Proceedings of the 2015 Winter Simulation Conference
L. Yilmaz, W. K. V. Chan, I. Moon, T. M. K. Roeder, C. Macal, and M. D. Rossetti, eds.

SUBSET SELECTION FOR SIMULATIONS ACCOUNTING FOR INPUT UNCERTAINTY

Canan G. Corlu

Metropolitan College
Boston University
Boston, MA 02215, USA

Bahar Biller

Software Sciences and Analytics
GE Global Research Center
Niskayuna, NY 12309, USA

ABSTRACT

We study a subset selection procedure – one of the well-known statistical methods of ranking and selection for stochastic simulations – in the presence of input parameter uncertainty; i.e., the parameters of the input distributions are unknown and there is only a limited amount of input data available for input parameter estimation. The goal is to present a new decision rule which identifies subsets of stochastic system designs including the best (i.e., the design with the largest or smallest expected performance measure) with a probability that exceeds some user-specified value. At WSC 2013, we studied this problem by restricting focus to the method of asymptotic normality approximation to represent input parameter uncertainty. Motivated by the limitations of the asymptotic normality approximation for simulations of complex systems with large numbers of input parameters, we revisit this problem with the simulation replication algorithm as an alternative method to capture input parameter uncertainty.

1 INTRODUCTION

An important step in the use of simulation for stochastic system design and analysis is to evaluate competing system designs. Despite the examples demonstrating the importance of capturing the uncertainty around the input parameter estimates obtained from limited input data (e.g., see Zouaoui and Wilson (2004) for a queuing simulation example and Biller et al. (2014) for an inventory simulation example), accounting for input parameter uncertainty in a stochastic simulation is still not the standard practice. Consequently, there is a lack of statistically-valid procedures that compare alternative simulated system designs by accounting for input parameter uncertainty. However, with the growing use of stochastic simulation in business system design and analysis, it is imperative to develop data-driven modeling support for the statistical comparison of alternative system designs in the absence of full knowledge about the systems' input distribution parameters. Our paper addresses this limitation of the simulation design and analysis by focusing on subset selection, which is one of the well-known statistical methods of ranking and selection for stochastic simulations.

When the input distributions and their parameters are unknown and the historical input data available for their estimation are limited, two of the main sources of uncertainty to represent in simulation output analysis are stochastic uncertainty (i.e., the uncertainty that is due to the dependence of the simulation output on random input processes) (Helton 1997) and parameter uncertainty (i.e., the uncertainty that is due to the estimation of the input-model parameters from limited data) (Raftery et al. 1996). Model uncertainty also arises due to the selection of a single input model from a set of alternative models. However, we assume that the functional forms of all input models are known; thus, we do not consider the model uncertainty.

Stochastic uncertainty is inherent in every simulation and controlled by the number of simulation replications. The parameter uncertainty is, on the other hand, often ignored as a result of driving simulations with the parameters of the probability distributions estimated from input data of finite length. This practice of simulation design and analysis leads to not only inconsistent estimates for the mean performance measures

but also inconsistent coverage of the confidence intervals (Barton 2012). In this paper, we account for the stochastic and parameter uncertainties in the analysis of simulation output data with the objective of selecting a subset of stochastic system designs, which includes the best with a probability that exceeds a user-specified value.

Goldsman and Nelson (1998) describe ranking and selection procedures as statistical methods specifically developed to select the best system, or a subset of systems that includes the best system, from among a collection of (competing) alternative system designs, while guaranteeing the probability of a correct selection to be at least some user-specified value under certain assumptions. A close look at the existing literature reveals that the development of the ranking and selection procedures dates back to the 1950s with Bechhofer (1954) establishing the indifference-zone formulation and Gupta (1956, 1965) formulating the subset selection. Both approaches are known to compensate for the limited inference provided by the hypothesis tests in identifying the system design with the best (largest or smallest) performance measure. Over the years, there has appeared various extensions of these procedures. We focus on the well-known subset-selection formulation and refer the reader to Kim (2013) for a comprehensive review of this literature.

Subset selection is a screening device that attempts to select a subset of system designs including the best with at least a pre-specified probability. Specifically, we consider k alternative stochastic system designs with input random vectors $\mathbf{Y}_i$, $i = 1,2,\ldots,k$ as the sources of uncertainty driving their stochastic simulations. Defining $X_{i\ell}$ for the ℓth simulation output from system i, we let $\mathbf{X}_i := \{X_{i\ell}; \ell = 1,2,\ldots,n_i\}$ denote the output sequence of n_i replications from the simulation. Under the assumption of complete knowledge about the distributions of the input random vectors $\mathbf{Y}_i$, $i = 1,2,\ldots,k$ (i.e., for the given functional forms of the input distributions, the input parameter vectors Ψ_i, $i = 1,2,\ldots,k$ are known), we consider $X_{i\ell}$, $\ell = 1,2,\ldots,n_i$ to be independent and normally distributed random variables each with a mean of $\mu_i := \mathbb{E}[X_{i\ell}]$ and a variance of $\sigma_i^2 := \mathbb{V}[X_{i\ell}]$ for $i = 1,2,\ldots,k$. This is often a reasonable assumption when $X_{i\ell}$ is the average of a large number of output variables or $X_{i\ell}$ is a batch mean in a steady-state simulation. We investigate the sensitivity of the performance of our procedures to the departure of the simulation output data from being normally distributed in Biller and Corlu (2015) where we discuss when it becomes important to couple our procedures with the method of batching to ensure the normality of the batched simulation output data. We also assume that $\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_k$ are mutually independent and compared based on their true means. Finally, we let the best design have the largest true mean and denote the ordered means by $\mu_{[k]} \geq \mu_{[k-1]} \geq \cdots \geq \mu_{[1]}$.

The goal of a subset selection procedure is to obtain a (random) subset $I \subseteq \{1,2,\ldots,k\}$ of k alternatives such that the probability of including the best system design in this subset, $\mathbb{P}\{[k] \in I\}$ is at least $1-\alpha$ with $\alpha \in (0, 1-k^{-1})$. Under the assumptions of known output variances σ_i^2, $i = 1,2,\ldots,k$ and n equal simulation replications across all of the k system designs, the subset I is characterized by Gupta (1965) as

$$I := \left\{ i : 1 \leq i \leq k,\ 1 \leq j \leq k,\ i \neq j, \text{ and } \bar{X}_i \geq \bar{X}_j - z_{(1-\alpha)^{1/(k-1)}} \left(\frac{\sigma_i^2 + \sigma_j^2}{n} \right)^{1/2} \right\}. \tag{1}$$

$\bar{X}_i := \sum_{\ell=1}^{n} X_{i\ell}/n$ is the sample mean of the simulation output data collected for system i and $z_{(1-\alpha)^{1/(k-1)}}$ is the $(1-\alpha)^{1/(k-1)}-$ quantile of the standard normal cumulative distribution function. Nelson et al. (2001) extend this single-stage subset selection procedure to permit unequal and unknown variances while Wilson (2001) enhances the lower bound on the probability of correct selection with a decomposition lemma for the screen-to-the-best procedure of Nelson et al. (2001). Boesel et al. (2003) further extend this procedure to allow unknown output variances σ_i^2, $i = 1,2,\ldots,k$ with unequal sample sizes n_i, $i = 1,2,\ldots,k$. The following rule is presented with $\bar{X}_i := \sum_{\ell=1}^{n_i} X_{i\ell}/n_i$ for $i = 1,2,\ldots,k$ and all values of $j \neq i \in \{1,2,\ldots,k\}$:

$$\bar{X}_i \geq \bar{X}_j - \left(t^2_{(1-\alpha)^{1/(k-1)},n_i-1} \frac{S_i^2}{n_i} + t^2_{(1-\alpha)^{1/(k-1)},n_j-1} \frac{S_j^2}{n_j} \right)^{1/2} \tag{2}$$

with $S_i^2 := \sum_{\ell=1}^{n_i} (X_{i\ell} - \bar{X}_i)^2/(n_i-1)$ as the unbiased sample variance estimator of σ_i^2 and $t_{(1-\alpha)^{1/(k-1)},n_i-1}$ as the $(1-\alpha)^{1/(k-1)}-$ quantile of the Student's t distribution with n_i-1 degrees of freedom for $i = 1,2,\ldots,k$.

Both the decision rule in (1) and the decision rule in (2) are obtained under the assumption of complete knowledge about the input distribution parameter vectors Ψ_i, $i = 1,2,\ldots,k$. In this paper, we relax this assumption and answer the following two questions: (1) When the input parameter vectors are estimated from the limited amounts of input data, how to capture the input parameter uncertainty in the simulation output data? (2) Given the characterization of the input parameter uncertainty in the simulation output data, how to redesign the subset-selection procedure to deliver a subset which still includes the best still with a probability that exceeds some user-specified value? We answer the first question in Section 2.1 where we present the simulation replication algorithm as a method of capturing the input parameter uncertainty. We answer the second question in Section 2.2 where we introduce a subset selection procedure accounting for the input uncertainty. Then, we illustrate the use of the procedure in Section 3 to find the inventory system designs maximizing the service level in the presence of limited amounts of historical demand data. Finally, we conclude with a summary of findings in Section 4.

2 SUBSET SELECTION UNDER PARAMETER UNCERTAINTY

We present the simulation replication algorithm to represent the input parameter uncertainty in Section 2.1 and redesign the subset-selection decision rule (2) to account for the input parameter uncertainty in Section 2.2.

2.1 Representing Input Parameter Uncertainty

In this section, we review the simulation replication algorithm for the decomposition of the simulation output variance into terms related to stochastic uncertainty and parameter uncertainty. We refer the reader to Barton (2012) for a comprehensive review of the methods of capturing parameter uncertainty in simulations.

Chick (2001) is the first to propose a simulation replication algorithm that has allowed simulation analysts to capture parameter uncertainty by sampling input distribution parameters from their density functions before each simulation replication. Focusing on the *i*th system design, we provide the simulation replication algorithm in Figure 1 as presented in Zouaoui and Wilson (2003) due to its ability to separately quantify the amounts of stochastic uncertainty and input parameter uncertainty in the simulation output data. Later in Section 2.2, we build on this output decomposition to reformulate the subset-selection procedure.

It is important to note that Zouaoui and Wilson (2003) present the simulation replication algorithm in Figure 1 as Bayesian with the ability to incorporate expert opinion into the representation of the input parameter uncertainty in simulation. This may potentially be a point of confusion for the reader because our decision rule in Section 2.2 and evaluations in Section 3 are frequentist. It is important to clarify that a Bayesian simulation replication algorithm can be readily replaced by its frequentist counterpart via the use of sampling density functions instead of posterior density functions for capturing the input parameter uncertainty. More specifically, the probability distribution functions $p_i(\Psi_i|\mathbf{d}_i)$, $i = 1,2,\ldots,R_i$ from which we generate the unknown input parameters Ψ_{ir} in Figure 1 can be the sampling distributions for the maximum likelihood estimates of these unknown input parameters. A sampling density function tells us which values the input parameters, which are the random variables, can have and how likely it is for the parameters to assume those values (Rohatgi 1976). Thus, the use of the sampling density functions in the simulation replication algorithm allows us to account for the uncertainty around the input-parameter estimates in the simulation outputs. Alternatively, the input parameter uncertainty can be captured by the posterior density functions of the unknown input parameters with the ability to account for the opinions

for $r = 1,2,\ldots,R_i$ replications **do**

generate the rth input parameter vector Ψ_{ir} from $p_i(\Psi_i|\mathbf{d}_i)$;

for $\ell = 1,2,\ldots,n_i$ **do**

generate the random-number input $\mathbf{u}_{i\ell r}$;

perform the ℓth simulation run using $\mathbf{u}_{i\ell r}$ and Ψ_{ir};

calculate the output response $X_{i\ell r} := \mathscr{G}_i(\mathbf{u}_{i\ell r}, \Psi_{ir})$;

end for

compute $\bar{X}_{ir} := \sum_{\ell=1}^{n_i} X_{i\ell r}/n_i$;

end for

compute $\bar{X}_i := \sum_{r=1}^{R_i} \bar{X}_{ir}/R_i$ as an estimate of $\mathbb{E}[X_i|\mathbf{d}_i]$.

Figure 1: Simulation replication algorithm.

of the experts, as described in Zouaoui and Wilson (2003). Whether a Bayesian or frequentist approach is used to represent parameter uncertainty is not a requirement of our subset selection procedure. Independent of which one of these two approaches is followed to quantify the parameter uncertainty in the simulation output data, we build on the underlying output variance decomposition to reformulate the subset-selection procedure in the following section.

In Figure 1, we denote the ith system parameter vector sampled in the rth (outer) replication of the simulation replication algorithm by Ψ_{ir}. In addition, we use $X_{i\ell r}$ for representing the output response from the ℓth (inner) simulation run, which is driven by the sampled values of the random-number input $\mathbf{u}_{i\ell r}$ and the input parameter vector Ψ_{ir}. The outer loop of the simulation replication algorithm estimates the uncertainty around the parameter vector Ψ_i by independently sampling new values of the input parameters from their joint density function $p_i(\Psi_i|\mathbf{d}_i)$ across R_i different simulation runs. The inner loop estimates the stochastic uncertainty by performing n_i independent runs conditional on these parameter values. Hence, this algorithm can be considered as an uncertainty decomposition algorithm (Zouaoui and Wilson 2003).

Next, we formulate the output process of the simulation replication algorithm for $X_{i\ell r}$, $\ell = 1,2,\ldots,n_i$, $r = 1,2,\ldots,R_i$ to obtain a characterization of the output response variance in terms of the stochastic uncertainty σ_i^2 and the input parameter uncertainty τ_i^2. Following Zouaoui and Wilson (2003), we express the output response from the rth simulation run of system i as $X_{i\ell r} = \mathscr{G}_i(\mathbf{u}_{i\ell r}, \Psi_{ir}) = g_i(\Psi_{ir}) + \sigma_i Z_{i\ell r}$. Under the hierarchical normal model, we assume that $Z_{i\ell r}$, $\ell = 1,2,\ldots,n_i$ are independent and identically distributed normal random variables, each with a mean of zero and a standard deviation of one. We use $\sigma_i Z_{i\ell r}$ to represent the deviation of the simulation output from the response surface $g_i(\Psi_{ir})$ due to the stochastic uncertainty whose source is the random-number input $\mathbf{u}_{i\ell r}$ for that run. Under the assumptions of $\mathbb{E}[\sigma_i Z_{i\ell r} \mid \mathbf{d}_i, \Psi_{ir}] = 0$ and $\mathbb{V}[\sigma_i Z_{i\ell r} \mid \mathbf{d}_i, \Psi_{ir}] = \sigma_i^2$, it holds that $\mathbb{E}[X_{i\ell r} \mid \mathbf{d}_i, \Psi_{ir}] = g_i(\Psi_{ir})$ and $\mathbb{V}[X_{i\ell r} \mid \mathbf{d}_i, \Psi_{ir}] = \sigma_i^2$ for $\ell = 1,2,\ldots,n_i$ and $r = 1,2,\ldots,R_i$. Furthermore, we assume that $g_i(\Psi_{ir}) = \beta_i + \eta_i(\Psi_{ir})$, where $\beta_i := \mathbb{E}_{\Psi_{ir}}[g_i(\Psi_{ir})|\mathbf{d}_i]$ is an unbiased estimator of the mean output response, $\mathbb{E}_{\Psi_{ir}}[\eta_i(\Psi_{ir}) \mid \mathbf{d}_i]$ is equal to zero, and $\mathbb{V}_{\Psi_{ir}}[\eta_i(\Psi_{ir}) \mid \mathbf{d}_i] := \tau_i^2$ stands for the uncertainty associated with the unknown input parameters of the ith system design. This well-known response-surface representation (Rao 1997) has been also suggested by Freimer and Schruben (2002), Ankenman and Nelson (2012), and Song and Nelson (2013) for the characterization of the input uncertainty in the simulation output processes. We use this representation in a similar manner to decompose

ith design response variance into components of stochastic uncertainty (σ_i^2) and parameter uncertainty (τ_i^2):

$$\begin{aligned}
\mathbb{V}[X_{i\ell r}|\mathbf{d}_i] &= \mathbb{V}_{\Psi_{ir}}\left[\mathbb{E}\left[\mathscr{G}_i(\mathbf{u}_{i\ell r},\Psi_{ir})|\mathbf{d}_i,\Psi_{ir}\right]|\mathbf{d}_i\right]+\mathbb{E}_{\Psi_{ir}}\left[\mathbb{V}\left[\mathscr{G}_i(\mathbf{u}_{i\ell r},\Psi_{ir})|\mathbf{d}_i,\Psi_{ir}\right]|\mathbf{d}_i\right] \\
&= \mathbb{V}_{\Psi_{ir}}\left[g_i(\Psi_{ir})|\mathbf{d}_i\right]+\mathbb{E}_{\Psi_{ir}}\left[\mathbb{V}\left[\sigma_i Z_{i\ell r}|\mathbf{d}_i,\Psi_{ir}\right]|\mathbf{d}_i\right] \\
&= \mathbb{V}_{\Psi_{ir}}\left[\beta_i+\eta_i(\Psi_{ir})|\mathbf{d}_i\right]+\mathbb{E}_{\Psi_{ir}}\left[\sigma_i^2|\mathbf{d}_i\right] \\
&= \tau_i^2+\sigma_i^2. \qquad (3)
\end{aligned}$$

Zouaoui and Wilson (2003) present procedures to estimate β_i, τ_i, and σ_i^2 using the output data $X_{i\ell r}$, $\ell = 1,2,\ldots,n_i$, $r = 1,2,\ldots,R_i$. Our purpose is, however, not the estimation of these parameters of the simulation response model. We develop subset selection procedures by building on the variance decomposition in (3) as we account for the stochastic and parameter uncertainties in the simulation outputs.

2.2 Subset Selection with Simulation Replication Algorithm

In this section, we represent the simulation output processes of the simulation replication algorithm by $X_{i\ell r} = \beta_i + \eta_{ir}(\Psi_{ir}) + \sigma_i Z_{i\ell r}$, $r = 1,2,\ldots,R_i$, $\ell = 1,2,\ldots,n_i$, $i = 1,2,\ldots,k$. We also use SR_i^2 for denoting $R_i^{-1}(R_i-1)^{-1}\sum_{r=1}^{R_i}(\bar{X}_{ir}-\bar{X}_i)^2$. Building on the structure of the simulation replication algorithm presented in Section 2.1, we consider a decision rule of the form $\bar{X}_{[k]} \geq \bar{X}_{[i]} - \mathscr{L}_{[k][i]}$, $i = k-1,k-2,\ldots,1$ with $\mathscr{L}_{[k][i]}^2 = t_{[k]}^2 SR_{[k]}^2 + t_{[i]}^2 SR_{[i]}^2$ such that $\mathbb{P}\{\bar{X}_{[k]} \geq \bar{X}_{[i]} - \mathscr{L}_{[k][i]}, i = k-1,k-2,\ldots,1\} \geq 1-\alpha$. First, we note that

$$\bar{X}_{[i]r} \sim \mathscr{N}\left(\beta_{[i]}, \tau_{[i]}^2 + \frac{\sigma_{[i]}^2}{n_{[i]}}\right) \text{ and } \bar{X}_{[i]} \sim \mathscr{N}\left(\beta_{[i]}, \frac{\tau_{[i]}^2}{R_{[i]}} + \frac{\sigma_{[i]}^2}{R_{[i]}n_{[i]}}\right), \quad i = 1,2,\ldots,k.$$

Defining $\chi^2(a)$ as a chi-squared random variable with a degrees of freedom, we represent $\mathscr{L}_{[k][i]}^2$ as follows:

$$\begin{aligned}
\mathscr{L}_{[k][i]}^2 &= \sum_{j\in\{[k],[i]\}} t_j^2 \frac{1}{R_j(R_j-1)} \sum_{r=1}^{R_j}(\bar{X}_{jr}-\bar{X}_j)^2 \\
&= \sum_{j\in\{[k],[i]\}} t_j^2 \frac{1}{R_j(R_j-1)} \left(\sum_{r=1}^{R_j}(\bar{X}_{jr}-\beta_j)^2 - R_j(\bar{X}_j-\beta_j)^2\right) \\
&= \sum_{j\in\{[k],[i]\}} t_j^2 \frac{1}{R_j(R_j-1)} \left(\left(\tau_j^2+\frac{\sigma_j^2}{n_j}\right)\chi^2(R_j) - \left(\tau_j^2+\frac{\sigma_j^2}{n_j}\right)\chi^2(1)\right) \\
&= \sum_{j\in\{[k],[i]\}} t_j^2 \left(\frac{\tau_j^2}{R_j}+\frac{\sigma_j^2}{R_j n_j}\right)\left(\frac{\chi^2(R_j-1)}{R_j-1}\right).
\end{aligned}$$

This leads to the following characterization for the probability of including the best system design $[k]$ in the subset $I := \{i : 1 \leq i \leq k,\ 1 \leq j \leq k,\ i \neq j,\ \text{and } \bar{X}_i \geq \bar{X}_j - \mathscr{L}_{ij}\}$ with $i = k-1,k-2,\ldots,1$:

$$\mathbb{P}\left\{Z_{[k][i]}^2 \leq \sum_{j\in\{[k],[i]\}} t_j^2 \left(\frac{\frac{\tau_j^2}{R_j}+\frac{\sigma_j^2}{R_j n_j}}{\frac{\tau_{[k]}^2}{R_{[k]}}+\frac{\sigma_{[k]}^2}{R_{[k]}n_{[k]}}+\frac{\tau_{[i]}^2}{R_{[i]}}+\frac{\sigma_{[i]}^2}{R_{[i]}n_{[i]}}}\right)\left(\frac{\chi^2(R_j-1)}{R_j-1}\right)\right\} \quad \geq \quad 2(1-\alpha)^{\frac{1}{k-1}}-1.$$

$t_j := t_{(1-\alpha)^{1/(k-1)},R_j-1}$ for $j \in \{[k],[i]\}$ and $i = k-1,k-2,\ldots,1$, ensure the inclusion of the best $[k]$ in the subset I with probability at least $1-\alpha$ (Banerjee 1961). We are now ready to present our decision rule:

Decision Rule When the stochastic simulation of each of the k system designs is conducted with the simulation replication algorithm outlined in Figure 1, the use of the following decision rule for $i \in \{1,2,\ldots,k\}$ and all values of $j \in \{1,2,\ldots,k\}$ which are different from i ensures the inclusion of the best design in the selected subset with a probability of at least $1-\alpha$:

$$\bar{X}_i \geq \bar{X}_j - \left(t^2_{(1-\alpha)^{1/(k-1)},R_i-1} \sum_{r=1}^{R_i} \frac{(\bar{X}_{ir}-\bar{X}_i)^2}{R_i(R_i-1)} + t^2_{(1-\alpha)^{1/(k-1)},R_j-1} \sum_{r=1}^{R_j} \frac{(\bar{X}_{jr}-\bar{X}_j)^2}{R_j(R_j-1)} \right)^{1/2}. \quad (4)$$

As long as we independently sample the input distribution parameters Ψ_{ir}, $r = 1,2,\ldots,R_i$, $i = 1,2,\ldots,k$ and the random-number inputs $\mathbf{u}_{i\ell r}$, $r = 1,2,\ldots,R_i$, $\ell = 1,2,\ldots,n_i$, $i = 1,2,\ldots,k$, this decision rule remains the same independent of whether we assume identical or different sources of uncertainty for the input random variables. Furthermore, this decision rule does not require any restrictive conditions among the output response functions unlike the case in which parameter uncertainty is captured by the method of asymptotic normality approximation (Corlu and Biller 2013).

3 APPLICATION TO STOCHASTIC INVENTORY SYSTEM SIMULATIONS

This section investigates the performance of subset-selection rules with applications to inventory simulations. We provide the design of experiments in Section 3.1 and present the results and insights in Section 3.2.

3.1 Design of Experiments

We consider $k := 40$ different stochastic inventory system designs, each of which is composed of two items. The objective is to identify the inventory system with the maximum joint demand fulfillment probability; i.e., the probability of fully satisfying all item demands. In the inventory system design $i \in \{1,2,\ldots,k\}$, we assume normally distributed demands with a mean of $\mu_{i,1} := 500-(200/(k-1))(i-1)$ and a standard deviation of $\sigma_{i,1} := 0.25\mu_{i,1}$ for the first item as well as a mean of $\mu_{i,2} := 600+(200/(k-1))(i-1)$ and a standard deviation of $\sigma_{i,2} := 0.15\mu_{i,2}$ for the second item. Using $\Phi^{-1}(\cdot)$ for representing the inverse cumulative distribution function of the standard normal random variable, we set the base-stock levels to $\mu_{i,1}+\Phi^{-1}((0.98-(0.40/(k-1))(i-1))^{1/2})\sigma_{i,1}$ and $\mu_{i,2}+\Phi^{-1}((0.98-(0.40/(k-1))(i-1))^{1/2})\sigma_{i,2}$ for item 1 and item 2, respectively. Therefore, if the parameters of the normal demand distributions were known with certainty, then the inventory system design 1 would have the highest fulfillment probability of 0.98 and system design k would have the minimum fulfillment probability of 0.58 (i.e., $0.98-0.40$). Hence, the index of the best system design $[k]$ with the largest true mean performance measure is 1 and the index of the system design $[1]$ with the smallest true mean performance measure is k. Furthermore, the true mean performance measures of the stochastic inventory system designs $[i]$ and $[i+1]$ differ from each other by $0.40/(k-1)$ for $i = 1,2,\ldots,k-1$. Therefore, it becomes more difficult to screen out the inferior system designs with the increasing number of system designs under consideration.

In the absence of known demand parameters, we assume the availability of $m \in \{10,30\}$ observations in each historical data set for the estimation of the unknown demand parameters. Our goal is to identify the pair of items which delivers the maximum joint demand fulfillment probability under demand parameter uncertainty. First, we follow the practice of accounting for only the stochastic uncertainty in the inventory system simulation. Therefore, we drive the inventory system simulation with the demand parameter estimates which are obtained from a single realization of the demand history, and we apply the subset selection decision rule in (2) to the simulation output data for $\alpha = 0.10$. Letting the number of simulation replications, n take the values of 10, 30, or 100, we present our findings of the following three performance measures, each of which is computed across 1000 macro-replications: The average probability of including

the best system design in the selected subset (AP of including the best), the average computational time (in seconds), and the mean of the subset size. We present the results in Table 1. Next, we follow the practice of accounting for both the stochastic uncertainty and the parameter uncertainty in the simulation output process, and apply the subset selection procedure in (4) to the simulation output data. We present our computational findings in Tables 2 and 3.

3.2 Results and Insights

Our results demonstrate the importance of accounting for parameter uncertainty in subset selection to improve the probability of including the best system design in the selected subset especially when the demand history is short and the subset selection procedure is used to screen a large number of competing system designs. Ignoring the parameter uncertainty in the application of the screen-to-the-best procedure provides 0.748 (Table 1) as the average probability of including the best system design in the selected subset when we perform 10 simulation replications ($n = 10$), and assume the availability of 10 observations in the historical data set ($m = 10$).

Table 1: Parameter uncertainty is ignored and screen-to-the-best procedure is used for subset selection.

	AP of Including the Best		Computational Time		Mean of the Subset Size	
n	$m = 10$	$m = 30$	$m = 10$	$m = 30$	$m = 10$	$m = 30$
10	**0.748**	0.938	**2.511**	2.528	**2.514**	2.302
30	**0.618**	0.912	6.617	6.505	1.586	1.500
100	**0.588**	0.808	20.792	20.611	1.298	1.280

We also identify 2.511 seconds as the average computational time and 2.514 as the average size of the selected subset. Biller and Corlu (2015) further identify the standard deviation of the subset size as 1.235. However, when we account for both the stochastic uncertainty and the parameter uncertainty in subset selection and use the method of simulation replication algorithm to represent the parameter uncertainty in the simulation output data, we identify the average probability of including the best system design in the selected subset as 0.785 with an average computational time of 1045.8 seconds and an expected subset size of 1.242 (and a standard deviation of 0.518 for the subset size). In comparison to the practice of ignoring input parameter uncertainty in subset selection, this corresponds to a relative improvement of 4.95% (i.e., from 0.748 in Table 1 to 0.785 in Table 2) in the average probability of including the best system design in the selected subset. Despite its computational intensity, we further decrease the average subset size from 2.511 to 1.242. Hence, our subset selection procedure building on the use of the simulation replication algorithm to represent parameter uncertainty not only improves the average probability of including the best system design in the selected subset but also reduces the average subset size in comparison to the subset selection procedure which ignores the parameter uncertainty. The improvement in the average probability of including the best system design in the selected subset becomes even more evident with the increasing number of simulation replications. When $n = 30$, we increase the average probability of having the best system design in the subset by 13.3% (i.e., from 0.618 in Table 1 to 0.700 in Table 2), accompanied by a reduction of 21.7% in the average subset size. We further increase this probability by 14.6% (i.e., from 0.588 in Table 1 to 0.674 in Table 2) with a reduction of 6.4% in the average subset size when $n = 100$.

Table 2: Parameter uncertainty is captured by the simulation replication algorithm.

	AP of Including the Best		Computational Time		Mean of the Subset Size	
n	$m = 10$	$m = 30$	$m = 10$	$m = 30$	$m = 10$	$m = 30$
10	**0.785**	0.792	**1045.8**	1048.5	**1.242**	1.108
30	**0.700**	0.798	3108.9	3112.4	**1.242**	1.116
100	**0.674**	0.813	10369.6	10469.7	**1.215**	1.111

An observation we make from the results reported in Table 1 is that both the average probability of including the best system design in the selected subset and the mean of the subset size decrease with the number of replications performed for the simulation of each stochastic inventory system design. This observation can be explained by the effect of the simulation replication number, n to decrease the magnitude of $\mathscr{L}_{ij}^2 := t^2_{(1-\alpha)^{1/(k-1)},n-1}(S_i^2+S_j^2)/n$, which arises in the decision rule comparing $\bar{X}_i$ to $\bar{X}_j$ in (2) with $n = n_i = n_j$. Although we make a similar observation for $m = 10$ in Table 2, this observation does not extend to $m = 30$. More specifically, we observe the average probability of including the best system design in the selected subset to increase with the number of simulation replications. This is because the parameter density functions provide more information about the unknown demand parameters when there are 30 observations in the historical data set in comparison to having 10 historical data observations, and increasing the number of replications allows us to incorporate more information about the distributions of the input random variables, conditional on the density function of the unknown demand parameters, into the simulation outputs.

Table 3: The average probabilities of the selected subsets to contain only the best system design (i.e., $I_1 = \{[k]\}$), the best two system designs (i.e., $I_2 = \{[k],[k-1]\}$), and the best three system designs (i.e., $I_3 = \{[k],[k-1],[k-2]\}$) when the number of simulation replications is 10; i.e., $n_i = 10$ for $i = 1,2,\ldots,k$.

	Ignoring Parameter Uncertainty		Accounting for Parameter Uncertainty	
Subset	$m = 10$	$m = 30$	$m = 10$	$m = 30$
I_1	**0.138**	0.220	**0.536**	0.704
I_2	**0.264**	0.476	**0.778**	0.908
I_3	**0.420**	0.716	**0.866**	0.976

Finding significant discrepancies among the mean subset sizes reported in Tables 1 and 2 motivates us to further investigate the contents of the selected subsets delivered by the subset selection procedures. In Table 3 where we perform 10 replications for the simulation of each inventory system design, we present first the average probability that the selected subset contains only the best inventory system design (i.e., $I = \{[k]\}$), then the average probability of including only the best two system designs in the selected subset (i.e., $I = \{[k],[k-1]\}$), and finally the average probability of containing only the best three system designs in the selected subset (i.e., $I = \{[k],[k-1],[k-2]\}$). By focusing on the same experimental design with $m = 10$ and $n = 10$, we find that there is more to understand about the performance of our subset selection procedures than the insights gained by comparing Table 1 to Table 2. In the case of implementing the subset selection procedure without taking the input parameter uncertainty into consideration, we correctly identify the best inventory system design with the maximum joint demand fulfillment probability among 40 competing system designs only for 13.8% of the time. However, accounting for input parameter uncertainty via the use of our subset selection procedure increases this frequency to 53.6% with an improvement of almost four fold. Furthermore, redesigning the subset selection procedure to account for input parameter uncertainty increases the average probability of identifying the best two system designs from 26.4% to 77.8% and the average probability of identifying the best three system designs from 42.0% to 86.6%. Therefore, there is considerable benefit to the use of a subset selection procedure which accounts for input parameter uncertainty to effectively screen out the inferior system designs.

4 CONCLUSION

We study a subset selection procedure when the parameters of the input distributions are unknown and there is only a limited amount of historical data available for parameter estimation. We redesign the subset selection procedure to account for parameter uncertainty and apply this procedure to stochastic inventory system simulations with the purpose of finding the inventory system designs maximizing the service level in the presence of limited amounts of historical demand data. Computational results demonstrate the importance

of accounting for parameter uncertainty in subset selection to improve the probability of including the best system design in the selected subset especially when the demand history is short and the subset selection procedure is used to screen a large number of competing system designs. We refer the reader to Biller and Corlu (2015), where the potential effects of batching and common random numbers for enhancing the performance of our subset selection procedure are investigated.

REFERENCES

Ankenman, B. E. and B. L. Nelson. 2012. "A Quick Assessment of Input Uncertainty". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 1 – 10. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Banerjee, S. 1961. "On Confidence Interval for Two-Means Problem Based on Separate Estimates of Variances and Tabulated Values of *t* Table". *Sankhya* A23: 359 – 378.

Barton, R. R. 2012. "Tutorial: Input Uncertainty in Output Analysis". In *Proceedings of the 2012 Winter Simulation Conference*, edited by C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, and A. M. Uhrmacher, 162 – 172. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Bechhofer, R. E. 1954. "A Single-Sample Multiple Decision Procedure for Ranking Means of Normal Populations with Known Variances". *Annals of Mathematical Statistics* 25: 16 – 39.

Biller, B. and C. G. Corlu. 2015. "Subset Selection Procedures for Stochastic Simulations with Unknown Input Parameters". Working Paper, Metropolitan College, Boston University, Boston, MA.

Biller, B., C. G. Corlu, and S. Tayur. 2014. "Service-Level Estimation in Stochastic Inventory System Simulations: The Case of Limited Historical Demand Data". Working Paper, Metropolitan College, Boston University, Boston, MA.

Boesel, J., B. L. Nelson, and S.-H. Kim. 2003. "Using Ranking and Selection to "Clean Up" After Simulation Optimization". *Operations Research* 51: 814 – 825.

Chick, S. E. 2001. "Input Distribution Selection for Simulation Experiments: Accounting for Input Uncertainty". *Operations Research* 49: 744 – 758.

Corlu, C. G. and B. Biller. 2013. "A Subset Selection Procedure Under Input Parameter Uncertainty". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S. -H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 463 – 473. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Freimer, M. and L. W. Schruben. 2002. "Collecting Data and Estimating Parameters for Input Distributions". In *Proceedings of the 2002 Winter Simulation Conference*, edited by E. Yucesan, C. Chen, J. L. Snowdon, and J. M. Charnes, 392 – 399. Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Goldsman, D. and B. L. Nelson. 1998. "Comparing Systems via Simulation". In *Handbook of Simulation: Principles, Methodology, Advances, Applications, and Practice*, edited by Jerry Banks, 273 – 306. John Wiley & Sons, Inc.

Gupta, S. S. 1956. *On a Decision Rule for a Problem in Ranking Means*. Doctoral Dissertation, Institute of Statistics, University of North Carolina, Chapel Hill, NC.

Gupta, S. S. 1965. "On Some Multiple Decision (Ranking and Selection) Rules". *Technometrics* 7: 225 – 245.

Helton, J. C. 1997. "Uncertainty and Sensitivity Analysis in the Presence of Stochastic and Subjective Uncertainty". *Journal of Statistical Computation and Simulation* 57: 3 – 76.

Kim, S. H. 2013. "Statistical Ranking and Selection". In *Encyclopedia of Operations Research and Management Science*, edited by S. I. Gass and M. C. Fu, New York: Springer.

Nelson, B. L., J. Swann, D. Goldsman, and W. -M. T. Song. 2001. "Simple Procedures for Selecting the Best System when the Number of Alternatives is Large". *Operations Research* 49: 950 – 963.

Raftery, A. E., D. Madigan, and C. T. Volinsky. 1996. "Accounting for Model Uncertainty in Survival Analysis Improves Predictive Performance (with discussion)". In *Bayesian Statistics*, edited by J. M. Bernardo, J. O. Berger, A. P. David, and A. F. M. Smith, 323 – 349. Oxford Press.

Rao, P. S. R. S. 1997. *Variance Components Estimation: Mixed Models, Methodologies, and Applications*. Chapman and Hall, London, UK.

Rohatgi, V. K. 1976. *An Introduction to Probability Theory and Mathematical Statistics*. Wiley, New York.

Song, E. and B. L. Nelson. 2013. "A Quicker Assessment of Input Uncertainty". In *Proceedings of the 2013 Winter Simulation Conference*, edited by R. Pasupathy, S. -H. Kim, A. Tolk, R. Hill, and M. E. Kuhl, 474 – 485, Piscataway, New Jersey: Institute of Electrical and Electronics Engineers, Inc.

Wilson, J. R. 2001. "A Multiplicative Decomposition Property of the Screening-and-Selection Procedures of Nelson et al". *Operations Research* 49: 964 – 966.

Zouaoui, F. and J. R. Wilson. 2003. "Accounting for Parameter Uncertainty in Simulation Input Modeling". *IIE Transactions* 35: 781 – 792.

Zouaoui, F. and J. R. Wilson. 2004. "Accounting for Input-Model and Input-Parameter Uncertainties in Simulation. *IIE Transactions* 36: 1135 – 1151.

AUTHOR BIOGRAPHIES

CANAN GUNES CORLU is an Assistant Professor of Administrative Sciences Department at Boston University. Her research interest is in the area of the design and analysis of stochastic simulations with applications to operations management. Her email address is canan@bu.edu.

BAHAR BILLER is a Research Scientist at Software Sciences and Analytics organization of the General Electric Global Research Center. Her research lies in the area of quantitative risk management and advances the design and analysis of large-scale business system simulations to aid decision-making under uncertainty. Her email address is bahar.biller@ge.com.