and T^* defined in* (A.1) *and* (A.6), *respectively,*

$$\Pr\{\mathscr{A}(T^*; \delta, \sigma_{ki}^2, 0, m) < 0\} \leqslant \Pr\{\mathscr{B}(T; \delta, \sigma_{ki}^2) < 0\}, \tag{A.8}$$

and

$$\Pr\{\mathscr{A}(T^*; \delta, \sigma_{ki}^2, m, 0) < 0\} \leqslant \Pr\{\mathscr{B}(T; \delta, \sigma_{ki}^2) < 0\}. \tag{A.9}$$

Lemma A.2 shows that, for the cases $n_k = 0$, $n_i = m$ and $n_k = m$, $n_i = 0$, the PICS of SSM is bounded by the probability that the corresponding Brownian-motion process exits the continuation region. In Lemma A.3, we show that this is also true for the remaining cases.

Lemma A.3. *For the Brownian-motion process $\mathscr{B}(t; \delta, \sigma_{ki}^2)$ with $\delta > 0$, $m > 0$, $\lambda > 0$, and the stopping times T and T^* defined in* (A.1) *and* (A.6), *respectively,*

$$\Pr\{\mathscr{A}(T^*; \delta, \sigma_{ki}^2, n_k, n_i) < 0\} \leqslant \Pr\{\mathscr{B}(T; \delta, \sigma_{ki}^2) < 0\}. \tag{A.10}$$

In Lemma A.4, we show that, with all other parameters being equal, the probability of incorrectly exiting the triangular region is non-increasing in the positive drift parameter of the process.

Lemma A.4. *For $n_k = 0$, $n_i = m$, with δ_{ki} defined in (A.4) and $\delta \leqslant \delta_{ki}$, we have that*

$$\Pr\{\mathscr{A}(T^*; \delta_{ki}, \sigma_{ki}^2, 0, m) < 0\} \leqslant \Pr\{\mathscr{A}(T^*; \delta, \sigma_{ki}^2, 0, m) < 0\}. \tag{A.11}$$

We use Lemmas A.1–A.4 to prove Theorem 1.

Proof of Theorem 1. Define the discrete exit time

$$r^* \equiv \inf\left\{r : |r\overline{Y}_{kn_{kr}} - r\overline{Y}_{in_{ir}}| \geqslant a_{ki} - r\lambda, r = 1, 2\ldots, \left\lfloor\frac{a_{ki}}{\lambda}\right\rfloor\right\},$$

and rewrite Definition (A.6) as

$$T^* = \inf\{t : |\mathscr{A}(t; \delta, \sigma_{ki}^2, n_k, n_i)| \geqslant a_{ki} - t\lambda\}.$$

SSM eliminates solutions (i.e., we inspect the process $\{r\overline{Y}_{kn_{kr}} - r\overline{Y}_{in_{ir}}; r \geqslant 1\}$) only at discrete steps $r \in \{n_0, n_0 + 1, \ldots, \lfloor a_{ki}/\lambda \rfloor\}$, whereas the process $\mathscr{A}(t; \delta_{ki}, \sigma_{ki}^2, n_k, n_i)$ can leave the continuation region at any continuous time t, $0 \leqslant t \leqslant a_{ki}/\lambda$. Using Jennison, Johnstone and Turnbull's result [13], we have that

$$\Pr\{r^*\overline{Y}_{kn_{kr^*}} - r^*\overline{Y}_{in_{ir^*}} < 0 \mid S_{ik}^2\} \leqslant \Pr\{\mathscr{A}(T^*; \delta_{ki}, \sigma_{ki}^2, n_k, n_i) < 0\}$$

(conditioning on S_{ik}^2 is needed to fix the parameter a_{ik}). Now we prove the PCS guarantee:

$$\Pr\{\text{incorrect selection}\} = \Pr\{\mathbf{x}_k \text{ is eliminated}\}$$

$$\leqslant \sum_{i=1}^{k-1} \Pr\{\mathbf{x}_i \text{ eliminates } \mathbf{x}_k\} \quad \text{(by Boole's inequality)}$$

$$= \sum_{i=1}^{k-1} E[\Pr\{\mathbf{x}_i \text{ eliminates } \mathbf{x}_k \mid S_{ik}^2\}]$$

$$\leqslant \sum_{i=1}^{k-1} E[\Pr\{\mathscr{A}(T^*; \delta_{ki}, \sigma_{ki}^2, n_k, n_i) < 0\}]$$

$$\leqslant \sum_{i=1}^{k-1} E[\Pr\{\mathscr{A}(T^*; \delta, \sigma_{ki}^2, n_k, n_i) < 0\}] \quad \text{(by Lemma A.4)}$$

$$\leqslant \sum_{i=1}^{k-1} E[\Pr\{\mathscr{B}(T; \delta, \sigma_{ki}^2) < 0\}] \quad \text{(by Lemma A.3)}$$

$$= \sum_{i=1}^{k-1} E\left[\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1 - \frac{1}{2}\mathscr{I}(\ell = c)\right)\exp\left\{\frac{(2c-\ell)\ell}{2c-1}\left(\frac{-2a_{ik}}{\sigma_{ik}^2}\right)(\delta - \lambda)\right\}\right] \tag{A.12}$$

$$= \sum_{i=1}^{k-1}\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1 - \frac{1}{2}\mathscr{I}(\ell = c)\right)E\left[\exp\left\{-\frac{(2c-\ell)\ell}{2c-1}\frac{\eta}{2}\left(\frac{f_{ik}S_{ik}^2}{\sigma_{ik}^2}\right)\right\}\right] \quad \text{(by (4))} \tag{A.13}$$

$$= \sum_{i=1}^{k-1}\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1 - \frac{1}{2}\mathscr{I}(\ell = c)\right)\left[1 + \frac{(2c-\ell)\ell\eta}{2c-1}\right]^{-f_{ik}/2} \tag{A.14}$$

$$\quad (\text{since } f_{ik}S_{ik}^2/\sigma_{ik}^2 \sim \chi_{f_{ik}}^2 \quad \text{and} \quad E[e^{t\chi_f^2}] = (1 - 2t)^{-f/2})$$

$$= (k-1)\alpha/(k-1) \quad \text{(by (5))}$$

$$= \alpha.$$

We get (A.12) by using Fabian's [8] result which shows that when $\lambda = \delta/(2c)$ and $m = a_{ik}/\lambda$, the expression on the right of (A.2) is given by

$$\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)\exp\left(-2\frac{\lambda^2}{\sigma^2}m(2j-\ell)\ell\right). \qquad \square$$

Proof of Corollary 1. Suppose $\mu_k \geqslant \mu_{k-1} \geqslant \cdots \geqslant \mu_1$. The statement is trivially true if $\mu_k - \mu_1 < \delta$, as we can select any system to satisfy the claim. Thus, let $t > 1$ be the first system index within δ of μ_k, i.e., $\mu_k - \mu_t < \delta$, and $\mu_k - \mu_{t-1} \geqslant \delta$. The correct selection in this context is when one of $\{\mathbf{x}_t, \mathbf{x}_{t+1}, \ldots, \mathbf{x}_k\}$ is selected.

$$
\begin{aligned}
\Pr\{\text{correct selection}\} &= \Pr\{\mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{t-1} \text{ are eliminated}\} \\
&\geqslant \Pr\{\mathbf{x}_k \text{ eliminates } \mathbf{x}_1, \mathbf{x}_2, \ldots, \mathbf{x}_{t-1}\} \\
&= 1 - \Pr\{\mathbf{x}_k \text{ is eliminated by } \mathbf{x}_i \text{ for some } i, 1 \leqslant i \leqslant t-1\} \\
&\geqslant 1 - \sum_{i=1}^{t-1}\Pr\{\mathbf{x}_k \text{ is eliminated by } \mathbf{x}_i\} \\
&\geqslant 1 - \frac{t-1}{k-1}\alpha \geqslant 1-\alpha.
\end{aligned}
$$

The second-to-last inequality follows from the way a_{ik} and λ are chosen in (5). $\square$

We need Lemma A.5 in the proof of Theorem 2.

Lemma A.5. *The function* $f(x) = (1 + a/x)^{-x}$, *where* $a > 0$ *and* $x > 0$, *is decreasing in* x.

Proof of Theorem 2. Define

$$\tau = \frac{(2c-\ell)\ell}{2c-1}\frac{\eta}{2}. \tag{A.15}$$

From (6) and (A.13), we have that

$$
\begin{aligned}
\Pr\{\text{incorrect selection}\} &\leqslant \sum_{i=1}^{k-1}\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)E\left[\exp\left\{-\frac{\tau f_{ik}(S_i^2+S_k^2)}{\sigma_i^2+\sigma_k^2}\right\}\right] \\
&= \sum_{i=1}^{k-1}\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)E\left[\exp\left\{-\frac{\sigma_i^2}{\sigma_i^2+\sigma_k^2}\frac{\tau f_{ik}S_i^2}{\sigma_i^2}\right\}\right] \\
&\quad \times E\left[\exp\left\{-\frac{\sigma_k^2}{\sigma_i^2+\sigma_k^2}\frac{\tau f_{ik}S_k^2}{\sigma_k^2}\right\}\right]
\end{aligned}
\tag{A.16}
$$

by independence of S_i^2 and S_k^2. First, we consider the case where $n_{k0} \geqslant n_{i0}$, and by (3), $f_{ik} = n_{i0} - 1$. We have that

$$
\begin{aligned}
\text{(A.16)} &= \sum_{i=1}^{k-1}\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)E\left[\exp\left\{-\frac{\sigma_i^2}{\sigma_i^2+\sigma_k^2}\frac{\tau(n_{i0}-1)S_i^2}{\sigma_i^2}\right\}\right] \\
&\quad \times E\left[\exp\left\{-\left(\frac{n_{i0}-1}{n_{k0}-1}\right)\frac{\sigma_k^2}{\sigma_i^2+\sigma_k^2}\frac{\tau(n_{k0}-1)S_k^2}{\sigma_k^2}\right\}\right] \\
&= \sum_{i=1}^{k-1}\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathscr{I}(\ell=c)\right)\left[1+2\tau\frac{\sigma_i^2}{\sigma_i^2+\sigma_k^2}\right]^{-(n_{i0}-1)/2}\left[1+2\tau\left(\frac{n_{i0}-1}{n_{k0}-1}\right)\frac{\sigma_k^2}{\sigma_i^2+\sigma_k^2}\right]^{-(n_{k0}-1)/2} \\
&\qquad (\text{because } (n_{i0}-1)S_i^2/\sigma_i^2 \sim \chi_{n_{i0}-1}^2 \text{ and } (n_{k0}-1)S_k^2/\sigma_k^2 \sim \chi_{n_{k0}-1}^2)
\end{aligned}
$$

$$
\leqslant \sum_{i=1}^{k-1}\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathcal{I}(\ell=c)\right)\left[1+2\tau\frac{\sigma_i^2}{\sigma_i^2+\sigma_k^2}\right]^{-(n_{i0}-1)/2}\left[1+2\tau\left(\frac{n_{i0}-1}{n_{k0}-1}\right)\frac{\sigma_k^2}{\sigma_i^2+\sigma_k^2}\right]^{-(n_{i0}-1)/2}
$$

(by Lemma A.5 and $n_{i0} \leqslant n_{k0}$)

$$
= \sum_{i=1}^{k-1}\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathcal{I}(\ell=c)\right)\left[1+2\tau+4\tau^2\left(\frac{\sigma_i\sigma_k}{\sigma_i^2+\sigma_k^2}\right)^2\right]^{-(n_{i0}-1)/2}
$$

$$
\leqslant \sum_{i=1}^{k-1}\sum_{\ell=1}^{c}(-1)^{\ell+1}\left(1-\frac{1}{2}\mathcal{I}(\ell=c)\right)[1+2\tau]^{-(n_{i0}-1)/2}
$$

$$\tag{A.17}$$

because $\tau > 0$. Notice that (A.17) is identical to (A.14); therefore, the rest of the proof proceeds similar to that of Theorem 1. The case where $n_{i0} > n_{k0}$ can also be proven in the same manner as above since (A.16) is unchanged when the role of i and k are reversed. The proof that SSM selects a near-best solution with probability at least $1 - \alpha$ immediately follows from Theorem 2 and Corollary 1.

Appendix B. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.ejor.2004.12.010.

References

[1] M.H. Alrefaei, S. Andradóttir, A simulated annealing algorithm with constant temperature for discrete stochastic optimization, Management Science 45 (1999) 748–764.

[2] M.H. Alrefaei, S. Andradóttir, A modification of the stochastic ruler method for discrete stochastic optimization, European Journal of Operational Research 133 (2001) 160–182.

[3] S. Andradóttir, Simulation optimization, in: J. Banks (Ed.), Handbook of Simulation, John Wiley, New York, 1998.

[4] S. Andradóttir, Accelerating the convergence of random search methods for discrete stochastic optimization, ACM Transactions on Modeling and Computer Simulation 9 (1999) 349–380.

[5] R.E. Bechhofer, C.W. Dunnett, D.M. Goldsman, M. Hartmann, 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 (1990) 971–1006.

[6] J.A. Buzacott, J.G Shantikumar, Stochastic Models of Manufacturing Systems, Prentice-Hall, Englewood Cliffs, NJ, 1993.

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

[8] V. Fabian, Note on Anderson's sequential procedures with triangular boundary, Annals of Statistics 2 (1974) 170–176.

[9] M.C. Fu, Optimization for simulation: Theory vs. practice, INFORMS Journal on Computing 14 (2002) 192–215.

[10] F. Glover, M. Laguna, Tabu Search, Kluwer Academic Publishers, Norwell, MA, 1998.

[11] M. Hartmann, An improvement on Paulson's sequential ranking procedure, Sequential Analysis 7 (1988) 363–372.

[12] M. Hartmann, 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 (1991) 1–16.

[13] C. Jennison, I.M. Johnstone, B.W. Turnbull, Asymptotically optimal procedures for sequential adaptive selection of the best of several normal means. Technical report No. 463, School of ORIE, Cornell University, Ithaca, New York, 1980.

[14] S.-H. Kim, B.L. Nelson, A fully sequential procedure for indifference-zone selection in simulation, ACM Transactions on Modeling and Computer Simulation 11 (2001) 251–273.

[15] L.W. Koenig, A.M. Law, 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 (1985) 719–734.

[16] B.L. Nelson, D. Goldsman, Comparisons with a standard in simulation experiments, Management Science 47 (2001) 449–463.

[17] E. Paulson, A sequential procedure for selecting the population with the largest mean from k normal populations, Annals of Mathematical Statistics 35 (1964) 174–180.

[18] J. Pichitlamken, A combined procedure for optimization via simulation, PhD dissertation, Department of Industrial Engineering and Management Sciences, Northwestern University, Evanston, IL, 2002.

[19] J. Pichitlamken, B.L. Nelson, A combined procedure for optimization via simulation, ACM Transactions on Modeling and Computer Simulation 12 (2) (2003) 155–179.

[20] L. Shi, S. Ólafsson, Nested partitions method for global optimization, Operations Research 48 (3) (2000) 390–407.

[21] L. Shi, S. Ólafsson, Nested partitions method for stochastic optimization, Methodology and Computing in Applied Probability 2 (3) (2000) 271–291.

[22] J.R. Swisher, P.D. Hyden, E. Yücesan, Discrete-event simulation optimization using ranking, selection, and mulitple comparison procedures: A survey, ACM Transactions on Modeling and Computer Simulation 13 (2) (2003) 134–154.

[23] P.J.M. van Laarhoven, E.H.L. Aarts, Simulated Annealing: Theory and Applications, D. Reidel Publishing, Dordrecht, The Netherlands, 1987.

[24] B.L. Welch, The generalization of 'student's' problem when several different population variances are involved, Biometrika 34 (1947) 28–35.