Reliability Analysis under Integrated Input Variable and Metamodel Uncertainty based on Bayesian Approach

Dawn An, Jooho Choi, and Junho Won
School of Aerospace & Mechanical Engineering, Korea Aerospace University
skal34@nate.com, jhchoi@kau.ac.kr, openworldsm@hanmail.net

Abstract: A reliability analysis procedure is proposed based on a Bayesian framework, which can address the uncertainty in the input variables and the metamodel uncertainty of the response function in an integrated manner. The input uncertainty includes the statistical uncertainty due to the lack of knowledge or insufficient data, which is often the case in the design practice. A method of posterior prediction is used to evaluate the influence of this uncertainty. The metamodel uncertainty is accounted for, which arises due to the surrogate approximation to reduce the costly computation of the response function. Gaussian process model, also known as Kriging model, is employed to assess the associated uncertainty in the form of prediction band. Posterior distributions are obtained by Markov Chain Monte Carlo (MCMC) method, which is an efficient simulation method to draw random sequence of parameters that samples the given distribution. Mathematical and engineering examples are used to demonstrate the proposed method.

Keywords: Reliability analysis, Epistemic uncertainty, Statistical uncertainty, Metamodel uncertainty, Bayesian approach

1. Introduction

In the modern industrial society, many efforts are directed to reduce the failure rate and costs of the product. As part of such efforts, reliability analysis for dealing with the input variable uncertainty in the design has been studied, which is to evaluate failure probability or safety levels of mechanical systems. A great number of researches have been conducted to efficiently handle these problems, which can be classified into sampling based methods such as Monte Carlo Simulation (MCS), Most Probable Point (MPP) based methods including First/ Second Order Reliability Method (FORM/SORM) (Rackwitz, 2001), and more recently moment based integration methods with the Dimension Reduction Method (DRM) as being the most noteworthy (Rahman and Xu, 2004; Won et al., 2009; Lee et al., 2008).

In the previous developments, the uncertainty has mostly been considered as aleatory uncertainty which is irreducible and related with inherent physical randomness that is completely described by a suitable probability model (Rahman and Xu, 2004; Won et al., 2009; Lee et al., 2008; Haldar and Mahadevan, 2000). Epistemic uncertainty, however, is prevalent in the real industrial world, which makes the existing methods less useful since it results from the lack of data or subjective knowledge. There have been recent studies to handle this uncertainty by using non-probabilistic methods, which includes the interval analysis (Qiu et al., 2004), the possibility theory (Mourelatos and Zhou, 2005; Du et al., 2006) and evidence theory (Helton et al., 2007; Mourelatos and Zhou, 2006). The weakness of these methods, however, is that the uncertainty is modeled more or less based on the subjective expert opinions. In engineering design practice, the uncertainty is often given by a small number of samples from historical data or actual experiment, which is

4th International Workshop on Reliable Engineering Computing (REC 2010)
Edited by Michael Beer, Rafi L. Muhanna and Robert L. Mullen
Copyright © 2010 Professional Activities Centre, National University of Singapore.
ISBN: 978-981-08-5118-7. *Published by* Research Publishing Services.
doi:10.3850/978-981-08-5118-7_042

too few to infer the probability distributions. This is called as statistical uncertainty (Noor, 2005; Kiureghian and Ditlevsen, 2009). The Bayesian approach can be a useful method in this case due to the advantages that it easily represents the insufficiency of the data in terms of the probability, it provides a unified way for aleatory and epistemic uncertainty in a single framework; and it can conveniently update the degree of uncertainty by adding more data to the prior information (Zhang and Mahadevan, 2001; Gunawan and Papalambros, 2006; Mahadevan and Rebba, 2005; Cruse and Brown, 2007).

In the Bayesian approach, the probability itself is treated as a random variable which quantifies our degree of belief on the probability in light of the observed data. In the recent literature, two-stage nested or double loop of reliability analysis were proposed in order to implement this procedure, in which the distribution parameters are treated as the unknown random variable. In this approach, the outer loop determines the CDF of probability, whereas the inner loop solves conventional reliability analysis problem given the distribution parameters. Several methods, e.g., FORM (Zhang and Mahadevan, 2001), Markov Chain Monte Carlo (MCMC) (Cruse and Brown, 2007) and the DRM by the authors were used for this purpose. The nested procedure, however, calls for substantial increase of computations, and is not tractable for the practical design purpose (Cruse and Brown, 2007; Eduard et al., 2002). In this study, a method using posterior prediction is proposed to resolve this problem, which requires only a single reliability analysis step.

In the evaluation of structural response function, common practice is to employ metamodel that approximates the original response in an effort to save computational cost. This causes another uncertainty, which we call metamodel uncertainty, since the model is constructed using only a finite number of responses and hence is unknown at the untried points (Gelman et al., 2004; O'Hagan, 2006). In this paper, Gaussian process model, also known as Kriging, is employed for the response approximation. The uncertainties of the associated parameters are investigated conditional on the finite number of responses.

The final goal of this study is the integration of all the uncertainties in a single Bayesian framework. Information of the posterior distribution, whether they are the model parameters of the input variables or the parameters of the metamodel, are obtained by employing MCMC simulation, which is a modern computational method to draw random sequence of parameters that samples the given distribution (Andrieu et al., 2003). Once the posterior samples of the parameters are available, predictive samples are drawn from the associated distribution given each value of the parameters. The uncertainty of the response information is then estimated using the drawn data. Mathematical examples and engineering problems are given to demonstrate that proposed method is feasible and practical.

2. Bayesian Reliability Analysis

2.1. Reliability analysis

In general, structural reliability analysis is to calculate the probability of a certain event, typically given in the form

$$p_g = P\left[G<0\right] = \int_{g(\mathbf{X})<0} f_{\mathbf{X}}\left(\mathbf{x}\right) d\mathbf{x} \text{ or } \int_{g<0} f_G\left(g\right) dg \tag{1}$$

where $\mathbf{X}=(X_1,X_2,...,X_n)$ is a vector of random input variables, $f_{\mathbf{X}}(\mathbf{x})$ is the joint PDF of $\mathbf{X}$, g is the response function or so-called limit state function, $f_G(g)$ is the PDF of g, and G is the probabilistic representation of g. The event $g(\mathbf{X})<0$ can be a failure or safety depending on the problem. Each of the random input variable X_i has its own statistical distribution described by a set of model parameters $\boldsymbol{\theta}$. If all the members of $\mathbf{X}$ are aleatory, i.e. if the values of model parameter $\boldsymbol{\theta}$ are completely known either through an infinite amount of data or well-established knowledge, then the problem becomes an ordinary reliability prediction, from which a deterministic value of the probability p_g is calculated using the existing methods such as the MCS, FORM or DRM. Figure 1 gives a schematic picture of the analysis procedure.

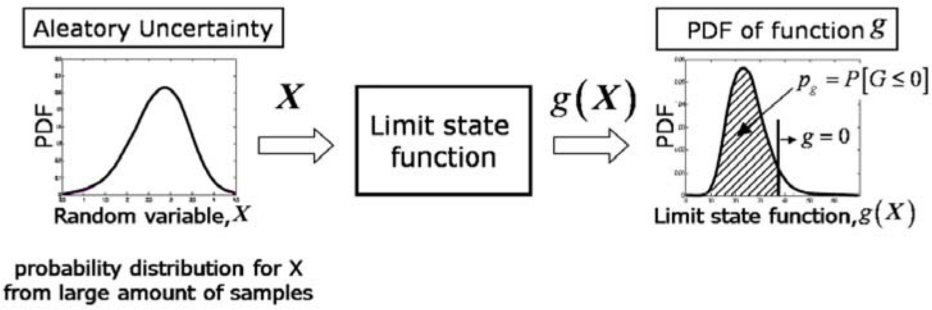


Figure 1. Conventional reliability analysis procedure.

2.2. Bayesian approach in case of input uncertainty

If part or all of the input variables $\mathbf{X}$ show epistemic or statistical uncertainty due to the insufficient data or knowledge, the corresponding model parameters $\boldsymbol{\theta}$ become uncertain, which leads to the uncertainty in the reliability prediction. In this case, the model parameters are assumed to be random, and Bayes' theorem is used to estimate the probabilistic behavior of the parameters (Gelman et al., 2004):

$$f(\boldsymbol{\theta}\,|\,\mathbf{x}) \propto f(\mathbf{x}\,|\,\boldsymbol{\theta})f(\boldsymbol{\theta}) \tag{2}$$

where $f(\boldsymbol{\theta}\,|\,\mathbf{x})$ is the posterior PDF of $\boldsymbol{\theta}$ conditional on the observed data $\mathbf{x}$, $f(\mathbf{x}\,|\,\boldsymbol{\theta})$ is the likelihood of the observed data $\mathbf{x}$ given the parameters $\boldsymbol{\theta}$, and $f(\boldsymbol{\theta})$ is the prior PDF of $\boldsymbol{\theta}$.

Since the model parameters are not constant but follow a statistical distribution, the probability p_g, as defined in Eq. (1), is no longer a deterministic value but behaves as a random variable, which is depicted as capital letter P_g. This is again another reliability analysis problem called the outer-loop, in which the input random variables $\mathbf{X}$ are replaced by $\boldsymbol{\theta}$, and the response function g is replaced by the probability P_g. During this step, the reliability analysis stated previously, which is called the inner-loop, should be conducted in order to obtain the realization of P_g. Therefore, the Bayesian approach constitutes a nested reliability analyses in which the inner loop constructs the usual PDF of the response function g given the model parameters $\boldsymbol{\theta}$, and the outer loop constructs the PDF of P_g due to the distribution of $\boldsymbol{\theta}$. Considered all together, the probability distribution of P_g can be expressed in the form

$$F_{P_g}\left(p_g\right)=P\left[P_g<p\right]=\int_{p_g\le p} f_{P_g}\left(p_g\right)dp_g \tag{3}$$

where $F_{P_g}\left(p_g\right)$ and $f_{P_g}\left(p_g\right)$ are the CDF and PDF of P_g, respectively. The procedure is summarized in Figure 2. In this figure, the model parameters $\boldsymbol{\theta}$ are expressed by the posterior PDF $f\left(\boldsymbol{\theta}\,|\,\mathbf{x}\right)$ conditional on the given samples $\mathbf{x}$. Once a set of $\boldsymbol{\theta}$ values are drawn from this random distribution, the probability distribution for $\mathbf{X}$ is established. Reliability analysis, as stated in Figure 1, is carried out to obtain the probability value p_g, which is depicted as the inner loop. The probability is evaluated for each drawn values of $\boldsymbol{\theta}$ in order to construct the PDF of the probability, which is depicted as the outer loop. The obtained PDF represents a degree of belief on the probability conditional on the sample data. Since the probability is always bounded between 0 and 1, the PDF can be conveniently modeled by Beta distribution (Gelman et al., 2004), which is

$$f_{P_g}\left(p\right)=\frac{1}{B\left(\alpha,\beta\right)}p^{\alpha-1}\left(1-p\right)^{\beta-1} \tag{4}$$

where α,β are associated model parameters. Using the Beta PDF of P_g, Maximum Likelihood Estimation (MLE) and lower and upper confidence bounds can be estimated easily.

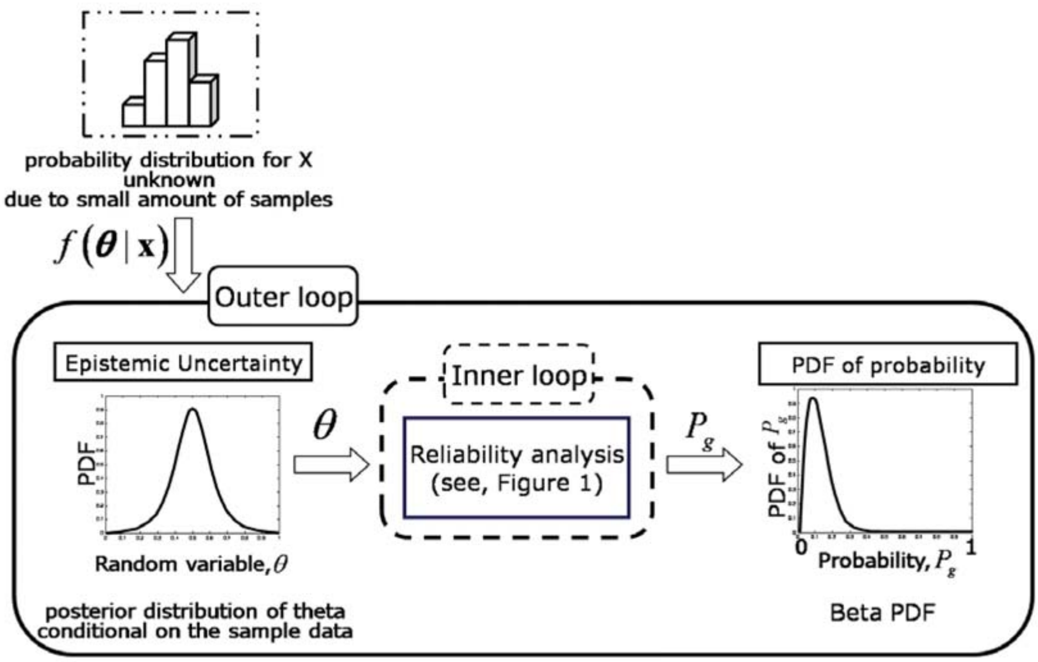


Figure 2. Bayesian reliability analysis procedure.

As was noted in the introduction, the authors have recently studied this problem using the nested approach. In the study, Dimension Reduction Method (DRM) (Rahman and Xu,2004) was used for each step of reliability analysis. The method determines the PDF of the response function due to the random input variables by efficiently calculating statistical moments using a small number of function calls. Then the whole step requires multiple runs of this, e.g., $(4q+1)(4r+1)$ with q and r being the number of uncertain model parameters $\boldsymbol{\theta}$ and number of input variables $\mathbf{X}$. The drawback of this method is that the computational cost is greatly increased due to the nested feature. Besides, the method works only for the parameters that follow standard distributions, which is not possible for the posterior distribution with non-conjugate prior as is often the case in the Bayesian procedure.

In this study, Markov Chain Monte Carlo (MCMC) method is employed, which is a simulation method to draw random sequence of parameters that samples the given distribution. The advantage of MCMC is that it easily produces posterior distribution with any complexity including no closed form expressions (Andrieu et al., 2003).

In order to reduce the computational cost incurred by the nested feature, predictive approach is proposed in this study, which is to use posterior predictive distribution given in the following form (Gelman et al., 2004)

$$f(\tilde{\mathbf{x}} \mid \mathbf{x}) = \int f(\tilde{\mathbf{x}} \mid \boldsymbol{\theta}) f(\boldsymbol{\theta} \mid \mathbf{x}) d\boldsymbol{\theta} \tag{5}$$

where the symbol $\sim$ represents the prediction, $f(\boldsymbol{\theta} \mid \mathbf{x})$ is the posterior distribution obtained by Eq. (2), and $f(\tilde{\mathbf{x}} \mid \boldsymbol{\theta})$ is the probability distribution of the prediction $\tilde{\mathbf{X}}$ conditional on the parameters $\boldsymbol{\theta}$. The predictive distribution can be obtained by integration of the two terms on the right. In practice, however, predictive samples $\tilde{\mathbf{x}}$ are drawn from the conditional probability distribution $f(\tilde{\mathbf{x}} \mid \boldsymbol{\theta})$, given the values of $\boldsymbol{\theta}$ drawn from the posterior distribution $f(\boldsymbol{\theta} \mid \mathbf{x})$ via the MCMC method. Once the samples of $\tilde{\mathbf{x}}$ are available, one can proceed to produce the data of response function by evaluating it at each sample point. Probability of the event given by Eq. (1) is then calculated from the resulting data.

2.3. Bayesian approach in case of metamodel uncertainty

Metamodel is commonly exploited in the modern simulation-based engineering analysis in order to reduce the computational cost, by approximating the original response to a surrogate function using a finite set of samples. One of popular choices receiving greatest attention is the Kriging model. Until recently, however, the Kriging was studied mostly from deterministic viewpoint, i.e., used just as a fitting or interpolating purpose while ignoring the uncertainty arising from not knowing the output of the computer simulation code, except at a finite set of sampling points. In (Kennedy and O'Hagan, 2001), they referred to this type of uncertainty as code uncertainty. We call this as metamodel uncertainty. Numerous efforts have been made in the statistical community to quantify this uncertainty (Kennedy and O'Hagan, 2001; Currin et al., 1991; Handcock and Stein, 1993). In one of the most popular approaches, the Kriging model is viewed as a realization of a Gaussian random process and Bayesian methods are used to quantify the associated uncertainties by calculating its posterior distribution of unknown parameters given the response values at the computer experiment points.

Let us now consider a case that the response function is approximately interpolated by the outputs $\mathbf{g} = [g_1, \ldots, g_m]'$ at a set of DOE points $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_m]'$ with m number. For this purpose, Gaussian random function is introduced as follows.

$$\hat{g}(\mathbf{x}) = \mathbf{f}(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}),\ Z \sim N\left(0\mathbf{I}_m, \sigma^2\mathbf{R}\right),\ \mathbf{R} = R\left(\mathbf{x}_i, \mathbf{x}_j\right),\ i, j = 1, \ldots, m \tag{6}$$

where ^ denotes the surrogate representation, $\mathbf{f}(\mathbf{x})\boldsymbol{\beta}$ is the normal linear model, $\mathbf{f} = [f_1, ..., f_k]$ and $\boldsymbol{\beta} = [\beta_1, ..., \beta_k]'$ are k number of the trial functions and associated parameters, respectively, Z is a Gaussian stochastic process with zero mean and variance σ^2, $\mathbf{I}_m$ is the $m \times m$ identity matrix, and R is a correlation function between $\mathbf{x}_i$ and $\mathbf{x}_j$ which is represented by

$$R(\mathbf{x}_i, \mathbf{x}_j) = \exp\left\{ -\left(\frac{\|\mathbf{x}_i - \mathbf{x}_j\|}{h} \right)^2 \right\} \tag{7}$$

where h is a correlation parameter that controls the degree of smoothness of the function. If the h gets higher, the model becomes smoother, but the singularity is encountered in the correlation matrix if it is too high. In most Kriging studies, h is determined by the method of maximum likelihood estimate (MLE). According to (Etman, 1994) and (Sasena et al., 2002), however, MLE method is not only computationally expensive which requires additional optimization process, but also the quality of the obtained parameter is questionable. In this study, h is considered as an unknown parameter to avoid this.

Based on the Eq. (6), the computer outputs $\mathbf{g}$ follow multivariate normal distribution:

$$\mathbf{g} \mid \boldsymbol{\beta}, \sigma^2, h, \mathbf{F} \sim N\left(\mathbf{F}\boldsymbol{\beta},\, \sigma^2 \mathbf{R}_{(\mathrm{X})}\right) \tag{8}$$

where

$$\mathbf{F} = \left[\mathbf{f}(\mathbf{x}_1), ..., \mathbf{f}(\mathbf{x}_m)\right]' \tag{9}$$

is m dimensional vector of $\mathbf{f}$ at $\mathbf{X}$, and the subscript $(\mathbf{X})$ in Eq. (8) denotes the correlation matrix in terms of $\mathbf{X}$. In this procedure, the parameters $\boldsymbol{\beta}$, σ and h are the unknowns to be determined. Assuming non-informative prior, the joint prior distribution for the unknown parameters is defined as

$$f\left(\boldsymbol{\beta}, \sigma^2, h\right) \propto \sigma^{-2} \tag{10}$$

Then the joint posterior distribution of the parameters is given by

$$\boldsymbol{\beta}, \sigma^2, h \mid \mathbf{g} \propto \left(\sigma^2\right)^{-\frac{m+2}{2}} \left|\mathbf{R}_{(\mathrm{X})}\right|^{-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2} (\mathbf{g} - \mathbf{F}\boldsymbol{\beta})' \mathbf{R}_{(\mathrm{X})}^{-1} (\mathbf{g} - \mathbf{F}\boldsymbol{\beta}) \right) \tag{11}$$

Once we obtain the distributional information on the parameters, we can proceed to draw the posterior predictive distribution of $\tilde{g}$ at a new set of untried points $\tilde{\mathbf{x}}$ using the drawn samples of parameters, which is given by the multivariate normal distribution:

$$\tilde{g} \mid \boldsymbol{\beta}, \sigma^2, h \sim N\left(\mathbf{f}(\tilde{\mathbf{x}})\boldsymbol{\beta} + \mathbf{R}_{(\tilde{\mathrm{x}},\mathrm{X})} \mathbf{R}_{(\mathrm{X})}^{-1} (\mathbf{g} - \mathbf{F}\boldsymbol{\beta}), \left(\mathbf{R}_{(\tilde{\mathrm{x}})} - \mathbf{R}_{(\tilde{\mathrm{x}},\mathrm{X})} \mathbf{R}_{(\mathrm{X})}^{-1} \mathbf{R}_{(\mathrm{X},\tilde{\mathrm{x}})} \right) \sigma^2 \right) \tag{12}$$

2.4. General procedure of Bayesian approach

The Bayesian procedure holds for both the cases of input uncertainty and metamodel uncertainty, which is the reason that the approach can be generalized to integrate all the uncertainties in a single Bayesian framework. The steps are summarized in Table I.

Table I. Procedure of integrated Bayesian reliability analysis		
	Input uncertainty	Metamodel (Kriging) uncertainty
Step 1 observation	$\mathbf{x}_i,\ i=1,\ldots,n$	$(\mathbf{x}_j, g_j),\ j=1,\ldots,m$
Step 2 posterior distribution	$f(\theta\mid\mathbf{x})\propto f(\mathbf{x}\mid\theta)f(\theta)$	$f(\beta,\sigma^2,h\mid\mathbf{g})$ $\sim f(\mathbf{g}\mid\beta,\sigma^2,h)f(\beta,\sigma^2,h)$ (see Eq. (11))
Step 3 posterior distribution sampling	$\theta_i,\ i=1,\ldots,N$	$\beta_j,\sigma_j^{\ 2},h_j,\ j=1,\ldots,N$
Step 4 predictive distribution	$f(\tilde{\mathbf{x}}\mid\mathbf{x})=\int f(\tilde{\mathbf{x}}\mid\theta)f(\theta\mid\mathbf{x})d\theta$	$\tilde{g}\sim N(\mathrm{M},\Sigma)$ where $\mathrm{M}=\mathbf{f}(\tilde{\mathbf{x}})\beta+\mathbf{R}_{(\tilde{\mathbf{x}},\mathbf{x})}\mathbf{R}_{(\mathbf{x})}^{\ -1}(\mathbf{g}-\mathbf{F}\beta)$ $\Sigma=\left(\mathbf{R}_{(\tilde{\mathbf{x}})}-\mathbf{R}_{(\tilde{\mathbf{x}},\mathbf{x})}\mathbf{R}_{(\mathbf{x})}^{\ -1}\mathbf{R}_{(\mathbf{x},\tilde{\mathbf{x}})}\right)\sigma^2$
Step 5 predictive distribution sampling	draw $\tilde{\mathbf{x}}_i$ from each $\theta_i,\ i=1,\ldots,N$	draw $\tilde{g}_j(\tilde{\mathbf{x}})$ from each $\beta_j,\sigma_j^{\ 2},h_j,\ j=1,\ldots,N$
Step 6 integrated reliability analysis	$\tilde{g}_i(\tilde{\mathbf{x}}_i)\sim$ $N\left(\mathbf{f}(\tilde{\mathbf{x}}_i)\beta_i+\mathbf{R}_{i(\tilde{\mathbf{x}}_i,\mathbf{x})}\mathbf{R}_{i(\mathbf{x})}^{\ -1}(\mathbf{g}-\mathbf{F}\beta_i),\left(\mathbf{R}_{i(\tilde{\mathbf{x}}_i)}-\mathbf{R}_{i(\tilde{\mathbf{x}}_i,\mathbf{x})}\mathbf{R}_{i(\mathbf{x})}^{\ -1}\mathbf{R}_{i(\mathbf{x},\tilde{\mathbf{x}}_i)}\right)\sigma_i^{\ 2}\right)$ Draw $\tilde{g}_i(\tilde{\mathbf{x}}_i)$ from each $\beta_i,\sigma_i^{\ 2},h_i$ at each $\tilde{\mathbf{x}}_i$ respectively $i=1,\ldots,N$ from this, calculate $P[G<0]$	

3. Input Variable Uncertainty

In this section, effect of response function due to the input variable uncertainty is studied using two mathematical examples. First is a function of single variable as follows (O'Hagan, 2006):

$$g(X)=X+3\sin(X/2) \tag{13}$$

The variable X is assumed as aleatory with normal distribution $N(3.5,1.2)$. The PDF distribution of $g(X)$ due to the randomness of X is obtained by using the classical MCS with $N=10^6$, which is plotted in Figure 3. Assuming $g(X)<3.5$ as a failure event, the probability $P[G<3.5]$ is calculated as 0.0462.

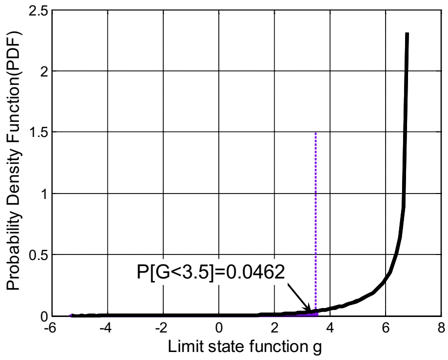


Figure 3. PDF of response function $g(X)$ of example 1.

Now the input variable is changed to the statistical uncertainty due to the limited amount of data with n numbers. Model parameters $\boldsymbol{\theta}$, which are the mean μ and standard deviation σ, become unknown accordingly. Let us assume that still the same values for the sample mean and standard deviation are observed from the data, i.e., $\bar{x}$=3.5 and s=1.2. Using the Bayes' theorem under a non-informative prior, the joint posterior PDF of the model parameters is given by (Gelman et al., 2004)

$$p\left(\mu,\sigma^2 \mid \mathbf{x}\right) \propto \sigma^{-n-2} \exp\left(-\frac{1}{2\sigma^2}\left[(n-1)s^2 + n(\bar{x}-\mu)^2\right]\right) \tag{14}$$

where the prior for μ and σ is based on the assumption that μ and σ is uniform on $(\mu,\ \log\sigma)$, or equivalently

$$p\left(\mu,\sigma^2\right) \propto \left(\sigma^2\right)^{-1} \tag{15}$$

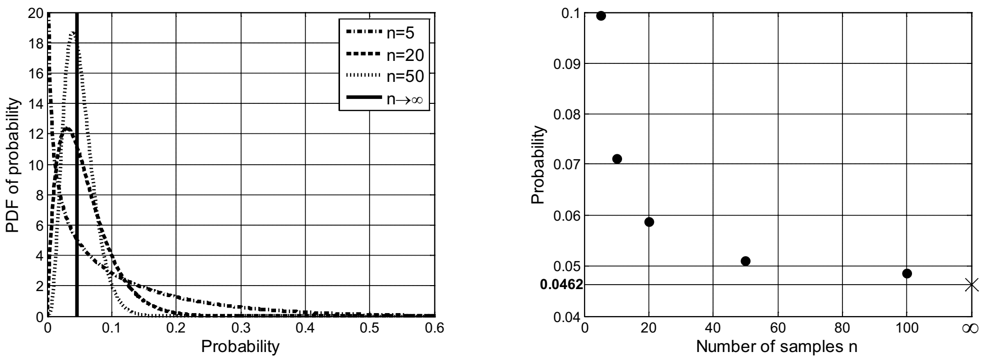
In this case, the marginal posterior distribution for μ and σ takes a standard form, which are t- and χ^2 (chi-square) distribution, respectively, owing to the employment of non-informative priors. If there exists a specific model on the prior for $\boldsymbol{\theta}$ from the previous experience, non-conjugate prior should be employed, which is one of the key feature of Bayesian approach.

As a result of Bayesian reliability analysis, the PDF of probability is given in Figure 4(a), which represents the degree of belief on the probability conditional on the provided data. As the number of data increases, the PDF gets narrower, and converges to a single value that was obtained with aleatory uncertainty. In Table II, the 90% confidence bounds are listed at each number n. In the design practice, natural choice is the upper bound values of the probability for the sake of safety.

In the predictive reliability analysis, predictive samples $\tilde{x}$ are drawn from the following conditional probability distribution at each value of theta obtained from the outer loop:

$$\tilde{x} \mid \mu,\sigma^2 \sim N\left(\mu,\sigma^2\right) \tag{16}$$

In this case, analytic form of the posterior predictive distribution for $\tilde{x}$ is available, which is t-distribution with the mean $\bar{x}$, standard deviation $s\sqrt{1+1/n}$ and $n-1$ degrees of freedom. Finally, response functions are obtained at each point of $\tilde{x}$, from which the probability $P[G<3.5]$ is calculated. The results are listed in Table II, and are also plotted in Figure 4(b). The probability converges to the aleatory value as in the nested case.



(a) nested approach (b) predictive approach

Figure 4. Bayesian reliability analysis results under statistical uncertainty.

Comparing with the upper bound values of nested approach, the predictive values are much lower, hence, are less conservative. Recall here that Eq. (5) represents a marginal distribution of $\tilde{x}$ conditional on x, i.e., average of the conditional prediction over the posterior distribution of $\boldsymbol{\theta}$. This means that the predictive values of the probability agree with the mean of the PDF of the probability in the nested approach. This can be evidenced by comparing the predictive values to the mean values in the Table II. The error magnitudes between the predictive p and the nested mean are found negligibly small.

Table II. Bayesian reliability analysis results of example 1

	Nested approach			Predictive approach (aleatory p =0.0462)	Error (%) $\frac{\lvert p-mean \rvert}{mean}\times 100$
	5%	mean	95%	p	
n =5	0.0024	0.0997	**0.3364**	**0.1016**	1.88
n =10	0.0068	0.0722	**0.2081**	**0.0712**	1.34
n =20	0.0114	0.0576	**0.1374**	**0.0566**	1.71
n =50	0.0207	0.0509	**0.0932**	**0.0519**	1.95

Consider next a function of two variables as follows (Youn and Wang, 2008):

$$g(\mathbf{X}) = 1 - X_1^2 X_2 / 20 \tag{17}$$

where both variables are normally distributed with $X_1 \sim N(2.9, 0.2)$ and $X_2 \sim N(2.8, 0.2)$. In this case, the probability $P[G<0]$ is denoted as reliability. The value is calculated as 0.8412 under the aleatory uncertainty. The results when only X_1 changes to statistical uncertainty with limited number of data while X_2 is still the aleatory uncertainty, are listed in Table III. The same feature is observed as the previous example except that the probability represents the reliability. The confidence bounds of nested approach as well as the predictive values converge in common toward the aleatory value. The lower bound values should be chosen to accommodate conservatism in this case. The predictive values are close to the mean values of the PDF as was found previously.

Table III. Bayesian reliability analysis results of example 2

	Nested approach			(aleatory p =0.8412) Predictive approach	Error (%)
	5%	mean	95%	p	$\frac{\lvert p - mean \rvert}{mean} \times 100$
n =5	**0.5404**	0.7968	0.9561	**0.7916**	0.65
n =10	**0.6541**	0.8179	0.9314	**0.8217**	0.46
n =20	**0.7219**	0.8309	0.9135	**0.8266**	0.52
n =50	**0.7740**	0.8383	0.8931	**0.8360**	0.28

From the results of two examples, it can be concluded that the nested approach provides much more information, i.e., the full PDF of probability at the expense of increased computation. On the other hand, the predictive approach provides just a point estimated value which is the mean of the PDF. Nevertheless, the latter approach also accounts for the degree of statistical uncertainty in the probability prediction in a cost-effective manner, which is more advantageous in the design practice.

4. Metamodel Uncertainty

In this section, metamodel uncertainty is studied for the same two examples, which arises by approximating the response function via Kriging model. In the first example, response values are computed at the equally spaced DOE points to construct metamodel. The trials functions and associated parameters are $\mathbf{f} = [1, x]$ and $\boldsymbol{\beta} = [\beta_1, \beta_2]$ respectively. The unknown parameters are then β_1, β_2, σ, and h which were given in Eq. (6) and (7). The posterior distributions of these parameters are obtained by using 30,000 iterations of MCMC technique. Note that h is implicitly expressed within the correlation matrix $\mathbf{R}$ in the posterior distribution Eq. (11), which justifies the use of the MCMC. The result in the case of 6 points which are from 0 to 15 with the increment 3 is shown in Figure 5. The distributions represent the uncertainty of the parameters due to the employment of metamodel. As was noted previously, common practice for the correlation parameter h was either to assume an appropriate value or to determine by using maximum likelihood method. Whichever method is used, the use of constant h has left difficulties such as arbitrariness of the value or additional computational burden. In this study, h is treated as uncertain parameter which avoids this problem.

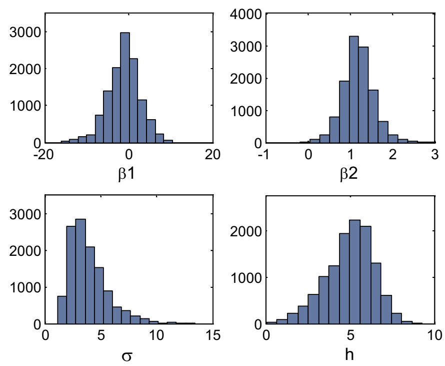


Figure 5. Posterior distribution of metamodel parameters of example 1.

After obtaining the posterior distribution of the parameters as given by Eq. (11) using MCMC method, the predictive sampling is carried out using the drawn samples of the parameters using Eq. (12). The 90% prediction intervals due to the metamodel uncertainty are shown in Figure 6 for the various cases. In the table, upper and lower row represent the results with the number of DOE points m =4 and 6 respectively. The first, second and third column represent the results of two constant h's with the value 0.5 and 5, and the uncertain h, respectively. In the figure, the solid curve is the actual model, the dashed curve is the estimated mean of the metamodel, and the dotted curve is its 90% prediction interval, respectively. From the figure, it is observed that overall, the prediction interval is narrowed down as m is increased. Depending on a choice of the point, however, this may not be the case as can be seen, e.g., at x=5 or 10, which shows wider interval at m=6 than m=4 case. In terms of h, the uncertainty is reduced as the value becomes higher which makes the function more smooth. In particular, in the Figure 6(e) of m=6 and h=5, the metamodel is very close to the actual model. The results of uncertain h indicate that their intervals are a little wider than those of h=5. In fact, this can be interpreted as the average of the predicted samples in terms of the posterior distribution of h. From these observations, it is concluded that the Bayesian approach efficiently quantifies the uncertainty of the Kriging metamodel in the form of prediction interval. The effect of uncertain h is also incorporated during the procedure, which is another advantage since it does not suffer from the arbitrary assignment of constant h.

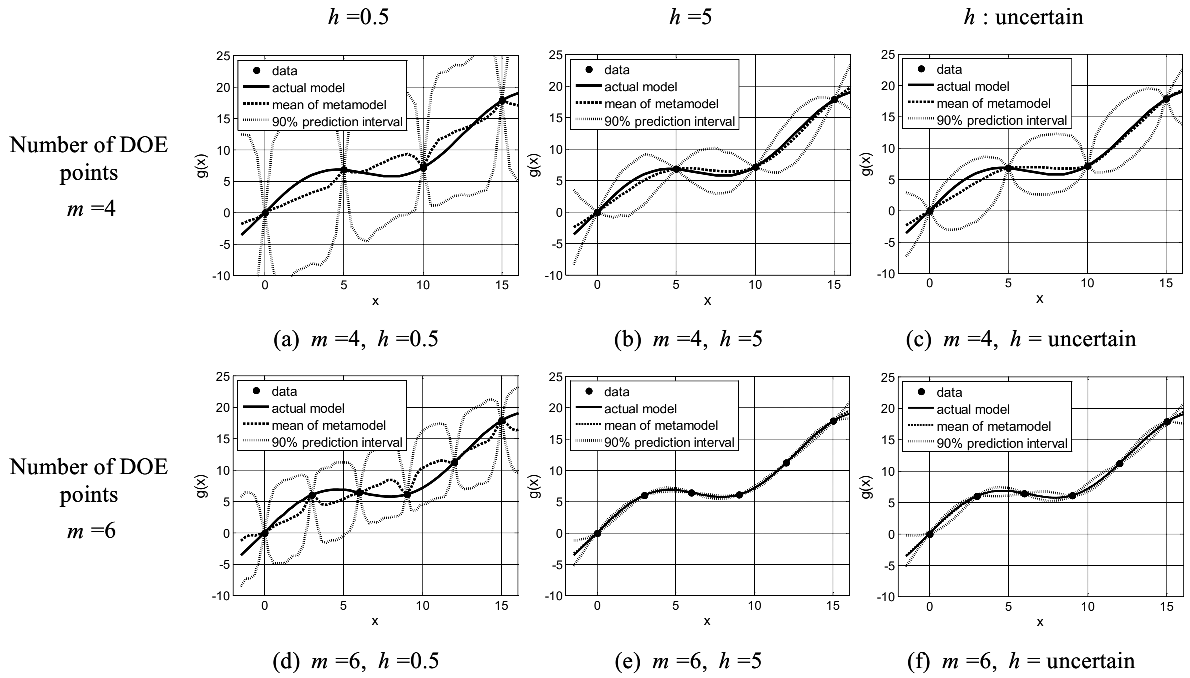


Figure 6. 90% prediction intervals due to the metamodel uncertainty for example 1.

Next, second example is considered, which is two variables function. In this case, the trials functions and associated parameters are $\mathbf{f} = [1, x_1, x_2]$ and $\boldsymbol{\beta} = [\beta_1, \beta_2, \beta_3]$ respectively. Then total of five parameters are considered including σ, h and $\boldsymbol{\beta}$. The posterior distributions of the unknown parameters in the case of

10 DOE points made by Latin Hypercube Sampling (LHS) (Cheng and Druzdzel, 2000) in the range of $\mathbf{x} = [1, 5]$ are shown in Figure 7.

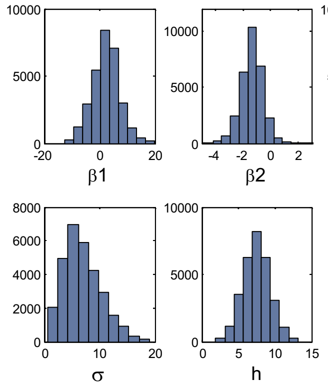


Figure 7. Posterior distribution of metamodel parameters of example 2.

As a result of Bayesian procedure, the upper and lower bounds of the predictive interval incorporating the uncertain h are given by surface plot as shown in Figure 8. In this figure, the dots at the bottom plane denotes the LHS points.

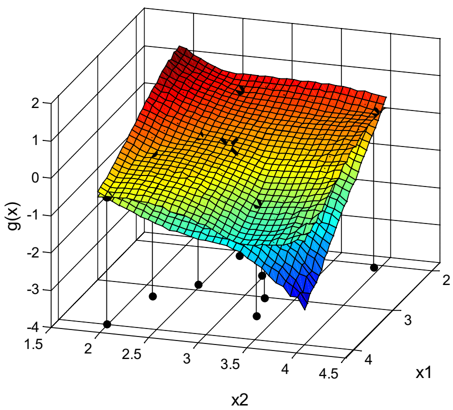


Figure 8. 90% prediction interval due to the metamodel uncertainty for example 2.

5. Integrated Uncertainty

The final goal in this study is to implement reliability analysis under the integrated input variable and metamodel uncertainty. As was noted in the general procedure, the predictive distribution of response function are obtained by drawing samples from conditional distribution of g at each point of the two set of samples which are predictive distribution for $\mathbf{x}$ and posterior distribution for $\boldsymbol{\beta}, \sigma^2, h$. Probability of an event for g then is calculated from the drawn samples.

In the first example, the predicted probabilities $P[G < 3.5]$ under various cases are shown in Figure 9 with the parameter h being uncertain. At each case of DOE points m, probability converges from above to a value as the number of data n increases. As the number of DOE points m increases, the probability also converges, but much more rapidly. After all, the probability converges to the aleatory value 0.0462.

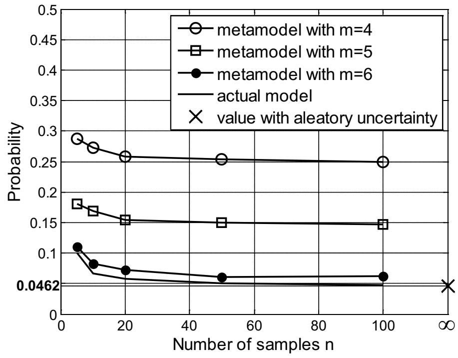


Figure 9. Integrated reliability analysis results of example 1.

In the second example, the probabilities are computed in the same manner under various cases. In this example, LHS is applied to create DOE points of the two input variables for metamodel construction. The results are given in Figure 10. Overall, similar pattern is observed in this case except that the probability denotes reliability. Of noteworthy is that a significant jump of the probability is observed from m =6 to 8 in Figure 10, which indicates that there may be a least necessary number of DOE points to obtain plausible result, and advises that the metamodel be constructed in adaptive manner.

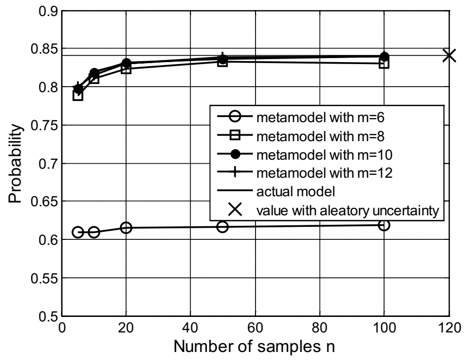


Figure 10. Integrated reliability analysis results of example 2.

6. Engineering Problems

6.1. I-BEAM EXAMPLE

Consider a limit state function that describes the allowable bending stress condition of simply supported I-beam as given in Figure 11 in the following form

$$g(\mathbf{X}) = \frac{27000X_2(X_1 - X_2)}{0.0822 \times 59^3 X_1} - 1170 \tag{18}$$

In the case of aleatory uncertainty, the variables X_1 and X_2 are normally distributed with $N(3000, 25)mm$ and $N(1830, 25)mm$ respectively. In the case of epistemic uncertainty, the X_1 becomes uncertain with only 10 number of data having the same mean and standard deviation.

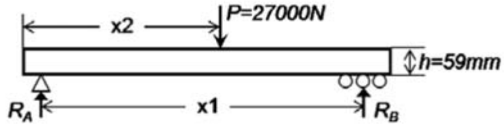


Figure 11. I-beam model illustration.

In the Table IV, the reliability $P[G < 0]$ is calculated for the three cases which are the input uncertainty only, and input plus metamodel uncertainty with m =10 and m =5, respectively. In the metamodel, the DOE points are made by LHS in the domain $2800 \le x_1 \le 3200$ and $1700 \le x_2 \le 2000$. As shown in Table IV, the reliability increases as the associated uncertainties are reduced, with the highest reliability being 0.9570 which is the value under the aleatory assumption.

Table IV. Integrated reliability analysis results of I-beam example

Uncertainty classification	Epistemic with n =10	Aleatory
Input + Metamodel (m =5)	0.9191	0.9388
Input + Metamodel (m =10)	0.9329	0.9494
Input variable only	0.9368	0.9570

6.2. FORTINI'S CLUTCH EXAMPLE

In this problem, the contact angle g is given in terms of four component variables as follows (Greenwood and Chase, 1990):

$$g(\mathbf{X}) = \arccos\left(\frac{X_1 + 1/2(X_2 + X_3)}{X_4 - 1/2(X_2 + X_3)}\right) \tag{19}$$

The variables are defined in Figure 12, and the design requirement is that g should lie between 5° and 9°.

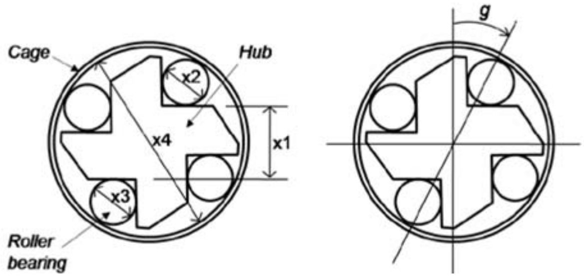


Figure 12. Fortini's clutch model illustration.

In Table V, the distribution types, the means and standard deviations of the variables are listed. The ranges of each variable are also given for the purpose of generating DOE points in the metamodel. In the case of epistemic uncertainty, only X_1 is chosen as uncertain by providing only 10 number of data.

Table V. Distribution definitions of input variable in Fortini's clutch example

Variables	Range of variable for metamodel	Distribution type	Mean (mm)	Std. (mm)
X_1	$54 \le x_1 \le 56$	Beta	55.29	0.0793
X_2	$21 \le x_2 \le 23$	Normal	22.86	0.0043
X_3	$21 \le x_3 \le 23$	Normal	22.86	0.0043
X_4	$100 \le x_4 \le 103$	Rayleigh	101.60	0.0793

The reliability analysis results are listed in Table VI. Again, the values show the same trend as the previous problem.

Table VI. Integrated reliability analysis results of Fortini's clutch example

Uncertainty classification	Epistemic with n =10	Aleatory
Input + Metamodel (m =10)	0.7732	0.7813
Input + Metamodel (m =30)	0.9704	0.9765
Input + Metamodel (m =50)	0.9813	0.9859
Input variable only	0.9917	0.9981

7. Concluding Remarks

In this paper, an integrated reliability analysis procedure is developed based on a Bayesian framework, which can address both the epistemic uncertainties arising from the limited data of input variables and construction of the metamodel. In the input uncertainty, a method of posterior prediction is proposed to efficiently evaluate the failure probability or reliability. From the study, it is found that the probability by the predictive approach corresponds to the mean of PDF of probability in the nested approach. Though the information of full PDF is lost in the predictive approach, it is indeed a more cost-effective method since it catches the influence of the input uncertainties using only a single pass of reliability analysis. In the metamodel uncertainty, the uncertainty due to the metamodel construction at a finite number of DOE points is quantified in the form of predictive interval by employing Gaussian process or Kriging model. During the procedure, the uncertainty of the correlation parameter is also incorporated, which has been taken as a constant value in most of the previous engineering literatures. A general procedure to integrate these uncertainties is presented based on a Bayesian framework. For an efficient evaluation of the posterior distribution in the procedure, Markov Chain Monte Carlo (MCMC) method is employed. The feasibility of the proposed method is validated by the mathematical and engineering problems.

Acknowledgements

This research was supported by Basic Science Research Program Through the National Research Foundation of Korea (NRF) funded by the Ministry of Education, Science and Technology (2009-0081438, 2008-02-010)

References

Andrieu, C., N. de Freitas, A. Doucet, and M. Jordan, An introduction to MCMC for Machine Learning. *Machine Learning*, 50:5-43, 2003.

Cheng, J. and M. J. Druzdzel. Latin hypercube sampling in Bayesian networks. *Proceedings of the 13th International Florida, Artificial Intelligence Research Symposium Conference (FLAIRS-2000)*, Orlando, Florida, May, 2000.

Cruse, T. A. and J. M. Brown. Confidence interval simulation for systems of random variables. *ASME Journal of Engineering for Gas Turbines and Power*, 129:836-842, 2007.

Currin, C., T. Mitchell, M. Morris, and D. Ylvisaker. Bayesian prediction of deterministic functions, with applications to the design and analysis of computer experiments, *Journal of the American Statistical Association*, 86:953-963, 1991.

Du, L., K. K. Choi, and B. D. Youn. Inverse possibility analysis method for possibility based design optimization. *AIAA Journal*, 44:2682-2690, 2006.

Eduard, H., K. Martina, K. H. Bernard, P. Jörg, and W. Martin. An approximate epistemic uncertainty analysis approach in the presence of epistemic and aleatory uncertainties. *Reliability Engineering and System*, 77:229-238, 2002.

Etman, L. F. P. Design and Analysis of Computer Experiments: The Method of Sacks et al. Engineering Mechanics report WFW 94.098, Eindhoven University of Technology, Netherlands, 1994.

Gelman, A., J. B. Carlim, H. S. Stern, and D. B. Rubin. *Bayesian Data Analysis (second ed.)*. New York, Chapman & Hall/CRC, 2004.

Greenwood, W. and K. Chase. Root sum squares tolerance analysis with nonlinear problems. *ASME Journal of Engineering for Industry*, 112:232-235, 1990.

Gunawan, S. and P. Y. Papalambros. A Bayesian approach to reliability-based optimization with incomplete information. *The 2006 ASME International Design Engineering Technical Conferences & Computers and Information in Engineering Conference*, Philadelphia, Pennsylvania, September 10-13, 2006.

Haldar, A. and S. Mahadevan. *Probability, Reliability, and Statistical Methods in Engineering Design*. New York, John Wiley & Sons, 2000.

Handcock, M. S. and M. L. Stein. A Bayesian analysis of Kriging. *Technometrics*, 35:403-410, 1993.

Helton, J. C., J. D. Johnson, W. L. Oberkampf, and C. B. Storlie. A sampling-based computational strategy for the representation of epistemic uncertainty in model predictions with evidence theory. *Computer Methods in Applied Mechanics and Enginerring*, 196:3980-3998, 2007.

Kennedy, M. C. and A. O'Hagan. Bayesian calibration of computer models (with discussion). *Journal of the Royal Statistical Society, Series B*, 63:425-464, 2001.

Kiureghian, A. D. and O. Ditlevsen. Aleatory or epistemic? does it matter?. *Structural Safety*, 31:105-112, 2009.

Lee, I., K. K. Choi, L. Du, and D. Gorsich. Inverse analysis method using MPP-based dimension reduction for reliability-based design optimization of nonlinear and multi-dimensional systems, *Computer Methods in Applied Mechanics and Enginerring*, 198:14-27, 2008.

Mahadevan, S. and R. Rebba. Validation of reliability computational models using Bayes networks. *Reliability Engineering & System Safety*, 87:223-232, 2005.

Mourelatos, Z. P. and J. Zhou. Reliability estimation and design with insufficient data based on possibility theory. *AIAA Journal*, 43:1696-1705, 2005.

Mourelatos, Z. P. and J. Zhou. A design optimization method using evidence theory. *ASME Journal of Mechanical Design*, 128:901-908, 2006.

Noor, A.K. Perspectives on nondeterministic approaches. In E. Nikolaidis, D. M. Ghiocel, and S. Singhal, editors, *Engineering Design Reliability Handbook*. New York, CRC Press, 2005.

O'Hagan, A. Bayesian analysis of computer code outputs: a tutorial. *Reliability Engineering and System Safety*, 91:1290-1300, 2006.

Qiu, Z., Y. Ma, and X. Wang. Comparison between non-probabilistic interval analysis method and probabilistic approach in static response problem of structures with uncertain-but-bounded parameters. *Computer Methods in Applied Mechanics and Enginerring*, 20:279-290, 2004.

Rackwitz, R. X. Reliability analysis—a review and some perspectives. *Structural Safety*, 23:365-395, 2001.

Rahman, S. and H. Xu. A Univariate Dimension-reduction method for multi-dimensional integration in stochastic mechanics. *Probabilistic Engineering Mechanics*, 19:393-408, 2004.

Sasena, M. J., P. Papalambros, and P. Goovaerts. Exploration of metamodeling sampling criteria for constrained global optimization. *Engineering Optimization*, 34:263-278, 2002.

Won, J. H., C. H. Choi, and J. H. Choi. Improving the dimension reduction method (DRM) in the uncertainty analysis and application to the reliability based design optimization. *Journal of Mechanical Science and Technology*, 23:1249-1260, 2009.

Youn, B. D. and P. Wang. Bayesian reliability-based design optimization using eigenvector dimension reduction (EDR) method. *Structural and Multidisciplinary Optimization*, 36:107-123, 2008.

Zhang, R. and S. Mahadevan. Integration of computation and testing for reliability estimation. *Reliability Engineering & System Safety*, 74:13-21, 2001.