Input uncertainty in stochastic simulations in the presence of dependent discrete input variables

DOI:
10.1057/s41273-017-0051-3

Document status and date:
Published: 02/10/2018

Document Version:
Publisher's PDF, also known as Version of Record (includes final page, issue and volume numbers)

Journal of Simulation

ISSN: 1747-7778 (Print) 1747-7786 (Online) Journal homepage: http://www.tandfonline.com/loi/tjsm20

Input uncertainty in stochastic simulations in the presence of dependent discrete input variables

Alp Akcay & Bahar Biller

To cite this article: Alp Akcay & Bahar Biller (2018) Input uncertainty in stochastic simulations in the presence of dependent discrete input variables , Journal of Simulation, 12:4, 295-306, DOI: 10.1057/s41273-017-0051-3

To link to this article: https://doi.org/10.1057/s41273-017-0051-3

Published online: 14 Dec 2017.

JOURNAL OF SIMULATION
2018, VOL. 12, NO. 4, 295–306
https://doi.org/10.1057/s41273-017-0051-3

OPEN ACCESS Check for updates

Input uncertainty in stochastic simulations in the presence of dependent discrete input variables

Alp Akcay[a] and Bahar Biller[b]

[a]School of Industrial Engineering, Eindhoven University of Technology, Eindhoven, The Netherlands; [b]Global Research Center, General Electric, Niskayuna, NY, USA

ABSTRACT
This paper considers stochastic simulations with correlated input random variables having NORmal-To-Anything (NORTA) distributions. We assume that the simulation analyst does not know the marginal distribution functions and the base correlation matrix of the NORTA distribution but has access to a finite amount of input data for statistical inference. We propose a Bayesian procedure that decouples the input model estimation into two stages and overcomes the problem of inconsistently estimating the base correlation matrix of the NORTA distribution in the presence of discrete input variables. It further allows us to estimate the variability of the simulation output data that are attributable to the input uncertainty due to not knowing the NORTA distribution. Using this input uncertainty estimate, we introduce a simple yet effective method to obtain input uncertainty adjusted credible intervals. We illustrate our method in an assemble-to-order production system with a correlated demand arrival process.

ARTICLE HISTORY
Received 6 July 2015
Accepted 10 January 2017

KEYWORDS
Multivariate input modeling; input uncertainty; simulation output analysis

1. Introduction

The use of simulation to estimate the performance of a stochastic system often requires the modeling of input processes with components that are not necessarily independent of each other. For example, the demand of a product may depend on the demands for other products. The processing times of a workpiece across several machines in a job shop and the exchange rates in a global supply chain could be other examples. Ignoring the dependence among input random variables often leads to poor estimates of the system performance (Biller & Corlu, 2011; Xie, Nelson, & Barton, 2014c).

While the full joint distribution of dependent random variables is an ideal input model, it is rarely used in simulation due to certain difficulties that arise as the number of random variables increases; we refer the reader to Ghosh (2004) for a detailed discussion. An alternative is to construct an input model that matches partial specifications of the multivariate input process. In particular, the multivariate input process is commonly specified in terms of the marginal (univariate) distributions and a measure of dependence that captures the dependence structure (Biller & Ghosh, 2006). Cario and Nelson (1997) introduce the NORmal-To-Anything (NORTA) method based on this view for generating random vectors with a prescribed correlation matrix. The NORTA method generates a standard multivariate normal vector (also called a base vector) and then performs a component-wise transformation to obtain the desired input vector, which is referred as having a NORTA distribution. In this paper, we consider a Bayesian approach to make inference on the Pearson product-moment correlation matrix of the base vector and the nonparametric marginal distributions of the input random variables.

A simulation analyst faces two challenges in this setting: The first challenge is the presence of discrete-valued variables in the input process. When an estimator of the marginal cumulative distribution function (cdf) is applied to discrete input data, a subsequent application of the inverse standard normal cdf does not lead to normally distributed input data. The direct use of the transformed input data, thus, leads to an inaccurate estimation of the base correlation matrix associated with the multivariate normal distribution from which the base vectors are generated in the NORTA framework. We illustrate this problem in Section 1.1. The second challenge is the need to account for the uncertainty in the marginal cdfs and the NORTA base correlation matrix in simulation output analysis. In Section 1.2, we review the previous work on the so-called *input uncertainty* and

CONTACT Alp Akcay a.e.akcay@tue.nl
This paper has been re-typeset by Taylor & Francis from the manuscript originally provided to the previous publisher.

discuss how we position the contributions of our paper within this literature.

1.1. The problem of discrete input random variables in NORTA estimation

We illustrate the problem of discrete input random variables in NORTA estimation by using a toy example with two input random variables in Section 1.1.1. and within the context of a discrete event simulation of an assemble-to-order production system in Section 1.1.2.

1.1.1. A simplified illustration

Suppose that X_1 is a standard uniform random variable, and X_2 is either zero or one with probability one-half each. Let F_1 and F_2 denote the cdf of X_1 and X_2, respectively. The NORTA method aims to generate samples of X_1 and X_2 with a specified correlation. To this end, the base random variables Z_1 and Z_2 are sampled from the standard bivariate normal distribution with the corresponding base correlation, say θ. The desired input variables are then obtained via the transformations $X_1 = F_1^{-1}(\Phi(Z_1))$ and $X_2 = F_2^{-1}(\Phi(Z_2))$, where $\Phi(\cdot)$ is the standard normal cdf. However, this method is *not* directly implementable in practice since the marginal cdfs F_1 and F_2 and the base correlation θ are not known.

A common approach to avoid making assumptions on the functional form of the cdf is to use the empirical cdf $\hat{F}_i(x) = (1/m)\sum_{t=1}^{m} \mathbf{1}\left(x_{t,i} \leq x\right)$ for $i = 1, 2$, where $\{(x_{t,1}, x_{t,2}); t = 1, 2, \ldots, m\}$ are the real-world input data and $\mathbf{1}(\cdot)$ is the indicator function. Then, the input data are transformed into $\hat{z}_{t,i} = \Phi^{-1}((m/(m+1))\hat{F}_i(x_{t,i})) = \Phi^{-1}(\tilde{F}_i(x_{t,i}))$ for $i = 1, 2$ and $t = 1, 2, \ldots, m$, and the transformed data are used to estimate θ (the scaling factor $m/(m+1)$ is commonly used to avoid infinities). When both marginal cdfs are continuous, Genest, Ghoudi, and Rivest (1995) show that the maximum-likelihood estimator $\hat{\theta}_{\text{MLE}}$ (i.e., the value of θ that maximizes the likelihood function $\sum_{t=1}^{m} \log \phi\left(\tilde{z}_{1,t}, \tilde{z}_{2,t};\theta\right)$with $\varphi(\cdot, \cdot;\theta)$ the standard bivariate normal density with correlation θ) is consistent. However, this does not hold if some of the marginals are discrete; notice that the transformed binary data $\{\tilde{z}_{t,2};\ t = 1, 2, \ldots, m\}$ are far from being normal. In fact, $\tilde{Z}_{t,2}$ approximately takes the values of $\Phi^{-1}((1/2)(m/(m+1)))$ and $\Phi^{-1}(m/(m+1))$ with equal probabilities. That is, the transformation $\Phi^{-1}(\tilde{F}_2(\cdot))$ on the discrete input data $\{x_{t,2};\ t = 1, 2, \ldots, m\}$ does not change its distribution into normal but only changes the sample space. Figure 1 plots the sampling distribution of $\hat{\theta}_{\text{MLE}}$ obtained from 1000 bootstrap samples for $\theta = 0.6$. Not surprisingly, the variance of $\hat{\theta}_{\text{MLE}}$ decreases as the length of the real-world input data increases. However, its mean converges to an incorrect value of θ; this is clearly seen when the length of the input data reaches 1000 at which the mean value of $\hat{\theta}_{\text{MLE}}$ is about 0.47 but not 0.6.

When there is no input uncertainty, the existence of discrete input variables in a NORTA framework has been of interest in simulation input modeling literature while solving the *pairwise correlation-matching problem*, i.e., searching for the base correlation that would produce the *specified* correlation for each pair of input variables. In this paper, we directly estimate the base correlation matrix from the input data instead of solving the correlation-matching problem to obtain the corresponding base correlation matrix. Therefore, we also avoid the well-known problem of ending up with a nonpositive semidefinite base correlation matrix after solving the correlation-matching problem; see Ghosh and Henderson (2002) for a detailed discussion of this problem.

1.1.2. Illustration in an Assemble-To-Order (ATO) system simulation

We consider the ATO system in Hong and Nelson (2006) where components are made to stock to meet random demands for finished products. Each product requires a set of key components and a set of nonkey components. Components bring revenue of s and have a holding cost of h for each time unit in inventory. If all key components are in stock, the customer buys the product assembled from all the key components and available non key components; otherwise, the customer leaves. The production time of components is normal with mean a and variance b^2. We tabulate all the parameters in Tables 1 and 2. The system is operated under a

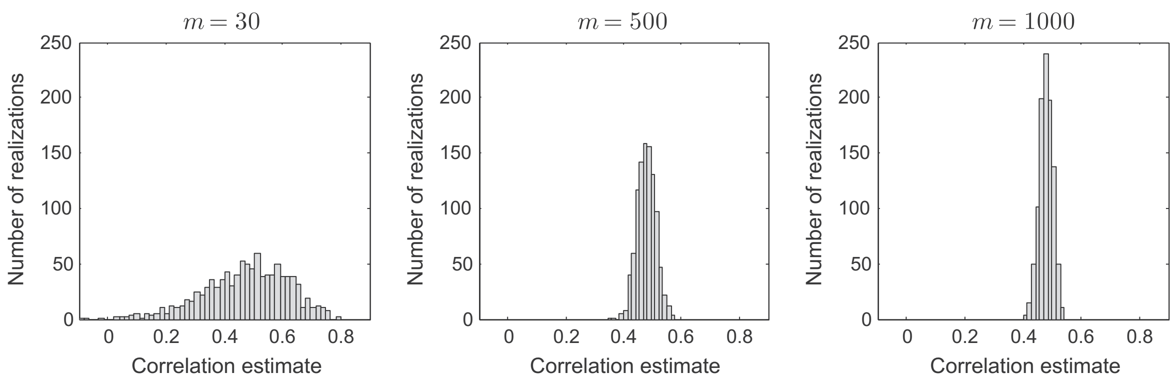

Figure 1. Sampling distribution of the NORTA base correlation estimator when its true value is 0.6.

Table 1. Parameters related to the components.

Component	s	h	a	b
I	1	2	0.15	0.0225
II	2	2	0.40	0.0600
III	3	2	0.25	0.3750
IV	4	2	0.15	0.0225
V	5	2	0.25	0.0375
VI	6	2	0.08	0.0120

Table 2. The number of key and nonkey components used by each product.

Product	Keys				Nonkeys	
	I	II	III	IV	V	VI
1	1	0	0	1	0	1
2	1	1	0	0	1	0
3	0	0	1	1	1	0

continuous-review base-stock policy. Simulation is used to estimate the mean profit per unit time by performing independent replications until the half-width of the 95% confidence interval is 5% of the average profit.

A customer arrives at random times and declares how many units of each product will be purchased. Let the demand have a *discrete* uniform distribution on the integers {1, …, 5} for each finished product and the time between customer arrivals is exponentially distributed with a mean of one time unit. There are two types of dependence in the input process: (i) demand random variables are correlated, (ii) there is a relationship between the time since the last customer arrival and the demand random variables. That is, there are four input variables in the demand arrival process, and we let the NORTA distribution model their underlying dependence structure with base correlation matrix Σ such that $\Sigma_{i,j} = \rho$ for $i \neq j$, $i, j \in \{1, \ldots, 4\}$ and diagonal elements equal to one. Suppose that Σ is unknown by the simulation analyst and estimated with maximum-likelihood estimation. The confidence interval is then *conditional* on the maximum-likelihood estimate of Σ, and hence, ignores the uncertainty around it. Table 3 presents the confidence interval coverage (i.e., the probability of trapping the true mean profit), average confidence interval length (CIL), and the standard deviation of the CIL calculated from 500 macroreplications. It also reports the estimates of the mean absolute percentage error (MAPE) and the mean square error (MSE) to assess the quality of the point estimate of the mean simulation output.

Table 3 shows that the coverage may decrease with increasing length of the input data—which is counter-intuitive as one expects to have less uncertainty around the maximum likelihood estimator of an unknown parameter with more input data. For example, when $\rho = 0.9$, we observe that coverage decreases from 45.8 to 13.2% as m, the length of the real-world input data, increases from 50 to 500. If the input variables were not discrete, it is well documented in the literature that the coverage increases when more input data are used in the

Table 3. Properties of the mean simulation output and the accompanying confidence interval when NORTA base correlation matrix is estimated with maximum-likelihood method; base-stock level is 10 for each component.

ρ	m	Coverage (%)	Avg. (CIL)	Std. (CIL)	MAPE	MSE
0.9	50	45.8	4.12	0.14	5.37	7.31
	100	35.0	4.09	0.13	5.86	8.11
	250	23.2	4.06	0.12	6.73	10.06
	500	13.2	4.02	0.10	7.54	12.05
0.6	50	44.4	3.15	0.25	6.91	8.15
	100	45.8	3.15	0.18	5.79	5.41
	250	46.8	3.13	0.14	5.57	4.71
	500	42.0	3.12	0.12	5.48	4.22
0.3	50	29.0	2.28	0.28	9.92	8.48
	100	40.0	2.28	0.20	7.40	4.73
	250	51.0	2.28	0.14	5.26	2.34
	500	58.2	2.27	0.11	4.52	1.72

maximum-likelihood estimation; see Biller and Corlu (2011) and Akcay, Biller, and Tayur (2012). Table 3 also shows for $\rho = 0.9$ that the MAPE and MSE both increase as m increases, meaning that the estimate of the mean simulation output moves away from its true value with more input data. This highlights the importance of handling discrete input random variables while estimating the NORTA base correlation matrix especially when the input process is highly correlated.

1.2. Input uncertainty in simulation output: background and contributions

The input uncertainty in stochastic simulations has been addressed in various ways; we refer the reader to Barton (2012) for a detailed survey. The main research streams can be classified based on whether (i) the method is Bayesian (e.g., Chick, 2001; Biller & Corlu, 2011) or frequentist (e.g., Xie, Nelson, & Barton,2014b; Lin, Song, & Nelson, 2015) and (ii) sampled values of the unknown input model components are fed into simulation directly (e.g., Ankenman & Nelson, 2012; Song and Nelson,) or by means of a simulation metamodel (e.g., Barton, Nelson, & Xie, 2014; Xie, Nelson, & Barton, 2014a). We position our work as Bayesian with direct resampling from the posterior distributions of multivariate input distributions.

In the Bayesian approaches, the input uncertainty is characterized by the posterior distributions of the parameters and the family of the input distributions. Subsequently, either the direct simulation or metamodeling is used to propagate the input uncertainty to the output mean. Chick (2001) performs Bayesian model averaging with direct simulation to obtain a point estimate and a confidence interval for the mean simulation output by averaging the output data obtained under posterior samples of the input parameters. Zouaoui and Wilson (2003) use a similar idea to quantify the contributions of input parameter uncertainty and the intrinsic simulation uncertainty by taking multiple simulation runs at the sampled input parameters. Zouaoui and Wilson (2004) extend the work of Zouaoui and Wilson (2003) to account for the uncertainty around the families of the input distributions as well. Direct sampling from the posterior distributions followed by subsequent simulations can be time-consuming and may lead to large confidence intervals. Ng and Chick (2006) aim to minimize the variance of the posterior simulation output distribution by representing the input parameter uncertainty in a linear metamodel and using large-sample approximations. Xie et al. (2014a) propose a global metamodel based on stochastic kriging with simulations performed at well-chosen design points, and find that the resulting confidence intervals are shorter than direct sampling when intrinsic simulation uncertainty is large. However, none of the papers above explicitly model and estimate the stochastic dependence between the input random variables.

To the best of our knowledge, there are two papers that consider the dependence between input random variables in a NORTA framework. Biller and Corlu (2011) adopt the Bayesian model average approach of Chick (2001) and propose a method based on the copula-vine representation of the input model to separately estimate and sample the correlation matrix and the parameters of known input distributions. More recently, Xie et al. (2014c) extend the work of Barton et al. (2014) in a NORTA setting by proposing an approach that identifies the NORTA base correlation as a function of the input-distribution moments. However, this approach uses a transformation between rank correlations and product-moment correlations that are only valid when all the marginal distributions are continuous. Similarly, the copula-vine representation of Biller and Corlu (2011) fails to work when at least one of the marginal distributions is discrete. Our paper fills this gap by allowing both continuous and discrete distributions to construct the input random vector.

The remainder of the paper is organized as follows: Section 2 proposes a Bayesian approach to separately estimate and sample the marginal cdfs and base correlation matrix of the NORTA distribution in the presence of discrete input variables and without the problem illustrated in Section 1.1. Section 3 quantifies the input uncertainty in this approach and discusses how it can be used to improve the simulation output analysis. Section 4 revisits the assemble-to-order system simulation to illustrate this improvement, and Section 5 concludes with a summary of results.

2. A Bayesian approach to estimate the NORTA distribution

We let $\mathbf{X} = (X_1, X_2, \dots, X_d)$ denote a multivariate input process composed of d correlated input random variables each of which can take either continuous or discrete values. Cario and Nelson (1997) introduce the NORTA method to generate random samples of $\mathbf{X}$ as follows: (i) Generate a multivariate normal vector $\mathbf{Z} = (Z_1, Z_2, \dots, Z_d)$ with mean vector zero and the base correlation matrix Σ. (ii) Compute the input variables via $X_i = F_i^{-1}(\Phi(Z_i))$ for $i = 1, 2, \dots, d$, where $F_i(\cdot)$ is the *unknown* marginal distribution of X_i. The input vector $\mathbf{X}$ generated by this procedure is said to have a d-dimensional NORTA distribution. Notice that the input variable $F_i^{-1}(\Phi(Z_i))$ has the required marginal distribution since $\Phi(Z_i)$ has standard uniform distribution. The base correlation matrix Σ is determined by solving the pairwise correlation-matching problem so that it induces the underlying correlation structure of $\mathbf{X}$.

When all the marginal cdfs are continuous, the NORTA distribution is of the form

$$\begin{aligned}\mathbb{P}(X_1 \le x_1, \dots, X_d \le x_d) &= \mathbb{P}(F_1(X_1) \le F_1(x_1), \dots, \\ &\qquad F_d(X_d) \le F_d(x_d)) \\ &= \mathbb{P}(Z_1 \le \Phi^{-1}(F_1(x_1)), \dots, Z_d \le \Phi^{-1}(F_d(x_d))) \\ &= \Phi_d(\Phi^{-1}(F_1(x_1)), \dots, \Phi^{-1}(F_d(x_d)); \Sigma), \end{aligned} \tag{1}$$

where $\Phi_d(\Phi^{-1}(\cdot), \dots, \Phi^{-1}(\cdot); \Sigma)$ is a normal copula (Nelsen, 2006). Equation (1) shows that the normal copula represents a unique NORTA distribution for the input vector $\mathbf{X}$ (Biller & Corlu, 2011; Corollary 1). That is, the dependence structure of a d-dimensional NORTA distribution is represented by a d-dimensional normal copula. This follows from the fact that the NORTA distribution is based on transformation from a multivariate normal distribution. If at least one of the input random variables is discrete, the normal copula representation of the NORTA distribution is not necessarily unique, but Equation (1) remains a well-defined joint distribution function for $\mathbf{X}$ (Smith & Khaled, 2012).

The NORTA method is not directly implementable since the marginal cdfs and the base correlation matrix are unknown. We present a Bayesian procedure to account for the uncertainty around them in the presence of the real-world input data $\underline{\mathbf{x}} = (\mathbf{x}_1, \mathbf{x}_2, \dots, \mathbf{x}_m)'$ of length m, where $\mathbf{x}_t = (x_{t,1}, x_{t,2}, \dots, x_{t,d})$ and $x_{t,i}$ is the tth realization of the input random variable X_i. The procedure consists of two steps: (i) Constructing a posterior distribution for the marginal cdf of each input random variable and (ii) approximating the posterior distribution of the base correlation matrix.

Our motivation for decoupling the estimation of marginal cdfs from the base correlation matrix stems from its widespread use in estimating the copula functions. In the copula literature, the estimation of the marginal distribution parameters and copula parameters in separate stages is known as the method of inference for margins (Cherubini, Luciano, & Vecchiato, 2004), and Joe (1997) proves that the resulting estimators verify the properties of strong consistency for continuous marginal cdfs. Biller and Corlu (2011) adapt the method of inference for margins within the context of Bayesian estimation and sampling the parameters of the NORTA distribution for continuous marginal cdfs. In this paper, we allow the input random variables to be discrete. We also note that our method in Section 2.2 is robust against the estimation errors of the marginal cdfs as it only consider the ranking of the real-world input data rather than using the transformed data $z_{t,i} = \Phi^{-1}(F(x_{t,i}))$ for directly estimating the NORTA base correlation matrix.

2.1. Inference for the marginal distributions

The parametric distribution families with a fixed number of parameters (e.g., normal, exponential, Poisson, lognormal) are often used as input models to generate the random variates in simulation. However, this approach can suffer from over- or underfitting of input data when there is a misalignment between the complexity of the model (represented by the number of parameters) and the amount of available input data. Thus, we adopt a *nonparametric* Bayesian approach for modeling the unknown marginal cdfs of the NORTA distribution. Our motivation is based on the idea that a nonparametric model with an unbounded complexity mitigates underfitting, while the Bayesian approach of computing the posterior over a set of distributions mitigates overfitting (Teh, 2010).

2.1.1. Discrete input variables

Suppose that the input random variable X_i takes values from a set with cardinality K_i. While this assumption seems restrictive, K_i can be taken arbitrarily big to model input variables with large sample spaces. Let $\mathbf{p}_i = (p_{i,1}, \dots, p_{i,K_i})$ denote the corresponding probability mass function (pmf), which is *unknown* to the simulation analyst. We capture the uncertainty around $\mathbf{p}_i$ by representing it with a Dirichlet distribution. The probability density function (pdf) of $\mathbf{p}_i$ is then given by

$$\mathrm{Dir}\,D(\mathbf{p}_i|\alpha_{i,1}, \dots, \alpha_{i,K_i}) = \frac{\Gamma(\sum_{k=1}^{K_i} \alpha_{i,k})}{\prod_{k=1}^{K_i} \alpha_{i,k}} \prod_{k=1}^{K_i} p_{i,k}^{\alpha_k - 1} \mathbf{1}(\mathbf{p}_j \in S_{K_i}),$$

where $\Gamma(\cdot)$ is the gamma function, and $S_{K_i} = \{\mathbf{p}_i : 0 \le p_{i,k} \le 1,\ k = 1, \dots, K_i, \text{ and } \sum_{k=1}^{K_i} \mathbf{p}_{i,k} = 1\}$. We note that the concentration parameter $\alpha_{i,k} \ge 0$ controls where the peak in the pdf of $\mathbf{p}_i$ occurs and $\sum_{k=1}^{K_i} \alpha_{i,k} > 0$ controls the amount of peakedness in this pdf. More specifically, the expectation of the distribution of $\mathbf{p}_i$ is given by $\left(\alpha_{i,1}, \dots, \alpha_{i,K_i}\right) / \sum_{k=1}^{k_i} \alpha_{i,k}$. Therefore, the simulation analyst can incorporate a prior knowledge on the distribution of an input random variable via the concentration parameters. For example, setting $\alpha_{i,k} = 1$ for $k = 1, \dots, K_i$ represents a flat prior, implying no prior knowledge or expert opinion. Let $(f_{i,1}, \dots, f_{i,K_i})$ denote the sample counts of the input random variable X_i in the historical data $\{x_{t,i}: t = 1, \dots, m\}$, i.e., $\sum_{k=1}^{K_i} f_{i,k} = m$. Since the Dirichlet distribution is a conjugate prior for p_i (Gelman et al., 2013), its density can be updated as $\mathrm{Dir}\,D(\mathbf{p}_i|\alpha_{i,1}+f_{i,1}, \dots, \alpha_{i,K_i} + f_{i,K_i})$ after observing the input data $\{x_{t,i} : t = 1, \dots, m\}$.

It is important to note that our ultimate goal is to account for the uncertainty around the marginal cdfs of the NORTA distribution in the simulation output data. For the discrete input random variables, the ease in generating the posterior samples of the marginal cdfs from a Dirichlet distribution will, therefore, be an important factor in the implementation of Algorithms 2, 3, and 4 in Section 3. One can conveniently sample the random vector $\mathbf{p}_i$ from the K_i-dimensional Dirichlet distribution with parameters $(\alpha_{i,1}, \cdots, \alpha_{i,K_i})$ by first drawing random samples w_{ij} from Gamma $(\alpha_{i,j}, 1)$ distribution each with density $w_{i,j}^{\alpha_{i,j}-1}\exp(-w_{i,j})/\Gamma(\alpha_{i,j})$ and then setting $p_{i,j} = w_{i,j} / \sum_{j=1}^{K_i} w_{i,j}$ for $j = 1, \dots, K_i$ (Gelman et al., 2013).

2.1.2. Continuous input variables

The approach in Section 2.1.1. can be generalized to continuous input random variables by using the Dirichlet process instead of the Dirichlet distribution. The Dirichlet process can be intuitively thought as a distribution over a set of probability distributions (Ferguson, 1973). Thus, a sample from the Dirichlet process is itself a probability distribution. The Dirichlet process prior $\mathrm{DirP}(F_i^0, \beta_i)$ is described by two quantities, a base measure F_i^0, and a concentration parameter β_i. In a Bayesian view, the quantities F_i^0 and β_i reflect the prior belief of the simulation practitioner on the cdf of the input random variable X_i: F_i^0 can be regarded as an a priori best guess or expert opinion for the cdf X_i, while the parameter β_i controls how tightly concentrated the prior is around F_i^o, reflecting the simulation analyst's confidence in F_i^0.

After observing the input data $\{x_{t,i} : t = 1, \dots, m\}$, the posterior distribution of F_i is again characterized by a Dirichlet process, i.e., $F_i \sim \mathrm{DirP}(\bar{F}, \beta_i + m)$, where

$$\bar{F} = \frac{m}{\beta_i + m}\hat{F} + \frac{\beta_i}{\beta_i + m}F_i^o \tag{2}$$

with $\hat{F}_i$ the empirical cdf of the input random variable X_i (Ferguson, 1973). Equation (2) shows that the base measure of the posterior $\bar{F}_i$ is a weighted average of the prior base F_i^0 and the empirical distribution $\hat{F}_i$. Notice that the concentration parameter β_i is determined in accordance with the simulation analyst's prior belief on the correctness of the prior base F_i^o relative to the empirical cdf $\hat{F}_i$. Specifically, if the simulation analyst considers to place $100\gamma\%$ weight to the prior base F_i^0 and $100(1-\gamma)\%$ to the empirical cdf, the value of β_i is set to $m\gamma/(1-\gamma)$. As β_i approaches zero, the base measure of posterior becomes the empirical distribution and it reduces to a noninformative prior. For a given value of β_i, as the length of the input data m increases, the posterior is dominated by the empirical distribution, while the empirical distribution becomes a closer approximation of the true input distribution. Thus, the convergence of the Dirichlet process to the true input distribution also-justifies its use for modeling the uncertainty in the cdf of an input random variable (Lam & Zhou, 2015).

Sethuraman (1994) shows that a Dirichlet process with base distribution $\bar{F}_i$ and concentration parameter $\beta + m$ can be constructed according to the stick-breaking process $\mathcal{SB}(x) = \sum_{k=1}^{s} \pi_{k,i} \mathbf{1}(x = x_{k,i})$ for $S = \infty$, where $\pi_{k,i} = v_{k,i} \prod_{\tau=1}^{k-1}(1 - v_{k,i}), v_{k,i} \overset{\text{iid}}{\sim} \text{Beta}(1, \beta_i + m)$ and $x_{k,i} \overset{\text{iid}}{\sim} \bar{F}_i$. We adopt this stick-breaking process to draw a cdf from the Dirichlet process. In any practical implementation of this algorithm, it is not possible to choose $S = \infty$. However, if S is set to a sufficiently large value, the cumulative probability $\sum_{k=1}^{S} \pi_{k,i}$ is close to one. In particular, one can continue to the stick-breaking process until the convergence criterion $1 - \sum_{k=1}^{S} \pi_{i,k} < \epsilon$ is satisfied for an arbitrarily small positive ϵ. It is important to note that the support of the cdf sampled from a Dirichlet process is a finite set of real numbers. A commonly used approach in simulation to generate continuous random variates from distributions with finite support is to use the linearly interpolated cdf as the input model (Barton & Schruben, 2001).

2.2. Inerence for the base correlation matrix

The input vector $\mathbf{X}$ that has a NORTA distribution means that the real-world input data are generated by the transformation $x_{t,i} = F_i^{-1}(\Phi(z_{t,i}))$ for $t = 1, 2, \ldots, m$ and $i = 1, 2, \ldots, d$. Let $\underline{\mathbf{Z}}$ denote the m by d matrix $(\mathbf{Z}_1, \mathbf{Z}_2, \ldots, \mathbf{Z}_m)^{\mathrm{T}}$, which is composed of m independent samples of $\mathbf{Z}_t = (Z_{t,1}, Z_{t,2}, \ldots, Z_{t,d})$. Notice that $\mathbf{Z}_t$ has a d-dimensional normal distribution with mean zero and correlation matrix Σ. The objective of this section is to capture the uncertainty in Σ *without* making a parametric assumption on the functional form of the marginal cdfs.

It is important to note that the realization of the matrix $\underline{\mathbf{Z}}$ is *not* observed in practice, but the simulation analyst has only access to the real-world input data $\underline{x}$. In fact, if $\underline{\mathbf{Z}}$ were observed, it could be used to directly estimate the base correlation matrix Σ. However, the input data $\underline{\mathbf{X}}$ provide some information about the base random variables even when the marginal distribution functions are unknown. We illustrate it with a simple example as follows: Suppose that the last five historical realizations of X_i are given by $(x_{1,i}, \ldots, x_{5,i}) = (1, 1, 2, 3, 2)$. Based on this input data, we can conclude that $z_{3,i} > z_{1,i}$; $z_{3,i} > z_{2,i}$; $z_{4,i} > z_{3,i}$; $z_{4,i} > z_{5,i}$; $z_{5,i} > z_{1,i}$; and $z_{5,i} > z_{2,i}$. This information can be combined with similar order relationships from the other input variables to capture the uncertainty in $\underline{Z}$ conditional on the base correlation matrix Σ.

More generally, since the marginal cdfs are nondecreasing functions, observing $x_{t,i} < x_{t',i}$ implies $z_{t,i} < z_{t',i}$ for $1 \le t \neq t' \le m$. It follows that the realization of $\underline{\mathbf{Z}}$, which we denote with $\underline{\mathbf{z}}$, is constrained to belong to the set

$$\mathcal{Z}(x) = \{\mathbf{z} \in \mathbb{R}^{m \times d} : L_{t,i}(\mathbf{z}) < z_{t,i} < U_{t,i}(\mathbf{z})\},$$

where $L_{t,i}(\mathrm{z}) := \max\{z_{t',i} : x_{t',i} < x_{t,i}\}$ and $U_{t,i}(\mathrm{z}) := \min\{z_{t',i} : x_{t,i} < x_{t',i}\}$. The set $\mathcal{Z}(\underline{x})$ contains the possible realizations of $\mathbf{Z}_t$, $t = 1, 2, \ldots, m$, consistent with the ordering of the observed input data $\underline{\mathbf{x}}$. Let $\underline{\mathbf{X}}$ denote the collection of NORTA input vectors $(\mathbf{X}_1, \mathbf{X}_2, \ldots, \mathbf{X}_m)^T$ whose realization is the input data $\underline{x}$ It follows that

$$\begin{aligned}
\mathbb{P}(\underline{\mathbf{X}}|\Sigma, F_1, \ldots, F_d) &= \mathbb{P}(\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}}), \underline{\mathbf{X}}|\Sigma, F_1, \ldots, F_d) \\
&= \mathbb{P}(\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}})|\Sigma, F_1, \ldots, F_d)\, \mathbb{P}(\underline{\mathbf{X}}|\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}}), \Sigma, F_1, \ldots, F_d) \\
&= \mathbb{P}(\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}})|\Sigma)\, \mathbb{P}(\underline{\mathbf{X}}|\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}}), \Sigma, F_1, \ldots, F_d)
\end{aligned} \tag{3}$$

The first equality follows because the event $\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}})$ occurs whenever $\underline{\mathrm{X}}$ is observed. The last equality is based on the fact that the probability of the event $\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}})$ does not depend on the marginal cdfs. Hoff (2007) proposes to drop the second term $\mathbb{P}(\underline{\mathbf{X}}|\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}}), \Sigma, F_1, \ldots, F_d)$ on the right-hand side of Equation (3) and use the first term $\mathbb{P}(\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}})|\Sigma)$ as the likelihood of the historical data. The intuition is that $\mathbb{P}(\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}})|\Sigma)$ includes the most of the information about Σ in the full likelihood of the historical data. Murray, Dunson, Carin, and Lucas (2013) provide empirical evidence on this, and prove that the use of extended rank likelihood leads to the asymptotically consistent estimation of the base correlation matrix Σ.

We use the extended rank likelihood to capture the uncertainty around the NORTA base correlation matrix Σ. However, instead of working with Σ directly as in Hoff (2007) and Murray et al. (2013), we focus on the precision matrix $\mathbf{K}$ of a multivariate normal distribution with correlation matrix Σ. In particular, we perform Bayesian inference for Σ by constructing a Markov chain having a stationary distribution equal to $\mathbb{P}(\mathbf{K}|\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}})) \propto \mathbb{P}(\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}})|\mathbf{K})\mathbb{P}(\mathbf{K})$ To this end, we construct a Gibbs sampler in terms of the precision matrix K, and this is more convenient to implement than had we focused on Σ, noting that the distribution of Z_i conditional on the other base random variables is normal with mean $-K_{i,i}^{-1}\mathbf{K}_{i,-i}\mathbf{z}_{t,-i}$ and with variance $K_{i,i}^{-1}$ (Gelman et al., 2013). In this representation, $\mathbf{K}_{i,-i}$ is the ith row of the precision matrix $\mathbf{K}$ with ith column removed. Similarly, $\mathbf{z}_{t,-i}$ is the base random vector $(z_{t,1}, z_{t,2}, \ldots, z_{t,d})$ with ith entry removed. Given the current state $\mathbf{K}^s$ of the chain, its next state $\mathbf{K}^{s+1}$ is generated by sequentially performing the two steps in Algorithm 1.

Algorithm 1 Approximating the posterior distribution of the precision matrix $\mathbf{K}$

1: **Step 1:** Resample the base-variable data $\underline{\mathbf{z}}$:
2: **for** $i = 1$ to d **do**
3:     **for** each $x \in \text{unique}\{x_{1,i}, \ldots, x_{m,i}\}$ **do**
4:         **for** each t such that $x_{t,i} = x$ **do**
5:             $\lambda_i \leftarrow -\mathbf{K}^s_{i,-i}\mathbf{z}_{t,-i}/K^s_{i,i}$ and $\sigma_i^2 \leftarrow 1/K^s_{i,i}$
6:             Calculate the bounds z_{lb} and z_{ub} in accordance with the set $\mathcal{Z}(\underline{\mathbf{x}})$:
7:                 $z_{lb} \leftarrow \max\{z_{t,i} : x_{t,i} < x\}$ and $z_{ub} \leftarrow \min\{z_{t,i} : x < x_{t,i}\}$
8:             Sample $u_{t,i}$ uniformly from the interval $\left[\Phi\left(\frac{z_{lb}-\lambda_i}{\sigma_i}\right), \Phi\left(\frac{z_{ub}-\lambda_i}{\sigma_i}\right)\right]$
9:             $z_{t,i} \leftarrow \lambda_i + \Phi^{-1}(u_{t,i})\sigma_i$
10:         **end for**
11:     **end for**
12: **end for**
13: **Step 2:** Resample the precision matrix using the updated base-variable data:
14: Sample $\mathbf{K}^{s+1}$ from a Wishart$(\kappa + m, \mathbf{K}_o + \underline{\mathbf{z}}'\underline{\mathbf{z}})$ distribution

We specify the prior distribution $\mathbb{P}(\mathbf{K})$ as a Wishart $(\kappa, \mathbf{K}_o)$. The parameters κ and $\mathbf{K}_o$ can be specified to incorporate any prior knowledge on the correlation between input random variables. For example, if there is no available information, setting $\kappa = d + 2$ and $\mathbf{K}_o$ as the identity matrix implies that the input random variables are considered independent before observing any historical data (Gelman et al., 2013). The simulation analyst performs the Gibbs sampler iteration outlined in Algorithm 1 until a stationary posterior distribution is obtained for the precision matrix $\mathbf{K}$ conditional on the observed event $\underline{\mathbf{z}} \in \mathcal{Z}(\underline{\mathbf{x}})$; we provide the implementation details in Section 4. The posterior is again a Wishart distribution (i.e., line 14 in Algorithm 1), and the samples of the base correclation matrix can be easily generated by applying the transformation $\sum_{i,j} K^{-1}_{i,j} \Big/ \sqrt{K^{-1}_{i,i}K^{-1}_{j,j}}$ for $i, j = 1, \ldots, d$. Notice that the simulation analyst does *not* need to approximate a new posterior distribution of the base correlation matrix from scratch for each set of the marginal distribution functions sampled during input uncertainty quantification (see Section 3). This is because we use the rank likelihood instead of the full likelihood of the real-world input data, and therefore, the event $\underline{\mathbf{Z}} \in \mathcal{Z}(\underline{\mathbf{X}})$ does not depend on the marginal cdfs. In addition, the posterior samples of the base correlation matrix are already on hand after enough number of iterations of Algorithm 1, facilitating the sampling of the base correlation matrix while quantifying the input uncertainty.

3. Accounting for input uncertainty in simulation output

The objective of this section is to address: (i) how the uncertainties around the marginal cdfs and the NORTA base correlation matrix translate into simulation output data variability, and (ii) how the quantification of the so-called *input uncertainty* can be used to improve output data analysis. We consider simulations for which independent replications are performed. The output of interest from a single simulation run, say $Y = Y(\mathbf{u}, \mathbf{F}, \Sigma)$, is a function of a random number stream u, the set of marginal cdfs $\mathbf{F} = (F1, \ldots, F_d)$, and the base correlation matrix R of the NORTA distribution. We let $\eta(\mathbf{F}, \Sigma)$ denote the expected value of Y conditional on the unknown information F and Σ, i.e., $\eta(\mathbf{F}, \Sigma) = \int Y(\mathbf{u}, \mathbf{F}, \Sigma)d\mathrm{u}$. The objective of a classical simulation experiment is to estimate $[q_L, q_U]$, where the unknown $\mathbf{F}$ and Σ are estimated from real-world input data before the simulation experiments. Since we take a Bayesian approach, $\mathbf{F}$ and Σ are random variables whose distributions represent the belief of the simulation analyst about their likely values.

We propagate the uncertainty around $\mathbf{F}$ and Σ to the simulation output data, and hence, implicitly induce a posterior distribution for $\eta(\mathbf{F}, \Sigma)$. More specifically, we construct a$(1 - \alpha)$ 100% credible interval $[q_L, q_U]$ for $\eta(\mathbf{F}, \Sigma)$ such that the *probability content*, which is defined as $\mathbb{P}(\eta(\mathbf{F}, \Sigma) \leq q_U|\underline{\mathrm{x}}) - \mathbb{P}(\eta(\mathbf{F}, \Sigma) \leq q_L|\underline{\mathrm{x}})$, is equal to $1 - \alpha$. Notice that there is not a unique credible interval meeting this requirement. We follow the convention of Xie et al. (2014a) and use two-sided, equal-tail probability $(1 - \alpha)$100% credible intervals in the remainder of the paper. More specifically, we let $q_L{:} = q_{\alpha/2} = \min\{q{:}\ \mathbb{P}(\eta(\mathbf{F}, \Sigma) \leq q) \geq \alpha/2\}$ and $q_U{:} = q_{1-\alpha/2} = \min\{q{:}\ \mathbb{P}(\eta(\mathbf{F}, \Sigma) \leq q) \geq 1 - \alpha/2\}$. As we cannot directly evaluate the posterior distribution of $\eta(\mathrm{F}, \Sigma)$, we obtain a Monte Carlo estimate of this credible interval as outlined in Algorithm 2.

Xie et al. (2014a) refer to $[q_{\alpha/2}, q_{1-\alpha/2}]$ and its Monte Carlo approximate $[\hat{q}_{\alpha/2}, \hat{q}_{1-\alpha/2}]$ as the *perect fidelity* credible interval since the mean-response function $\eta(\cdot, \cdot)$ is assumed to be known in Algorithm 2. The perfect fidelity credible interval represents the case with no intrinsic simulation uncertainty (which is due to the use of random numbers), and it is the benchmark against which we compare our proposed credible interval later in this section.

Since $\eta(\cdot, \cdot)$ is in fact unknown, an immediate approach is to use simulation to estimate $\eta(\mathbf{F}_r, \Sigma_r)$ for

Algorithm 2 Approximating the credible interval $[q_{\alpha/2}, q_{1-\alpha/2}]$.

for $r = 1$ to R **do**
  Sample the marginal cdfs $\mathbf{F}_r$ from their posterior distributions.
  Sample a correlation matrix $\mathbf{\Sigma}_r$ from its posterior distribution obtained via Algorithm 1.
  Calculate the mean simulation response $\eta(\mathbf{F}_r, \mathbf{\Sigma}_r)$.
end for
Set $\hat{q}_{\alpha/2} = \eta_{(\lceil R\alpha/2 \rceil)}$ and $\hat{q}_{1-\alpha/2} = \eta_{(\lceil R(1-\alpha/2)\rceil)}$, where $\eta_{(r)}$ denotes the rth smallest mean response in the set $\{\eta(\mathbf{F}_r, \mathbf{\Sigma}_r) : r = 1, 2, \ldots, R\}$.
Return the credible interval $[\hat{q}_{\alpha/2}, \hat{q}_{1-\alpha/2}]$ as the approximate of $[q_{\alpha/2}, q_{1-\alpha/2}]$.

all $r \in \{1, 2, \ldots, R\}$. In Algorithm 3, we present a revised version of Algorithm 2 where the mean simulation output is estimated with L independent simulation replications at each of R posterior samples of the marginal cdfs and the base correlation matrix. Xie et al. (2014a) refer Algorithm 3 as the *direct simulation method*, and it is based on the Bayesian simulation replications algorithm which is proposed by Chick (2001) for $L = 1$ and extended by Zouaoui and Wilson (2003).

Algorithm 3 Approximating the credible interval $[\hat{q}_{\alpha/2}, \hat{q}_{1-\alpha/2}]$ for unknown $\eta(\cdot,\cdot)$.

for $r = 1$ to R **do**
  Sample the marginal cdfs $\mathbf{F}_r$ from their posterior distributions.
  Sample a correlation matrix $\mathbf{\Sigma}_r$ from its posterior distribution obtained via Algorithm 1.
  for $l = 1$ to L **do**
    Run the simulation to obtain the output data $y(\mathbf{u}_l, \mathbf{F}_r, \mathbf{\Sigma}_r)$.
  end for
  Calculate the mean $\bar{y}_r := (1/L)\sum_{l=1}^{L} y(\mathbf{u}_l, \mathbf{F}_r, \mathbf{\Sigma}_r)$.
end for
Set $\hat{\hat{q}}_{\alpha/2} = \bar{y}_{(\lceil R\alpha/2 \rceil)}$ and $\hat{\hat{q}}_{1-\alpha/2} = \bar{y}_{(\lceil R(1-\alpha/2)\rceil)}$, where $\bar{y}_{(r)}$ denotes the rth smallest mean response in the set $\{\bar{y}_r : r = 1, 2, \ldots, R\}$.
Return the credible interval $[\hat{\hat{q}}_{\alpha/2}, \hat{\hat{q}}_{1-\alpha/2}]$ as the approximate of $[\hat{q}_{\alpha/2}, \hat{q}_{1-\alpha/2}]$.

Since Algorithm 3 builds on the nonparametric representation of the distribution of mean simulation output, the number of outer replications R needs to be large enough so that the posterior distributions of the marginal cdfs and the base correlation matrix are well approximate. For example, Xie et al. (2014a) recommend that R should be at least 1000 and Biller and Corlu (2011) perform the Bayesian simulation replication algorithm with values of R varying from 1000 to 10,000 in the presence of correlated input random variables. Furthermore, a large number of L is required to eliminate the intrinsic simulation uncertainty, and hence, to achieve a close approximation of $[\hat{\hat{q}}_{\alpha/2}, \hat{\hat{q}}_{1-\alpha/2}]$ to $[q_{\alpha/2}, q_{1-\alpha/2}]$. Consequently, a total budget of RL simulation replications is required in Algorithm 3. The demand for a large computational budget is one of the key factors that makes capturing the input uncertainty difficult in stochastic simulations (Barton et al., 2014).

In our preliminary numerical experiments with N independent replications of an assemble-to-order inventory system, we observed that an accurate estimate of the *total* simulation output variability $\hat{\sigma}^2_{Tot}$, which takes the form

$$\frac{1}{N-1}\sum_{n=1}^{N}(y(\mathbf{u}_n,\ \mathrm{F}_n,\ \Sigma_n) - \bar{\bar{y}}_{Tot})^2 \tag{4}$$

with $\bar{\bar{y}}_{Tot} = \sum_{n=1}^{N} y(\mathbf{u}_n, \mathbf{F}_n, \Sigma_n)/N$, is often reached with not very large N values, leading us to question whether it is possible to propose an alternative technique that approximates the perfect fidelity credible interval in a computationally less demanding way.

Our alternative method to approximate the perfect fidelity credible interval builds on the input uncertainty quantification method of Ankenman and Nelson (2012). Adopting a random effects output data model, Ankenman and Nelson (2012) assume that the total simulation output data variability σ^2_{Tot} is the sum of input uncertainty v^2 and the intrinsic simulation uncertainty σ^2, i.e., $\sigma^2_{Tot} = v^2 + \sigma^2$. Under a total computing budget of N_0 diagnostic simulation replications, the authors estimate the input uncertainty ratio $\tau = v^2/\sigma^2$, a measure of how large input uncertainty is relative to the intrinsic simulation uncertainty, with

$$\hat{\tau} = \frac{1}{J}\left[\left(\frac{JK-K-2}{K-1}\right)\frac{J\sum_{k=1}^{K}(\bar{Y}(\mathbf{F}_k,\Sigma_k) - \bar{\bar{Y}})^2}{\sum_{j=1}^{J}\sum_{k=1}^{K}(Y(\mathbf{u}_j,\mathbf{F}_k,\Sigma_k) - \bar{Y}(\mathbf{F}_k,\Sigma_k))^2} - 1\right] \tag{5}$$

with $\bar{Y}(\mathbf{F}_k, \Sigma_k) = \sum_{j=1}^{J} Y(\mathbf{u}_j, \mathbf{F}_k, \Sigma_k)/J$, $\bar{\bar{Y}} = \sum_{k=1}^{K} \bar{Y}(F_k, \Sigma_k)/K$ and $N_0 = JK$. For the derivation of the estimator in (5), we refer the reader to Dean and Voss (1999). We note that a simplifying approximation in this setup is that σ^2, the intrinsic simulation uncertainty, does not depend on which NORTA distribution is used as the input model. While this assumption may not hold in practice, it will be acceptable to get an approximation of the effect of input uncertainty (Zouaoui & Wilson, 2003; Ankenman & Nelson, 2012; Song & Nelson, 2013, 2015).

Algorithm 4 outlines our method which consists of three steps: First, we estimate the input uncertainty ratio $\hat{\tau}$ with N_0 diagnostic experiments. We allocate the diagnostic simulation replication budget N_0 as $J = 2$ and $K = N_0/2$, as suggested by Dean and Voss (1999) to minimize the variance of $\hat{\tau}$. Second, we estimate total variability σ^2_{Tot} in simulation output data by (4) with

N simulation replications. Since, $\sigma^2_{Tot} = \nu^2 + \sigma^2$ and $\nu^2 = \tau\sigma^2$, we can estimate the intrinsic simulation uncertainty as $\hat{\sigma}^2 = \hat{\sigma}^2_{Tot}/(\hat{\tau}+1)$, which is the only portion of the output data variability that more simulation replications can reduce. Thus, eliminating this portion in the credible interval construction would help us to mimic a situation with ample computing resources as shown in Algorithm 2. In particular, we replace the variance estimator $\hat{\sigma}^2_{Tot}$ with $\hat{\sigma}^2_{Tot}\hat{\tau}/(\hat{\tau}+1)$ to account for only the simulation output data variability that is attributable to the finiteness of the real-world input data, leading to the *input uncertainty adjusted credible interval* provided in the last step of Algorithm 4.

Algorithm 4 Approximating the input uncertainty adjusted credible interval.

Step 1: Estimate the input uncertainty ratio τ with a simulation budget of N_0 replications:
Specify J and K such that $JK = N_0$.
for $k = 1$ to K **do**
  Sample the marginal cdfs $\mathbf{F}_k$ from their posterior distributions.
  Sample a correlation matrix $\mathbf{\Sigma}_k$ from its posterior distribution obtained via Algorithm 1.
  for $j = 1$ to J **do**
    Run the simulation to obtain the output $y(\mathbf{u}_j, \mathbf{F}_k, \mathbf{\Sigma}_k)$.
  end for
  Calculate the mean $\bar{y}_k(\mathbf{F}_k, \mathbf{\Sigma}_k) := (1/J)\sum_{j=1}^{J} y(\mathbf{u}_j, \mathbf{F}_k, \mathbf{\Sigma}_k)$.
end for
Calculate the grand mean $\bar{\bar{y}} = \sum_{k=1}^{K} \bar{y}_k(\mathbf{F}_k, \mathbf{\Sigma}_k)/K$.
Estimate the input uncertainty ratio $\hat{\tau}$ by using Equation 5.
Step 2: Estimate the total variability σ^2_{Tot} of the simulation output data with a simulation budget of N replications by using Equation 4.
Step 3: Return the input uncertainty adjusted credible intervals as $\bar{\bar{y}} \pm z_{1-\alpha/2}\hat{\sigma}_{Tot}\sqrt{\hat{\tau}/(\hat{\tau}+1)}$, where $z_{1-\alpha/2}$ is the standard normal random variable such that $\Phi(z_{1-\alpha/2}) = 1-\alpha/2$.

The input uncertainty adjusted credible interval reflects only the contribution of the input uncertainty but not the contribution of the intrinsic simulation uncertainty, representing a scenario which is only possible if the simulation analyst has infinite amount of computing resources to perform the simulation experiments (i.e., $L \to \infty$ in Algorithm 3). Furthermore, it uses $N_0 + N$ simulation replications to approximate the perfect fidelity credible interval, while the direct simulation method in Algorithm 3 requires RL simulation replications. Since one can estimate τ accurately for K and J values that are much smaller than the values of R and L suggested for Algorithm 3, the approximate perfect fidelity credible interval from Algorithm 4 often takes closer values to $[q_{\alpha/2}, q_{1-\alpha/2}]$ under the same simulation budget. In the next section, we compare the input uncertainty adjusted credible interval and its counterpart from Algorithm 3 in terms of their width and closeness to the target probability content, and find that the input uncertainty adjusted credible interval leads to remarkably better performance by using significantly less computing effort.

The construction of the input uncertainty adjusted credible interval assumes that the posterior distribution of the mean simulation output is normally distributed. If this assumption does not hold, the mean simulation output can be represented with the Johnson system of distributions, a highly flexible family of distributions that captures a wide variety of distributional shapes. Step 3 of Algorithm 4 then returns the $\alpha/2$th and $(1-\alpha/2)$th quantiles of this distribution. We identify the appropriate Johnson distribution family and the corresponding parameters by using the moment-matching method (Equation (8) in DeBrota et al., 1988) where the second central moment of the mean simulation output is equal to the variance component that is attributable to the input uncertainty.

4. Numerical results

In Section 1.1.2, we illustrated the poor performance of simulation confidence intervals when the uncertainty in the NORTA base correlation manix is ignored, and showed how the performance becomes even worse when the discreteness of input data is ignored while estimating the NORTA base correlation manix. In this section, we continue to consider the ATO system introduced in Section 1.1.2, and now assume that not only the NORTA base correlation matrix but also the marginal cdfs are unknown by the simulation analyst. Our objective is to show how the proposed Bayesian approach for estimating the base correlation manix (Algorithm 1) and the use of input uncertainty adjusted credible intervals (Algorithm 4) lead to an improved assessment of system performance. In particular, we compare the performance of the input uncertainty adjusted credible intervals with the performance of the credible intervals obtained from the implementation of the direct simulation method outlined in Algorithm 3. As a natural benchmark, we also provide the perfect fidelity credible intervals that represent the case with no inninsic simulation uncertainty. We calculate the perfect fidelity credible interval with a total computing budget of 500,000 simulation replications, i.e., Algorithm 3 is applied with $R = 1000$ and $L = 500$ (our side experiments show that taking L equal to 500 is sufficient enough to drive the intrinsic uncertainty close to zero.)

To sample from the posterior distribution of the NORTA base correlation manix, we assume independence as the prior information and iterate the Gibbs sampling scheme outlined in Algorithm 1 for 25,000 times and save every 10 iterations so that the autocorrelation across these saved iterations of the Markov chain are negligible. We see that the convergence to stationarity occurs almost certainly within the first 5000 iterations. Dropping these iterations to allow for burn-in, we are left with 2000 samples from the posterior distribution of the NORTA base correlation matrix. For the marginal cdf of each demand random variable, we use a discrete uniform prior, i.e., the simulation analyst believes that each of the five demand realizations are equally likely before observing any demand realizations. Furthermore, we let the prior distribution of the demand interarrival time be uniform between zero and two time units.

For the direct simulation, we consider that the simulation analyst has a total computing budget of 2000 simulation replications with $R = 1000$ and $L = 2$. That is, two replications are performed at each posterior sample of the NORTA distribution. This experimental design is the same as in Xie et al. (2014a). Alternatively, we consider a total computing budget of *half* size, namely 1000 simulation replications, for the construction of input uncertainty adjusted credible intervals. In particular, we assume the availability of a computing budget of $N_0 = 500$ replications to estimate the input uncertainty ratio $\hat{\tau}$ via diagnostic experiments (i.e., Dean and Voss (1999) suggests $J = 2$ and $K = 250$ to minimize the variance of $\hat{\tau}$) and use $N = 500$ replications to estimate the total variability $\hat{\sigma}^2_{Tot}$ of simulation output data, leading to a total computing budget of 1000 simulation replications. We choose $N = 500$ because our preliminary numerical experiments show that $\hat{\sigma}^2_{Tot}$ does not change considerably after 500 replications.

We present the equal-tail 95% perfect fidelity credible intervals and the 95% direct simulation and input uncertainty adjusted credible intervals (IUA-CI) for $m = 50$, 100, and 250 in Tables 4, 5, and 6, respectively. Specifically, the lower bound (LCI) and the upper bound (UCI) correspond to the $(\alpha/2)100\%$ and the $(1—\alpha/2)100\%$ quantiles of the posterior distribution of the mean simulation output with $\alpha = 0.05$. Similar to Xie et al. (2014a), we define the percentage of perfect fidelity mean simulation output values that fall between the LCI and UCI values of a given credible interval as the probability content (PC) of that credible interval. The LCI mean, UCI mean, and PC mean columns presented in each table are the averages from 500 macroreplications. As a measure of precision, we also report the standard deviation (std) of LCI, UCI and PC values.

Tables 4, 5, and 6 show that the probability content of the IUA-CIs is closer to the nominal value 95%, while the probability content of the direct simulation credible interval is much larger. Furthermore, the LCI and UCI values are remarkably closer to their perfect fidelity counterparts, while the credible intervals obtained by direct simulation are much wider. To put in another way, the direct simulation credible intervals overestimate the overall uncertainty around the mean simulation output. The so-called overcoverage problem is well known in the literature (Barton, 2012), and also illustrated by Xie et al. (2014a) in the simulation study of a critical care facility. The overcoverage problem implies that the credible intervals from the direct simulation might not be practical to use because of their large size. For example, Table 4 shows that the average [LCI, UCI] is [17.98, 75.04] for the direct simulation with $\rho = 0.9$, while it takes the value of [25.73, 64.15] for the IUA-CI. Notice that the latter is much closer to the perfect fidelity credible interval which is only possible with enough computational resources to

Table 4. Credible interval [LCI, UCI] and probability content (PC) estimates when $m = 50$ and $\alpha = 0.05$.

		LCI mean	LCI std	UCI mean	UCI std	PC mean	PC std
$\rho = 0.9$	Perfect fidelity	24.95	6.19	63.49	6.09	–	–
	Direct simulation	17.98	6.47	75.04	7.13	0.995	0.002
	IUA-CI	25.73	7.59	64.15	7.17	0.943	0.084
$\rho = 0.6$	Perfect fidelity	14.57	7.89	53.94	7.53	–	–
	Direct simulation	2.53	8.02	69.87	8.46	0.998	0.001
	IUA-CI	14.93	8.29	54.12	8.82	0.952	0.083
$\rho = 0.3$	Perfect fidelity	5.34	9.13	44.29	9.21	–	–
	Direct simulation	−9.52	9.02	66.28	9.31	0.999	0.010
	IUA-CI	6.35	10.11	43.53	10.23	0.941	0.109

Table 5. Credible interval [LCI, UCI] and probability content (PC) estimates when $m = 100$ and $\alpha = 0.05$.

		LCI mean	LCI std	UCI mean	UCI std	PC mean	PC std
$\rho = 0.9$	Perfect fidelity	29.94	5.75	57.89	5.49	–	–
	Direct simulation	21.18	5.69	72.01	5.38	0.999	0.001
	IUA-CI	30.04	6.83	58.61	6.20	0.945	0.091
$\rho = 0.6$	Perfect fidelity	20.96	5.89	49.55	5.63	–	–
	Direct simulation	7.03	5.01	66.91	5.81	0.999	0.000
	IUA-CI	21.30	7.84	48.39	7.58	0.941	0.098
$\rho = 0.3$	Perfect fidelity	16.56	5.92	43.22	6.02	–	–
	Direct simulation	−2.13	5.38	63.28	5.61	0.999	0.000
	IUA-CI	16.29	6.11	43.63	7.27	0.939	0.095

Table 6. Credible interval [LCI, UCI] and probability content (PC) estimates when $m = 250$ and $\alpha = 0.05$.

		LCI mean	LCI std	UCI mean	UCI std	PC mean	PC std
$\rho = 0.9$	Perfect fidelity	37.13	3.88	54.12	3.52	–	–
	Direct simulation	26.22	2.83	70.17	3.76	1.000	0.000
	IUA-CI	36.92	4.11	55.81	4.21	0.940	0.071
$\rho = 0.6$	Perfect fidelity	26.64	3.80	45.13	3.82	–	–
	Direct simulation	11.09	3.71	65.51	4.09	1.000	0.000
	IUA-CI	25.56	5.01	45.91	5.13	0.956	0.091
$\rho = 0.3$	Perfect fidelity	19.21	3.28	37.92	4.91	–	–
	Direct simulation	−0.91	3.60	62.41	3.95	1.000	0.000
	IUA-CI	18.64	5.54	38.51	5.47	0.946	0.101

make the intrinsic uncertainty negligible in a stochastic simulation. It is worth noting that IUA-CIs achieve this in our numerical experiments by using only half of the computing budget allocated for the construction of the direct simulation credible intervals.

5. Conclusion

Motivated by the separate estimation of the marginal distribution functions and the dependence measures in copula functions, we decouple the estimation of the NORTA distribution—a widely used input model in stochastic simulations with correlated input variables—into two stages when the input random variables can take discrete values. We present a Markov chain Monte Carlo algorithm that only uses the rankings of the real-world input data to approximate the posterior distribution of the NORTA base correlation matrix, making this posterior distribution independent of any particular choice of the marginal distribution functions.

The proposed Bayesian estimation approach lends itself to the use of the Bayesian simulation replications to capture the input uncertainty in simulation output data. The independence of the posterior distribution of the NORTA base correlation matrix from the marginal distribution functions frees the simulation analyst to obtain a new posterior distribution for the NORTA base correlation matrix for every posterior sample of the marginal distribution functions. This alleviates the computational burden while capturing the input uncertainty in simulation output data. Furthermore, we quantify the ratio of the input uncertainty to the intrinsic simulation uncertainty via a random effects model. We show that this quantification comes with a practical benefit to build credible intervals for the mean simulation output. More specifically, our numerical analysis shows that approximating the perfect fidelity credible intervals by eliminating the estimated contribution of the intrinsic simulation uncertainty leads to a close approximation of the perfect fidelity credible intervals with less computational effort.

We note that a normal copula is unable to capture nonlinear dependence relationships that are typical in areas such as risk management and finance. The selection of an appropriate copula function that can capture dependence measures beyond linear correlation (e.g., tail dependence) is an important problem for a simulation analyst. A possible extension of our paper is to incorporate the uncertainty in the parametric form of the copula function into the construction of the input uncertainty adjusted credible intervals.

Statement of contribution

The use of simulation to estimate the performance of a stochastic system often requires the modeling of correlated input random variables. Examples include the product demands in an inventory system, the processing times of a workpiece across several machines in a job shop, and the exchange rates in a global supply chain. In this paper, we consider that the input process is represented with the NORmal-To-Anything (NORTA) distribution (Cario & Nelson, 1997). However, the simulation practitioner does not know the marginal distributions and the base correlation matrix of this input model. We consider two main challenges relevant in this setting:

The presence of the discrete-valued random variables in the input process. When an estimator of the marginal cumulative distribution function (cdf) is applied to discrete input data, a subsequent application of the inverse standard normal cdf does not lead to normally distributed input data. The direct use of the transformed input data, then, leads to an inaccurate estimation of the base correlation matrix.

Input uncertainty. The errors associated with incorrectly estimating the marginal distribution functions and the base correlation matrix of the NORTA distribution lead to inaccurate performance-measure estimates.

We summarize the contributions of our paper as follows:

We propose a Markov chain Monte Carlo algorithm to sample from the posterior distribution of the base correlation matrix in the presence of discrete input random variables.

Having the posterior samples from the marginal cdfs and the base correlation matrix of the NORTA distribution on hand, we propose a practical method to incorporate the assessment of input uncertainty into a credible interval constructed from simulation output data.

We illustrate our method in an assemble-to-order production system with a correlated demand arrival process including both discrete and continuous input random variables. Our numerical analysis shows that the proposed input uncertainty adjusted credible intervals have probability contents closer to the target value and achieve shorter widths compared to the credible intervals obtained with direct simulation.

Disclosure statement

No potential conflict of interest was reported by the authors.

References

Akcay, A., Biller, B., & Tayur S. (2012). A simulation-based approach to capturing autocorrelated demand parameter uncertainty in inventory management. In C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, & A. M. Uhrmacher (Eds.), *Proceedings of the 2012 Winter Simulation Conerence* (pp. 3213–3224). Pittsburgh, PA: Carnegie Mellon University.

Ankenman, B. E., & Nelson, B. L. (2012). A quick assessment of input uncertainty. *Proceedings of the 2012 Winter Simulation Conference* (pp. 241–250).

Barton, R. R. (2012). Tutorial: Input uncertainty in output analysis. In: C. Laroque, J. Himmelspach, R. Pasupathy, O. Rose, & A. M. Uhrmacher (Eds.), *Proceedings of the 2012 Winter Simulation Conference* (pp. 67–78).

Barton, R. R., Nelson, B. L., & Xie, W. (2014). Quantifying input uncertainty via simulation confidence intervals. *INFORMS Journal on Computing, 26*, 74–87.

Barton, R. R., & Schruben, L. W. (2001). Resampling methods for input modeling. In D. J. Medeiros, B. A. Peters, J. S. Smith, & M. W. Rohrer (Eds.), *Proceedings of the 2001 Winter Simulation Conerence* (pp. 372–378).

Biller, B., & Corlu, C. G. (2011). Accounting for parameter uncertainty in large-scale stochastic simulations with correlated inputs. *Operations Research, 59*, 661–673.

Biller, B., & Ghosh, S. (2006). Multivariate input processes. In S. G. Henderson & B. L. Nelson (Eds.), *Handbooks in Operations Research and Management Science 13* (pp. 1230–153). Amsterdam: Elsevier.

Cario, M. C., & Nelson, B. L. (1997). *Modeling and generating random vectors with arbitrary marginal distributions and correlation matrix*. Evanston: Northwestern University.

Cherubini, U., Luciano, E., & Vecchiato, W. (2004). *Copula methods in finance*. Hoboken: Wiley.

Chick, S. E. (2001). Input distribution selection for simulation experiments: accounting for input uncertainty. *Operations Research, 49*, 744–758.

Dean, A., & Voss, D. (1999). *Design and analysis of experiments*. New York, NY: Springer.

DeBrota, D. J., Roberts, S. D., Swain, J. J., Dittus, R. S., Wilson, J. R., & Venkatraman, S. (1988). Input modeling with the Johnson system of distributions. In P. Haigh, M. Abrams, & J. Comfort (Eds.), *Proceedings of the 1988 Winter Simulation Conference* (pp. 165–179).

Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *The Annals of Statistics, 1*, 209–230.

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A., & Rubin, D. B. (2013). *Bayesian Data Analysis*. Boca Raton: CRC Press.

Genest, C., Ghoudi, K., & Rivest, L. P. (1995). A semiparametric estimation procedure of dependence parameters in multivariate families of distributions. *Biometrika, 82*, 543–552.

Ghosh, S. (2004). *Dependence in stochastic simulation models* PhD thesis. Cornell University.

Ghosh, S., & Henderson, S. G. (2002). Chessboard distributions and random vectors with specified marginals and covariance matrix. *Operations Research, 50*, 820–834.

Hoff, P. (2007). Extending the rank likelihood for semiparametric copula estimation. *The Annals of Applied Statistics, 1*, 265–283.

Hong, L. J., & Nelson, B. L. (2006). Discrete optimization via simulation using COMPASS. *Operations Research, 54*, 115–129.

Joe, H. (1997). *Multivariate models and dependence concepts*. London: Chapman & Hall.

Lam, H., & Zhou, E. (2015). Quantifying uncertainty in sample average approximation. In *Proceedings of the 2015 Winter Simulation Conference* (pp. 3846–3857).

Lin, Y., Song, E., & Nelson, B. L. (2015). Single-experiment input uncertainty. *Journal of Simulation, 9*, 249–259.

Murray, J. S., Dunson, D. B., Carin, L., & Lucas, J. E. (2013). Bayesian Gaussian copula factor models for mixed data. *Journal of the American Statistical Association, 108*, 656–665.

Nelsen, R. B. (2006). *An introduction to copulas* (2nd ed.). New York, NY: Springer.

Ng, S. H., & Chick, S. E. (2006). Reducing parameter uncertainty for stochastic systems. *ACM Transactions on Modeling and Computer Simulation, 16*, 26–51.

Sethuraman, J. (1994). A constructive definition of Dirichlet priors. *Statistica Sinica, 4*, 639–650.

Smith, M. S., & Khaled, M. A. (2012). Estimation of copula models with discrete margins via Bayesian data augmentation. *Journal of the American Statistical Association, 107*, 290–303.

Song, E., & Nelson, B. L. (2013). A quicker assessment of input uncertainty. In R. Pasupathy, S.-H. Kim, A. Tolk, R. Hill, & M. E. Kuhl (Eds.), *Proceedings of the 2013 Winter Simulation Conference* (pp. 474–485).

Song, E., & Nelson, B. L. (2015). Quickly assessing contributions to input uncertainty. *IIE Transactions, 47*, 893–909.

Teh, Y. W. (2010). Dirichlet processes. In C. Sammut & G. I. Webb (Eds.), *Encyclopedia of Machine Learning*. New York, NY: Springer.

Xie, W., Nelson, B. L., & Barton, R. R. (2014a). A Bayesian framework for quantifying uncertainty in stochastic simulation. *Operations Research, 62*, 1439–1452.

Xie, W., Nelson, B. L., & Barton, R. R. (2014b). *Statistical uncertainty analysis for stochastic simulation* (Working paper). Department of Industrial Engineering and Management Sciences, Northwestern University.

Xie, W., Nelson, B. L., & Barton, R. R. (2014c). Statistical uncertainty analysis for stochastic simulation with dependent input models. In I. O. Ryzhov, L. Yilmaz, S. Buckley, A. Tolk, S. Y. Diallo, & J. A. Miller (Eds.), *Proceedings of the 2014 Winter Simulation Conference*.

Zouaoui, F., & Wilson, J. R. (2003). Accounting for parameter uncertainty in simulation input modeling. *IIE Transactions, 35*, 781–792.

Zouaoui, F., & Wilson, J. R. (2004). Accounting for input-model and input-parameter uncertainties in simulation. *IIE Transactions, 36*, 1135–1151.