Proceedings of the 2014 Winter Simulation Conference
A. Tolk, S. Y. Diallo, I. O. Ryzhov, L. Yilmaz, S. Buckley, and J. A. Miller, eds.

STATISTICAL UNCERTAINTY ANALYSIS FOR STOCHASTIC SIMULATION WITH DEPENDENT INPUT MODELS

Wei Xie

Rensselaer Polytechnic Institute
Troy, NY 12180-3590, USA

Barry L. Nelson

Northwestern University
Evanston, IL 60208, USA

Russell R. Barton

The Pennsylvania State University
University Park, PA 16802, USA

ABSTRACT

When we use simulation to estimate the performance of a stochastic system, lack of fidelity in the random input models can lead to poor system performance estimates. Since the components of many complex systems could be dependent, we want to build input models that faithfully capture such key properties. In this paper, we use the flexible NORmal To Anything (NORTA) representation for dependent inputs. However, to use the NORTA representation we need to estimate the marginal distribution parameters and a correlation matrix from real-world data, introducing input uncertainty. To quantify this uncertainty, we employ the bootstrap to capture the parameter estimation error and an equation-based stochastic kriging metamodel to propagate the input uncertainty to the output mean. Asymptotic analysis provides theoretical support for our approach, while an empirical study demonstrates that it has good finite-sample performance.

1 INTRODUCTION

Stochastic simulation is used to characterize the behavior of complex systems that are driven by random input models. The choice of input models directly impacts the system performance estimates. It is common practice to treat the input models as a collection of independent univariate distributions. However, these simple input models do not always faithfully represent the physical input processes. For example, in a manufacturing system the processing times for a single workpiece at a series of machining stations could be dependent due to characteristics of that particular workpiece; and in a supply chain system the customer demands over different products from multi-product warehouses could be related. Ignoring such dependence can lead to poor estimates of system performance measures (Livny et al. 1993). Thus, it is desirable to build input models that faithfully capture the dependence.

Considering the amount of information needed to construct full joint distributions, in this paper we focus on input models characterized by marginal distributions and correlation matrix. The marginal distributions have known parametric families with unknown parameter values. The dependence between different components of the input models can be measured by various criteria (Biller and Ghosh 2006); we focus on the Spearman rank correlation. Even though product-moment correlation is widely used in engineering applications, it cannot capture complex nonlinear dependence and its definition needs the variances of the components to be finite. The Spearman rank correlation can avoid these problems.

Since the input models are estimated from finite samples of real-world data, a complete statistical characterization of stochastic system performance requires quantifying both simulation and input estimation error. We build on Xie et al. (2014a), which used a metamodel-assisted bootstrapping approach to form

978-1-4799-7486-3/14/$31.00

a confidence interval (CI) accounting for the impact of input and simulation uncertainty on the system's mean performance estimates. However, their study assumes that the input distributions are univariate and mutually independent.

To efficiently and correctly account for the dependence between various components of input models, we introduce a more general metamodel-assisted bootstrapping framework that can quantify the impact of dependent input models and simulation estimation error on system performance estimates, while simultaneously reducing the influence of simulation estimation error compared with direct simulation methods. Specifically, we estimate the key properties of a flexible input model with real-world data. The bootstrap is then used to quantify the estimation error of input marginal distributions and dependence measures. An equation-based stochastic kriging metamodel propagates the input uncertainty to the output mean. From this, we can derive a CI that accounts for both simulation and input uncertainty. Notice that we are interested in the dependence between different components of the input distributions instead the dependence among the estimated input-model parameter values.

There are two central contributions of this paper: First, we generalize the metamodel-assisted bootstrap framework of Xie et al. (2014a) to stochastic simulations with dependent input models. Second, we propose a rigorous analysis for cases where the dependence is measured by Spearman rank correlation.

The next section reviews some previous work for input uncertainty in simulation. This is followed by a formal description of the problem of interest in Section 3. In Section 4, we propose a generalized metamodel-assisted bootstrapping framework and provide an algorithm to build a CI accounting for both input and simulation estimation error on the system mean performance estimates. Our approach is supported with asymptotic analysis. We then report results of finite-sample behavior from an empirical study in Section 5 and conclude the paper in Section 6.

2 BACKGROUND

Various approaches to account for input uncertainty in stochastic simulations have been proposed, see Barton (2012) for a review. The metamodel-assisted bootstrapping approach was introduced in Barton et al. (2014). In that paper the input models are a collection of independent univariate parametric distributions with known parametric families and unknown parameters values. The input parameters are estimated by real-world data and the bootstrap is used to quantify the input-parameter uncertainty. Based on the simulation outputs at a few design points, a flexible stochastic kriging (SK) metamodel is built to propagate the input uncertainty to the output mean. A CI is derived to quantify the impact of input uncertainty. Compared with the direct simulation approach, the metamodel can reduce the impact of simulation estimation error. However, Barton et al. (2014) assumed that the simulation budget is not tight and the metamodel uncertainty can be ignored.

If the true mean response surface is complex, especially for high-dimensional problems with many input distributions, and the computational budget is tight, then the impact of metamodel uncertainty can no longer be ignored. The metamodel-assisted bootstrapping approach was improved in Xie et al. (2014a) to build a CI accounting for the impact from both input and metamodel uncertainty on the system mean estimates. Further, a variance decomposition was proposed to estimate the relative contribution of input to overall uncertainty, which is very useful for decision makers to determine where to put more effort to reduce the system uncertainty. The metamodel-assisted bootstrapping approach demonstrated robust performance even when there is a tight computational budget and simulation estimation error is large. However, Xie et al. (2014a) also assumed that the input models are mutually independent univariate distributions.

To faithfully capture the dependence between different components of input models, Cario and Nelson (1997) proposed a flexible NORmal To Anything (NORTA) distribution to represent and generate random vectors with almost arbitrary marginal distributions and correlation matrix. Since NORTA representations are estimated from finite samples of real-world data, Biller and Corlu (2011) proposed a Bayesian approach to account for parameter uncertainty. Specifically, the uncertainty about the NORTA distribution parameter estimates is quantified by posterior distributions. For complex stochastic systems with a large number of correlated inputs, a fast algorithm was proposed to draw samples from these posterior distributions to quantify the input uncertainty. Then, the direct simulation method was used to propagate the input uncertainty to the output mean. However, when the simulated system is complex and the computational budget is tight, the direct simulation method can not efficiently use the computational budget and incurs substantial simulation estimation error.

The good performance of metamodel-assisted bootstrapping for stochastic simulations with independent univariate input distributions in Barton et al. (2014) and Xie et al. (2014a) motivates us to extend it to more complex cases with dependence in the input models. Therefore, in this paper, we use flexible NORTA representations for unknown dependent inputs. Since the NORTA method is based on estimating marginal distribution parameters and a correlation matrix from real-world data, metamodel-assisted bootstrapping is generalized to quantify the impact of both NORTA parameters and simulation estimation error while simultaneously reducing the influence of simulation estimation error due to output variability.

3 PROBLEM STATEMENT

The stochastic simulation output is a function of random numbers and the input models denoted by F, where F is composed of a collection of input distributions used to drive the simulation. For notation simplification, we do not explicitly include the random numbers. The output from the jth replication of a simulation with input models F can be written as

$$Y_j(F) = \mu(F) + \varepsilon_j(F)$$

where $\mu(F) = \mathrm{E}[Y_j(F)]$ denotes the unknown output mean and $\varepsilon_j(F)$ represents the simulation error with mean zero. Notice that the simulation output depends on the choice of input models.

Let F^c denote the true "correct" input models, which are unknown and estimated from finite samples of real-world data. Our goal is to quantify the impact of the statistical error on system mean performance estimates by finding a $(1-\alpha)100\%$ CI $[Q_L, Q_U]$ such that

$$\Pr\{\mu(F^c) \in [Q_L, Q_U]\} = 1 - \alpha.$$

Let $F = \{F_1, F_2, \ldots, F_L\}$ with $F_1, F_2, \ldots, F_L$ mutually independent; F could be composed of univariate and multivariate joint distributions. In this paper, we do not consider time-series input processes. Let F_ℓ be a d_ℓ dimensional distribution having marginal distributions denoted by $\{F_{\ell,1}, F_{\ell,2}, \ldots, F_{\ell,d_\ell}\}$ with $d_\ell \geq 1$. For the ℓth distribution F_ℓ with $d_\ell > 1$, let $\mathbf{X}_\ell \sim F_\ell$ be a $d_\ell \times 1$ random vector having a $d_\ell \times d_\ell$ Spearman rank correlation matrix denoted by $R_{\mathbf{X}_\ell}$ with elements

$$R_{\mathbf{X}_\ell}(i,j) = \mathrm{cor}(F_{\ell,i}(X_{\ell,i}), F_{\ell,j}(X_{\ell,j})) = \frac{\mathrm{E}[F_{\ell,i}(X_{\ell,i})F_{\ell,j}(X_{\ell,j})] - \mathrm{E}[F_{\ell,i}(X_{\ell,i})]\mathrm{E}[F_{\ell,j}(X_{\ell,j})]}{\sqrt{\mathrm{Var}(F_{\ell,i}(X_{\ell,i}))\mathrm{Var}(F_{\ell,j}(X_{\ell,j}))}}$$

with $i, j = 1, 2, \ldots, d_\ell$. Since the correlation matrix is symmetric and diagonal terms are 1, we can view a $d_\ell \times d_\ell$ correlation matrix as an element of a $d_\ell^* \equiv d_\ell(d_\ell - 1)/2$ dimensional vector space. Then, the Spearman rank correlation matrix can be uniquely specified by a $d_\ell^* \times 1$ vector denoted by $\mathbf{V}_{\mathbf{X}_\ell}$. For F_ℓ with $d_\ell = 1$, $\mathbf{V}_{\mathbf{X}_\ell}$ is empty and $d_\ell^* = 0$.

For F_ℓ with $\ell = 1, 2, \ldots, L$, we assume that the families of marginal distributions $\{F_{\ell,1}, F_{\ell,2}, \ldots, F_{\ell,d_\ell}\}$ are known, but not their parameter values. Let an $h_{\ell,i} \times 1$ vector $\boldsymbol{\theta}_{\ell,i}$ denote the unknown parameters for the ith marginal distribution $F_{\ell,i}$. By stacking $\boldsymbol{\theta}_{\ell,i}$ with $i = 1, 2, \ldots, d_\ell$ together, we have a $d_\ell^\dagger \times 1$ dimensional parameter vector $\boldsymbol{\theta}_\ell^\top \equiv (\boldsymbol{\theta}_{\ell,1}^\top, \boldsymbol{\theta}_{\ell,2}^\top, \ldots, \boldsymbol{\theta}_{\ell,d_\ell}^\top)$ with $d_\ell^\dagger \equiv \sum_{i=1}^{d_\ell} h_{\ell,i}$.

Suppose the input model is characterized by marginal distributions and correlation matrix, which are specified by $\boldsymbol{\vartheta} \equiv \{(\boldsymbol{\theta}_\ell; \mathbf{V}_{\mathbf{X}_\ell}), \ell = 1, 2, \ldots, L\}$ that includes $d \equiv \sum_{\ell=1}^{L} d_\ell^\dagger + \sum_{\ell=1}^{L} d_\ell^*$ elements. We call $\boldsymbol{\vartheta}$ the *input model parameters*. By abusing notation, we can rewrite $\mu(F)$ as $\mu(\boldsymbol{\vartheta})$. The true input parameters $\boldsymbol{\vartheta}^c$ are unknown and estimated from finite samples of real-world data. Thus, our goal can be restated as finding a $(1-\alpha)100\%$ CI $[Q_L, Q_U]$ such that

$$\Pr\{\mu(\boldsymbol{\vartheta}^c) \in [Q_L, Q_U]\} = 1 - \alpha. \tag{1}$$

Let m_ℓ denote the number of i.i.d. real-world observations available from the ℓth input process $\mathscr{X}_{\ell,m_\ell} \equiv \left\{\mathbf{X}_\ell^{(1)}, \mathbf{X}_\ell^{(2)}, \ldots, \mathbf{X}_\ell^{(m_\ell)}\right\}$ with $d_\ell \times 1$ random vector $\mathbf{X}_\ell^{(i)} \overset{i.i.d}{\sim} F_\ell^c$ for $i = 1,2,\ldots,m_\ell$. Let $\mathbf{X_m} = \{\mathscr{X}_{\ell,m_\ell}, \ell = 1,2,\ldots,L\}$ be the collection of samples from all L input distributions in F^c, where $\mathbf{m} = (m_1, m_2, \ldots, m_L)$. The estimators for the input model parameters are denoted by $\widehat{\boldsymbol{\vartheta}}_\mathbf{m} \equiv \{(\widehat{\boldsymbol{\theta}}_{\ell,m_\ell}; \widehat{\mathbf{V}}_{\mathbf{X}_{\ell,m_\ell}}), \ell = 1,2,\ldots,L\}$ and are a function of $\mathbf{X_m}$. The real-world data are a particular realization of $\mathbf{X_m}$, say $\mathbf{x}_\mathbf{m}^{(0)}$.

Input uncertainty is quantified by the sampling distribution of $\mu(\widehat{\boldsymbol{\vartheta}}_\mathbf{m})$. Further, since the underlying response surface $\mu(\cdot)$ is unknown, for any $\boldsymbol{\vartheta}$ let $\widehat{\mu}(\boldsymbol{\vartheta})$ denote the corresponding mean response estimator. Thus, there are both input and simulation estimation errors in the system mean performance estimates. *This paper focuses on estimating the performance of complex stochastic systems with input models including multivariate distributions. Our objective is to propose an approach to quantify the overall impact of both input and simulation estimation error on system mean performance estimates and then build a CI satisfying Equation (1). Further, since each simulation run could be computationally expensive and we have computational budget denoted by N, we want to reduce the influence of simulation estimation error due to output variability.*

4 METAMODEL-ASSISTED BOOTSTRAPPING FRAMEWORK

For simulations with parametric input distributions that are univariate and mutually independent, a metamodel-assisted bootstrapping framework was used to account for the impact of both input and simulation estimation errors in Xie et al. (2014a). In this section, we generalize the metamodel-assisted bootstrapping approach to allow NORTA input models. A procedure to build a CI for the system mean performance that accounts for overall uncertainty is described and supported by asymptotic analysis.

4.1 Bootstrap for Input Uncertainty

The way we choose to represent input models plays an important role in the implementation of metamodel-assisted bootstrapping. For the ℓth input distribution F_ℓ, instead of using the natural parameters $\boldsymbol{\theta}_\ell$ to characterize the marginal distributions, we can use moments; see Barton et al. (2014) for an explanation. Suppose that the parametric marginal distribution $F_{\ell,i}$ can be uniquely characterized by its first $h_{\ell,i}$ finite moments denoted by the $h_{\ell,i} \times 1$ vector $\boldsymbol{\psi}_{\ell,i}$ for $i = 1,2,\ldots,d_\ell$. By stacking $\boldsymbol{\psi}_{\ell,i}$ with $i = 1,2,\ldots,d_\ell$ together, we have a $d_\ell^\dagger \times 1$ dimensional vector of marginal moments $\boldsymbol{\psi}_\ell^\top \equiv (\boldsymbol{\psi}_{\ell,1}^\top, \boldsymbol{\psi}_{\ell,2}^\top, \ldots, \boldsymbol{\psi}_{\ell,d_\ell}^\top)$. Therefore, the input models can be characterized by the moments $\mathscr{M} \equiv \{(\boldsymbol{\psi}_\ell; \mathbf{V}_{\mathbf{X}_\ell}), \ell = 1,2,\ldots,L\}$. Abusing notation, we rewrite $\mu(\boldsymbol{\vartheta})$ as $\mu(\mathscr{M})$.

The true moments of dependent input models, denoted by $\mathscr{M}^c$, are unknown and estimated using $\mathbf{X_m}$. We use moment estimators for marginal distributions, denoted by $\widehat{\boldsymbol{\psi}}_{\ell,m_\ell}$. The estimation error of input models can be quantified by the sampling distribution of $\widehat{\mathscr{M}_\mathbf{m}} \equiv \{(\widehat{\boldsymbol{\psi}}_{\ell,m_\ell}; \widehat{\mathbf{V}}_{\mathbf{X}_{\ell,m_\ell}}), \ell = 1,2,\ldots,L\}$, denoted by $F^c_{\mathscr{M}_\mathbf{m}}$. Therefore, the input uncertainty can be measured by the sampling distribution of $\mu(\widehat{\mathscr{M}_\mathbf{m}})$ with $\widehat{\mathscr{M}_\mathbf{m}} \sim F^c_{\mathscr{M}_\mathbf{m}}$.

Since it is hard to derive the sampling distribution $F^c_{\mathscr{M}_\mathbf{m}}$, we use bootstrap resampling to approximate it (Shao and Tu 1995). Since $F_1, F_2, \ldots, F_L$ are mutually independent, we can do bootstrapping for each distribution separately. For distribution F_ℓ, let $A_\ell \equiv \{1,2,\ldots,m_\ell\}$ be the index set of the real-world observations. The procedure to quantify the input uncertainty by the bootstrap is as follows:

1. For the ℓth distribution F_ℓ with $\ell = 1,2,\ldots,L$, draw m_ℓ samples with replacement from set A_ℓ and obtain indexes $\{i_1, i_2, \ldots, i_{m_\ell}\}$; choose corresponding samples from real-world data $\mathbf{x}_\mathbf{m}^{(0)}$ and get $\widetilde{\mathbf{x}}_{\ell,m_\ell}^{(1)} \equiv \{\mathbf{x}_\ell^{(i_1)}, \mathbf{x}_\ell^{(i_2)}, \ldots, \mathbf{x}_\ell^{(i_{m_\ell})}\}$. Denote the collection of bootstrap samples by $\widetilde{\mathbf{x}}_\mathbf{m}^{(1)} =$

$\{\widetilde{\mathbf{x}}^{(1)}_{\ell,m_\ell}, \ell = 1,2,\ldots,L\}$ and use it to calculate the bootstrapped moment estimates, denoted by $\widetilde{\mathscr{M}}^{(1)}_{\mathbf{m}} \equiv \left\{(\widetilde{\boldsymbol{\psi}}^{(1)}_{\ell,m_\ell}; \widetilde{\mathbf{V}}^{(1)}_{\mathbf{X}_\ell,m_\ell}), \ell = 1,2,\ldots,L\right\}$.

2. Repeat the previous step B times to generate $\widetilde{\mathscr{M}}^{(b)}_{\mathbf{m}}$ for $b = 1,2,\ldots,B$.

The bootstrap resampled moments are drawn from the bootstrap distribution denoted by $\widetilde{F}_{\mathscr{M}_{\mathbf{m}}}(\cdot|\mathbf{x}^{(0)}_{\mathbf{m}})$ with $\widetilde{\mathscr{M}}^{(b)}_{\mathbf{m}} \sim \widetilde{F}_{\mathscr{M}_{\mathbf{m}}}(\cdot|\mathbf{x}^{(0)}_{\mathbf{m}})$. For estimation of a CI, B is recommended to be a few thousand. In this paper, a $\widehat{}$ denotes a quantity estimated from real-world data, while a $\widetilde{}$ denotes a quantity estimated from bootstrapped data.

4.2 NORTA Representation

In this paper, multivariate input models are characterized by their marginal distributions and correlation matrix. Since this partial characterization does not uniquely determine the joint distributions except in special cases, we use the NORTA representation introduced by Cario and Nelson (1997); NORTA vectors are completely specified by their marginal distributions and correlation matrix. In this section we describe how to apply NORTA in our metamodel-assisted bootstrapping approach.

Since $F = \{F_1, F_2, \ldots, F_L\}$ with $F_1, F_2, \ldots, F_L$ mutually independent, we only need to apply NORTA separately for the distributions F_ℓ with $d_\ell > 1$. Specifically, we represent $\mathbf{X}_\ell$ as a transformation of a d_ℓ dimensional standard multivariate normal (MVN) vector $\mathbf{Z}_\ell = (Z_{\ell,1}, Z_{\ell,2}, \ldots, Z_{\ell,d_\ell})^\top$ with product-moment correlation matrix $\rho_{\mathbf{Z}_\ell}$,

$$\mathbf{X}_\ell^\top = \left(F^{-1}_{\ell,1}[\Phi(Z_{\ell,1}); \boldsymbol{\theta}_{\ell,1}], F^{-1}_{\ell,2}[\Phi(Z_{\ell,2}); \boldsymbol{\theta}_{\ell,2}], \ldots, F^{-1}_{\ell,d_\ell}[\Phi(Z_{\ell,d_\ell}); \boldsymbol{\theta}_{\ell,d_\ell}]\right). \tag{2}$$

If the marginal distribution families are given, as we assume here, then the NORTA representation for F_ℓ is $(\boldsymbol{\theta}_\ell, \rho_{\mathbf{Z}_\ell})$. Let $R_{\mathbf{Z}_\ell}$ denote a $d_\ell \times d_\ell$ Spearman rank correlation matrix for $\mathbf{Z}_\ell$. Notice that there is a closed form relationship between product-moment and Spearman rank correlations for the standard multivariate normal distribution (Clemen and Reilly 1999): For any $i, j = 1,2,\ldots,d_\ell$, we have

$$R_{\mathbf{Z}_\ell}(i,j) = \frac{6}{\pi} \sin^{-1}\left(\frac{\rho_{\mathbf{Z}_\ell}(i,j)}{2}\right). \tag{3}$$

In this paper, we focus on continuous marginal distributions with strictly increasing cdfs $\{F_{\ell,1}, F_{\ell,2}, \ldots, F_{\ell,d_\ell}\}$ for $\ell = 1,2,\ldots,L$. Since the Spearman rank correlation is invariant under the monotone one-to-one transformation $F^{-1}_{\ell,i}[\Phi(\cdot)]$, we have $R_{\mathbf{X}_\ell} = R_{\mathbf{Z}_\ell}$. Given $\boldsymbol{\vartheta} = \{(\boldsymbol{\theta}_\ell; \mathbf{V}_{\mathbf{X}_\ell}), \ell = 1,2,\ldots,L\}$, the procedure to generate the NORTA random variates is as follows:

1. From $\mathbf{V}_{\mathbf{X}_\ell}$, get the Spearman rank correlation matrix $R_{\mathbf{Z}_\ell} = R_{\mathbf{X}_\ell}$.
2. Calculate $\rho_{\mathbf{Z}_\ell}(i,j) = 2\sin(\pi R_{\mathbf{Z}_\ell}(i,j)/6)$ for $i, j = 1,2,\ldots,d_\ell$.
3. Generate $\mathbf{Z}_\ell \overset{i.i.d.}{\sim} \mathrm{MVN}(\mathbf{0}, \rho_{\mathbf{Z}_\ell})$ and obtain $\mathbf{X}_\ell$ by using Equation (2).
4. Repeat Steps 1–3 for all F_ℓ with $d_\ell > 1$.

Notice that when we use Spearman rank correlation to measure the input model dependence, the choice of marginal distributions and correlation is separable. Thus, for any feasible $\boldsymbol{\vartheta}$, we can find the corresponding NORTA representations.

4.3 Stochastic Kriging Metamodel

Instead of running simulations for systems with input models represented by bootstrap moment samples $\{\widetilde{\mathscr{M}}^{(1)}_{\mathbf{m}}, \widetilde{\mathscr{M}}^{(2)}_{\mathbf{m}}, \ldots, \widetilde{\mathscr{M}}^{(B)}_{\mathbf{m}}\}$, we can choose a small number of design points and run simulations there to build a metamodel; see Xie et al. (2014a). We then propagate the input uncertainty to the output mean by using the metamodel. In this paper, we use a flexible stochastic kriging (SK) metamodel introduced by Ankenman et al. (2010). We briefly review it in this section. For more detailed information, see Ankenman et al. (2010).

Suppose that the underlying true (but unknown) response surface is a realization of a stationary Gaussian Process (GP). We model the simulation output Y by

$$Y_j(\mathbf{x}) = \beta_0 + W(\mathbf{x}) + \varepsilon_j(\mathbf{x}). \tag{4}$$

This model includes two sources of uncertainty: the simulation output uncertainty $\varepsilon_j(\mathbf{x})$ and the mean response uncertainty $W(\mathbf{x})$.

Since stochastic systems with dependent input models having similar key properties tend to have mean responses close to each other, SK uses a mean-zero, second-order stationary GP $W(\cdot)$ to account for this spatial dependence of the response surface. The uncertainty about the unknown true response surface $\mu(\mathbf{x})$ is represented by a GP $M(\mathbf{x}) \equiv \beta_0 + W(\mathbf{x})$ (note that β_0 can be replaced by a more general trend term $\mathbf{f}(\mathbf{x})^\top \boldsymbol{\beta}$). For many, but not all, simulation settings the output is an average of a large number of more basic outputs, so a normal approximation can be applied: $\varepsilon(\mathbf{x}) \sim \mathrm{N}(0, \sigma^2_\varepsilon(\mathbf{x}))$.

In SK, the covariance between $W(\mathbf{x})$ and $W(\mathbf{x}')$ quantifies how knowledge of the surface at one location affects the prediction at another location. In this paper, a parametric form of the covariance is used to capture this spatial dependence, $\Sigma(\mathbf{x},\mathbf{x}') = \mathrm{Cov}[W(\mathbf{x}), W(\mathbf{x}')] = \tau^2 r(\mathbf{x}-\mathbf{x}')$, where τ^2 denotes the variance and $r(\cdot)$ is a correlation function that depends only on the distance $\mathbf{x}-\mathbf{x}'$. Using prior information about the smoothness of $\mu(\cdot)$, we can choose the form of correlation function. Based on previous study (Xie et al. 2010), we use the product-form Gaussian correlation function

$$r(\mathbf{x}-\mathbf{x}') = \exp\left(-\sum_{j=1}^{d} \phi_j (x_j - x'_j)^2 \right) \tag{5}$$

for the empirical evaluation in Section 5. Let $\boldsymbol{\phi} = (\phi_1, \phi_2, \ldots, \phi_d)$ represent the correlation function parameters. So in summary, before having any simulation results the uncertainty about $\mu(\mathbf{x})$ is represented by a Gaussian process $M(\mathbf{x}) \sim \mathrm{GP}(\beta_0, \tau^2 r(\mathbf{x}-\mathbf{x}'))$.

To reduce the uncertainty about $\mu(\mathbf{x})$ we choose an experiment design consisting of pairs $\mathscr{D} \equiv \{(\mathbf{x}_i, n_i), i = 1,2,\ldots,k\}$ at which to run simulations, where $(\mathbf{x}_i, n_i)$ denotes the location and the number of replications, respectively, at the ith design point. The simulation outputs at $\mathscr{D}$ are $\mathbf{Y}_{\mathscr{D}} \equiv \{(Y_1(\mathbf{x}_i), Y_2(\mathbf{x}_i), \ldots, Y_{n_i}(\mathbf{x}_i)); i = 1,2,\ldots,k\}$ and the sample mean at design point $\mathbf{x}_i$ is $\bar{Y}(\mathbf{x}_i) = \sum_{j=1}^{n_i} Y_j(\mathbf{x}_i)/n_i$. Let the sample means at all k design points be $\bar{\mathbf{Y}}_{\mathscr{D}} = (\bar{Y}(\mathbf{x}_1), \bar{Y}(\mathbf{x}_2), \ldots, \bar{Y}(\mathbf{x}_k))^T$. Since the use of common random numbers is detrimental to prediction (Chen et al. 2012), the simulations at different design points are independent and the variance of $\bar{\mathbf{Y}}_{\mathscr{D}}$ is represented by a $k \times k$ diagonal matrix $C = \mathrm{diag}\left\{\sigma^2_\varepsilon(\mathbf{x}_1)/n_1, \sigma^2_\varepsilon(\mathbf{x}_2)/n_2, \ldots, \sigma^2_\varepsilon(\mathbf{x}_k)/n_k\right\}$.

Let Σ be the $k \times k$ spatial covariance matrix of the design points and let $\Sigma(\mathbf{x}, \cdot)$ be the $k \times 1$ spatial covariance vector between each design point and a fixed prediction point $\mathbf{x}$. If the parameters $(\tau^2, \boldsymbol{\phi}, C)$ are known, then the metamodel uncertainty can be characterized by a refined GP $M_p(\mathbf{x})$ that denotes the conditional distribution of $M(\mathbf{x})$ given all simulation outputs,

$$M_p(\mathbf{x}) \sim \mathrm{GP}(m_p(\mathbf{x}), \sigma^2_p(\mathbf{x})) \tag{6}$$

where $m_p(\cdot)$ is the minimum mean squared error (MSE) linear unbiased predictor

$$m_p(\mathbf{x}) = \widehat{\beta}_0 + \Sigma(\mathbf{x},\cdot)^\top (\Sigma + C)^{-1} (\bar{\mathbf{Y}}_{\mathscr{D}} - \widehat{\beta}_0 \cdot 1_{k\times 1}), \tag{7}$$

and the corresponding variance is

$$\sigma^2_p(\mathbf{x}) = \tau^2 - \Sigma(\mathbf{x},\cdot)^\top (\Sigma + C)^{-1} \Sigma(\mathbf{x},\cdot) + \eta^\top [1^\top_{k\times 1} (\Sigma + C)^{-1} 1_{k\times 1}]^{-1} \eta \tag{8}$$

where $\widehat{\beta}_0 = [1_{k\times 1}^\top(\Sigma + C)^{-1}1_{k\times 1}]^{-1}1_{k\times 1}^\top(\Sigma + C)^{-1}\bar{\mathbf{Y}}_{\mathscr{D}}$ and $\eta = 1 - 1_{k\times 1}^\top(\Sigma + C)^{-1}\Sigma(\mathbf{x},\cdot)$ (Ankenman et al. 2010).

Since in reality the spatial correlation parameters τ^2 and $\boldsymbol{\phi}$ are unknown, maximum likelihood estimates are typically used for prediction, and the sample variance is used as an estimate for the simulation variance at design points C (Ankenman et al. 2010). By plugging these into Equations (7) and (8) we can obtain the estimated mean $\widehat{m}_p(\mathbf{x})$ and variance $\widehat{\sigma}_p^2(\mathbf{x})$. Thus, the metamodel we use is $\widehat{\mu}(\mathbf{x}) = \widehat{m}_p(\mathbf{x})$ with variance estimated by $\widehat{\sigma}_p^2(\mathbf{x})$. Notice that since accounting for the parameter estimation error is intractable, this plug-in estimator is commonly used in the kriging literature. In the metamodel-assisted bootstrapping approach, the dependent input model characterized by moments $\mathscr{M}$ can be interpreted as a location $\mathbf{x}$ in a d dimensional space.

4.4 Procedure to Build CI

Since there are both input and simulation estimation errors in the system mean performance estimates, in this section, we propose a procedure to build a CI quantifying the overall uncertainty by using our generalized metamodel-assisted bootstrapping approach. Asymptotic analysis shows that as $\mathbf{m}, B \to \infty$, the CI is consistent.

Based on a hierarchical approach, we propose the following procedure to build a $(1-\alpha)100\%$ bootstrap percentile CI:

1. Identify a design space E for input model moments $\mathscr{M}$ over which to fit the metamodel. Since the metamodel is used to propagate the input uncertainty measured by the bootstrapped moments $\widetilde{\mathscr{M}}_{\mathbf{m}}$ to the output mean, the design space is chosen to be the smallest ellipsoid covering most likely bootstrapped moments. See Barton et al. (2014) for more detailed information.
2. To obtain an experiment design $\mathscr{D} = \{(\mathscr{M}_i, n_i), i = 1, 2, \ldots, k\}$, use a Latin hypercube sample to embed k design points into the design space E and assign equal replications to these points to exhaust N.
3. For $i = 1$ to k (loop through k design points)
   (a) Use moment matching to calculate the marginal parameters $\boldsymbol{\theta}_\ell^{(i)}$ for $\ell = 1, 2, \ldots, L$.
   (b) By following the description in Section 4.2, find the NORTA representation with parameters $\big(\boldsymbol{\theta}_\ell^{(i)}, \rho_{\mathbf{Z}_\ell}^{(i)}\big)$ for F_ℓ with $\ell = 1, 2, \ldots, L$ and $d_\ell > 1$.

   Next i
4. At all design points, generate samples of $\mathbf{X}_\ell$ by using NORTA representations for F_ℓ with $d_\ell > 1$ and using standard approaches (Nelson 2013) for F_ℓ with $d_\ell = 1$, $\ell = 1, 2, \ldots, L$. Use these samples to drive simulations (see Section 4.2) and obtain outputs $\mathbf{y}_{\mathscr{D}}$. Compute the sample average $\bar{y}(\mathscr{M}_i)$ and sample variance $s^2(\mathscr{M}_i)$ of the simulation outputs, $i = 1, 2, \ldots, k$. Fit a SK metamodel to obtain $\widehat{m}_p(\cdot)$ and $\widehat{\sigma}_p^2(\cdot)$ using $\big(\bar{y}(\mathscr{M}_i), s^2(\mathscr{M}_i), \mathscr{M}_i\big)$, $i = 1, 2, \ldots, k$.
5. For $b = 1$ to B
   (a) Generate bootstrap moments $\widetilde{\mathscr{M}}_{\mathbf{m}}^{(b)}$ by following the procedure in Section 4.1.
   (b) Draw $\widehat{M}_b \sim \mathrm{N}\left(\widehat{m}_p(\widetilde{\mathscr{M}}_{\mathbf{m}}^{(b)}), \widehat{\sigma}_p^2(\widetilde{\mathscr{M}}_{\mathbf{m}}^{(b)})\right)$.

   Next b
6. Report CI: $\left[\widehat{M}_{(\lceil B\frac{\alpha}{2}\rceil)}, \widehat{M}_{(\lceil B(1-\frac{\alpha}{2})\rceil)}\right]$ where, $\widehat{M}_{(1)} \le \widehat{M}_{(2)} \le \cdots \le \widehat{M}_{(B)}$ are the sorted values.

If SK parameters $(\tau^2, \boldsymbol{\phi}, C)$ are known and we replace $\widehat{M}_b$ in Step 5(b) of the CI procedure with $M_b \sim \mathrm{N}\big(m_p(\widetilde{\mathscr{M}}_{\mathbf{m}}^{(b)}), \sigma_p^2(\widetilde{\mathscr{M}}_{\mathbf{m}}^{(b)})\big)$, then we can show that the CI obtained $[M_{(\lceil B\frac{\alpha}{2}\rceil)}, M_{(\lceil B(1-\frac{\alpha}{2})\rceil)}]$ is asymptotically consistent by Theorem 1. Further, this CI characterizes the impact from both input and metamodel uncertainty on the system performance estimate.

Theorem 1 Suppose that the following assumptions hold.

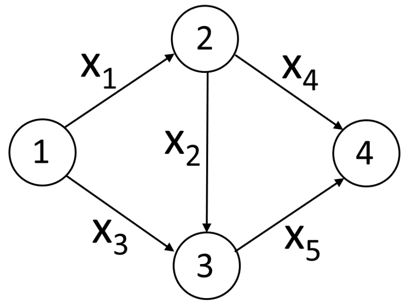

Figure 1: A stochastic activity network.

1. We have i.i.d. observations $\mathbf{X}_\ell^{(j)} \stackrel{i.i.d}{\sim} F_\ell^c$ for $j=1,2,\ldots,m_\ell$ and $\ell=1,2,\ldots,L$.
2. The $\varepsilon_j(\mathbf{x}) \stackrel{i.i.d.}{\sim} \mathrm{N}(0,\sigma_\varepsilon^2(\mathbf{x}))$ for any $\mathbf{x}$, and $M(\mathbf{x})$ is a stationary, separable GP with a continuous correlation function satisfying

$$1-r(\mathbf{x}-\mathbf{x}') \leq \frac{c}{|\log(\parallel \mathbf{x}-\mathbf{x}' \parallel_2)|^{1+\gamma}} \text{ for all } \parallel \mathbf{x}-\mathbf{x}' \parallel_2 \leq \delta$$

for some $c>0$, $\gamma>0$ and $\delta<1$, where $\parallel \mathbf{x}-\mathbf{x}' \parallel_2 = \sqrt{\sum_{j=1}^d (x_j - x_j')^2}$.
3. The input processes, simulation noise $\varepsilon_j(\mathbf{x})$ and GP $M(\mathbf{x})$ are mutually independent and the bootstrap process is independent of all of them.

As $m\to\infty$, we have $m_\ell/m \to c_\ell$ with $\ell=1,2,\ldots,L$ for a constant $c_\ell>0$. Then the interval $[M_{(\lceil B\frac{\alpha}{2}\rceil)}, M_{(\lceil B(1-\frac{\alpha}{2})\rceil)}]$ is asymptotically consistent, meaning the iterated limit

$$\lim_{m\to\infty}\lim_{B\to\infty} \Pr\{M_{(\lceil B\alpha/2\rceil)} \leq M_p(\mathscr{M}^c) \leq M_{(\lceil B(1-\alpha/2)\rceil)}\} = 1-\alpha. \tag{9}$$

Proof: By Theorem 1 in Xie et al. (2014b), we have $\widetilde{\mathscr{M}_{\mathbf{m}}} \stackrel{a.s.}{\to} \mathscr{M}^c$ as $m\to\infty$. Then, following similar steps as in proving Theorem 1 in Xie et al. (2014a), we can show Equation (9). □

5 EMPIRICAL STUDY

Since $F_1, F_2, \ldots, F_L$ are mutually independent, without loss of generality we consider an example with $L=1$ and suppress the subscript for the ℓth input distribution. We use a stochastic activity network as shown in Figure 1 to examine the finite-sample performance of our metamodel-assisted bootstrapping approach. Suppose that the time required to complete task (arc) i is denoted by X_i for $i=1,2,\ldots,5$ and $\mathbf{X}^\top = (X_1, X_2, \ldots, X_5)$. We wish to compute the time to complete the project, which is the longest path through the network, $Y = \max\{X_1+X_2+X_5, X_1+X_4, X_3+X_5\}$. We are interested in the mean response $\mathrm{E}[Y]$.

We assume that F^c is NORTA. The marginal distributions are $X_i \sim \exp(\theta_i^c)$ for $i=1,2,\ldots,5$ with means $\boldsymbol{\theta}^c = (10,5,12,11,5)^\top$. The Spearman rank correlation matrix is

$$R_{\mathbf{X}}^c = \begin{pmatrix} 1 & 0.5 & 0.5 & 0.3 & 0 \\ & 1 & 0.5 & 0 & 0.3 \\ & & 1 & 0 & 0.3 \\ & & & 1 & 0.1 \\ & & & & 1 \end{pmatrix}.$$

Therefore, the number of parameters characterizing the input model F is $d = d^{\dagger} + d^{*} = 5 + (5 \times 4)/2 = 15$. Since the true mean response $\mu(\boldsymbol{\vartheta}^c)$ is unknown, we run 10^7 replications and obtain the estimated true mean response 27.549 with standard error 0.005.

To evaluate our metamodel-assisted bootstrapping approach, we pretend that the input-model parameters $(\boldsymbol{\theta}^c, R^c_{\mathbf{X}})$ are unknown and they are estimated by m i.i.d. observations from F^c; this represents obtaining "real-world data." The goal is to build a CI quantifying the impact of both input and simulation estimation error on the system mean response estimate.

We compare metamodel-assisted bootstrapping to the conditional CI and direct bootstrapping. For the conditional CI, we fit the input distribution to the real-world data by moment matching and allocate all computational budget N replications to simulating the resulting system. In direct bootstrapping, we run N/B replications of the simulation at each bootstrap moment $\widetilde{\mathscr{M}}_m^{(b)}$, record the average simulation output $\bar{Y}_b = \bar{Y}(\widetilde{\mathscr{M}}_m^{(b)})$, and report the percentile CI $\left[\bar{Y}_{(\lceil B\frac{\alpha}{2}\rceil)}, \bar{Y}_{(\lceil B(1-\frac{\alpha}{2})\rceil)}\right]$.

Table 1 shows the statistical performance of conditional and direct bootstrapping CIs and metamodel-assisted bootstrapping with $k = 80$ design points, $m = 100, 500, 1000$ real-world observations, and computational budget of $N = 10^3$, 10^4, and 10^5 replications. We ran 1000 macro-replications of the entire experiment. In each macro-replication, we first generate m multivariate observations by using NORTA with parameters $(\boldsymbol{\theta}^c, R^c_{\mathbf{X}})$. Then, for the conditional CI, we run N replications at the estimated parameters $(\widehat{\boldsymbol{\theta}}_m, \widehat{R}_{\mathbf{X},m})$ and build CIs with nominal 95% coverage of the response mean. For direct bootstrapping and metamode-assisted bootstrapping, we use bootstrapping to generate $B = 1000$ samples moments to quantify the input uncertainty. Since $\mu(\cdot)$ is unknown, we use the fixed computational budget N to propagate the input uncertainty either via direct simulation or the SK metamodel to build percentile CIs with nominal 95% coverage.

From Table 1 we observe that under the same computational budget N, the conditional CIs that only account for the simulation uncertainty tend to have undercoverage. The CIs obtained by direct simulation are much wider and they typically have obvious over-coverage. The CIs obtained by metamodel-assisted bootstrapping have coverage much closer to the nominal level of 95%. As N increases and simulation estimation error decreases, the undercoverage problem for the conditional CI becomes worse. Since direct bootstrapping and metamodel-assisted bootstrapping use the same set of bootstrapped samples to quantify input uncertainty, the overcoverage for the direct bootstrap represents the additional uncertainty introduced while propagating the input uncertainty to the output mean. Table 1 shows that the metamodel can effectively use the computational budget and reduce the impact from simulation estimation error. Further, as the computational budget increases, the difference between the CIs obtained by the two methods diminishes.

Figure 2 shows a scatter plot of conditional CIs and CIs obtained by direct simulation and metamodel-assisted bootstrapping when $m = 500$, $k = 80$ and $N = 10^4$. It includes results from 100 macro-replications. The horizontal axis represents $(Q_L + Q_U)/2$ that gives a point estimate of system mean performance, where Q_L and Q_U are the lower and upper bounds of the CIs. The vertical axis is $(Q_U - Q_L)/2$ that gives half width of the CIs. Region 1 contains points that correspond to CIs having underestimation and Region 3 contains points corresponding to overestimation, while Region 2 contains CIs that cover $\mu(F^c)$ (Kang and Schmeiser 1990). Conditional CIs have short width and their centers have large variance. Therefore, they have serious undercoverage. For results from metamodel-assisted bootstrapping, the proportion of CIs in Region 2 is close to 95% and CIs outside tend to have underestimation.

6 CONCLUSIONS

In this paper, we extended the metamodel-assisted bootstrapping approach to cases with dependence in the input models. The input models are characterized by their marginal distributions and Spearman rank correlation, which are estimated from real-world data. Metamodel-assisted bootstrapping uses the bootstrap to quantify the input uncertainty and propagates it to the output mean by using a SK metamodel. This

Table 1: Results for CIs when $m = 100, 500, 1000$.

	$m = 100$	$N = 10^3$	$N = 10^4$	$N = 10^5$
conditional CI	coverage	43%	14.1%	4.4%
	CI width (mean)	2.2	0.7	0.2
	CI width (SD)	0.19	0.1	0.02
direct simulation	coverage	100%	100%	99.4%
	CI width (mean)	67.3	22.8	9.7
	CI width (SD)	6	1.9	0.8
metamodel-assisted	coverage	97.6%	96.1%	94.8%
	CI width (mean)	12.4	8	7.4
	CI width (SD)	2.5	1.2	0.9
	$m = 500$	$N = 10^3$	$N = 10^4$	$N = 10^5$
conditional CI	coverage	76.9%	32.1%	12%
	CI width (mean)	2.2	0.7	0.2
	CI width (SD)	0.1	0.02	0.01
direct simulation	coverage	100%	100%	100%
	CI width (mean)	66.4	21.9	7.5
	CI width (SD)	3.6	1	0.3
metamodel-assisted	coverage	96.3%	98.7%	97.2%
	CI width (mean)	10.3	4.5	3.5
	CI width (SD)	2.9	0.9	0.4
	$m = 1000$	$N = 10^3$	$N = 10^4$	$N = 10^5$
conditional CI	coverage	83.4%	42.2%	16.7%
	CI width (mean)	2.2	0.7	0.2
	CI width (SD)	0.09	0.02	0.01
direct simulation	coverage	100%	100%	100%
	CI width (mean)	66.5	21.7	7.2
	CI width (SD)	3.2	0.9	0.3
metamodel-assisted	coverage	97.1%	94%	96.6%
	CI width (mean)	10.2	3.6	2.6
	CI width (SD)	2.9	1.1	0.3

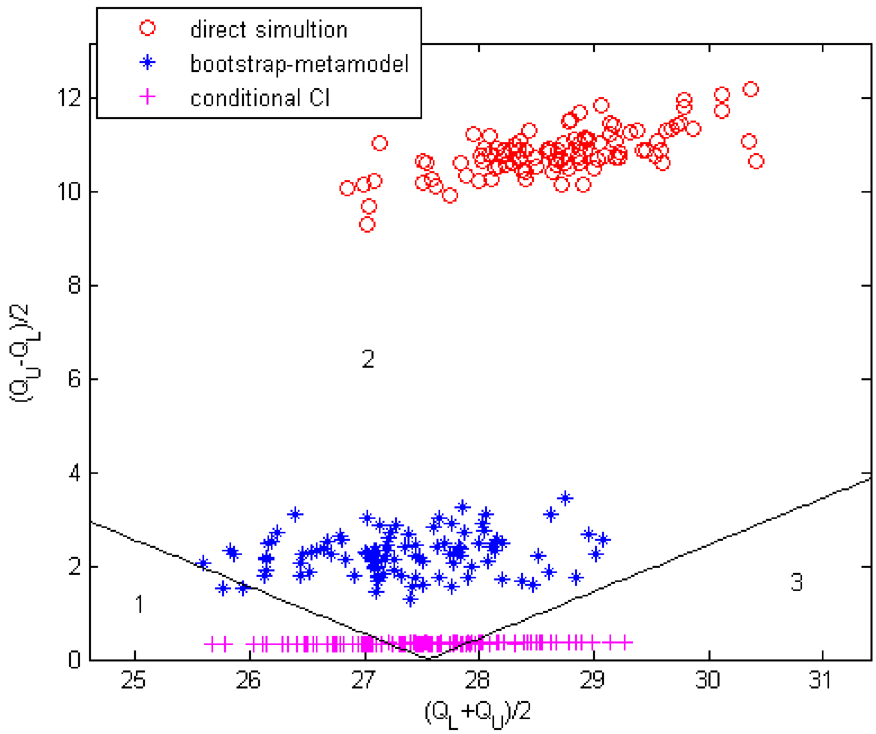

Figure 2: Scatter plot of conditional CIs and CIs obtained by direct simulation and metamodel-assisted bootstrapping when $m = 500$, $k = 80$ and $N = 10^4$.

approach delivers a CI quantifying the overall uncertainty of the system performance estimate. Compared with the direct bootstrap, the metamodel can make effective use of the simulation budget. An empirical study demonstrated that our metamodel-assisted bootstrap has good finite-sample performance under various quantities of real-world data and simulation budget.

ACKNOWLEDGMENTS

This paper is based upon work supported by the National Science Foundation under Grant No. CMMI-0900354.

REFERENCES

Ankenman, B. E., B. L. Nelson, and J. Staum. 2010. "Stochastic Kriging for Simulation Metamodeling". *Operations Research* 58:371–382.

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, 67–78: IEEE.

Barton, R. R., B. L. Nelson, and W. Xie. 2014. "Quantifying Input Uncertainty via Simulation Confidence Intervals". *Informs Journal on Computing* 26:74–87.

Biller, B., and C. G. Corlu. 2011. "Accounting for Parameter Uncertainty in Large-Scale Stochastic Simulations with Correlated Inputs". *Operations Research* 59:661–673.

Biller, B., and S. Ghosh. 2006. "Multivariate Input Processes". In *Handbooks in Operations Research and Management Science: Simulation*, edited by S. Henderson and B. L. Nelson, Chapter 5. Elsevier.

Cario, M. C., and B. L. Nelson. 1997. "Modeling and Generating Random Vectors with Arbitrary Marginal Distributions and Correlation Matrix". Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University.

Chen, X., B. E. Ankenman, and B. L. Nelson. 2012. "The Effect of Common Random Numbers on Stochastic Kriging Metamodels". *ACM Transactions on Modeling and Computer Simulation* 22:7.

Clemen, R. T., and T. Reilly. 1999. "Correlations and Copulas for Decision and Risk Analysis". *Management Science* 45:208–224.

Kang, K., and B. Schmeiser. 1990. "Graphical Methods for Evaluation and Comparing Confidence-Interval Procedures". *Operations Research* 38 (3): 546–553.

Livny, M., B. Melamed, and A. K. Tsiolis. 1993. "The Impact of Autocorrelation on Queueing Systems". *Management Science* 39:322–339.

Nelson, B. L. 2013. *Foundations and Methods of Stochastic Simulation: A First Course*. Springer-Verlag.

Shao, J., and D. Tu. 1995. *The Jackknife and Bootstrap*. Springer-Verlag.

Xie, W., B. L. Nelson, and R. R. Barton. 2014a. "Statistical Uncertainty Analysis for Stochastic Simulation". Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University.

Xie, W., B. L. Nelson, and R. R. Barton. 2014b. "Statistical Uncertainty Analysis for Stochastic Simulation with Dependent Input Models". Technical report, Department of Industrial Engineering and Management Sciences, Northwestern University.

Xie, W., B. L. Nelson, and J. Staum. 2010. "The Influence of Correlation Functions on Stochastic Kriging Metamodels". In *Proceedings of the 2010 Winter Simulation Conference*, edited by B. Johansson, S. Jain, J. Montoya-Torres, J. Hugan, and E. Yucesan, 1067–1078: IEEE.

AUTHOR BIOGRAPHIES

WEI XIE is an assistant professor in the Department of Industrial and Systems Engineering at Rensselaer Polytechnic Institute. Her research interests are in computer simulation, applied statistics and data analytics. Her email address is weixie2013@u.northwestern.edu.

BARRY L. NELSON is the Walter P. Murphy Professor and Chair of the Department of Industrial Engineering and Management Sciences at Northwestern University. He is a Fellow of INFORMS and IIE. His research centers on the design and analysis of computer simulation experiments on models of stochastic systems, and he is the author of *Foundations and Methods of Stochastic Simulation: A First Course*, from Springer. His e-mail address is nelsonb@northwestern.edu.

RUSSELL R. BARTON is Professor of Supply Chain and Information Systems and Professor of Industrial Engineering at the Pennsylvania State University. He currently serves as senior associate dean for research and faculty in the Smeal College of Business. He received a B.S. degree in Electrical Engineering from Princeton University and M.S. and Ph.D. degrees in Operations Research from Cornell University. Before entering academia, he spent twelve years in industry. He is a past president of the INFORMS Simulation Society, and currently serves as the I-Sim representative on the INFORMS Subdivisions Council and chairs the QSR Advisory Board. He is a senior member of IIE and IEEE and a Certified Analytics Professional. His research interests include applications of statistical and simulation methods to system design and to product design, manufacturing and delivery. His email address is rbarton@psu.edu.