

MEAN-VARIANCE BASED RANKING AND SELECTION

Demet Batur
F. Fred Choobineh

Department of Industrial and Management Systems Engineering
University of Nebraska-Lincoln
Lincoln, NE 68588, U.S.A.

ABSTRACT

The traditional approach in ranking and selection procedures is to compare simulated systems based on the *mean* performance of a metric of interest. The system with the largest (or smallest) mean performance is deemed as the best system. However, the system with the best mean performance may be an undesirable choice because of its large variance. Variance is a measure of risk. A highly variable system performance shows that the system is not under control. Both mean and variance of a performance metric need to be compared to determine the best system. We present a statistically valid selection procedure for comparing simulated systems based on a *mean-variance* dominance relationship. The system with the best mean and smallest variance is deemed as the best system. If there is not a unique best system, the procedure identifies a set of nondominant systems. In both cases, a prespecified probability of correct selection is guaranteed.

1 INTRODUCTION

Computer simulation is often used to estimate the performance of complex stochastic systems. In practical applications decision makers are often interested in choosing the best system from a set of alternative systems. For example, automated guided vehicle systems are compared based on their throughput. When simulations are expensive to run, determining the number of simulation replications needed for each alternative to make a statistically valid selection decision becomes a challenge. Sequential ranking and selection (R&S) procedures are statistical procedures developed to simulate each alternative system until a valid statistical selection decision can be made.

In the simulation literature the interest has often been on selecting the best system based on the mean performance of a metric of interest. The system with the largest (or smallest) mean is deemed as the best system. Refer to [Kim and Nelson \(2006\)](#) for a comprehensive review of the R&S procedures in simulation. However, mean is only a measure of the average behavior, and the system with the best mean may have the largest variance. A common misconception is to confuse *estimation error* with *risk* ([Henderson and Nelson 2006](#)). The confidence interval of a mean performance metric is only a measure of the estimation error. It is not related to the risk associated with the system's performance. As the simulation run length or the number of simulation replications increases, the estimation error decreases; however, the risk associated with the system's performance remains the same. A system with large risk is unpredictable and hard to control. One way of measuring risk is through the variance. Hence, a selection criterion more comprehensive than the mean measure is the mean-variance dominance selection criterion. In the mean-variance dominance criterion, systems are compared based on both mean and variance measures. Refer to [Batur and Choobineh \(2009\)](#) for a discussion of different selection criteria in ranking and selection.

In the mean-variance based comparison approach, the system with the best (largest or smallest) mean and the smallest variance is deemed as the best system. Suppose $E_A(X)$ and $E_B(X)$ are the mean values and $\text{Var}_A(X)$ and $\text{Var}_B(X)$ are the variance values of the performance metric of interest for Systems A and B, respectively. System A has *mean-variance* dominance over System B if $E_A(X) \geq$ (or \leq) $E_B(X)$ and $\text{Var}_A(X) \leq \text{Var}_B(X)$ in the case of larger (or smaller) is better and at least one inequality holds strictly. If there is not a unique best system, a set of nondominant systems exists. Here we develop a sequential R&S procedure where comparison of the simulated systems is based on the mean-variance dominance relationship.

Studies in the mean-variance based comparison in the ranking and selection literature is limited. The authors in [Santner and Tamhane \(1984\)](#) propose to specify two integers a and b satisfying $1 \leq a, b \leq p$, where p is the number

of alternative systems. A system with a mean value among the a largest ones and variance among the b smallest ones is defined as a *good* system. The objective is to select the system with the largest mean from the set of good systems if the set is nonempty.

In the proposed mean-variance dominance R&S procedure, the mean and variance comparison tests are performed simultaneously, and sampling is sequential. When a system is detected to be inferior to another system based on both mean and variance measures, that system is eliminated. However, if one system is detected to be better based on the mean but worse based on the variance to another system, then the two systems are declared nondominant.

Although there are many single-objective R&S procedures in the literature for mere mean-based or variance-based comparison, they cannot be simply combined to rank systems based on the mean-variance dominance relationship. In those single-objective procedures if two systems have close performance, one is randomly selected as the best one. However, in the mean-variance dominance approach if two systems have close performance (are in the indifference zone) based on the mean (or variance) measure, they are declared indifferent based on that measure, and the selection decision is made based on the variance (or mean) measure.

The proposed mean-variance R&S procedure compares K systems. The procedure recommends a set of systems which may include one or more systems. If the number of systems in the resulting set is one, then that system is declared the *best* system based on the mean and variance measures with a prespecified probability of correct selection (PCS). However, if there are more than one system in the set, then those systems are declared *nondominant* based on both measures, again with a prespecified PCS.

In the proposed procedure it is assumed that the simulation output from each system are independent and identically distributed (IID) and normal. This assumption does not hold in general; however, approximately IID normal data can be obtained through appropriately batched output data or sample averages of independent replications.

In Section 2 the mean-variance based selection (MVS) procedure is presented followed by experiments in Section 3 and conclusion in Section 4.

2 METHODOLOGY

The problem is to compare K systems through simulation. The focus is on the comparison of the simulated systems based on both the mean and variance of the performance metric of interest. Let the unknown mean and variance values of the performance metric of the K systems be $\mu_1, \mu_2, \dots, \mu_K$ and $\sigma_1^2, \sigma_2^2, \dots, \sigma_K^2$, respectively. Without loss of generality it is assumed that a *larger* mean performance value is better. Under the mean-variance dominance criterion if two systems have the same variance values but different mean values, then the system with the larger mean is better than the other. Similarly if two systems have the same mean values but different variance values, then the system with the smaller variance is better than the other. Two systems are nondominant if they have the same mean and variance values or one system has a larger mean while the other has a smaller variance. Finally, a system that has larger mean but smaller variance than all the other systems is the best system.

Let $X_{i\ell}$, $i = 1, 2, \dots, K$, $\ell = 1, 2, \dots$ be the ℓ th observation from the simulation of system i . We assume that the simulation output data of each system are IID normal with mean μ_i and variance σ_i^2 , for $i = 1, 2, \dots, K$,

$$\{X_{i\ell} : \ell = 1, 2, \dots\} \stackrel{\text{IID}}{\sim} \text{Normal}(\mu_i, \sigma_i^2).$$

Also, it is assumed that systems are simulated independently.

Since it is not possible to establish dominance with complete certainty with a finite number of simulated observations, indifference-zone parameters are introduced for both the mean and variance measures. The mean indifference-zone parameter is denoted by δ ; it is the smallest practical difference worth detecting between the mean measures of two systems. The mean indifference-zone parameter used here has a slightly different functionality than the indifference-zone parameter used in mean-based R&S procedures in the literature. In a mean-based R&S procedure the decision maker wants to decide if the mean measure of one system is larger than that of another system, i.e., $\mu_1 > \mu_2$ or $\mu_1 < \mu_2$. If the mean values are within the indifference-zone range, i.e., $|\mu_1 - \mu_2| < \delta$, then selection of either system is considered a correct selection. However, in the mean-variance dominance approach presented here, the decision maker wants to decide if

$$\mu_1 > \mu_2, \mu_1 = \mu_2, \text{ or } \mu_1 < \mu_2.$$

This way if two systems are indifferent based on the mean measure, the dominant system is determined based on the variance measure.

By utilizing the indifference-zone parameter if $0 < \mu_1 - \mu_2 < \delta$, then both $\mu_1 > \mu_2$ and $\mu_1 = \mu_2$ decisions are considered correct. Similarly, if $-\delta < \mu_1 - \mu_2 < 0$, then $\mu_1 = \mu_2$ and $\mu_1 < \mu_2$ decisions are considered correct. On the other hand, if $\mu_1 - \mu_2 \geq \delta$, then $\mu_1 > \mu_2$ is the only correct decision. Similarly, if $\mu_1 - \mu_2 \leq -\delta$, then $\mu_1 < \mu_2$ is the only correct decision.

The indifference-zone parameter for the ratio of variances is denoted by R^2 where $R > 1$. It is the smallest practical ratio worth detecting in the ratio of the variances of two systems. The decision maker wants to decide if

$$\sigma_1^2/\sigma_2^2 > 1, \sigma_1^2/\sigma_2^2 = 1, \text{ or } \sigma_2^2/\sigma_1^2 > 1.$$

If $1 < \sigma_1^2/\sigma_2^2 < R^2$, then both $\sigma_1^2/\sigma_2^2 > 1$ and $\sigma_1^2/\sigma_2^2 = 1$ decisions are considered correct. However, if $1 < \sigma_2^2/\sigma_1^2 < R^2$, then $\sigma_2^2/\sigma_1^2 > 1$ and $\sigma_1^2/\sigma_2^2 = 1$ decisions are considered correct. If $\sigma_1^2/\sigma_2^2 \geq R^2$, then $\sigma_1^2/\sigma_2^2 > 1$ is the only correct decision. The same holds for the σ_2^2/σ_1^2 ratio.

In the proposed selection procedure, comparison of simulated systems based on the mean and variance measures are performed simultaneously. For each pairwise comparison based on the mean measure, a three-way hypothesis test is performed. The null hypothesis is the equivalence of the mean measures, i.e., $H_0^\mu : \mu_1 - \mu_2 = 0$. The alternative hypotheses are one mean measure being strictly greater than the other mean measure, i.e., $H_1^\mu : \mu_1 - \mu_2 > 0$ and $H_2^\mu : \mu_1 - \mu_2 < 0$. A similar three-way hypothesis test is performed for the ratio of the variances, i.e., $H_0^\sigma : \sigma_1^2/\sigma_2^2 = 1$, $H_1^\sigma : \sigma_1^2/\sigma_2^2 > 1$, and $H_2^\sigma : \sigma_2^2/\sigma_1^2 > 1$.

Rejection of the null hypothesis when it is true is called a type I error. The three-way hypothesis test is developed such that the probability of committing a type I error is at least α , and α is distributed between the two alternatives, i.e., $\Pr(\text{Accepting } H_1|H_0 \text{ is true}) \leq \alpha/2$ and $\Pr(\text{Accepting } H_2|H_0 \text{ is true}) \leq \alpha/2$. Nonrejection of the null hypothesis when it is false is called a type II error. The three-way hypothesis test is developed such that the probability of committing a type II error is at least β , i.e., $\Pr(\text{Do not reject } H_0|H_0 \text{ is false}) \leq \beta$.

In general, in hypothesis testing the user has control over the type I error; however, the type II error is uncontrollable. The user can decrease both errors by increasing the number of observations, but it is not possible to control the probability of committing a type II error unless a specific alternative hypothesis exists. Here in order to be able to control the probability of committing a type II error, the alternative mean hypotheses are transformed into $H_1^\mu : \mu_1 - \mu_2 \geq \delta$ and $H_2^\mu : \mu_1 - \mu_2 \leq -\delta$, where δ is the mean indifference-zone parameter. According to the transformed three-way test, if the difference between the mean measures is larger than the indifference-zone value δ , the test detects the true difference selecting the correct alternative hypothesis with a probability of at least $1 - \beta$. For example, if $\mu_1 - \mu_2 \geq \delta$, then the decision is $H_1^\mu : \mu_1 - \mu_2 > 0$ with a probability of at least $1 - \beta$. However, if the difference between the mean measures is in the range of 0 to δ , the test detects the true difference selecting the correct alternative hypothesis or fails to reject the null hypothesis with a total probability of at least $1 - \alpha$. For example, if $0 < \mu_1 - \mu_2 < \delta$, then the decision will be $H_0^\mu : \mu_1 = \mu_2$ or $H_1^\mu : \mu_1 - \mu_2 > 0$ with a total probability of at least $1 - \alpha$.

Similarly the alternative variance hypotheses are transformed into $H_1^\sigma : \sigma_1^2/\sigma_2^2 \geq R^2$ and $H_2^\sigma : \sigma_2^2/\sigma_1^2 \geq R^2$ where $R^2 > 1$ is the variance indifference-zone parameter. Similar to the mean measure if the ratio of the variances is larger than R^2 , the test detects the correct alternative hypothesis with a probability of at least $1 - \beta$. For example if $\sigma_1^2/\sigma_2^2 \geq R^2$, then the decision is $H_1^\sigma : \sigma_1^2/\sigma_2^2 > 1$ with a probability of at least $1 - \beta$. However, if the ratio of the variances is in the range of 1 to R^2 , then the test detects the true ratio selecting the true alternative hypothesis or fails to reject the null hypothesis with a total probability of at least $1 - \alpha$. For example, if $1 < \sigma_1^2/\sigma_2^2 < R^2$, then the decision is $H_0^\sigma : \sigma_1^2/\sigma_2^2 = 1$ or $H_1^\sigma : \sigma_1^2/\sigma_2^2 > 1$ with a total probability of at least $1 - \alpha$.

Paulson (1964) presents sequential procedures for testing the three-way mean and variance tests presented above. These tests are shown to satisfy the prespecified probability of committing type I and II errors, i.e.,

$$\begin{aligned} \Pr(\text{Do not reject } H_0^\mu \mid |\mu_1 - \mu_2| \geq \delta) &\leq \beta, \\ \Pr(\text{Accept } H_1^\mu \mid \mu_1 - \mu_2 = 0) &\leq \alpha/2, \\ \Pr(\text{Accept } H_2^\mu \mid \mu_1 - \mu_2 = 0) &\leq \alpha/2. \end{aligned}$$

In the comparison of K systems for each measure, $K(K - 1)/2$ pairwise comparisons are performed. The probability of committing type I or II errors for each pairwise comparison and measure is set to $\alpha/[K(K - 1)]$ in the procedure. This assures that the overall probability of selecting the correct set of nondominant systems is at least $1 - \alpha$.

The proposed mean-variance based selection (MVS) procedure is presented in Figure 1. In the Setup step, the parameters of the procedure are specified: i) the probability $1 - \alpha$ of correctly selecting the set of nondominant systems and ii) the indifference-zone parameter δ for the mean and R^2 for the variance. Set I is the set of nondominated systems; it initially includes all K systems. Set N is the set of pairs of nondominant systems; it is initially an empty set. The probability of committing type I or II error α_K for each test is set to $\alpha/[K(K - 1)]$. The constant parameter d for the mean test and the constant parameter λ for the variance test are determined based on δ and R , respectively. Finally, the initial sample size n_0 is determined from α_K and λ .

In the Initialization step, n_0 observations are generated from every system $i \in I$, and the number of replications r is set to n_0 . In the Calculation step, the following statistics are calculated. The sample mean and variances are calculated for every system $i \in I$. The sample variance of the differences are calculated for every pair of systems $i, j \in I, i \neq j$. The $a_{ij}(r)$ and $b_{ij}(r)$ statistics are calculated for the mean hypothesis tests, and the $u_{ij}(r)$ and $v_{ij}(r)$ statistics are calculated for the variance hypothesis tests. In the Decision step, decisions of the hypothesis tests for each pair of systems are

Procedure MVS

Setup: Set the probability of selecting the correct set of nondominant systems to $1 - \alpha$. Specify the mean indifference-zone parameter δ and the variance indifference-zone parameter R^2 . Define I as the set of initial K systems. Define N as the set of pairs of nondominant systems, and set $N = \emptyset$. Set the probability of committing type I or II errors for each three-way hypothesis test to $\alpha_K = \alpha/[K(K - 1)]$. Set parameter d for the mean test to $3\delta/8$ and parameter λ for the variance test to $1 + (0.7)(R - 1)$. Set the initial sample size n_0 to $\lceil 1 - \ln(\alpha_K)/\ln(\lambda) \rceil + 1$.

Initialization: Generate n_0 observations from each system $i \in I$. Set the number of replications r to n_0 .

Calculation: For every $i \in I$ calculate the sample means $\bar{X}_i(r) = (1/r)\sum_{\ell=1}^r X_{i\ell}$ and the sample variances $S_i^2(r) = (1/(r-1))\sum_{\ell=1}^r (X_{i\ell} - \bar{X}_i(r))^2$. For every $i, j \in I, i \neq j$, calculate the sample variance of the differences

$$S_{ij}^2(r) = \frac{1}{r-1} \sum_{\ell=1}^r [X_{i\ell} - X_{j\ell} - (\bar{X}_i(r) - \bar{X}_j(r))]^2.$$

For the comparison of the means of system pairs $i, j \in I, i \neq j$, calculate

$$a_{ij}(r) = rd + \frac{S_{ij}^2(r) \left[(2/\alpha_K)^{\frac{2}{r-1}} - 1 \right] (r-1)}{4d} \quad \text{and} \quad b_{ij}(r) = r(\delta - d) - \frac{S_{ij}^2(r) \left[(1/\alpha_K)^{\frac{2}{r-1}} - 1 \right] (r-1)}{4d}.$$

For the comparison of the variances of system pairs $i, j \in I, i \neq j$, calculate

$$u_{ij}(r) = \frac{\lambda \left(S_i^2(r)/S_j^2(r) \right) \left(\lambda - \alpha_K^{\frac{1}{r-1}} \right)}{\left(\lambda \alpha_K^{\frac{1}{r-1}} - 1 \right)} \quad \text{and} \quad v_{ij}(r) = \frac{\left[S_i^2(r)/S_j^2(r) \right] \left[\lambda - (2/\alpha_K)^{\frac{1}{r-1}} \right]}{\lambda \left[\lambda (2/\alpha_K)^{\frac{1}{r-1}} - 1 \right]}.$$

Decision: For every $i, j \in I, i \neq j$:

If $-b_{ij}(r) < \sum_{\ell=1}^r (X_{i\ell} - X_{j\ell}) < b_{ij}(r)$, decide $\mu_i = \mu_j$.

If $\sum_{\ell=1}^r (X_{i\ell} - X_{j\ell}) > a_{ij}(r)$, decide $\mu_i > \mu_j$.

If $\sum_{\ell=1}^r (X_{i\ell} - X_{j\ell}) < -a_{ij}(r)$, decide $\mu_i < \mu_j$.

If $u_{ij}(r) < R^2$ and $u_{ji}(r) < R^2$, decide $\sigma_i^2 = \sigma_j^2$.

If $r > \lceil 1 - \ln(\alpha_K/2)/\ln(\lambda) \rceil$ and $v_{ij}(r) > 1$, decide $\sigma_i^2 > \sigma_j^2$.

If $r > \lceil 1 - \ln(\alpha_K/2)/\ln(\lambda) \rceil$ and $v_{ji}(r) > 1$, decide $\sigma_i^2 < \sigma_j^2$.

Elimination: For every $i, j \in I, i \neq j$:

If i) $\sigma_i^2 > \sigma_j^2$ and $\mu_i = \mu_j$, ii) $\sigma_i^2 = \sigma_j^2$ and $\mu_i < \mu_j$, or iii) $\sigma_i^2 > \sigma_j^2$ and $\mu_i < \mu_j$ decisions are made, eliminate system i and set $I = I/\{i\}$.

If i) $\sigma_i^2 < \sigma_j^2$ and $\mu_i = \mu_j$, ii) $\sigma_i^2 = \sigma_j^2$ and $\mu_i > \mu_j$, or iii) $\sigma_i^2 < \sigma_j^2$ and $\mu_i > \mu_j$ decisions are made, eliminate system j and set $I = I/\{j\}$.

If i) $\sigma_i^2 = \sigma_j^2$ and $\mu_i = \mu_j$, ii) $\sigma_i^2 > \sigma_j^2$ and $\mu_i > \mu_j$, or iii) $\sigma_i^2 < \sigma_j^2$ and $\mu_i < \mu_j$ decisions are made, declare systems i and j nondominant and set $N = N \cup \{(i, j)\}$.

Termination: If $|I| = 1$, then stop and declare system $i \in I$ as the best system. If set N includes every (i, j) pair where $i, j \in I, i \neq j$, then stop and declare I as the set of nondominant systems.

Otherwise take one more observation from every system $i \in I$, set $r = r + 1$, and go back to the Calculation step.

Figure 1: Algorithmic statement for mean-variance based selection.

made using these statistics and the indifference-zone parameters. For a pair of systems, if none of the conditions are satisfied, it means that the hypothesis test is inconclusive at stage r and more observations are needed.

In the Elimination step, according to the results of the hypothesis tests for the mean and variance, elimination and nondominance decisions are made. If a system is detected to be worse than another system in I based on both the mean and variance measures, then that system is eliminated and deleted from I . On the other hand, if two systems are detected to be nondominant, then that pair is added to the set of nondominant pairs N . In the Termination step, there

are two stopping rules. The first rule is to stop if set I includes only one system. In this case the system in set I is declared as the best system with probability $1 - \alpha$ based on both the mean and variance measures. The second rule is to stop if all pairs of the systems in set I is included in set N . This means that for every pair of systems that have not been eliminated, a nondominance decision is made. In this case, set I is declared the set of nondominant systems with probability $1 - \alpha$ based on both the mean and variance measures.

3 EXPERIMENTS

We tested the performance of the MVS procedure on a testbed of systems with different mean and variance configurations. The testbed consists of five configurations. The performance metric of interest from each system is assumed to be normally distributed. Also it is assumed that a *larger* mean is preferred.

The experiment configurations are presented in Table 1. The correct set of nondominant systems for each configuration is also shown in the table. In Configuration 1, the first system's mean and variance are both 1. The means and variances of the other systems are $1 - \delta$ and R^2 , respectively. In this configuration System 1 is the best system because it has the largest mean and smallest variance. These mean and variance configurations are called *slippage* configurations because the differences in the mean values between the best system and the other systems are all equal to the mean indifference-zone parameter δ . Similarly, the ratios of the variances of the inferior systems to the best system are all equal to the variance indifference-zone parameter R^2 .

Table 1: Configurations of the systems tested ($\delta > 0$ and $R > 1$).

		Sys. 1	Sys. 2	...	Sys. i	...	Sys. K	Correct N set
Configuration 1	Mean	1	$1 - \delta$...	$1 - \delta$...	$1 - \delta$	$\{1\}$
	Variance	1	R^2	...	R^2	...	R^2	
Configuration 2	Mean	1	$1 - \delta$...	$1 - \delta$...	$1 - \delta$	$\{1\}$
	Variance	1	1	...	1	...	1	
Configuration 3	Mean	1	1	...	1	...	1	$\{1\}$
	Variance	1	R^2	...	R^2	...	R^2	
Configuration 4	Mean	1	$1 - \delta$...	$1 - (i - 1)\delta$...	$1 - (K - 1)\delta$	$\{\text{All systems}\}$
	Variance	$R^{2(K-1)}$	$R^{2(K-2)}$...	$R^{2(K-i)}$...	1	
Configuration 5	Mean	1	$1 - \delta$...	$1 - (i - 1)\delta$...	$1 - (K - 1)\delta$	$\{1\}$
	Variance	1	R^2	...	$R^{2(i-1)}$...	$R^{2(K-1)}$	

In Configuration 2, the mean measures have the slippage configuration and the variances are all ones. In this configuration system 1 is the best because systems are equivalent based on the variance but system 1 is better based on the mean. In Configuration 3, the variance measures have the slippage configuration and the means are all ones. In this configuration system 1 is the best because systems are equivalent based on the mean but system 1 is better based on the variance.

In Configuration 4, both mean and variance values decrease as the number of systems K increases. Hence, if a system has larger mean, it also has a larger variance. So all systems are nondominant. In Configuration 5, the mean values decrease but the variance values increase as K increases. Hence, system 1 is the best system with the largest mean and smallest variance.

In the experiments the nominal PCS is set to $1 - \alpha = 0.95$. The mean indifference-zone value is set to $\delta = 0.1$, and the variance indifference-zone value is set to $R^2 = (1.1)^2$. We replicate each experiment 1000 times. The following statistics are reported: the estimated PCS and the sample average of total number of observations (SATO) that is required to select the set of nondominant systems. The SATO value provides a measure of the computational efficiency of the procedure.

In Table 2 the results of the experiments are presented when $K = 2$. As seen in the experiments all PCS values are above the prespecified level $1 - \alpha = 0.95$. The results of the experiments when $K = 5$ are presented in Table 3. When we compare the SATO values in Tables 2 and 3, we observe that Configuration 4 requires significantly more observations when $K = 5$ compared to $K = 2$ case. This is due to the fact this configuration includes systems which are nondominant such that the mean and variance measures of the neighboring systems only differ by the indifference-zone amount. Hence, as the number of these nondominant systems increases, it becomes more and more difficult for the procedure to make a decision. On the other hand, we observe that increasing the number of systems in Configuration 5 has a smaller increasing effect in SATO because the new systems with smaller means and larger variances are easier to eliminate by the best system.

In order to show the performance of the MVS procedure when systems have mean or variance values within the indifference zone, additional system configurations shown in Table 4 are tested. In Configuration 6 the mean performance measures are within the indifference zone. So the MVS procedure is expected to make $\mu_1 < \mu_2$ or $\mu_1 = \mu_2$ decisions. Both decisions are considered correct because the mean of system 2 is larger than the mean of system 1;

Table 2: Results of the experiments when $K = 2, \delta = 0.1$, and $R = 1.1$.

	SATO	PCS
Configuration 1	4,135	1.00
Configuration 2	3,877	0.97
Configuration 3	5,055	0.98
Configuration 4	4,127	0.97
Configuration 5	4,243	1.00

Table 3: Results of the experiments when $K = 5, \delta = 0.1$, and $R = 1.1$.

	SATO	PCS
Configuration 1	17,226	1.00
Configuration 2	16,068	0.99
Configuration 3	20,464	0.99
Configuration 4	34,569	0.98
Configuration 5	9,484	1.00

however, the difference is so small that the equality decision is also considered correct. When these two decisions are combined with the fact that the variance of system 2 is larger than the variance of system 1, the selection of system 1 as the best system or the selection of both systems as nondominant are considered correct.

Table 4: Configurations of the systems within the indifference zone ($\delta > 0$ and $R > 1$).

		Sys. 1	Sys. 2	Correct N set
Configuration 6	Mean	1	$1 + \delta/2$	$\{1\}$ or $\{1, 2\}$
	Variance	1	R^2	
Configuration 7	Mean	1	$1 + \delta$	$\{2\}$ or $\{1, 2\}$
	Variance	1	R_s^2 where $1 < R_s < R$	

In Configuration 7 the variance values are within the indifference zone. So both $\sigma_1 < \sigma_2$ and $\sigma_1 = \sigma_2$ decisions are considered correct. When these two decisions are combined with the fact that the mean of system 2 is larger than the mean of system 1, the selection of system 2 as the best system or the selection of both systems as nondominant are considered correct.

The results of the experiments for these configurations are presented in Table 5. The R_s value is set to $1.05 < R$. As seen in the results for both configurations the PCS value is larger than the prespecified PCS of 0.95.

4 CONCLUSION AND FUTURE WORK

The traditional approach in ranking and selection procedures in simulation is to compare systems based on a mean performance metric of interest. However, mean is only a measure of the average behavior of the system. A system with the best mean can have a very large variance. Variance is a measure of the risk associated with the system performance. A system with high variance is generally undesirable. Hence, systems must be compared based on both the mean and variance measures. In this paper we present a mean-variance dominance based ranking and selection procedure for comparing K systems based on a performance metric of interest. In this procedure simulated systems are compared based on both the mean and variance performance measures. The system with the largest (or smallest) mean and the smallest variance is defined as the best system. If some systems are better in terms mean but worse in terms of the variance, then the procedure selects a set of nondominant systems with a prespecified probability of correct selection.

As a future work we plan to extend this procedure to the comparison of systems based on multiple performance metrics, e.g., comparing alternative distribution networks based on the delivery times and unit delivery costs. Also, comparing systems in the existence of stochastic constraints involving mean or variance measures is a problem that can be tackled by the selection technique presented in this paper.

Table 5: Results of the experiments when $K = 5$.

	SATO	PCS
Configuration 6	6,430	0.99
Configuration 7	4,493	0.99

ACKNOWLEDGMENTS

This material was supported by the National Science Foundation under Grant Number EPS-0701892. We would like to thank Lina Wang for preparing the code of the procedure.

REFERENCES

- Batur, D., and F. Choobineh. 2009. Do mean-based ranking and selection procedures consider systems' risk? In *Proceedings of the 2009 Winter Simulation Conference*, ed. M. D. Rossetti, R. R. Hill, B. Johansson, A. Dunkin, and R. G. Ingalls, 423–433. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.
- Henderson, S. G., and B. L. Nelson. 2006. Stochastic computer simulation. In *Handbooks in Operations Research and Management Science: Simulation*, ed. S. G. Henderson and B. L. Nelson, 1–18. Oxford: Elsevier.
- Kim, S.-H., and B. L. Nelson. 2006. Selecting the best system. In *Handbooks in Operations Research and Management Science: Simulation*, ed. S. G. Henderson and B. L. Nelson, 501–534. Oxford: Elsevier.
- Paulson, E. 1964. Sequential estimation and closed sequential decision procedures. *The Annals of Mathematical Statistics* 35 (3): 1048–1058.
- Santner, T. J., and A. C. Tamhane. 1984. Designing experiments for selecting a normal population with a large mean and a small variance. In *Design of Experiments: Ranking and Selection*, ed. T. J. Santner and A. C. Tamhane 179–198. New York: Marcel Dekker.

AUTHOR BIOGRAPHIES

DEMET BATUR is a lecturer in Industrial and Management Systems Engineering at the University of Nebraska-Lincoln. She received her Ph.D. in Industrial and Systems Engineering from the Georgia Institute of Technology. Her research interests are in simulation output analysis and ranking and selection procedures. She is a member of INFORMS and IIE. Her email address is dbatur2@unl.edu.

F. FRED CHOOBINEH is a professor of Industrial and Management Systems Engineering and Milton E. Mohr Distinguished Professor of Engineering at the University of Nebraska-Lincoln where he also holds a courtesy appointment as a professor of Management. He received his B.S. in Electrical Engineering, Masters of Engineering, and Ph.D. in Industrial Engineering from Iowa State University in Ames, Iowa. He is a Fellow of the Institute of Industrial Engineers and a member of IEEE and INFORMS. His research interests include design and control of manufacturing systems and decision analysis. His research has been funded by NSF and industry. He is a registered Professional Engineer in Nebraska. His email address is fchoobineh1@unl.edu.