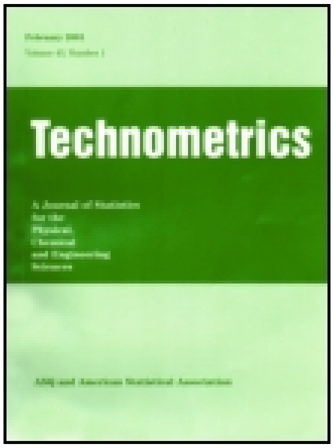


Technometrics
Publication details, including instructions for authors and subscription information:
http://www.tandfonline.com/loi/utch20

Importance Sampling for Reliability Evaluation With Stochastic Simulation Models

Youngjun Choe[a], Eunshin Byon[a] & Nan Chen[b]

[a] Department of Industrial and Operations Engineering, University of Michigan, Ann Arbor, MI 48109, (;)

[b] Department of Industrial and Systems Engineering, National University of Singapore, Singapore, 117576, ()

Accepted author version posted online: 13 Feb 2015.Published online: 17 Aug 2015.

To cite this article: Youngjun Choe, Eunshin Byon & Nan Chen (2015) Importance Sampling for Reliability Evaluation With Stochastic Simulation Models, Technometrics, 57:3, 351-361, DOI: 10.1080/00401706.2014.1001523

To link to this article: http://dx.doi.org/10.1080/00401706.2014.1001523

Importance Sampling for Reliability Evaluation With Stochastic Simulation Models

Youngjun CHOE and Eunshin BYON

Department of Industrial and Operations Engineering
University of Michigan
Ann Arbor, MI 48109
(yjchoe@umich.edu; ebyon@umich.edu)

Nan CHEN

Department of Industrial and Systems Engineering
National University of Singapore
Singapore 117576
(isecn@nus.edu.sg)

Importance sampling has been used to improve the efficiency of simulations where the simulation output is uniquely determined, given a fixed input. We extend the theory of importance sampling to estimate a system's reliability with stochastic simulations. Thanks to the advance of computing power, stochastic simulation models are employed in many applications to represent a complex system behavior. A stochastic simulation model generates stochastic outputs at the same input. Given a budget constraint on total simulation replications, we develop a new approach, which we call stochastic importance sampling, which efficiently uses stochastic simulations with unknown output distribution. Specifically, we derive the optimal importance sampling density and allocation procedure that minimize the variance of an estimator. Application to a computationally intensive aeroelastic wind turbine simulation demonstrates the benefits of the proposed approach. Supplementary materials for this article are available online.

KEY WORDS: Monte Carlo; Stochastic simulation; Variance reduction; Wind energy.

1. INTRODUCTION

With the rapid growth of computing power over the last decades, computer simulation modeling has become very popular in many applications where real experiments are expensive, or perhaps impossible. These simulation models are often used to evaluate the reliability of large-scale, complex systems. For example, safety evaluation of a nuclear power plant often employs complex computer simulations (D'Auria et al. 2006). The U.S. Department of Energy (DOE)'s National Renewable Energy Laboratory (NREL) has developed aeroelastic simulation tools to help wind turbine manufacturers design reliable systems (Jonkman and Buhl 2005; Jonkman 2009).

This study is concerned with reliability (or failure probability) evaluation using simulations. As simulation models become more realistic and their degrees of freedom increase, estimating reliability remains challenging, because each simulation replication is computationally expensive. Moreover, the reliability evaluation of a highly reliable system often requires a large number of simulation replications to accurately quantify a small failure event probability. For example, to evaluate the reliability of a wind turbine, Moriarty (2008) used grid computing with 60 desktops at NREL for 5 weeks, and Manuel, Nguyen, and Barone (2013) used cluster computing with 1024 cores at Sandia National laboratories. As a result, efficiency improvement of estimating a failure event probability is of great importance in a large-scale, complex simulation.

This study extends the theory of importance sampling (IS) to compute reliability using simulations. IS is one of the most popular variance reduction techniques to improve the estimation accuracy. The underlying concept is to control the sampling of input variables in the simulation so that the outputs are observed more frequently in the regions of interest. Various IS methods have been proposed for the cases where a simulation output is deterministic at a fixed input (De Boer et al. 2005; Cannamela, Garnier, and Iooss 2008; Dubourg, Sudret, and Deheeger 2013). In these simulation models, the randomness resides only in input variables, but the input and output relationship of the simulation model is deterministic. In this study, we call these simulation models deterministic simulation models.

Other simulation models, called stochastic simulation models, have one or more stochastic elements inside simulations. A unique input generates different outputs in multiple simulation replications, that is, stochastic outputs are observed at a fixed input. For instance, the NREL wind turbine simulators generate stochastic load outputs, given a wind condition (Jonkman and Buhl 2005; Jonkman 2009). Stochastic simulation models are popular owing to their flexibility to represent complex stochastic systems and to the increased computing power with advanced random number generating capability. However, the conventional IS method devised for deterministic simulation models is not applicable to the stochastic simulation model (to be detailed in Section 2).

This study develops a new approach, called stochastic importance sampling (SIS), which efficiently uses stochastic simulations with unknown output distribution. We propose two methods to estimate a failure probability. First, we use a failure probability estimator that allows multiple simulation replications at each input and derive the optimal IS density and allocation of simulation efforts at each sampled input for minimizing the estimator variance. Second, we propose another estimator that

TECHNOMETRICS, AUGUST 2015, VOL. 57, NO. 3
DOI: 10.1080/00401706.2014.1001523

allows one simulation replication at each sampled input and derive the optimal IS density. Both methods use variance decomposition (Kroese, Taimre, and Botev 2011) to account for different sources of output variability and functional minimization (Courant and Hilbert 1989) to find the optimal IS densities. We demonstrate the proposed methods using the NREL simulators to estimate wind turbine failure probabilities. The implementation results suggest that the SIS approach can produce estimates with smaller variances compared to alternative approaches when the total simulation budget is fixed.

In the remainder of the article, Section 2 reviews the problem background and relevant studies. Section 3 presents the proposed methodology. Section 4 discusses benchmark methods for comparison with the proposed approach. Section 5 describes numerical examples, and Section 6 applies the proposed methods to the wind turbine simulators. Section 7 concludes the article.

2. BACKGROUND AND LITERATURE REVIEW

We first give an overview of IS in a deterministic simulation model, which we call DIS approach. Let $\mathbf{X}$, an input vector, denote a random vector following a known density, f. Given $\mathbf{X}$, a simulator generates an output, $Y = g(\mathbf{X})$, via a deterministic performance function, $g(\cdot)$. The function, $g(\cdot)$, is not explicitly known, but we can evaluate it with a simulation model. In reliability analysis with a deterministic simulation model, we are interested in estimating the failure probability, $P(Y > l) = E[\mathbb{I}(g(\mathbf{X}) > l)]$, where l denotes the system's resistance level and $\mathbb{I}(\cdot)$ is the indicator function.

The crude Monte Carlo (CMC) method is one of the simplest methods to estimate the failure probability. In CMC, we independently draw $\mathbf{X}_i$, $i = 1, 2, \ldots, N_T$, from its density, f, and unbiasedly estimate the failure probability by

$$\hat{P}_{\mathrm{CMC}} = \frac{1}{N_T} \sum_{i=1}^{N_T} \mathbb{I}\left(g(\mathbf{X}_i) > l\right), \tag{1}$$

where N_T is the total number of simulation replications.

Alternatively, DIS uses the following estimator,

$$\hat{P}_{\mathrm{DIS}} = \frac{1}{N_T} \sum_{i=1}^{N_T} \mathbb{I}\left(g(\mathbf{X}_i) > l\right) \frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}, \tag{2}$$

where $\mathbf{X}_i, i = 1, 2, \ldots, N_T$, is independently sampled from q, called an IS density. Since $\mathbf{X}_i$ is sampled from q, we multiply the likelihood ratio, $f(\mathbf{X}_i)/q(\mathbf{X}_i)$, in (2) to obtain an unbiased estimator of $P(Y > l)$. Note that $\hat{P}_{\mathrm{DIS}}$ in (2) is unbiased under the condition that $q(\mathbf{x}) = 0$ implies that $\mathbb{I}(g(\mathbf{x}) > l) f(\mathbf{x}) = 0$ for any $\mathbf{x}$. An appropriately selected IS density reduces the estimator variance. It is well known that the following IS density renders $\mathrm{var}[\hat{P}_{\mathrm{DIS}}]$ zero (Kroese, Taimre, and Botev 2011):

$$q_{\mathrm{DIS}}(\mathbf{x}) = \frac{\mathbb{I}\left(g(\mathbf{x}) > l\right) f(\mathbf{x})}{P(Y > l)}. \tag{3}$$

Here, $q_{\mathrm{DIS}}(\mathbf{x})$ can be interpreted as the conditional density of $\mathbf{X}$, given that the failure event occurs. Since the denominator in (3) is the target quantity one wants to estimate and $\mathbb{I}(g(\mathbf{x}) > l)$ is unknown, $q_{\mathrm{DIS}}(\mathbf{x})$ is not implementable in practice. Several approximations have been developed, including the cross-entropy method (De Boer et al. 2005) and metamodel-based approximations (Dubourg, Sudret, and Deheeger 2013). These methods aim to find good IS densities that focus sampling efforts on the failure event region.

Existing IS studies consider the deterministic performance function, $g(\cdot)$. That is, for a fixed input, $\mathbf{x}$, the observed output, $Y = g(\mathbf{x})$, is always the same. This case corresponds to the simulation with a deterministic simulation model where the same input generates the same output. On the other hand, when a stochastic simulation model is used, the simulation output is random even at the same input. We can express the random output as $Y = g(\mathbf{X}, \boldsymbol{\epsilon})$, where $\boldsymbol{\epsilon}$ collectively denotes the uncontrollable randomness inside the simulator and $\mathbf{X}$ denotes a controllable random vector with its known density, f.

One might claim that in any simulations, both variables, $\mathbf{X}$ and $\boldsymbol{\epsilon}$, are controllable because some sampling distributions are specified for both variables to run the simulation. However, there are some cases where the DIS approach cannot be applied. First, to use DIS, the joint density function of $\mathbf{X}$ and $\boldsymbol{\epsilon}$, which needs to be biased in the IS method, should be explicitly defined. In many realistic simulations, the relationships among the elements of $\boldsymbol{\epsilon}$ (or between $\mathbf{X}$ and $\boldsymbol{\epsilon}$) are governed by physical rules and constraints, and thus finding an explicit form of the joint density function can be intractable. Second, even if we know the joint density function of $\mathbf{X}$ and $\boldsymbol{\epsilon}$ explicitly, when the dimension of $\boldsymbol{\epsilon}$ is extremely high, applying DIS becomes very difficult due to the curse of dimensionality (Au and Beck 2003). In addition, some third-party simulation software may not allow access and control for $\boldsymbol{\epsilon}$.

For example, with a specification we adopted from Moriarty (2008), the NREL simulators use over 8 million random variables for each simulation run to generate a three-dimensional stochastic wind profile at multiple grid points via the inverse Fourier transform (Jonkman 2009). The relationship of $\mathbf{X}$, the input wind condition, with $\boldsymbol{\epsilon}$, which collectively represents the 8 million *plus* random variables, is highly complicated due to the spatial and temporal dependence coupled with the inverse Fourier transform. Consequently, one cannot find the explicit joint density function of $\mathbf{X}$ and $\boldsymbol{\epsilon}$. Even if one were to find it, applying the DIS approach jointly to $\mathbf{X}$ and $\boldsymbol{\epsilon}$ is virtually impossible due to the curse of dimensionality as previously mentioned. In fact, this difficulty is typical for many realistic simulations of actual stochastic systems with high degrees of freedom.

Therefore, for the stochastic simulation models where we effectively do not have control over $\boldsymbol{\epsilon}$, the DIS density in (3), q_{DIS}, can no longer be optimal. In fact, q_{DIS} cannot be applied to the stochastic simulation model because given $\mathbf{x}$, $\mathbb{I}(g(\mathbf{x}) > l)$ in (3) is random.

Recently, stochastic simulation models that consider stochastic outputs given an input condition have also been studied in the literature (Huang et al. 2006; Ankenman, Nelson, and Staum 2010). Ankenman, Nelson, and Staum (2010) considered a queueing system simulation as an example of stochastic simulation models, where the arrival rate is the input, x, and the average number of customers in the system during specific time units, T, is an output, Y. Here, $\boldsymbol{\epsilon}$ collectively denotes the customer interarrival times and the service times. Huang et al. (2006) also considered stochastic simulation models and used the inventory system simulation where the output, a total cost per month, is stochastic, given the input including a reorder point and a maximal holding quantity. Even though these studies account for the intrinsic uncertainty in outputs, their focuses are different from our study. For example, Ankenman, Nelson, and Staum (2010) developed a stochastic simulation metamodeling method, extending the kriging methodology (Joseph 2006), and estimated an unknown quantity based on a metamodel. We note that this metamodeling-based approach is useful for estimating a mean response. However, this approach usually smooths a response function so that it loses its estimation accuracy in a tail probability estimation, as discussed in Cannamela, Garnier, and Iooss (2008).

Another well-known approach is "IS for stochastic simulations," which has been extensively studied (Heidelberger 1995) after the seminal article by Glynn and Iglehart (1989). This approach is proven effective if we can control stochastic processes inside a simulation. However, when a simulator involves complicated processes (e.g., wind turbine simulators), controlling these processes can be difficult, if not impossible. Therefore, our proposed approach treats a simulator as a black box model, and thus differs from the existing approach.

3. METHODOLOGY

This section devises optimal SIS methods for stochastic simulation models. We include the detailed derivations in the supplementary document.

3.1 Failure Probability Estimators

A stochastic simulation model generates a random variable, Y, given a realization of the input, $\mathbf{X} \in \mathbb{R}^p$. In this context, the failure probability is

$$P(Y > l) = E_f\,[P(Y > l \mid \mathbf{X})] = \int_{\mathcal{X}_f} P(Y > l \mid \mathbf{X} = \mathbf{x})\, f(\mathbf{x})\, d\mathbf{x}, \tag{4}$$

where f is the density of $\mathbf{X}$ with the support of $\mathcal{X}_f$, and the subscript f appended to the expectation operator in (4) indicates that the expectation is taken with respect to f. We call an estimator of $P(Y > l)$, $\hat{P}(Y > l)$, a probability of exceedance (POE) estimator.

A simple Monte Carlo (MC) estimator for $P(Y > l)$ in (4) is

$$\hat{P}_{\mathrm{MC}} = \frac{1}{M}\sum_{i=1}^{M} \hat{P}(Y > l \mid \mathbf{X}_i) = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)} > l\right)\right), \tag{5}$$

where $\mathbf{X}_i$, $i = 1, 2, \ldots, M$, is independently sampled from f. The number of sampled inputs, M, is called an input sample size. At each $\mathbf{X}_i$, we run simulations N_i times to obtain N_i outputs, $Y_j^{(i)}$, $j = 1, 2, \ldots, N_i$, where $Y_j^{(i)}$ denotes the output obtained in the jth replication. Note that the estimator in (5) allows multiple replications at each $\mathbf{X}_i$ to account for the stochastic outputs at the same input. We call the number of simulation replications at each $\mathbf{X}_i$, N_i, an allocation size. In (5), we call $\hat{P}(Y > l \mid \mathbf{X}_i)$ a conditional POE estimator. The total number of replications is $N_T = \sum_{i=1}^{M} N_i$. With deterministic simulation models, multiple replications at the same input are not necessary because the outcome is conclusively determined at the given input.

In the spirit of IS methods, we propose the following SIS estimator:

$$\hat{P}_{\mathrm{SIS1}} = \frac{1}{M}\sum_{i=1}^{M} \hat{P}(Y > l \mid \mathbf{X}_i)\,\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)} = \frac{1}{M}\sum_{i=1}^{M}\left(\frac{1}{N_i}\sum_{j=1}^{N_i}\mathbb{I}\left(Y_j^{(i)} > l\right)\right)\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}, \tag{6}$$

where $\mathbf{X}_i$ is drawn from q. $\hat{P}_{\mathrm{SIS1}}$ is unbiased if $q(\mathbf{x}) = 0$ implies $\hat{P}(Y > l \mid \mathbf{X} = \mathbf{x})\, f(\mathbf{x}) = 0$ for any $\mathbf{x}$. We assume that the total simulation budget, N_T, and the input sample size, M, are given. Note that since we treat the stochastic elements inside the simulator as an uncontrollable input, we apply the underlying idea of IS only to X and use the sample mean to estimate the conditional POE. Here, the conditional POE can be viewed as the success probability parameter in the binomial distribution, and the sample mean is the unique uniformly minimum-variance unbiased estimator for the binomial distribution (Casella and Berger 2002).

We also propose an alternative estimator that restricts N_i to be one:

$$\hat{P}_{\mathrm{SIS2}} = \frac{1}{N_T}\sum_{i=1}^{N_T}\mathbb{I}\,(Y_i > l)\,\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}, \tag{7}$$

where Y_i is an output at $\mathbf{X}_i$, $i = 1, 2, \ldots, N_T$. $\hat{P}_{\mathrm{SIS2}}$ is also an unbiased estimator of $P(Y > l)$ if $q(\mathbf{x}) = 0$ implies $\mathbb{I}\,(Y > l)\, f(\mathbf{x}) = 0$ for any $\mathbf{x}$. In the sequel, Sections 3.2 and 3.3 derive the minimum-variance solutions for the estimators in (6) and (7), respectively.

3.2 Stochastic Importance Sampling Method 1

We want to find the optimal allocation sizes and the optimal IS density that minimize the variance of the failure probability estimator in (6). Considering the two sources of randomness, namely, stochastic inputs and stochastic elements inside the stochastic simulation model, we decompose the estimator variance into two components as

$$\begin{aligned}
\mathrm{var}\left[\hat{P}_{\mathrm{SIS1}}\right] &= \mathrm{var}\left[\frac{1}{M}\sum_{i=1}^{M}\hat{P}(Y > l \mid \mathbf{X}_i)\,\frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)}\right] \\
&= \frac{1}{M^2}E_q\Bigg[\mathrm{var}\Bigg[\sum_{i=1}^{M}\hat{P}(Y > l \mid \mathbf{X}_i) \\
&\qquad \times \frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)} \mid \mathbf{X}_1, \ldots, \mathbf{X}_M\Bigg]\Bigg] \\
&\quad + \frac{1}{M^2}\mathrm{var}_q\Bigg[E\Bigg[\sum_{i=1}^{M}\hat{P}(Y > l \mid \mathbf{X}_i) \\
&\qquad \times \frac{f(\mathbf{X}_i)}{q(\mathbf{X}_i)} \mid \mathbf{X}_1, \ldots, \mathbf{X}_M\Bigg]\Bigg].
\end{aligned} \tag{8}$$

Let $s(\mathbf{X})$ denote the conditional POE, $P(Y > l \mid \mathbf{X})$. Using the fact that $\mathbf{X}_i \overset{\text{iid}}{\sim} q$ for $i = 1, 2, \ldots, M$, we simplify $\mathrm{var}\left[\hat{P}_{\mathrm{SIS1}}\right]$ in

(8) to

$$\mathrm{var}\left[\hat{P}_{\mathrm{SIS1}}\right] = \frac{1}{M^2} E_q \left[\sum_{i=1}^{M} \frac{1}{N_i} s(\mathbf{X}_i)\left(1 - s(\mathbf{X}_i)\right) \frac{f(\mathbf{X}_i)^2}{q(\mathbf{X}_i)^2} \right] + \frac{1}{M} \mathrm{var}_q \left[s(\mathbf{X}) \frac{f(\mathbf{X})}{q(\mathbf{X})} \right]. \tag{9}$$

To find the optimal allocation size and the optimal IS density function, we first profile out N_i and express the variance in (9) in terms of $q(\mathbf{X})$. Lemma 1 presents the optimal assignment of simulation replications to each $\mathbf{X}_i$ for any given q.

Lemma 1. Given q, the variance in (9) is minimized if and only if

$$N_i = \frac{\sqrt{s(\mathbf{X}_i)\left(1 - s(\mathbf{X}_i)\right) f(\mathbf{X}_i)/q(\mathbf{X}_i)}}{\sum_{j=1}^{M} \sqrt{s(\mathbf{X}_j)\left(1 - s(\mathbf{X}_j)\right) f(\mathbf{X}_j)/q(\mathbf{X}_j)}} \cdot N_T \quad \text{for } i = 1, 2, \ldots, M. \tag{10}$$

Next, we use the optimal allocation size in Lemma 1 to derive the optimal IS density for the estimator in (6). Plugging the N_i's in (10) into the estimator variance in (9) gives

$$\mathrm{var}\left[\hat{P}_{\mathrm{SIS1}}\right] = \frac{1}{M}\frac{1}{N_T}\left(E_f\left[s(\mathbf{X})\left(1 - s(\mathbf{X})\right) \frac{f(\mathbf{X})}{q(\mathbf{X})} \right] + (M-1)\left(E_f\left[\sqrt{s(\mathbf{X})\left(1 - s(\mathbf{X})\right)} \right] \right)^2 \right) + \frac{1}{M}\left(E_f\left[s(\mathbf{X})^2 \frac{f(\mathbf{X})}{q(\mathbf{X})} \right] - P\left(Y > l\right)^2 \right). \tag{11}$$

We minimize the functional in (11) using the principles of the calculus of variations (Courant and Hilbert 1989) and find the optimal IS density, q_{SIS1}. We also plug q_{SIS1} into (10) to attain the optimal allocation size, which leads to Theorem 1.

Theorem 1.

(a) The variance of the estimator in (6) is minimized if the following IS density and the allocation size are used:

$$q_{\mathrm{SIS1}}(\mathbf{x}) = \frac{1}{C_{q1}} f(\mathbf{x}) \sqrt{\frac{1}{N_T} s(\mathbf{x})\left(1 - s(\mathbf{x})\right) + s(\mathbf{x})^2}, \tag{12}$$

$$N_i = N_T \frac{\sqrt{\frac{N_T(1 - s(\mathbf{x}_i))}{1 + (N_T - 1)s(\mathbf{x}_i)}}}{\sum_{j=1}^{M} \sqrt{\frac{N_T(1 - s(\mathbf{x}_j))}{1 + (N_T - 1)s(\mathbf{x}_j)}}}, \quad i = 1, 2, \ldots, M, \tag{13}$$

where C_{q1} is $\int_{\mathcal{X}_f} f(\mathbf{x}) \sqrt{\frac{1}{N_T} s(\mathbf{x}) \cdot \left(1 - s(\mathbf{x})\right) + s(\mathbf{x})^2}\, d\mathbf{x}$ and $s(\mathbf{x})$ is $P(Y > l | \mathbf{X} = \mathbf{x})$.

(b) With q_{SIS1} and $N_i, i = 1, 2, \ldots, M$, the estimator in (6) is unbiased.

We call this approach stochastic importance sampling method 1 (SIS1). The optimal SIS1 density in (12) focuses its sampling efforts on the region where the failure event of interest likely occurs. On the other hand, the input condition, $\mathbf{x}_i$, with a smaller $s(\mathbf{x}_i)$ needs a larger accompanying N_i. In other words, among the important input conditions under which a system can possibly fail (i.e., the conditions that q_{SIS1} samples), SIS1 balances the simulation efforts by allocating a larger (smaller) number of replications in the area with a relatively small (large) $s(\mathbf{x})$.

We note that when applied to a deterministic simulation model, the proposed SIS1 method reduces to the DIS method with q_{DIS} in (3). Using $s(\mathbf{x}) = \mathbb{I}\left(Y > l\right), \forall \mathbf{x} \in \mathcal{X}_f$, in a deterministic simulation model where $Y = g(\mathbf{x})$ is the deterministic output of the simulator at an input, $\mathbf{x}$, we can see that q_{SIS1} in (12) is reduced to q_{DIS}. Also, when $s(\mathbf{x})$ is an indicator function, the first term in the variance in (9) vanishes, implying that we do not need the allocation step for SIS1 as we do not for DIS.

3.3 Stochastic Importance Sampling Method 2

This section derives the optimal IS density minimizing the variance of the failure probability estimator in (7), which restricts the allocation size to be one at each sampled input. This approach does not require the allocation of N_i. The SIS2 estimator in (7) essentially takes a similar form in (2) used for a deterministic simulation model. However, it is not possible to use q_{DIS} in (3) for stochastic simulation models since Y is not a deterministic function of $\mathbf{X}$. Theorem 2 presents the optimal IS density for the estimator in (7) with a stochastic simulation model.

Theorem 2.

(a) The variance of the estimator in (7) is minimized with the density,

$$q_{\mathrm{SIS2}}(\mathbf{x}) = \frac{1}{C_{q2}} f(x)\sqrt{s(x)}, \tag{14}$$

where C_{q2} is $\int_{\mathcal{X}_f} f(\mathbf{x})\sqrt{s(\mathbf{x})} d\mathbf{x}$ and $s(\mathbf{x})$ is $P(Y > l | \mathbf{X} = \mathbf{x})$.

(b) With q_{SIS2}, the estimator in (7) is unbiased.

We call this approach stochastic importance sampling method 2 (SIS2). Similar to SIS1, SIS2 focuses its sampling efforts on the input conditions under which the failure event likely occurs with a high probability, $s(\mathbf{x})$. Also, when applied to deterministic simulation models, q_{SIS2} in (14) is reduced to q_{DIS} in (3).

3.4 Implementation Guidelines

In implementing SIS1, we use rounded N_i. If the rounding yields zero, we use one to ensure the unbiasedness. Note that q_{SIS1}, N_i's, and q_{SIS2} require the conditional POE, $s(\mathbf{x})$, which is unknown. Therefore, the optimal solutions in (12)–(14) are theoretically optimal, but not implementable, which is a common problem encountered in any IS methods. In our implementation, we approximate the conditional POE using a parametric regression model (or metamodel). The estimators in (6) and (7) are still unbiased with this approximation.

We can consider several methods to approximate the conditional POE. In many studies, Gaussian regression or its variants have been used to approximate the simulation model (Seber and Lee 2003; Cannamela, Garnier, and Iooss 2008; Ankenman, Nelson, and Staum 2010). In particular, when the output, Y, is the average of the quantities generated from a stochastic process or system, Gaussian regression or its variants would provide good approximation. More generally, when Y tends to follow a distribution in the exponential family, generalized linear model (GLM; Green and Silverman 1994) or generalized additive model (GAM; Hastie and Tibshirani 1990) could be employed. When the distribution belongs to a nonexponential family, generalized additive model for location, scale, and shape (GAMLSS; Rigby and Stasinopoulos 2005) will provide a flexible modeling framework. For example, if Y is the maximum or minimum of the quantities during a specific time interval (e.g., maximum stress during 10 min operations), the Generalized Extreme Value (GEV) distribution (Coles 2001) can be employed for fitting the conditional distribution with the GAMLSS framework (to be detailed in Section 6).

While general regression models focus on capturing input-to-output relationships and are relatively straightforward to check the model accuracy, determining the metamodel accuracy for conditional POE imposes more challenges because not only is the regression relationship important, but selecting the appropriate distribution is also crucial. If the distribution fitting is not carefully conducted, the approximated POE might not help achieving the full potential of the proposed method. Provided that the primary purpose of the metamodel is to approximate the conditional POE, we recommend using goodness-of-fit tests for checking the metamodel accuracy (Stephens 1974). Different tests have their own pros and cons depending on the hypothesized distribution; thus, it is advisable to decide on the specific test based on the distribution of interest. Extensive studies have been conducted on the tests for specific distributions (e.g., Choulakian and Stephens 2001).

We summarize SIS1 and SIS2 procedures as follows:

Step 1. Approximate the conditional POE, $s(\mathbf{x})$, with a metamodel.
Step 2. Sample $\mathbf{x}_i$, $i = 1, 2, \ldots, M$, from q_{SIS1} in (12) for SIS1 or q_{SIS2} in (14) for SIS2 (note that $M = N_T$ for SIS2).
Step 3. Determine the allocation size, N_i for each $\mathbf{x}_i$, $i = 1, 2, \ldots, M$, using (13) for SIS1 or set $N_i = 1$, $i = 1, 2, \ldots, M$, for SIS2.
Step 4. Run simulation N_i times at each $\mathbf{x}_i$, $i = 1, 2, \ldots, M$.
Step 5. Estimate the failure probability using (6) for SIS1 or (7) for SIS2.

4. BENCHMARK METHODS

We compare our two methods, SIS1 and SIS2, with two benchmark methods. First, we use the CMC estimator in (1), which is an unbiased estimator of the failure probability even if the simulation model is stochastic. The variance is known as $P(Y > l)(1 - P(Y > l))/N_T$.

Second, we introduce a new IS density, q_{BIS}, which mimics q_{DIS} in (3). Recalling that it is not possible to use the IS density in (3) for stochastic simulation models, we simply replace the failure indicator function in (3), $\mathbb{I}(Y > l)$, with the conditional POE, $s(\mathbf{x})$, to obtain

$$q_{\mathrm{BIS}}(\mathbf{x}) = \frac{s(\mathbf{x}) f(\mathbf{x})}{P(Y > l)}. \tag{15}$$

With $q_{\mathrm{BIS}}(\mathbf{x})$, we use the failure probability estimator in (7). We call this approach benchmark importance sampling (BIS), since it emulates DIS.

5. NUMERICAL EXAMPLES

We investigate the performances of the SIS methods using numerical examples with various settings. We take a deterministic simulation example in Cannamela, Garnier, and Iooss (2008) and modify it to have stochastic elements. Specifically, we use the following data-generating structure:

$$X \sim N(0, 1), \qquad Y|X \sim N\big(\mu(X), \sigma^2(X)\big), \tag{16}$$

where the mean, $\mu(X)$, and the standard deviation, $\sigma(X)$, of the normal distribution are

$$\begin{aligned} \mu(X) &= 0.95\delta X^2 (1 + 0.5\cos(5X) + 0.5\cos(10X)), \\ \sigma(X) &= 1 + 0.7|X| + 0.4\cos(X) + 0.3\cos(14X). \end{aligned} \tag{17}$$

In practice, we do not know the conditional distribution for $Y|X$; thus, as a metamodel, we use the normal distribution with the following mean and standard deviation:

$$\begin{aligned} \hat{\mu}(X) &= 0.95\delta X^2 (1 + 0.5\rho\cos(5X) + 0.5\rho\cos(10X)), \\ \hat{\sigma}(X) &= 1 + 0.7|X| + 0.4\rho\cos(X) + 0.3\rho\cos(14X). \end{aligned} \tag{18}$$

Here, we include the parameters δ and ρ to control the similarity of the IS density to the original input density and the metamodeling accuracy, respectively. We set $N_T = 1000$ (with $M = 300$ for SIS1) and repeat the experiment 500 times to obtain the sample average and the standard error of each method's POE estimation. We use the following setup as a baseline and vary each parameter to see its effect on the performances of the proposed methods: $P_T = 0.01$, $\delta = 1$, and $\rho = 1$. We explain each parameter and summarize the experiment results as follows:

- P_T, the magnitude of target failure probability: We study how the proposed methods perform at different levels of $P_T = P(Y > l)$. The computational efficiency of each method is evaluated using the standard error or equivalently the relative ratio, $N_T / N_T^{(\mathrm{CMC})}$, where $N_T^{(\mathrm{CMC})}$ is the number of CMC simulation replications needed to achieve the same standard error of each method. Table 1 suggests that the computational gains of SIS1 and SIS2 against CMC generally increase as P_T gets smaller. Also, SIS1 and SIS2 always outperform BIS, providing more accurate estimates with lower standard errors.
- δ, the difference between the original input density and the optimal IS density: We consider δ of 1 or -1. The densities, f and q_{SIS1} (or q_{SIS2}), are more different from each other when $\delta = 1$ than when $\delta = -1$. Table 1 suggests that the computational gains of SIS1 and SIS2 are much more significant when $\delta = 1$ than when $\delta = -1$. Interestingly, when $\delta = -1$, BIS shows no advantage over CMC, whereas the proposed methods still lead to lower standard errors than CMC.
- ρ, the metamodeling accuracy: We vary ρ in $\hat{\mu}(X)$ and $\hat{\sigma}(X)$ in (18) to control the quality of the metamodel. Table 2 shows that the standard errors of all IS estimators increase as ρ decreases. However, the standard errors of both SIS1 and SIS2 increase more slowly than BIS. The fact that the increment of the SIS2's standard error is minimal indicates that SIS2 is less sensitive to the metamodel quality than SIS1. The performance of BIS differs sig-

Table 1. POE estimation results with different δ and P_T ($\rho = 1$)

		$\delta = 1$			$\delta = -1$		
			P_T			P_T	
		0.10	0.05	0.01	0.10	0.05	0.01
SIS1	Sample average	0.1004	0.0502	0.0100	0.1001	0.0500	0.0100
	Standard error	0.0068	0.0039	0.0005	0.0090	0.0062	0.0026
	Relative ratio	51%	32%	2.5%	90%	81%	68%
SIS2	Sample average	0.0999	0.0501	0.0100	0.1001	0.0500	0.0099
	Standard error	0.0069	0.0042	0.0006	0.0086	0.0064	0.0028
	Relative ratio	53%	37%	3.6%	82%	86%	79%
BIS	Sample average	0.1002	0.0505	0.0101	0.1009	0.0503	0.0101
	Standard error	0.0089	0.0068	0.0014	0.0095	0.0067	0.0031
	Relative ratio	88%	97%	20%	100%	95%	97%
CMC	Sample average	0.1005	0.0506	0.0100	0.1005	0.0498	0.0100
	Standard error	0.0092	0.0070	0.0030	0.0096	0.0071	0.0031

NOTE: The "Relative ratio" is $N_T/N_T^{(\mathrm{CMC})}$, where $N_T^{(\mathrm{CMC})} = P_T(1 - P_T)/(\mathrm{S.E.})^2$. S.E. denotes the standard error.

nificantly depending on the metamodel quality, and BIS generates an even higher standard error than CMC when $\rho = 0$.

Next, we investigate the impact of the variation of the randomness inside simulations. In Section 3, we noted that SIS1 and SIS2 are reduced to DIS when they are applied to a deterministic simulation model. Thus, we expect that if the uncontrollable randomness represented by ϵ has a small level of variation, the standard errors of SIS1 and SIS2 will be close to zero. To illustrate, we consider the same data-generating structure in (16) and (17), but with the constant variance, $\sigma^2(X) = \tau^2$. We use the optimal IS densities for SIS1 and SIS2 in simulations. Table 3 shows that as τ gets close to zero, so do the standard errors of SIS1 and SIS2. That is, the proposed methods practically reduce to DIS.

We conduct additional experiments with other parameter settings, which are detailed in the supplementary document: (a) experiment results with different M/N_T ratios suggest that the standard error of the SIS1 estimator is generally insensitive to the choice of M/N_T ratio; (b) in investigating the effects of the metamodeling inaccuracy for the global pattern and different locality levels of $\mu(X)$, we do not find any clear patterns for this specific example. We also devise numerical examples with a multivariate input vector and observe the similar patterns discussed above (detailed in the supplementary document).

In summary, SIS1 and SIS2 always outperform BIS and CMC in various settings. We obtain remarkable improvements of computational efficiency when the input density and SIS1 (or SIS2) density are different. Also, as the target failure probability gets smaller, the efficiencies of SIS1 and SIS2 increase. Overall, SIS1 yields smaller standard errors than SIS2 in most cases. However, when it is difficult to build a good-quality metamodel (e.g., due to complex response surface over the input space), SIS2 would provide robust estimations because it is less sensitive to the metamodel quality.

6. IMPLEMENTATION WITH WIND TURBINE SIMULATORS

We implement the proposed approach to evaluate the reliability of a wind turbine operated in dynamic wind conditions (Byon, Ntaimo, and Ding 2010), using the NREL simulators. Implementation details are provided in the supplementary document.

6.1 Description of NREL Simulations

Following wind industry practice and the international standard, IEC 61400-1 (International Electrotechnical Commission 2005), we use a 10 min average wind speed as an input, X, to

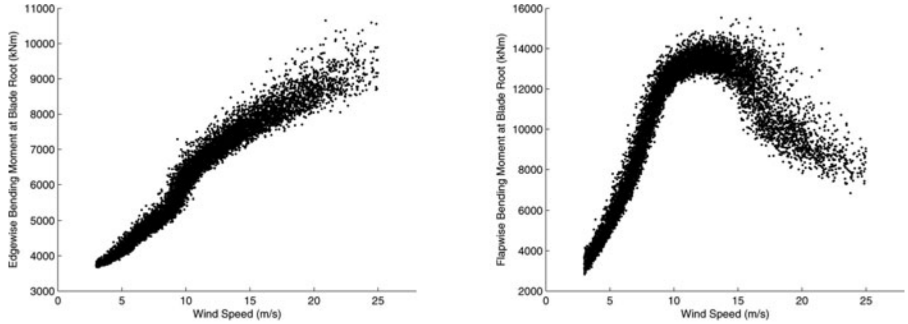
Table 2. POE estimation results with different ρ ($\delta = 1$)

			ρ	
		1.00	0.50	0
SIS1	Sample average	0.0100	0.0100	0.0101
	Standard error	0.0005	0.0008	0.0017
SIS2	Sample average	0.0100	0.0101	0.0100
	Standard error	0.0006	0.0007	0.0010
BIS	Sample average	0.0101	0.0100	0.0102
	Standard error	0.0014	0.0018	0.0063
CMC	Sample average	0.0099	0.0099	0.0099
	Standard error	0.0030	0.0030	0.0030

Table 3. POE estimation results with different τ ($\delta = 1$)

				τ		
		0.50	1.00	2.00	4.00	8.00
SIS1	Sample average	0.0102	0.0101	0.0101	0.0102	0.0100
	Standard error	0.0001	0.0001	0.0005	0.0021	0.0028
SIS2	Sample average	0.0102	0.0101	0.0101	0.0104	0.0100
	Standard error	0.0001	0.0002	0.0006	0.0023	0.0028

NOTE: SIS1's standard errors for $\tau = 0.50$ and $\tau = 1.00$ are 0.00007 and 0.00013, respectively, in more digits.



(a) Edgewise bending moment (b) Flapwise bending moment

Figure 1. Load outputs from the NREL simulators.

the NREL simulators. As the density of X, f, we use a Rayleigh density with a truncated support, following Moriarty (2008).

Given a 10 min average wind speed, X, the NREL simulators, including TurbSim (Jonkman 2009) and FAST (Jonkman and Buhl 2005), simulate the turbine's 10 min operations. Among the 10 min load responses, we take the maximum response of a load type as an output variable, Y. Specifically, we study two load types, edgewise and flapwise bending moments (hereafter, edgewise and flapwise moments) at a blade root, as they are of great concern in ensuring a wind turbine's structural reliability (Moriarty 2008). Hereafter, a simulation replication denotes the 10 min simulation that generates a 10 min maximum load (hereafter, a load, or response), given a 10 min average wind speed (hereafter, a wind speed). Figure 1 shows the load outputs in a range of wind conditions. High wind speed tends to cause large edgewise moments, which are dominated by gravity loading. Flapwise moments depend on the pitch regulation (Moriarty 2008; Yampikulsakul et al. 2014) that controls the blade pitch angles to reduce the loading on the blades when the wind speed is higher than the rated speed (11.5 m/s in Figure 1(b)).

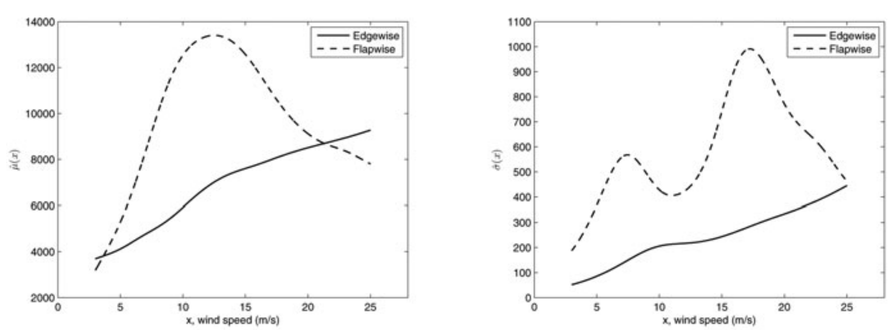
Table 4. Failure probability estimation by SIS1 method with different ratios of M to N_T

	Edgewise ($l = 8600$ kNm, $N_T = 1000$)		Flapwise ($l = 13{,}800$ kNm, $N_T = 2000$)	
M/N_T	Sample average	Standard error (95% bootstrap CI)	Sample average	Standard error (95% bootstrap CI)
10%	0.0486	0.0016 (0.0012, 0.0020)	0.0523	0.0034 (0.0026, 0.0041)
30%	0.0486	0.0018 (0.0014, 0.0022)	0.0514	0.0028 (0.0022, 0.0033)
50%	0.0487	0.0022 (0.0017, 0.0026)	0.0516	0.0032 (0.0024, 0.0039)
80%	0.0483	0.0022 (0.0017, 0.0025)	0.0527	0.0033 (0.0024, 0.0041)

6.2 Approximation of Conditional POE With a Metamodel

To implement SIS1, SIS2, and BIS, we need the conditional POE, $s(x)$, which is unknown in practice. We approximate it using a parametric regression model. Lee et al. (2013) model loads in wind turbine field data using a nonhomogenous GEV distribution. We apply a similar procedure for approximating the conditional POE.

To begin, we obtain a pilot sample of NREL simulations to build the metamodel. The pilot sample consists of 600 observations of (X, Y) pairs, where X is the wind speed uniformly sampled between 3 m/s and 25 m/s, and Y is the corresponding load response from the NREL simulators. In the metamodel, we use a nonhomogenous GEV distribution to approximate the conditional distribution of $Y|X = x$ and express the location and scale parameters as functions of wind speeds as in Lee et al. (2013). We also considered other parametric distributions including Weibull, Gamma, and lognormal distributions. How-



(a) Estimated location parameter function, $\hat{\mu}(x)$ (b) Estimated scale parameter function, $\hat{\sigma}(x)$

Figure 2. Estimated parameter functions for edgewise and flapwise moments.

ever, GEV provides the best fit for our chosen load types. The cumulative distribution function of GEV is expressed as follows, with the location parameter function, $\mu(x)$, the scale parameter function, $\sigma(x)$, and the shape parameter, ξ:

$$P(Y \leq y \mid X = x) = \begin{cases} \exp\left(-\left(1+\xi\left(\frac{y-\mu(x)}{\sigma(x)}\right)\right)^{-1/\xi}\right) \\ \quad \text{for } \xi \neq 0 \\ \exp\left(-\exp\left(-\frac{y-\mu(x)}{\sigma(x)}\right)\right) \\ \quad \text{for } \xi = 0\,. \end{cases}$$

We model the location and scale parameter functions with cubic smoothing spline functions. For the shape parameter, we use a constant, ξ, to avoid an overly complicated model as suggested in Lee et al. (2013). To estimate the spline function parameters and the shape parameter, we use the GAMLSS framework (Rigby and Stasinopoulos 2005). Specifically, we maximize the log-likelihood function penalized by the roughness of $\mu(x)$ and $\log \sigma(x)$ for fixed smoothing parameters, λ_μ and λ_σ:

$$\max \mathcal{L}_p = \mathcal{L} - \lambda_\mu \int_{-\infty}^{\infty} \mu''(x)^2\, dx - \lambda_\sigma \int_{-\infty}^{\infty} \left((\log\sigma)''(x)\right)^2\, dx,$$

where $\mathcal{L}$ is the log-likelihood function of the pilot data, (X_i, Y_i), $i = 1, 2, \ldots, 600$. The roughness penalties based on the second derivatives are commonly employed in the literature (Hastie and Tibshirani 1990; Green and Silverman 1994). We find the smoothing parameters, λ_μ and λ_σ, which minimize the Bayesian information criterion as suggested in Rigby and Stasinopoulos (2005). Figures 2(a) and 2(b) present the estimated location and scale parameter functions, $\hat{\mu}(x)$ and $\hat{\sigma}(x)$, respectively. The estimated shape parameters, $\hat{\xi}$, are -0.0359 and -0.0529 for the edgewise and flapwise moments, respectively.

Next, we conduct the Kolmogorov–Smirnov (KS) test to see the goodness-of-fit of the GEV distribution. We standardize the output, using the estimated location and scale functions shown in Figure 2, and perform the KS test on the standardized loads, Z_i, $i = 1, 2, \ldots, 600$, with the null hypothesis, $H_0 : Z \sim \text{GEV}(\mu = 0, \sigma = 1, \hat{\xi})$. The test results support the use of GEV distribution for the edgewise and flapwise moments with the p-values of 0.716 and 0.818, respectively. In the supplementary document, we include additional tests at important wind speeds, which also support the use of GEV distribution.

6.3 Sampling From IS Densities

To avoid difficulties in drawing samples from the IS densities whose normalizing constants are unknown, we use the following acceptance–rejection algorithm (Kroese, Taimre, and Botev 2011).

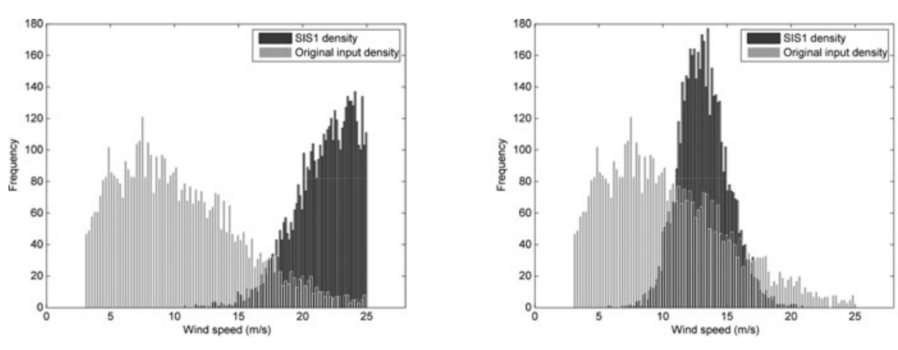
Acceptance–rejection algorithm

Step 1. Sample $\mathbf{x}$ from the input distribution, f.
Step 2. Sample u from the uniform distribution over the interval, $(0, f(\mathbf{x}))$.
Step 3. If $u \leq C_q \cdot q(\mathbf{x})$, return $\mathbf{x}$; otherwise, repeat from Step 1.

Here, C_q denotes the normalizing constant corresponding to the IS density, namely, C_{q1} for SIS1, C_{q2} for SIS2, and $P(Y > l)$ for BIS. Note that $C_q \cdot q(\mathbf{x})$ only involves $f(\mathbf{x})$ and $s(\mathbf{x})$. Thus, without a knowledge of C_q, this algorithm returns $\mathbf{x}$, which follows the target IS density, q. This algorithm exactly samples from q when the inequality condition, $f(\mathbf{x}) \geq C_q \cdot q(\mathbf{x})$, $\forall \mathbf{x} \in \mathcal{X}_f$ is satisfied. The IS densities, q_{SIS1}, q_{SIS2}, and q_{BIS}, satisfy this inequality condition. The acceptance rate of the algorithm is equal to C_q (Kroese, Taimre, and Botev 2011).

The acceptance–rejection method has several advantages. First, this method keeps the unbiasedness of the estimator because of its independent and exact sampling nature. Second, we can always use the original input distribution, f, as an auxiliary distribution. However, we can also use other sampling methods such as Markov chain Monte Carlo (MCMC). MCMC method can be useful if the input, $\mathbf{X}$, is high dimensional (Kroese, Taimre, and Botev 2011). The choice of sampling method is flexible in implementing SIS1 and SIS2. In practice, the computational cost of the sampling would be insignificant, for example, sampling thousands of inputs from the IS densities is a matter of seconds, whereas thousands of the NREL simulation replications can take days.

Figure 3 shows empirical SIS1 densities using the sampled wind speeds from the acceptance–rejection algorithm. For the edgewise moments in Figure 3(a), compared to the original input density, the SIS1 density has a higher mass at high wind speeds where high loads likely occur and high load variability is observed (see Figure 1(a)). Similarly, the SIS1 density for flapwise moments in Figure 3(b) centers around the rated speed, 11.5 m/s, where high loads and variability are observed (see Figure 1(b)). Using the same acceptance–rejection



(a) Edgewise moments with l =9,300 kNm (b) Flapwise moments with l =14,300 kNm

Figure 3. Comparison of empirical densities: original input density, f, versus SIS1 density, q_{SIS1}.

Table 5. Estimation results of the failure probability for edgewise moments

Method	$l = 8600$ kNm, $N_T = 1000$ Sample average	Standard error (95% bootstrap CI)	Relative ratio	$l = 9300$ kNm, $N_T = 3000$ Sample average	Standard error (95% bootstrap CI)	Relative ratio
SIS1	0.0486	0.0016 (0.0012, 0.0020)	5.5%	0.00992	0.00040 (0.00032, 0.00047)	4.9%
SIS2	0.0485	0.0020 (0.0016, 0.0024)	8.7%	0.01005	0.00044 (0.00036, 0.00051)	5.9%
BIS	0.0488	0.0029 (0.0020, 0.0037)	18%	0.00995	0.00056 (0.00042, 0.00068)	9.6%

algorithm, we also draw wind speeds from the SIS2 and BIS densities.

Even though we sample inputs from the IS densities without knowing the value of the normalizing constant, C_q, we still need to compute C_q for estimating the failure probability because the likelihood ratio in the IS estimators, $f(\mathbf{X})/q(\mathbf{X})$, needs to be evaluated to ensure the unbiasedness of the estimators. This issue has been studied in the literature (Hesterberg 1995). In this study, we employ numerical integration to compute C_q since a state-of-the-art numerical integration technique leads to an accurate evaluation of C_q (Shampine 2008). Our numerical studies in the supplementary document also show that the numerical integration does not affect the POE estimation accuracy.

6.4 Sensitivity Analysis With Different *M* in SIS1

Recall that in SIS1, we derived the optimal SIS1 density, q_{SIS1}, and the optimal allocation size, $N_i, i = 1, 2, \ldots, M$, for a given input sample size, M, and a total computational resource, N_T. To see the effect of the ratio of M to N_T on POE estimation, we consider the four ratios of M to N_T, 10%, 30%, 50%, and 80%. Table 4 summarizes the sample average and standard error based on 50 POE estimates. We also obtain the 95% confidence interval (CI) of the standard error by using the bootstrap percentile interval (Efron and Tibshirani 1993). Overall, the standard errors are comparable among different ratios.

Similar results are also observed in the extensive numerical studies where we have tested 10%, 30%, 50%, 70%, and 90% of M/N_T ratios for the univariate and multivariate examples (see the supplementary document). All of these results indicate that the estimation accuracy is not sensitive to the size of M, given N_T. In the subsequent implementations, we use the ratio of 10% and 30% for the edgewise and flapwise moments, respectively.

6.5 Implementation Results

Tables 5 and 6 summarize the implementation results for the edgewise and flapwise moments, respectively, using 50 POE estimates for SIS1, SIS2, and BIS. For each response type, we use two different values of the resistance level, l. In general, the SIS1's standard errors appear to be slightly smaller than the SIS2's. In all cases, SIS1 and SIS2 outperform BIS, which confirms the theoretical advantage of their variance reductions.

We also assess the computational gains of the IS methods over CMC. Let $N_T^{(\text{CMC})}$ denote the number of CMC simulation replications to achieve the same standard error of the corresponding method in each row of Tables 5 and 6. With $N_T^{(\text{CMC})}$ replications, the standard error of the CMC estimator is $\sqrt{P(1-P)/N_T^{(\text{CMC})}}$, where P is the true failure probability, $P(Y > l)$. Since P is unknown, we use the sample average of SIS1 for P because SIS1 generates the smallest standard error in all cases. With the estimated $N_T^{(\text{CMC})}$, we compute the relative ratio, $N_T/N_T^{(\text{CMC})}$, as shown in Tables 5 and 6. For the edgewise moment, the SIS methods need about 5% to 9% of the CMC efforts. In other words, for $l = 8600$ kNm, CMC needs about 11,000 to 18,000 replications to obtain the same accuracy achieved by SIS1 and SIS2 with 1000 replications. For $l = 9300$ kNm, CMC needs 51,000 to 61,000 replications compared to SIS1 and SIS2 with 3000 replications.

We explain the fact that the computational gains of the SIS methods for the flapwise moment are not as substantial as for the edgewise moment using Figure 3; the SIS1 density for the flapwise moment is not as different from the original input density

Table 6. Estimation results of the failure probability for flapwise moments

Method	$l = 13{,}800$ kNm, $N_T = 2000$ Sample average	Standard error (95% bootstrap CI)	Relative ratio	$l = 14{,}300$ kNm, $N_T = 9000$ Sample average	Standard error (95% bootstrap CI)	Relative ratio
SIS1	0.0514	0.0028 (0.0022, 0.0033)	32%	0.01070	0.00061 (0.00047, 0.00074)	32%
SIS2	0.0527	0.0032 (0.0025, 0.0038)	42%	0.01037	0.00063 (0.00046, 0.00078)	34%
BIS	0.0528	0.0038 (0.0030, 0.0044)	59%	0.01054	0.00083 (0.00055, 0.00110)	59%

as is the SIS1 density for the edgewise moment. We observe similar results for the SIS2 density. As a result, the computational gains by biasing the input distribution using the SIS methods become less obvious for the flapwise moment than the edgewise moment. Recall that we observed the similar pattern in the numerical studies discussed in Section 5, where the computational gains of SIS1 and SIS2 are less remarkable when the optimal IS density is similar to the original input density (with $\delta = -1$ in (17)).

7. SUMMARY

This study proposes an extended framework of IS for the reliability evaluation using a stochastic simulation model. The applicability of the existing IS methods is limited to simulations with deterministic simulation models where an output is uniquely determined at a given input.

By accounting for different sources of output variability in stochastic simulation models, we develop two methods for estimating a failure probability. For SIS1, which allows multiple replications at each sampled input, we derive the optimal IS density and allocation size that minimize the variance of the estimator. For SIS2, which uses one replication at each sampled input, we derive the optimal IS density. Since SIS2 imposes an additional restriction on the allocation size, SIS1 is more flexible. However, SIS2 does not need to determine the input sample size and the optimal allocation size. The implementation results suggest that the performance of SIS1 is comparable to SIS2 in most cases and that both SIS methods can significantly improve the estimation accuracy over the two benchmark methods, BIS and CMC. We observe that the computational gains of the SIS methods become larger when a smaller POE needs to be estimated and when the difference between the IS density and the original input density is larger.

In the future, we plan to look more closely at the cross-entropy method (Rubinstein 1999; De Boer et al. 2005), where the candidate IS density is selected among a parametric family. We also plan to investigate the methods that estimate a very small probability in the binomial distribution to improve the estimation of the conditional POE in stochastic simulation models. Also, to balance the cost of constructing the metamodel and the accuracy of reliability estimation, we will consider an iterative approach where the metamodel is sequentially refined with new simulation outputs. Finally, we plan to develop a new SIS method to optimize a simulation experiment for evaluating the reliability associated with multiple responses.

SUPPLEMENTARY MATERIALS

Derivations for the optimal solutions of SIS1 and SIS2 with the multivariate input vector, numerical examples with the univariate input variable and the multivariate input vector, and implementation details with the wind turbine simulators.

ACKNOWLEDGMENTS

The authors thank the editor, anonymous associate editor, and anonymous reviewers for their constructive and thoughtful comments on various aspects of this work. This work was partially supported by the National Science Foundation (grant no. CMMI-1362513).

[Received October 2013. Revised December 2014.]

REFERENCES

Ankenman, B., Nelson, B. L., and Staum, J. (2010), "Stochastic Kriging for Simulation Metamodeling," *Operations Research*, 58, 371–382. [352,354]

Au, S. K., and Beck, J. L. (2003), "Important Sampling in High Dimensions," *Structural Safety*, 25, 139–163. [352]

Byon, E., Ntaimo, L., and Ding, Y. (2010), "Optimal Maintenance Strategies for Wind Power Systems Under Stochastic Weather Conditions," *IEEE Transactions on Reliability*, 59, 393–404. [356]

Cannamela, C., Garnier, J., and Iooss, B. (2008), "Controlled Stratification for Quantile Estimation," *The Annals of Applied Statistics*, 2, 1554–1580. [351,353,354,355]

Casella, G., and Berger, R. L. (2002), *Statistical Inference*, Australia: Thomson Learning. [353]

Choulakian, V., and Stephens, M. A. (2001), "Goodness-of-Fit Tests for the Generalized Pareto Distribution," *Technometrics*, 43, 478–484. [355]

Coles, S. G. (2001), *An Introduction to Statistical Modeling of Extreme Values*, London: Springer. [355]

Courant, R., and Hilbert, D. (1989), *Methods of Mathematical Physics*, New York: Wiley. [352,354]

D'Auria, F., Bousbia-Salah, A., Petruzzi, A., and Del Nevo, A. (2006), "State of the Art in Using Best Estimate Calculation Tools in Nuclear Technology," *Nuclear Engineering and Technology*, 38, 11–32. [351]

De Boer, P., Kroese, D. P., Mannor, S., and Rubinstein, R. Y. (2005), "A Tutorial on the Cross-Entropy Method," *Annals of Operations Research*, 134, 19–67. [351,352,360]

Dubourg, V., Sudret, B., and Deheeger, F. (2013), "Metamodel-Based Importance Sampling for Structural Reliability Analysis," *Probabilistic Engineering Mechanics*, 33, 47–57. [351,352]

Efron, B., and Tibshirani, R. (1993), *An Introduction to the Bootstrap*, New York: Chapman & Hall. [359]

Glynn, P. W., and Iglehart, D. L. (1989), "Importance Sampling for Stochastic Simulations," *Management Science*, 35, 1367–1392. [353]

Green, P. J., and Silverman, B. W. (1994), *Nonparametric Regression and Generalized Linear Models*, London: Chapman and Hall. [355,358]

Hastie, T. J., and Tibshirani, R. J. (1990), *Generalized Additive Models*, London: Chapman and Hall. [355,358]

Heidelberger, P. (1995), "Fast Simulation of Rare Events in Queueing and Reliability Models," *ACM Transactions on Modeling and Computer Simulation (TOMACS)*, 5, 43–85. [353]

Hesterberg, T. C. (1995), "Weighted Average Importance Sampling and Defensive Mixture Distributions," *Technometrics*, 37, 185–194. [359]

Huang, D., Allen, T. T., Notz, W. I., and Zeng, N. (2006), "Global Optimization of Stochastic Black-Box Systems via Sequential Kriging Meta-Models," *Journal of Global Optimization*, 34, 441–466. [352]

International Electrotechnical Commission (2005), *Wind Turbines—Part 1: Design Requirements*, IEC/TC88, 61400-1 (3rd ed.), Geneva: IEC. [356]

Jonkman, B. J. (2009), "TurbSim User's Guide: Version 1.50," Technical Report NREL/TP-500-46198, National Renewable Energy Laboratory, Golden, Colorado. [351,352,357]

Jonkman, J. M., and Buhl, M. L., Jr. (2005), "FAST User's Guide," Technical Report NREL/EL-500-38230, National Renewable Energy Laboratory, Golden, Colorado. [351,357]

Joseph, V. R. (2006), "Limit Kriging," *Technometrics*, 48, 458–466. [353]

Kroese, D. P., Taimre, T., and Botev, Z. I. (2011), *Handbook of Monte Carlo Methods*, New York: Wiley. [352,358]

Lee, G., Byon, E., Ntaimo, L., and Ding, Y. (2013), "Bayesian Spline Method for Assessing Extreme Loads on Wind Turbines," *The Annals of Applied Statistics*, 7, 2034–2061. [357,358]

Manuel, L., Nguyen, H. H., and Barone, M. F. (2013), "On the Use of a Large Database of Simulated Wind Turbine Loads to Aid in Assessing Design Standard Provisions," in *Proceedings of the 51st AIAA Aerospace Sciences Meeting including the New Horizons Forum and Aerospace Exposition*, Grapevine, Texas, 1–9. [351]

Moriarty, P. (2008), "Database for Validation of Design Load Extrapolation Techniques," *Wind Energy*, 11, 559–576. [351,352,357]

Rigby, R. A., and Stasinopoulos, D. M. (2005), "Generalized Additive Models for Location, Scale And Shape," *Journal of the Royal Statistical Society*, Series C, 54, 507–554. [355,358]

Rubinstein, R. (1999), "The Cross-Entropy Method for Combinatorial and Continuous Optimization," *Methodology and Computing in Applied Probability*, 1, 127–190. [360]

Seber, G. A. F., and Lee, A. J. (2003), *Linear Regression Analysis*, Hoboken: Wiley. [354]

Shampine, L. F. (2008), "Vectorized Adaptive Quadrature in MATLAB," *Journal of Computational and Applied Mathematics*, 211, 131–140. [359]

Stephens, M. A. (1974), "EDF Statistics for Goodness of Fit and Some Comparisons," *Journal of the American Statistical Association*, 69, 730–737. [355]

Yampikulsakul, N., Byon, E., Huang, S., Sheng, S., and You, M. (2014), "Condition Monitoring of Wind Power System With Nonparametric Regression Analysis," *IEEE Transactions on Energy Conversion*, 29, 288–299. [357]