Rare failure event analysis of structures under mixed uncertainties

Pengfei Wei
Associate Professor, School of Mechanics, Civil Engineering and Architecture, Northwestern Polytechnical University, Xi'an, China

Humboldt Research Fellow, Institute for Risk and Reliability, Leibniz Universität Hannover, Hannover, Germany

Sifeng Bi
Humboldt Research Fellow, Institute for Risk and Reliability, Leibniz Universität Hannover, Hannover, Germany

Yi Zhang
Associate Professor, Department of Civil Engineering, Tsinghua University, Beijing, China

Michael Beer
Professor, Institute for Risk and Reliability, Leibniz Universität Hannover, Hannover, Germany

Institute for Risk and Uncertainty, University of Liverpool, Liverpool, UK

School of Civil Engineering & Shanghai Institute of Disaster Prevention and Relief, Tongji University, Shanghai, China

ABSTRACT: Two challenges may exist in the reliability analysis of highly reliable structures in, e.g., aerospace engineering. The first one is that, the failure probability may be extremely small (typically, smaller than 1e-6), which commonly prevents us from generating accurate estimation with acceptable computational costs by using the available methods. The second one is that, the available information for the input variables may be subject to incompleteness (e.g., sparse data) and/or imprecision (e.g., measuring error), which, makes it impossible to generate precise probability models for the input variables. To address the above two challenges, this work proposes two effective algorithms based on combining the sampling techniques (i.e., extended Monte Carlo simulation and subset simulation), active learning techniques and high-dimensional model representation decomposition. The proposed methods can effectively estimate the failure probability function w.r.t. the uncertain distribution parameters of the input variables with small number of training samples even when the failure event is extremely rare. A numerical test example is introduced to illustrate the proposed methods.

In structural engineering, evaluating the reliability of structures with the consideration of uncertainties has been a common and necessary task. However, two challenges commonly prevent us from learning the true values of the reliability. The first one is due to limited computational resources. With the increment of complexity of structural systems, large-scale computational models need to be developed to simulate their behavior, making each call of the computational models being time-consuming. Further, due to the requirement of high reliability for complex products such as aircraft, the structural failure events are commonly very rare. Both the time-consuming models and rare failure events lead to high computational cost when performing the reliability analysis with the traditional methods such as subset simulation (SS) (Au and Beck, 2001). The second challenge results from the limited information of the structural inputs such

as dimension sizes, material properties and loads. For correctly learning the degree of structural reliability, the amount of data for input parameters should also be large and accurate enough so as to learn their true probability distribution densities (PDFs). However, in real applications especially the design stage, the available information is commonly incomplete and/or imprecise. In this situation, the epistemic uncertainty emerges and prevents us from generating the true PDFs. The mixed uncertainties are commonly characterized by imprecise probability models such as *p*-box model (Sun et al., 2012), fuzzy probability model (Beer et al., 2013) and evidence theory (Sentz and Ferson, 2002). This work aims at developing effective methods for meeting the above two challenges based on the parameterized imprecise probability models.

Reliability analysis based on imprecise probability models has attracted wide attentions, and many traditional reliability analysis methods, such as first-order reliability method (FORM) (Zhang et al., 2015) and SS (Alvarez et al., 2018), have been extended to meet this requirement. There are also specific methods developed for uncertainty propagation of imprecise probability models, and the two popular ones are the extended Monte Carlo simulation (EMCS) (Wei et al., 2014) and Interval Monte Carlo simulation (IMCS) (Zhang et al., 2013). The EMCS is a non-intrusive sampling technique for efficiently propagating the imprecise probability models through any black-box computational models, while the IMCS method is an intrusive method which needs interval finite element analysis for each interval sample, thus, generally, the EMCS method has wider application. However, the current version of EMCS method cannot meet the two above-mentioned challenges which are common in practical applications.

In this work, we develop efficient algorithms for analyzing rare failure events subjected to parameterized imprecise probability models, by improving the EMCS methods with high-dimensional model representation (HDMR) decomposition, SS and active learning procedures.

The proposed methods can substantially improve the accuracy and efficiency of the EMCS method by adaptively generating a set of intermediate failure surfaces, and present the errors of estimations. The proposed method is demonstrated by a numerical test example.

1. A BRIEF REVIEW OF THE EMCS METHOD

Before the introduction of the methods, we firstly introduce some terminologies. Let $y = g(x)$ indicate the limit state function, where $x = (x_1, x_2, \ldots, x_n)^T$ refers to the *n*-dimensional uncertain input variables, and y indicates the performance variable of structure. Commonly, x are random variables characterized by precise probability models due to aleatory uncertainty. However, due to the incompleteness and/or imprecision of the available data, their distribution functions can not be precisely generated, and in this situation, the imprecise probability models such as p-box model and fuzzy probability model, are commonly introduced. In this work, we only consider the parametric imprecise probability models, which means that the epistemic uncertainty of one input variable is characterized by its distribution parameters. Let $f_X(x; \theta)$ denote the joint PDF of x, where $\theta = (\theta_1, \theta_2, \ldots, \theta_d)^T$ refers to the uncertain distribution parameters. If the fuzzy probability model is used, θ is a fuzzy vector; while the parametric p-box model is applied, it is characterized by interval/convex models; if the second-order probability model is utilized, θ is a subjective random vector. Let $f_\Theta(\theta)$ denote the attributed joint PDF of θ, and it is assumed that the distribution parameters are independent, i.e., $f_\Theta(\theta) = \prod_{i=1}^{d} f_{\Theta_i}(\theta_i)$, where $f_{\Theta_i}(\theta_i)$ refers to the marginal PDF of θ_i. One should note that the PDF of θ is just attributed for developing the new methods, but does not mean that θ have specific probability distributions.

The failure event is indicated by $y<0$. Thus, the failure domain is represented by $F = \{x : g(x) < 0\}$, and the indicator function of F is defined by

$$I_F(x) = \begin{cases} 1 & \text{if } x \in F \\ 0 & \text{else} \end{cases} \tag{1}$$

Then the failure probability function $P_f(\theta)$ is expressed as:

$$P_f(\theta) = \int I_F(x) f_X(x;\theta) \mathrm{d}x \tag{2}$$

The EMCS method for estimating the above failure probability is inspired by the importance sampling, and the EMCS estimator is

$$P_f(\theta) = \int I_F(x) \frac{f_X(x;\theta)}{f_X(x;\theta^*)} f_X(x;\theta^*) \mathrm{d}x$$

$$\hat{=} \frac{1}{N} \sum_{k=1}^{N} I_F(x^{(k)}) \frac{f_X(x^{(k)};\theta)}{f_X(x^{(k)};\theta^*)} \tag{3}$$

where θ^* refers to a pre-specified values for generating the sampling PDF $f_X(x;\theta^*)$ and $x^{(k)}$ ($k = 1,2,\ldots,N$) are a set of MCS samples generated from $f_X(x;\theta^*)$. The EMCS estimator in Eq. (3) is simple and easy to implement, but it has several disadvantages. For rare failure events, tremendous function calls are needed, making the implementation of the method computationally impractical. For problems with many uncertain distribution parameters (e.g., $d>6$), the variance of the estimator may be too large. In the next section, we introduce several techniques to overcome the above shortcomings.

2. THE METHODS

The estimators in Eq.(3) is unbiased, but for problems with many uncertain distribution parameters, the variance of the estimator may be too large, due to the large diversity of the samples of $I_F(x) f_X(x;\theta) / f_X(x;\theta^*)$. However, the failure probability function is generally governed by low-order component functions, and high order (≥ 3) interactions are usually non-influential. Thus,

we firstly propose to approximate the failure probability function with HDMR decomposition. Generally, there are two kinds of HDMR decomposition, i.e., the cut-HDMR (Li and Wang, 2001) and RS (Random sampling)-HDMR (Li et al., 2002). In this work, the cut-HDMR is utilized.

2.1. The cut-HDMR decomposition

The HDMR decomposition of the failure probability function is expressed as (Li and Wang, 2001):

$$P_f(\theta) = P_{f0} + \sum_{i=1}^{d} P_{fi}(\theta_i) + \sum_{1 \leq i < j \leq d} P_{fij}(\theta_{ij}) \\ + \ldots + P_{f12\ldots d}(\theta) \tag{4}$$

where θ_{ij} indicates the 2-dimensional vector consists of θ_i and θ_j, for cut-HDMR decomposition, the component functions on the right side of Eq. (4) are formulated as:

$$\begin{cases} P_{f\text{cut}0} = P_f(\theta^*) \\ P_{f\text{cut}i}(\theta_i) = P_f(\theta_i, \theta^*_{-i}) - P_{f\text{cut}0} \\ P_{f\text{cut}ij}(\theta_{ij}) = P_f(\theta_{ij}, \theta^*_{-ij}) - P_{f\text{cut}i}(\theta_i) - P_{f\text{cut}j}(\theta_j) - P_{f\text{cut}0} \\ \ldots \end{cases}$$

$$(5)$$

where θ_{-i} indicates the (d-1)-dimensional vector containing all the components of θ but θ_i, and θ^* refers to the expansion points around which one would like to decompose the failure probability function. In this work, we will estimate the HDMR component functions with advanced EMCS method, thus θ^* in Eq. (3) and Eq.(5) should be the same. Commonly, the former two order component functions are accurately enough for approximating the failure probability function, thus we need only to estimate the constant, first order and second order components in Eq. (5).

Given a set of samples $x^{(k)}$ ($k = 1,2,\ldots,N$) generated from $f_X(x;\theta^*)$, the EMCS estimators

for the former two order cut-HDMR component functions can be derived as:

$$\begin{cases} \hat{P}_{fcut0} = \dfrac{1}{N}\sum_{j=1}^{N} y^{(k)} \\ \hat{P}_{fcuti}(\theta_i) = \dfrac{1}{N}\sum_{k=1}^{N} y^{(k)} r_i\left(x^{(k)};\theta_i,\theta_{-i}^*\right) \\ \hat{P}_{fcutij}(\theta_{ij}) = \dfrac{1}{N}\sum_{k=1}^{N} y^{(k)} r_{ij}\left(x^{(k)};\theta_{ij},\theta_{-ij}^*\right) \end{cases} \quad (6)$$

where

$$\begin{cases} r_i\left(x^{(k)};\theta_i,\theta^*\right) = \dfrac{f\left(x^{(k)};\theta_i,\theta_{-i}^*\right)}{f\left(x^{(k)};\theta^*\right)} - 1 \\ r_{ij}\left(x^{(k)};\theta_{ij},\theta^*\right) = \dfrac{f\left(x^{(k)};\theta_{ij},\theta_{-ij}^*\right)}{f\left(x^{(k)};\theta^*\right)} - \dfrac{f\left(x^{(k)};\theta_i,\theta_{-i}^*\right)}{f\left(x^{(k)};\theta^*\right)} \\ \qquad\qquad - \dfrac{f\left(x^{(k)};\theta_i,\theta_{-i}^*\right)}{f\left(x^{(k)};\theta^*\right)} + 1 \end{cases}$$
$$(7)$$

Obviously, the above estimators are unbiased, and have smaller variances since only one or two distribution parameters vary. The variances of the estimators in Eq. (7) can be easily derived, and we don't give more details.

Both the component functions in Eq.(5) and their estimators in Eq. (6) possess the vanishing property, which means that, if any parameter θ_{i_a} contained in the component $E_{cutyi_1i_2\ldots,i_s}\left(\theta_{i_1i_2\ldots i_s}\right)$ is fixed at the expansion point $\theta_{i_a}^*$, then this component function equals to zero. The computational cost of the estimators in Eq.(6) is still unacceptable for real applications. In the next subsection, we improve the EMCS estimators with SS technique.

2.2. Extended subset simulation
The extended subset simulation (ESS) for estimating the component functions in Eq. (5) is directly derived from the classical SS procedure. Suppose now we need to estimate the constant

component P_{fcut0}. Let p_0 denotes the intermediate probability (commonly specified as 0.1~0.3), and N indicates the size of samples used for generating the intermediate failure thresholds. Based on this setting, we can perform the classical SS procedure for estimating P_{fcut0}. Suppose we obtain m intermediate failure surfaces, and the corresponding failure thresholds are denoted as $b_1 > b_2 > \ldots > b_m = 0$, and the Monte Carlo (MC) or Markov Chain Monte Carlo (MCMC) sample set for specifying the jth failure threshold is denoted as $S_j = \left\{\left(x_j^{(k)}, y_j^{(k)}\right) : k = 1, 2, \ldots, N\right\}$, then the ESS estimators for the former two order component functions are derived as:

$$\begin{cases} \hat{P}_{fcut0} = p_0^{m-1}\dfrac{1}{N}\sum_{k=1}^{m} I_{F_m}\left(x_m^{(k)}\right) \\ \hat{P}_{fcuti}(\theta_i) = p_0^{m-1}\left[\dfrac{1}{N}\sum_{k=1}^{N} I_{F_m}\left(x_m^{(k)}\right) r_i\left(x_m^{(k)};\theta_i,\theta^*\right)\right] \\ \hat{P}_{fcuti}(\theta_{ij}) = p_0^{m-1}\left[\dfrac{1}{N}\sum_{k=1}^{N} I_{F_m}\left(x_m^{(k)}\right) r_{ij}\left(x_m^{(k)};\theta_{ij},\theta^*\right)\right] \end{cases}$$
$$(8)$$

The statistical properties of the above estimators are analogous to the classical SS estimator, one can refer to Au and Beck (2001) or our subsequent full-length paper for details. Obviously, the estimators in Eq. (8) also possess the vanishing property. The total number of function calls of the above ESS procedure is mN, which is the same as the classical SS procedure. This cost is still too high for complex structures. In the next subsection, we further improve the ESS method with active learning procedure.

2.3. The AK-ESS methods
The active learning procedure, e.g., the AK-MCS method (Echard et al., 2011), has drawn the most attentions around these years in the area of structural reliability analysis due to their high efficiency. We use it here to further improve the ESS method, and we denote the new method as "AK-ESS". The steps for implementing the method is briefly summarized as follows.

Step 1: Generate a set of MC samples $\mathbf{S}_1 = \left\{ \boldsymbol{x}_1^{(k)} : k = 1, 2, \ldots, N \right\}$ from $f_X\left(\boldsymbol{x}; \boldsymbol{\theta}^* \right)$. Randomly generate N_0 (e.g., N_0=12) samples from $\mathbf{S}_1$, compute the corresponding limit state function values, and attribute these N_0 initial training data to the set $\mathbf{S}_T$. Let q=1.

Step 2: Train a Kriging surrogate model with $\mathbf{S}_T$.

Step 3: Predict the limit state function values for each non-training sample, and compute or update the value of b_q for the qth intermediate failure domain. If b_q <0, let b_q =0.

Step 4: Compute the U value for each non-training sample by $U_k = \left| \mu_g\left(\boldsymbol{x}^{(k)} \right) - b_q \right| \Big/ \sigma_g\left(\boldsymbol{x}^{(k)} \right)$, where $\mu_g\left(\boldsymbol{x}^{(k)} \right)$ and $\sigma_g\left(\boldsymbol{x}^{(k)} \right)$ indicate the Kriging prediction at the point $\boldsymbol{x}^{(k)}$ as well as the corresponding mean square error (MSE). If $\min U_k > U_0$, go to **Step 5**; else, find the non-training sample in $\mathbf{S}_q - \mathbf{S}_T$ with the minimum U value, compute the corresponding limit state function value, attribute this point to the training data set $\mathbf{S}_T$, and turn to **Step 2**.

Step 5: If b_q =0, turn to Step 6. If b_q>0, let q=q+1, and generate a conditional sample set $\mathbf{S}_q = \left\{ \boldsymbol{x}_q^{(k)} : k = 1, 2, \ldots, N \right\}$ following conditional PDF $f_X\left(\boldsymbol{x} \mid F_{q-1}, \boldsymbol{\theta}^* \right)$ based on any MCMC algorithm (Papaioannou et al., 2015) by calling the Kriging surrogate model. Specify the value of b_q. If b_q<0, let b_q =0, and compute the U value for each non-training sample contained in $\mathbf{S}_q - \mathbf{S}_T$. If $\min U_k > U_0$, turn to the beginning of this step, else find the non-training sample with the minimum U value, compute the corresponding limit state function value, add this point to the training data set, and turn to **Step 2**.

Step 6: Let m=q, and estimate the cut-HDMR component functions by the estimators in Eq. (8) with the sign of the limit state function value for each input sample is estimated by the Kriging surrogate models.

In Step 4 and 5, the U_0 is commonly set to be 2, which is similar to the classical AK-MCS procedure, and the principle $\min U_k > U_0$ promise that the probability of correctly predicting the sign of the limit state function for each non-training sample is no less than $\Phi(2)$, where $\Phi(\bullet)$ is the distribution function of standard normal distribution.

In the above procedure, the intermediate probability p_0 can be set to a very small one (e.g., 1e-3) by setting the sample size N to a large one (e.g. 1e5). Thus the number of intermediate failure surfaces is much less than the ESS procedure. Further, each intermediate failure surface is approximated by the surrogate model produced by the active learning procedure, thus only small number of function calls are needed. As the sign of the limit state function for each sample in each intermediate failure domain is promised by the principle $\min U_k > U_0$, the AK-ESS procedure is also accurate.

2.4. Test example

Consider a two-dimensional toy test example with limit state function represented by:

$$y = g(\boldsymbol{x}) = 1 - \frac{(x_1 - 1)^2}{25} - \frac{(x_2 - 1)^3}{36} \qquad (9)$$

where x_1 and x_2 are two normal variables with mean parameters (denoted as μ_1 and μ_2) bounded in [-0.2, 0.2] and standard deviations (SDs) (denoted as σ_1 and σ_2) bounded in [0.8, 1.2].

For ESS procedure, p_0 and N are set to be 0.1 and 1e4 respectively, while for AK-ESS procedure, these two parameters are set to be 1e-3 and 1e5 respectively. The ESS procedure produces five intermediate failure surfaces with failure thresholds being 0.8285, 0.5895, 0.3573, 0.1175 and 0, thus the total number of function calls is 5e4. The AK-ESS procedure adaptively produces two intermediate failure surfaces with failure thresholds being 0.3488 and 0, and the total number of function calls for learning these two failure surfaces are 22 and 11 respectively, thus

the total number of function calls is 33, much less than that of the ESS method. The constant cut-HDMR term estimated by the ESS and AK-ESS methods are 3.928e-5 and 4.029e-5, and the SDs of these two estimates are 2.4e-6 and 4.1e-6 respectively. The reference solution estimated by the crude MCS procedure is 4.230e-05 with SD being 2.06e-06. Thus, the constant components are accurately estimated by both methods.

The first-order component functions are then estimated by the ESS and AK-ESS methods with the same sets of samples used for computing the constant components, and the results are shown in Figure 1, together with the results computed by crude MCS procedure for comparison. As can be seen, all the four first order component functions are accurately estimated by both methods.
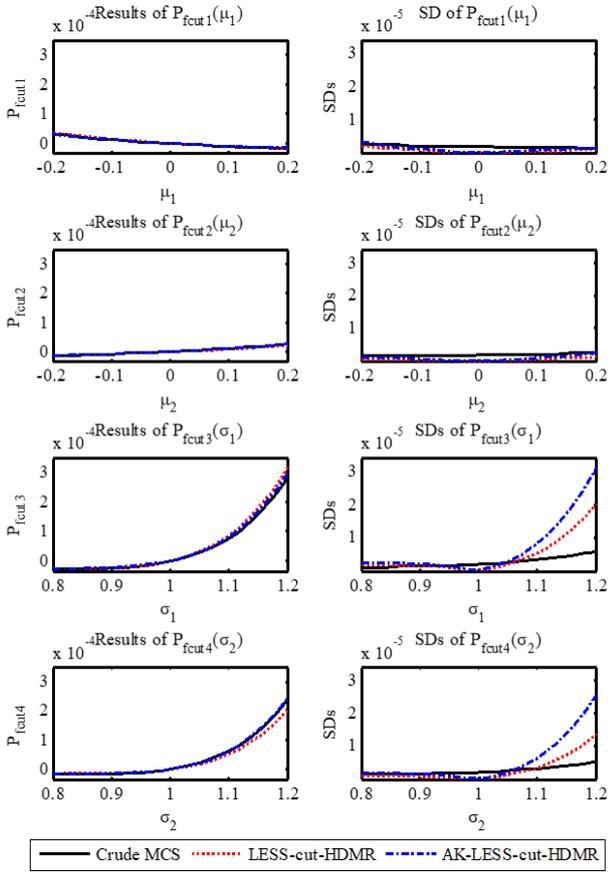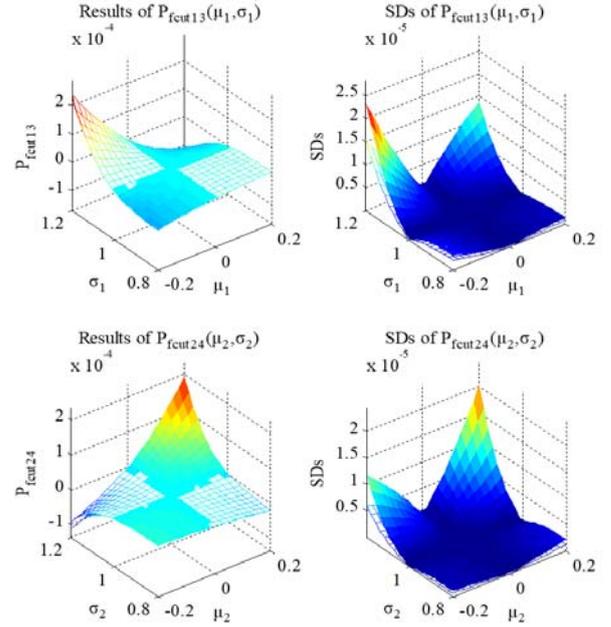

Figure 2: Results of the two influential second order component functions, where the surfaces with mesh indicate the results obtained by ESS, and those without mesh refer to the results obtained by AK-ESS.

It is also shown that the component functions of μ_1 and μ_2 are not influential, and for formulating the failure probability functions, these two components can be neglected. Qualitative measure of the importance of each component function can be achieved using global sensitivity analysis technique (Wei et al., 2015), which will be comprehensively investigated in the full-length paper.

There are totally six second order component functions, but only two of them are influential, which are displayed in Figure 2. It is shown that these two component functions computed by the two proposed methods match well, and their SDs are small enough. Since the estimators are asymptotically unbiased, it is believed both methods correctly estimate the second order component functions. Then the failure probability function can be approximated by

$$\hat{P}_f(\boldsymbol{\theta}) = \hat{P}_{fcut0} + \hat{P}_{fcut3}(\sigma_1) + \hat{P}_{fcut4}(\sigma_2) \\ + \hat{P}_{fcut12}(\mu_1, \sigma_1) + \hat{P}_{fcut34}(\mu_2, \sigma_2) \tag{10}$$

Figure 1: Results of the first order component functions.

3. CONCLUSIONS

This paper presents two efficient algorithms for estimating the structural failure probability function for structures with input variables being characterized by any parameterized imprecise probability models. The proposed methods combine the advantages of EMCS, cut-HDMR, SS and AK-MCS procedures, and are especially useful for structures with rare failure events. These methods are shown to have excellent local performance. In the future work, these methods will be extended to the non-parameterized imprecise probability models. The proposed methods are significantly useful for estimating the degree of reliability especially at the design stage when the available information is sufficient. The proposed methods also provide a solid foundation for further research on, e.g., sensitivity analysis and uncertainty-based design optimization, under mixed uncertain environments.

4. REFERENCES

Alvarez, D. A., Uribe, F., Hurtado, J. E. (2018). "Estimation of the lower and upper bounds on the probability of failure using subset simulation and random set theory." *Mechanical Systems and Signal Processing*, 100, 782-801

Au, S. K., and Beck, J. L. (2001). "Estimation of small failure probabilities in high dimensions by subset simulation." *Probabilistic Engineering Mechanics*, 16, 263-277

Beer, M., Ferson, S., Kreinovich, V. (2013). "Imprecise probabilities in engineering analyses." *Mechanical Systems and Signal Processing*, 37(1) 4-29

Echard, B., Gayton, N., Lemaire, M. (2011). "AK-MCS: an active learning reliability method combining kriging and Monte Carlo simulation. " *Structural Safety*, 33(2), 145-154.

Li, G. Y., Wang, S. W. (2001). "High dimensional model representations generated from low dimensional data samples. i. mp-cut-hdmr." *Journal of Mathematical Chemistry,* 30(1), 1-30.

Li, G.Y., Wang S.W., Rabitz H. (2002). "Practical approaches to construct RS-HDMR component functions." *Journal of Physical Chemistry A*, 106(37), 8721-8733.

Papaioannou, I., Betz, W., Zwirglmaier, K., Straub, D.(2015). "MCMC algorithms for subset simulation." *Probabilistic Engineering Mechanics*, 41, 89-103.

Sentz, K., Ferson, S. (2002). "Combination of evidence in Dempster-Shafer theory." Livermore, California: Sandia National Laboratories

Sun, S., Fu, G., Djordjević, S., Khu, S. T. (2012). "Separating aleatory and epistemic uncertainties: Probabilistic sewer flooding evaluation using probability box." Journal of hydrology, 420, 360-372.

Wei, P., Lu, Z., Song, J. (2014). "Extended Monte Carlo simulation for parametric global sensitivity analysis and optimization." AIAA Journal, 52(4), 867-878.

Wei, P., Lu, Z., Song, J. (2015). "Variable importance analysis: a comprehensive review." Reliability Engineering & System Safety, 142, 399-432.

Zhang, H., Dai, H., Beer, M., Wang, W. (2013). "Structural reliability analysis on the basis of small samples: an interval quasi-Monte Carlo method." Mechanical Systems and Signal Processing, 37(1-2), 137-151

Zhang, Z., Jiang, C., Wang, G. G., Han, X. (2015). "First and second order approximate reliability analysis methods using evidence theory." Reliability Engineering & System Safety, 137, 40-49