The Seventh Asian-Pacific Symposium on Structural Reliability and Its Applications (APSSRA2020)
October 4–7 2020, Tokyo, Japan
T. Takada, I. Yoshida & T. Itoi (editors)

Bayesian polynomial chaos expansion for structural reliability analysis

Y. Zhou[1], ZZ. Lu[2]

[1] *School of Aeronautics, Northwestern Polytechnical University, Xi'an, Shaanxi, China. Email: 494519858@mail.nwpu.edu.cn*

[2] *School of Aeronautics, Northwestern Polytechnical University, Xi'an, Shaanxi, China. Email: zhenzhoulu@nwpu.edu.cn*

Abstract: Polynomial chaos expansion (PCE) technique has been widely used to replace expensive simulation models in order to reduce computing costs. In this paper, we develop a new structural reliability analysis method based on PCE. To address that conventional regression method of sparse PCE cannot provide the surrogate error measure (predictive variance) which is employed to improve the sampling performance in reliability analysis, variational Bayesian inference, with state of the art performance in robust regression, is employed to build sparse PCE. Then, an active learning function is employed as the guideline to adaptively select new training points. This active learning procedure stops when the accuracy of the reliability estimate reaches a specific target. To assess the performance of the proposed method, it is compared detailedly with two well-established AK-MCS and APCK-MCS methods. The results show that the proposed method is superior to these two well-established in terms of efficiency and accuracy.

Keywords: Polynomial chaos expansion, structural reliability analysis, Variational Bayesian inference

1. Introduction

Reliability analysis aims at evaluating the safety level of structures or system. For time-demanding and complex performance functions of structural responses, the evaluation of small failure probability with fewer model evaluations is a challenging problem in engineering application. First-order Reliability method (FORM) and Second-order Reliability method (SORM) are two classical methods for estimating the failure probability (Rackwitz 1978, Der Kiureghian et al. 1991). These two methods approximate the performance function at the most probable point (MPP) in the failure domain with first-order and second-order polynomial, on which the failure probability is estimated (Konakli et al. 2016). SORM captures the nonlinearity near the verified point and partially improves the prediction accuracy of the failure probability of FORM. However, SORM is impossible to capture the local property of the performance function around the design checking point. For performance functions with highly nonlinear or multimodal behaviors, the accuracy of FORM and SORM may not be acceptable.

Finite element model is widely used for many structural systems with expensively time-consuming performance functions. The cost of a large number of repeated simulations is extremely expensive. In order to improve computational efficiency and avoid a large number of expensive finite element simulations, surrogate modeling techniques which allow one to develop a cheap-to-evaluated surrogate from a limited collection of model evaluations are widely used. Building a surrogate relies on the evaluation of the original model at a set of points in the input space. The efficiency of a surrogate modeling technique depends on its ability to provide sufficiently accurate representations of the exact performance function by using relatively small size of training points. This can be particularly challenging in cases that determining the tails of the distribution of model response with high accuracy is important for estimating the small failure probabilities required in reliability analysis. During the past decades, various surrogate modeling techniques have been proposed for reliability analysis. Kriging (also known as Gaussian process (Sack et al. 1989)) is one of the most popular, it is easy to implement to assess the failure probability of complex structures, since it is based on the hypothesis that the weight of the failure probability is located in the vicinity of the unique most possible failure point. Several reliability analysis methods based on the Kriging include the efficient global reliability analysis (EGRA) method (Bichon et al. 2008), the active learning reliability method combining Kriging and Monte Carlo simulation (AK-MCS) (Echard et al. 2011), active Kriging combined with subset simulation (AK-SS) (Huang et al. 2016) or (modified) importance sampling (AK-IS and AK-MIS) (Echard et al. 2013, Yun et al. 2018), Kriging-based quasi-optimal importance sampling (Meta-IS) (Dubourg et al. 2013), the global sensitivity analysis enhanced Kriging (GSAS) [Hu and Mahadevan 2016]. By selecting more training points near the limit state surface according to several active learning functions such as the U function (Echard et al. 2011) and the expected feasibility function (EFF) (Bichon et al. 2008), the accuracy and efficiency of reliability estimate can be improved. Schöbi et.al (2015) developed a new surrogate modeling technique called Polynomial-chaos-Kriging (PC-Kriging). It employs orthonormal multivariate polynomials as drift term of Kriging to convey the data mean trend. A modified version of the AK-MCS method, i.e., active PC-Kriging Monte-Carlo simulation (APCK-MCS), was then used to enhance the accuracy of reliability estimate or quantile estimate (Schöbi et al. 2016). Because of the lack of a prior knowledge about the output, leave-one-out cross validation (LOOCV) is employed as an object function to select optimal regression functions for PC-Kriging from the set of orthonormal multivariate polynomials. In fact, LOOCV is an approximation method used to estimate the actual generalization error. The overfitting phenomenon

could influence its basis selection.

Polynomial chaos expansion (PCE) established in the context of uncertainty quantification is a promising surrogate model for uncertainty quantification (Xiu and Karniadakis 2002). Alban et al. (2010) and Marelli S et al. (2018) combined the bootstrap resampling and PCE to construct an active learning method which adaptively approximates the limit state surface to perform a failure probability estimate. Pan and Dias (2017) combined sparse PCE and sliced inverse regression technique to achieve a dimension reduction for high-dimensional reliability analysis problems. However, the prediction variance which can be used to improve the sampling performance of PCE is difficult to be obtained for these techniques.

In this paper, we demonstrate the potential of Bayesian regression-based polynomial chaos expansion to provide surrogate appropriate for reliability analysis. Bayesian regression methods have received much attention in the compressed sensing problem since they generally achieve the best recovery performance (Zhang and Rao 2011). Variational Bayesian inference (VBI) with Student-t distribution proposed by Tipping and Lawrence (2005) is one important family. Based on VBI, we employ an automatic search algorithm to enhance sparsity. Since VBI is a Bayesian method, not only the prediction mean but also the prediction variance can be obtained. The predictive variance plays an important role to improve the accuracy of sparse polynomial chaos approximations for a fixed computational budget in the context of design of experiment. Further, the estimation of failure probability analysis is transformed into a classification problem. In this case, an active learning function that utilizes the space-filling and projective properties is used. It enables to select training points to improve the accuracy of the surrogate in the key regions which have a significant impact on the failure probability. A benchmark case is used for validating and assessing the performance of the proposed method, and it is detailedly compared with the well-established AK-MCS and APCK-MCS methods.

The rest of the work is organized as follows. In Section 2, the general PCE is recalled. In Section 3, a detailed derivation of the Bayesian modeling for PCE is presented. Section 4 discusses the implementation details of the proposed method for structural reliability analysis. The performance of the proposed method is assessed in Section 5. Section 6 presents conclusion.

2. A brief review of polynomial chaos expansion

Let $\boldsymbol{\Xi} = \{\Xi_1,...,\Xi_n\}$ be a vector of random variables which contains the component dimension sizes, material properties, etc.. The joint PDF of $\boldsymbol{\Xi}$ is denoted as $f_{\boldsymbol{\Xi}}(\boldsymbol{\xi})=\prod_{j=1}^{n} f_{\Xi_j}(\xi_j)$. Performance function $g(\boldsymbol{\Xi})$ is commonly an explicit function of random variables $\boldsymbol{\Xi}$. Let $\psi_l(\boldsymbol{\Xi})$ and ω_l respectively represent the multivariate orthogonal polynomial and the deterministic coefficient to be computed. The performance function $g(\boldsymbol{\Xi})$ can be written using the formulation of PCE,

$$g(\boldsymbol{\Xi})=\sum_{l=1}^{\infty}\omega_l\psi_l(\boldsymbol{\Xi}) \tag{1}$$

The expansion of Eq.(1) should be truncated for computation. Let $\boldsymbol{\omega}=\{\omega_1,...,\omega_P\}^{\mathrm{T}}$, so that a truncated PCE $g_p(\boldsymbol{\Xi})$ with total degree p is given by:

$$g_p(\boldsymbol{\Xi})=\boldsymbol{\omega}^{\mathrm{T}}\boldsymbol{\psi}(\boldsymbol{\Xi})=g(\boldsymbol{\Xi})-\varepsilon(\boldsymbol{\Xi}) \tag{2}$$

where $P=(n+p)!/n!p!$ and $\varepsilon(\boldsymbol{\Xi})$ represents the residual introduced by truncating the expansion to a finite number of terms. The basis function vector $\boldsymbol{\psi}(\boldsymbol{\Xi})=\{\psi_1(\boldsymbol{\Xi}),...,\psi_P(\boldsymbol{\Xi})\}^{\mathrm{T}}$ is constructed with respect to the distribution information of $\boldsymbol{\Xi}$ so that $\psi_l(\boldsymbol{\Xi})$'s are orthogonal. Often in practice, many of the terms are negligible and thus $g(\boldsymbol{\Xi})$ admits a sparse representation of the following form,

$$g(\boldsymbol{\Xi})\approx\sum_{l\in\mathcal{A}^p}\omega_l\psi_l(\boldsymbol{\Xi}) \tag{3}$$

where $\mathcal{A}^p$ is the truncation set with few elements.

3. Variational Bayesian inference for polynomial chaos expansion

3.1 Variation Bayesian inference using Student-t residual distribution

Suppose that we have generated the initial experimental design with N realizations of $\boldsymbol{\Xi}$, we then have the matrix $\boldsymbol{\xi}_D=\{\boldsymbol{\xi}^{(1)},...,\boldsymbol{\xi}^{(N)}\}^{\mathrm{T}}$ of training points. $g(\boldsymbol{\Xi})$ is then evaluated for each training point and yields a corresponding model evaluation vector $\boldsymbol{\mathcal{Y}}=\{\mathcal{Y}^{(1)},...,\mathcal{Y}^{(N)}\}^{\mathrm{T}}$ $=\left(g\left(\boldsymbol{\xi}^{(1)}\right),...,g\left(\boldsymbol{\xi}^{(N)}\right)\right)^{\mathrm{T}}$. The observed data can be denoted as the linear system,

$$\boldsymbol{\mathcal{Y}}=\boldsymbol{\Psi}\boldsymbol{\omega}+\boldsymbol{\varepsilon}_D \tag{4}$$

where $\boldsymbol{\Psi}=\{\boldsymbol{\Psi}^{(1)},...,\boldsymbol{\Psi}^{(N)}\}^{\mathrm{T}}$ and $\boldsymbol{\varepsilon}_D=\{\varepsilon^{(1)},...,\varepsilon^{(N)}\}^{\mathrm{T}}\left(\varepsilon^{(i)}=\varepsilon\left(\boldsymbol{\xi}^{(i)}\right)\right.$ $\left.,i=1,...,N\right)$.

The application of Bayesian inference in the formulation usually assumes that the residual $\varepsilon(\boldsymbol{\Xi})$ follows Gaussian distribution with zero mean and a constant variance κ^{-1} . The model evaluation $g(\boldsymbol{\xi})$ at an arbitrary point $\boldsymbol{\xi}$ is then a Gaussian variable with an conditional PDF $f\left(g(\boldsymbol{\xi})\,|\,\boldsymbol{\omega},\kappa\right)$. The corresponding posterior distribution of the model parameters $\boldsymbol{\omega}$ and κ is inferred by applying Bayes' rule as:

$$f(\boldsymbol{\omega},\kappa\,|\,\boldsymbol{\mathcal{Y}})=\frac{f(\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\omega},\kappa)\,f(\boldsymbol{\omega},\kappa)}{f(\boldsymbol{\mathcal{Y}})} \tag{5}$$

where $f(\boldsymbol{\mathcal{Y}}\,|\,\boldsymbol{\omega},\kappa)$ is the likelihood function. $f(\boldsymbol{\omega},\kappa)$ is the

prior PDF of the Bayesian model parameters and $f(\mathcal{Y})$ is the marginal likelihood which is expressed as,

$$f(\mathcal{Y})=\int f(\mathcal{Y}|\boldsymbol{\omega},\kappa)f(\boldsymbol{\omega},\kappa)d\boldsymbol{\omega}d\kappa \tag{6}$$

An acknowledged limitation of the Gaussian residual model is that it is not robust (Tipping and Lawrence 2005). Indeed, if the model response follows a heavy-tailed distribution, the accuracy of the predictor can be significantly compromised. To alleviate the problem, it could utilize a more robust noise distribution, such as a zero-mean Student-t distribution,

$$f\left(\varepsilon^{(i)}|\upsilon,\sigma\right)=\frac{\Gamma\left((\upsilon+1)/2\right)}{\Gamma(\upsilon/2)\sqrt{\pi\upsilon}\sigma}\left[1+\frac{1}{\upsilon}\left(\frac{\varepsilon^{(i)}}{\sigma}\right)^2\right]^{-\frac{(\upsilon+1)}{2}} \tag{7}$$

A joint Gaussian-Gamma distribution can be used to express the prior in the formulation given by:

$$\begin{aligned} f\left(\varepsilon^{(i)}|C_0,D_0\right)&=\int_0^\infty f\left(\varepsilon^{(i)}|\kappa_i\right)f\left(\kappa_i|C_0,D_0\right)d\kappa_i \\ &=\int_0^\infty \mathcal{N}\left(\varepsilon^{(i)}|0,\kappa_i^{-1}\right)\mathrm{Gam}\left(\kappa_i|C_0,D_0\right)d\kappa_i \end{aligned} \tag{8}$$

where $\mathrm{Gam}\left(\kappa_i|C_0,D_0\right)=D_0^{C_0}/\Gamma(C_0)\kappa_i^{C_0-1}\exp\left(-\kappa_i D_0\right)$, $\Gamma(\cdot)$ is the Gamma function. The equivalent distribution is obtained with $\upsilon=2C_0$ and $\sigma=\sqrt{D_0/C_0}$ (Tipping and Lawrence 2005). It is easy to see that as $C_0\to\infty$, $f\left(\varepsilon^{(i)}|C_0,D_0\right)$ tends to a Gaussian distribution. As a result, the likelihood function $f(\mathcal{Y}|\boldsymbol{\omega},\boldsymbol{\kappa})$ can be further modeled as the product of several Gaussian distribution varying variances,

$$\begin{aligned} f(\mathcal{Y}|\boldsymbol{\omega},\boldsymbol{\kappa})&=\prod_{i=1}^{N} f\left(\mathcal{Y}^{(i)}|\boldsymbol{\omega},\kappa_i\right) \\ &=\prod_{i=1}^{N}\mathcal{N}\left(\mathcal{Y}^{(i)}|\boldsymbol{\Psi}^{(i)}\boldsymbol{\omega},\kappa_i^{-1}\right) \end{aligned} \tag{9}$$

In order to maintain conjugacy, a similar hierarchical prior is aligned to $\boldsymbol{\omega}$ given by,

$$f(\boldsymbol{\omega}|\boldsymbol{\theta})=\prod_{l=1}^{P}\mathcal{N}\left(\omega_l|0,\theta_l^{-1}\right) \tag{10}$$

where the hyper-parameters in the vector $\boldsymbol{\theta}=\{\theta_1,...,\theta_P\}^{\mathrm{T}}$ control the prior variance on each polynomial chaos coefficient and follow independent Gamma distribution. The hyper-prior of θ_l is then given as:

$$\begin{aligned} f\left(\theta_l|A_0,B_0\right)&=\mathrm{Gam}\left(\theta_l|A_0,B_0\right) \\ &=\frac{B_0^{A_0}}{\Gamma\left(A_0\right)}\theta_l^{A_0-1}\exp\left(-\theta_l B_0\right) \end{aligned} \tag{11}$$

Overall, the posterior distribution over all unknowns is inferred as:

$$\begin{aligned} &f\left(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta}|\mathcal{Y},A_0,B_0,C_0,D_0\right) \\ &\quad=\frac{f(\mathcal{Y}|\boldsymbol{\omega},\boldsymbol{\kappa})\,f(\boldsymbol{\kappa}|C_0,D_0)\,f(\boldsymbol{\omega}|\boldsymbol{\theta})\,f(\boldsymbol{\theta}|A_0,B_0)}{f(\mathcal{Y})} \end{aligned} \tag{12}$$

where $f(\boldsymbol{\kappa}|C_0,D_0)=\prod_{i=1}^{N}\mathrm{Gam}\left(\kappa_i|C_0,D_0\right)$ and $f(\boldsymbol{\theta}|A_0,B_0)=\prod_{l=1}^{P}\mathrm{Gam}\left(\theta_l|A_0,B_0\right)$. The main objective of the Bayesian formulation is to compute the model parameters $\{\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta}\}$. Since the hyper-prior in the Bayesian formulation makes the inference in Eq.(12) intractable, VBI is adopted.

The VBI procedure is utilized for robust Bayesian modelling is inferred by approximating the marginal likelihood, which is made by constructing a variational lower bound (VLB) with respect to a variational PDF $q(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})$. The VLB is defined as,

$$\mathcal{L}\left(q(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})\right)=\int q(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})\log\left\{\frac{f(\mathcal{Y},\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})}{q(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})}\right\}d\boldsymbol{\omega}d\boldsymbol{\kappa}d\boldsymbol{\theta} \tag{13}$$

In order to have an appropriate solution, $q(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})$ should be specified. It is shown that $q(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})$ can be partitioned using a factorized distribution to maintain conjugacy (Waterhouse et al. 1996). The VLB $\mathcal{L}\left(q(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})\right)$ is then maximized with respect to each of the factorized distribution parameters for assessing an appropriate solution of the variational distribution. The factorized distribution is often called the expectation-maximization (EM) formulation, in which $q(\boldsymbol{\omega})$, $q(\boldsymbol{\kappa})$ and $q(\boldsymbol{\theta})$ are respectively given by,

$$q(\boldsymbol{\omega})=\mathcal{N}(\boldsymbol{\omega}|\boldsymbol{\mu}_\omega,\boldsymbol{\Sigma}_\omega) \tag{14}$$

where $\boldsymbol{\Sigma}_\omega=\left(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{\Psi}+\boldsymbol{\Theta}\right)^{-1}$, $\boldsymbol{\mu}_\omega=\boldsymbol{\Sigma}_\omega\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Theta}\mathcal{Y}$, $\boldsymbol{\Theta}=\mathrm{diag}\left(\langle\theta_1\rangle,...,\langle\theta_P\rangle\right)$ and $\boldsymbol{\Lambda}=\mathrm{diag}\left(\langle\kappa_1\rangle,...,\langle\kappa_N\rangle\right)$,

$$q(\boldsymbol{\kappa})=\prod_{i=1}^{N}\mathrm{Gam}\left(\kappa_i|\tilde{C},\tilde{D}_i\right) \tag{15}$$

where $\tilde{C}=C_0+1/2$ and $\tilde{D}_i=D_0+1/2\left[y^{(i)}-2y^{(i)}\boldsymbol{\Psi}^{(i)}\langle\boldsymbol{\omega}\rangle+\boldsymbol{\Psi}^{(i)}\left\langle\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}\right\rangle(\boldsymbol{\Psi}^{(i)})^{\mathrm{T}}\right]$,

$$q(\boldsymbol{\theta})=\prod_{l=1}^{P}\mathrm{Gam}\left(\theta_l|\tilde{A},\tilde{B}_l\right) \tag{16}$$

where $\tilde{A}=A_0+1/2$ and $\tilde{B}_l=B_0+\langle\omega_l\rangle/2$. The expectation terms required to be evaluated in Eqs.(14)-(16) are computed using the standard moments and given by,

$$\begin{aligned} \langle\theta_l\rangle&=\tilde{A}/\tilde{B}_l & \langle\kappa_i\rangle&=\tilde{C}/\tilde{D}_i \\ \langle\boldsymbol{\omega}\rangle&=\boldsymbol{\mu}_\omega & \left\langle\boldsymbol{\omega}\boldsymbol{\omega}^{\mathrm{T}}\right\rangle&=\boldsymbol{\mu}_\omega\boldsymbol{\mu}_\omega^{\mathrm{T}}+\boldsymbol{\Sigma}_\omega \end{aligned} \tag{17}$$

The VLB $\mathcal{L}\left(q(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})\right)$ can then be computed from Eqs.(12)-(17). It is also worth mentioning that in the form of $q(\boldsymbol{\omega})$, the inversion of the $P\times P$ matrix $\boldsymbol{\Sigma}_\omega$ requires an $\mathcal{O}(P^3)$ operation. This can be problematic since P can be quite large. To alleviate this problem, $\boldsymbol{\Sigma}_\omega$ can be computed as:

$$\begin{aligned} \boldsymbol{\Sigma}_\omega&=\left(\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Lambda}\boldsymbol{\Psi}+\boldsymbol{\Theta}\right)^{-1} \\ &=\boldsymbol{\Theta}^{-1}-\boldsymbol{\Theta}^{-1}\boldsymbol{\Psi}^{\mathrm{T}}\left(\boldsymbol{\Psi}\boldsymbol{\Theta}^{-1}\boldsymbol{\Psi}^{\mathrm{T}}+\boldsymbol{\Lambda}^{-1}\right)^{-1}\boldsymbol{\Psi}\boldsymbol{\Theta}^{-1} \\ &=\boldsymbol{\Theta}^{-1}-\boldsymbol{\Theta}^{-1}\boldsymbol{\Psi}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_y\right)^{-1}\boldsymbol{\Psi}\boldsymbol{\Theta}^{-1} \end{aligned} \tag{18}$$

We then need only invert the $N\times N$ matrix $\boldsymbol{\Sigma}_y$, reducing the operation to $\mathcal{O}(N^3)$. This procedure becomes more efficient for highly nonlinear problems, since P which reflects the complexity of the polynomial chaos

approximation raises polynomially with the total degree p.

3.2 Automatic search algorithm for sparse PCE

The structural form of the factorized distribution makes it very convenient to implement the iterative cycle procedure to estimate $\{\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta}\}$. Upon convergence, we find that many of the model parameters in the vector $\boldsymbol{\theta}$ are driven to zero. If $\theta_l = 0$, then the posterior probability satisfies $\Pr(\omega_l \mid \boldsymbol{\mathcal{Y}}, \theta_l = 0) = 0$, where $\Pr(\cdot)$ represents a probability measure (Wipf and Rao 2004). Therefore, we should apply a detection rule to detect which hyper-parameters should be negligible. The detection rule is to choose between the two hypotheses:

$$\begin{cases} \mathcal{H}_0: & \theta_l = 0 \\ \mathcal{H}_1: & \theta_l > 0 \end{cases} \qquad l = 1,...,P \tag{19}$$

Choosing $\mathcal{H}_0$ as the true hypotheses means that the coefficient ω_l=0 and the lth column $\boldsymbol{\Psi}_{*l}$ of $\boldsymbol{\Psi}$ is pruned from the linear system in Eq.(4). To compare the two hypotheses, we apply the test used for sparse signal reconstruction (Hurtado et al. 2013). Let $S_{\mathcal{Y}}$=$\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^{\mathrm{T}}$ denote the sample estimate of the matrix $\boldsymbol{\Sigma}_{\mathcal{Y}}$ and $\boldsymbol{\gamma}_l = \boldsymbol{\Sigma}_{\mathcal{Y}}^{-1}\boldsymbol{\Psi}_l$, the test is denoted as follows,

$$\begin{aligned} \mathcal{H}_0: \quad & \mathcal{T}_0 = \frac{2\boldsymbol{\gamma}_l^{\mathrm{T}} S_{\mathcal{Y}} \boldsymbol{\gamma}_l}{\boldsymbol{\gamma}_l^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathcal{Y}}^{(\sim l)} \boldsymbol{\gamma}_l} \sim \chi_2^2 \\ \mathcal{H}_1: \quad & \mathcal{T}_1 = \frac{2\boldsymbol{\gamma}_l^{\mathrm{T}} S_{\mathcal{Y}} \boldsymbol{\gamma}_l}{\boldsymbol{\gamma}_l^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathcal{Y}} \boldsymbol{\gamma}_l} \sim \chi_2^2 \end{aligned} \tag{20}$$

where $\boldsymbol{\Sigma}_{\mathcal{Y}}^{(\sim l)}$ and $\boldsymbol{\Sigma}_{\mathcal{Y}}$ are the covariance of $\boldsymbol{\mathcal{Y}}$ under the respective hypotheses $\mathcal{H}_0$ and $\mathcal{H}_1$. They have the relationship $\boldsymbol{\Sigma}_{\mathcal{Y}} = \boldsymbol{\Sigma}_{\mathcal{Y}}^{(\sim l)} + \langle\theta_l\rangle^{-1}\boldsymbol{\Psi}_{*l}\boldsymbol{\Psi}_{*l}^{\mathrm{T}}$. The quadratic form of $S_{\mathcal{Y}}$ follows a chi-squared distribution with two degrees of freedom as shown in Eq.(20). $\mathcal{H}_0$ is true if the test statistic in Eq.(20) falls below the threshold α, which is set to meet the probability of false alarm (Hurtado et al. 2013):

$$P_{\mathrm{FA}} = \Pr(\mathcal{T}_0 > \alpha; \mathcal{H}_0) = 1 - F_{\chi_2^2}(\alpha) \tag{21}$$

where $F_{\chi_2^2}(\cdot)$ represents the cumulative distribution function of the distribution χ_2^2. The probability of detecting a significant component can be denoted as:

$$\begin{aligned} P_{\mathrm{SC}} &= \Pr(\mathcal{T}_0 > \alpha; \mathcal{H}_1) \\ &= \Pr(\mathcal{T}_1 > \frac{\boldsymbol{\gamma}_l^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathcal{Y}}^{(\sim l)} \boldsymbol{\gamma}_l}{\boldsymbol{\gamma}_l^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathcal{Y}} \boldsymbol{\gamma}_l} \alpha; \mathcal{H}_1) \\ &= \Pr(\mathcal{T}_1 > \left(1 - \langle\theta_l\rangle^{-1} \boldsymbol{\Psi}_{*l}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathcal{Y}}^{-1} \boldsymbol{\Psi}_{*l}\right)\alpha; \mathcal{H}_1) \\ &= 1 - F_{\chi_2^2}\left[\left(1 - \langle\theta_l\rangle^{-1} \boldsymbol{\Psi}_{*l}^{\mathrm{T}} \boldsymbol{\Sigma}_{\mathcal{Y}}^{-1} \boldsymbol{\Psi}_{*l}\right)\alpha\right] \end{aligned} \tag{22}$$

Based on the decision test, an automatic search algorithm is then proposed to for induce sparse. The flowchart of the automatic search algorithm is sketched in Fig. 1.

To start the algorithm, the parameters A_0, B_0, C_0 and D_0 of the Gamma distributions should be specified. To produce an uninformative prior, the initial values of the parameters are chosen as $A_0 = B_0 = 10^{-6}$ and $C_0 = D_0 = 10^{-2}$. It is worth mentioning that the probability P_{FA} of false alarm in the detection test works as a tuning parameter of the algorithm. Once P_{FA} is set, the detection test is completely set and needs no further adjustment. A low value of P_{FA} will promote sparser solutions. In the paper, we set the probability of false alarm to P_{FA}=10^{-4} for the sparsity process step of our algorithm. A stop criterion is also required to be imposed for converging the algorithm. If the change in the value of VLB between two iterations is less than ε_T=10^{-6}, the algorithm is ultimately to have a good convergence. At convergence of the parameter estimation and sparsity process steps, we can make predictions based on the posterior distribution over the polynomial chaos coefficients. The predictive distribution for an arbitrary point ξ is then computed as,

$$\begin{aligned} f\left(g_p(\xi)|\kappa\right) &= \int f\left(g_p(\xi) \mid \boldsymbol{\omega}, \kappa\right) q(\boldsymbol{\omega}) d\boldsymbol{\omega} \\ &= \mathcal{N}\left(g_p(\xi) \mid \mu_{\mathcal{A}^p}(\xi), v_{\mathcal{A}^p}^2(\xi)\right) \end{aligned} \tag{23}$$

where prediction mean $\mu_{\mathcal{A}^p}(\xi) = \boldsymbol{\mu}_{\boldsymbol{\omega}}^{\mathrm{T}} \boldsymbol{\psi}(\xi)$ and prediction variance $v_{\mathcal{A}^p}^2(\xi) = \kappa^{-1} + \boldsymbol{\psi}(\xi)^{\mathrm{T}} \boldsymbol{\Sigma}_{\boldsymbol{\omega}} \boldsymbol{\psi}(\xi)$.

Since many of the expectancy terms $\langle\theta_l\rangle (l = 1,...,P)$ are set to zero, the sparsity of the prediction mean $\boldsymbol{\mu}_{\boldsymbol{\omega}}$ is guaranteed. However, κ is undermined since nothing is known about the true variability at an unknown model response $g(\xi)$. To overcome the limitation, we consider the Voronoi partition of the space $\boldsymbol{\mathcal{D}}$ according to the training set and have the cell $\mathcal{V}_i$ as

$$\mathcal{V}_i = \left[\{\xi\} \in \boldsymbol{\mathcal{D}}, \left|\{\xi\} - \{\xi^{(i)}\}\right| \le \left|\{\xi\} - \{\xi^{(j)}\}\right|, \forall i \ne j\right], 1 \le i \le j \le N \tag{24}$$

Therefore, the inverse variance κ at an arbitrary point ξ is estimated as $\kappa = \kappa_i\left(\{\xi^{(i)}\} \in \mathcal{V}_i\right)$.

4. Active learning method for reliability analysis

4.1 Active learning function

In structural reliability analysis, the failure event is defined by $g(\boldsymbol{\Xi}) \le 0$, so that the failure probability P_F is denoted as,

$$P_F = \mathrm{Prob}\left(g(\boldsymbol{\Xi}) < 0\right) = \int_{D_F} f_{\boldsymbol{\Xi}}(\boldsymbol{\xi}) d\boldsymbol{\xi} \tag{25}$$

where $D_F = \{\boldsymbol{\xi} : g(\boldsymbol{\xi}) \le 0\}$ is the failure domain and S_F = $\{\boldsymbol{\xi} : g(\boldsymbol{\xi}) > 0\}$ is the safe domain. The limit state surface is defined by $g(\boldsymbol{\Xi})$=0 which lies at the boundary between the two domains. The prediction mean $\mu_{\mathcal{A}^p}(\xi)$ is used to estimate P_F by means of Monte Carlo sampling method. The failure probability can be estimated as follows:

$$\hat{P}_F = \frac{1}{N_{MCS}} \sum_{m=1}^{N_{MCS}} I_F\left[\mu_{\mathcal{A}^p}\left(\boldsymbol{\xi}^{(m)}\right)\right] \tag{26}$$

where N_{MCS} is the total number of samples and $I_F\left[\mu_{\mathcal{A}^p}\left(\boldsymbol{\xi}^{(m)}\right)\right]$ is the indicator function of failure, $I_F\left[\mu_{\mathcal{A}^p}\left(\boldsymbol{\xi}^{(m)}\right)\right]=1$, if $\mu_{\mathcal{A}^p}\left(\boldsymbol{\xi}^{(m)}\right)\le 0$, otherwise $I_F\left[\mu_{\mathcal{A}^p}\left(\boldsymbol{\xi}^{(m)}\right)\right]=0$.

In order to reduce the computational cost during surrogate modeling, PCE is built and refined adaptively. This refinement procedure is usually based on active learning functions, which determine the location of a new training point. The most widely used one is U function (Echard et al. 2011). In the paper, the U function for the refinement of the performance function is given by:

$$U(\boldsymbol{\xi}) = \frac{\left|\mu_{\mathcal{A}^p}(\boldsymbol{\xi})\right|}{v_{\mathcal{A}^p}(\boldsymbol{\xi})} \tag{27}$$

1. **Input** p, $\boldsymbol{\Psi}$, $\boldsymbol{\mathcal{Y}}$, A_0, B_0, C_0, D_0, P_{FA} and ε_T
2. **Initialization**
3. Set k=0, $\boldsymbol{\Theta}^{(k)}=\boldsymbol{I}_P$, $\boldsymbol{\Lambda}^{(k)}=\boldsymbol{I}_N$, $S_y=\boldsymbol{\mathcal{Y}}\boldsymbol{\mathcal{Y}}^{\mathrm{T}}$, $\tilde{A}=A_0+1/2$, $\tilde{C}=C_0+1/2$
4. **Repeat**
5. – Update Step
6. $\quad\boldsymbol{\Sigma}_{\mathcal{Y}}^{(k)}=\boldsymbol{\Psi}(\boldsymbol{\Theta}^{(k)})^{-1}\boldsymbol{\Psi}^{\mathrm{T}}+(\boldsymbol{\Lambda}^{(k)})^{-1}$
7. $\quad\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{(k)}=\boldsymbol{\Theta}^{-1}-\boldsymbol{\Theta}^{-1}\boldsymbol{\Psi}^{\mathrm{T}}\left(\boldsymbol{\Sigma}_{\mathcal{Y}}^{(k)}\right)^{-1}\boldsymbol{\Psi}\boldsymbol{\Theta}^{-1}$
8. $\quad\boldsymbol{\mu}_{\boldsymbol{\omega}}^{(k)}=\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{(k)}\boldsymbol{\Psi}^{\mathrm{T}}\boldsymbol{\Theta}^{(k)}\boldsymbol{\mathcal{Y}}$
9. $\quad\tilde{B}_l^{(k)}=B_0+1/2\,\mu_{\omega,l}^{(k)} \quad (l=1,\ldots,P)$
10. $\quad\tilde{D}_l^{(k)}=D_0+1/2\left[y^{(i)}-2y^{(i)}\boldsymbol{\Psi}^{(i)}\boldsymbol{\mu}_{\boldsymbol{\omega}}^{(k)}+\boldsymbol{\Psi}^{(i)}\left[\boldsymbol{\mu}_{\boldsymbol{\omega}}^{(k)}(\boldsymbol{\mu}_{\boldsymbol{\omega}}^{(k)})^{\mathrm{T}}+\boldsymbol{\Sigma}_{\boldsymbol{\omega}}\right](\boldsymbol{\Psi}^{(i)})^{\mathrm{T}}\right]$
11. – Sparsity Process Step
12. $\quad$Set $\theta_l=\left\langle\theta_l^{(k)}\right\rangle=0$ if
13. $$\frac{2(\boldsymbol{\Psi}_l)^{\mathrm{T}}(\boldsymbol{\Sigma}_{\mathcal{Y}}^{(k)})^{-1}S_y(\boldsymbol{\Sigma}_{\mathcal{Y}}^{(k)})^{-1}\boldsymbol{\Psi}_l}{\boldsymbol{\Psi}_l(\boldsymbol{\Sigma}_{\mathcal{Y}}^{(k)})^{-1}\boldsymbol{\Psi}_l\left[1-\left\langle\theta_l^{(k)}\right\rangle^{-1}\boldsymbol{\Psi}_{\omega l}^{\mathrm{T}}\boldsymbol{\Sigma}_{\mathcal{Y}}^{-1}\boldsymbol{\Psi}_{\omega l}\right]}\le F_{\chi_1^2}^{-1}(1-P_{\mathrm{FA}})$$
14. $\quad$Set $k=k+1$, $\boldsymbol{\Theta}^{(k)}=\mathrm{diag}\left(\tilde{A}/\tilde{B}_l^{(k-1)},\ldots,\tilde{A}/\tilde{B}_l^{(k-1)}\right)$ and $\boldsymbol{\Lambda}^{(k)}=\mathrm{diag}\left(\tilde{C}/\tilde{D}_1^{(k-1)},\ldots,\tilde{C}/\tilde{D}_N^{(k-1)}\right)$.
15. **Until**
16. $k>1$ and $$\frac{\left|\mathcal{L}\left[q(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})\right]^{(k)}-\mathcal{L}\left[q(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})\right]^{(k-1)}\right|}{\mathcal{L}\left[q(\boldsymbol{\omega},\boldsymbol{\kappa},\boldsymbol{\theta})\right]^{(k)}}\le\varepsilon_T$$
17 **Output** $\boldsymbol{\Sigma}_{\boldsymbol{\omega}}=\boldsymbol{\Sigma}_{\boldsymbol{\omega}}^{(k-1)}$, $\boldsymbol{\mu}_{\boldsymbol{\omega}}=\boldsymbol{\mu}_{\boldsymbol{\omega}}^{(k-1)}$ and $\kappa_l=\left\langle\kappa_l^{(k-1)}\right\rangle(i=1,\ldots,N)$

Figure 1.A automatic search algorithm for sparse PCE

However, the candidate point with the smallest value of $U(\boldsymbol{\xi})$ sometimes tends to cluster with existing training points. To address the issues, the aggregate of the Intersite and Projected (Intersite-Proj) distance which consider both space-filling and projective properties is used to ensure that the new points reside far away from the existing training points. The Intersite-Proj distance between the candidate point $\boldsymbol{\xi}$ and the exiting design point $\boldsymbol{\xi}^{(i)}$ is calculated by (Crombecq et al. 2011),

$$\left(\sqrt[n]{N+1}-1\right)\left\|\boldsymbol{\xi}-\boldsymbol{\xi}^{(i)}\right\|_2+(N+1)\left\|\boldsymbol{\xi}-\boldsymbol{\xi}^{(i)}\right\|_{-\infty} \tag{28}$$

where $\left\|\boldsymbol{\xi}-\boldsymbol{\xi}^{(i)}\right\|_2$ is the L_2 norm of $\boldsymbol{\xi}-\boldsymbol{\xi}^{(i)}$, it defines the space-filling property, and $\left\|\boldsymbol{\xi}-\boldsymbol{\xi}^{(i)}\right\|_{-\infty}$ is the minus infinity norm of $\boldsymbol{\xi}-\boldsymbol{\xi}^{(i)}$, it defines the projective property. Therefore, an objective function using minimal Intersite-Proj distance is defined as

$$\mathrm{IP}(\boldsymbol{\xi}) = \frac{\begin{array}{l}\left(\sqrt[n+1]{N+1}-1\right)\min\limits_{\boldsymbol{\xi}^{(i)}\in\boldsymbol{\xi}_D}\left\|\boldsymbol{\xi}-\boldsymbol{\xi}^{(i)}\right\|_2\\ +(N+1)\min\limits_{\boldsymbol{\xi}^{(i)}\in\boldsymbol{\xi}_D}\left\|\boldsymbol{\xi}-\boldsymbol{\xi}^{(i)}\right\|_{-\infty}\end{array}}{\begin{array}{l}\left(\sqrt[n+1]{N+1}-1\right)\max\limits_{\boldsymbol{\xi}^{(i)},\boldsymbol{\xi}^{(j)}\in\boldsymbol{\xi}_D}\left\|\boldsymbol{\xi}^{(i)}-\boldsymbol{\xi}^{(j)}\right\|_2\\ +(N+1)\max\limits_{\boldsymbol{\xi}^{(i)},\boldsymbol{\xi}^{(j)}\in\boldsymbol{\xi}_D}\left\|\boldsymbol{\xi}^{(i)}-\boldsymbol{\xi}^{(j)}\right\|_{-\infty}\end{array}} \tag{29}$$

The active learning function for the selection of new training points at each iteration is developed as,

$$UI(\boldsymbol{\xi}) = \frac{U(\boldsymbol{\xi})}{\mathrm{IP}(\boldsymbol{\xi})} \tag{30}$$

Therefore, the candidate point with the minimum value of $UI(\boldsymbol{\xi})$ is selected to refine the surrogate model.

4.2 Degree selection and stop condition

Besides defining the active learning function for design of experiment, another main concern is to determine the most suitable degree p for PCE. A quantitative measure that quantifies the uncertainty of reliability approximations is then considered as,

$$\begin{aligned}\mathrm{UF}(p) &= \int\Phi\left(\frac{-\left|\mu_{\mathcal{A}^p}(\boldsymbol{\xi})\right|}{v_{\mathcal{A}^p}(\boldsymbol{\xi})}\right)f_{\boldsymbol{\Xi}}(\boldsymbol{\xi})\,d\boldsymbol{\xi}\\ &\approx\sum_{m=1}^{N_{MCS}}\Phi\left(\frac{-\left|\mu_{\mathcal{A}^p}\left(\boldsymbol{\xi}^{(m)}\right)\right|}{v_{\mathcal{A}^p}\left(\boldsymbol{\xi}^{(m)}\right)}\right)\quad p\in\{1,\ldots,p_{\max}\}\end{aligned} \tag{31}$$

where $\Phi(\cdot)$ denotes the standard normal cumulative distribution function and $p_{\max}$ defines the maximum element in the candidate set $\{1,\ldots,p_{\max}\}$. A bigger value $\mathrm{UF}(p)$ represents more uncertainty in reliability approximation (Sun et al. 2017). Therefore, a candidate degree p with the smallest value of Eq.(31) is selected as the most suitable degree and the corresponding surrogate is then used to construct the active learning function for selecting the next training point.

4.3 Stop condition

In principle, the relative bias between P_F and $\hat{P}_F$ is able to assess the quality of the surrogate model in order to see how accurate it is to estimate the failure probability . However, the bias is undetermined because of the unknown true failure probability . Alternatively, we employ the *K*-fold cross-validation based method to estimate the bias term.

According to the principle of *K*-fold cross-validation (*K* is often set to 5–10 for the consideration of computation cost), the exiting training point set $\boldsymbol{\xi}_D$ are divided into *K* subsets and the cross-validation sample subsets $\boldsymbol{\xi}_D^{(-s)}(s=1,\ldots,K)$ are obtained where one of the subsets $\boldsymbol{\xi}_D^{(s)}$ is taken out of $\boldsymbol{\xi}_D$. Based on $\boldsymbol{\xi}_D^{(-s)}(s=1,\ldots,K)$, the surrogate model is built *K* time and corresponding

constructed surrogate model is denoted by $\mu_{\mathcal{A}^p}^{(-s)}$ $(s=1,...,K)$. We then obtain the K_s estimates $\hat{P}_F^{(-s)}$ $(s=1,...,K)$ of failure probability and the stopping condition based on K fold cross-validation is defined as:

$$\varepsilon_F = \frac{\max_{s=\{1,...,K\}} \left| P_F - \hat{P}_F^{(-s)} \right|}{P_F} \leq \varepsilon_F^{\mathrm{Target}} \tag{32}$$

The target accuracy criterion $\varepsilon_F^{\mathrm{Target}}$ is set to 0.005 in the paper.

5. Numerical example

In this section, the proposed method is applied to a structural reliability problem for the sake of illustration. In the following, we will refer to the proposed method as active polynomial chaos expansion combined with Monte-Carlo simulation (APCE-MCS). It will be compared with AK-MCS and APCK-MCS in the numerical problem. The AK-MCS and APCK-MCS employ the U function and the best sparse set of multivariate polynomials for APCK-MCS is selected by LAR algorithm. The results of them are obtained by UQLAB (Marelli et al. 2015), an uncertainty quantification software.

As shown in Figure 2, a nonlinear undamped one-degree-of-freedom system (Rajashekhar and Ellingwood 1993) is taken as the problem, the performance function of the system is given as

$$g(\boldsymbol{\Xi}) = 3r - \left| \frac{2F}{m\omega_0^2} \sin\left(\frac{\omega_0 t_l}{2} \right) \right|$$

where $\boldsymbol{\Xi} = \{m, c_1, c_2, r, F, t_l\}$ and $\omega_0 = \sqrt{c_1 + c_2/m}$. The distributions and parameters of the six input variables are given in Table 1.

Since input variables are Gaussian, multivariate Hermite polynomials are selected for polynomial chaos approximation. The initial experimental design for the surrogate model is obtained with experimental design generated by quasi-Sobol' sequence consisting of 10 training points. Figure 3 gives a convergence rate of the estimated failure probability for the proposed method, AK-MCS and APCK-MCS. It illustrates the major benefit of using the proposed method, as a faster convergence is achieved with relatively small training sample size. AK-MCS requires almost 78 design points to be close to a convergent value, whereas the proposed method needs approximately 50% less points to achieve a similar level accuracy (see Table 2).

Table 1. Input variables of the numerical example

Input	Distribution	Mean	Standard deviation
m	Gaussian	1	0.05
c_1	Gaussian	1	0.1
c_2	Gaussian	0.1	0.01
r	Gaussian	0.5	0.05
F	Gaussian	1	0.2
t_l	Gaussian	1	0.2

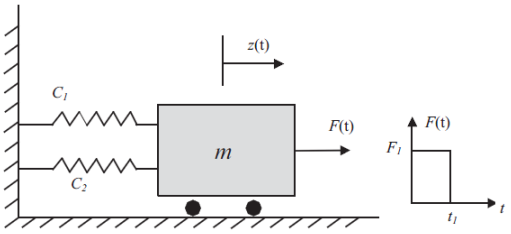

Figure 2. A non-linear oscillator

Table 2. Results of the numerical example

Input	N	P_f	Relative error (%)
MCS	2×10^7	0.0284	-
AK-MCS	78	0.0285	0.04
APCK-MCS	86	0.0294	1.75
The proposed method	42	0.0284	0

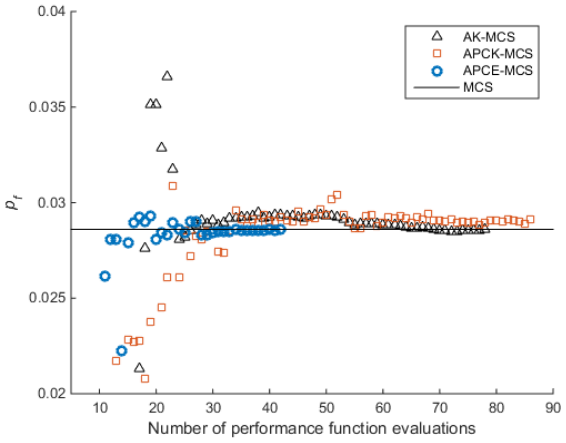

Figure 3. Convergence curve of the estimated failure probability .

6. Conclusions

This paper develops a new method to perform structural reliability analysis by PCE with a limited additional cost. The proposed method relies on variational Bayesian inference, which incorporates a parameterized prior to encourage the model with a sparse structure. The sparse PCE is built with an initial experimental design and updated according to a new active learning function progressively.

This method has shown comparable performance with regard to the well-established AK-MCS and APCK-MCS techniques on an analytical benchmark function. In fact, the computational efficiency of the proposed method can be further improved by combining it with importance sampling or subset simulation. A future area of research will be to perform further validations on relevant industrial applications.

References

Alban N., Nicolas G., Lucdulong J., Maurice L., Pierre V. and Haidar J. 2010. Rpcm: a strategy to perform

reliability analysis using polynomial chaos and resampling. European Journal of Computational Mechanics, 19(8): 795-830.

Bichon B.J., Eldred M.S., Swiler L.P., Mahadevan S. and McFarland J.M. 2008. Efficient global reliability analysis for nonlinear implicit performance functions. AIAA Journal, 46:2459–2468.

Crombecq e., Laermans E. and Dhaene T. 2011. Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling. European Journal of Operational Research, 214(3): 683-696.

Dubourg V., Sudret B. and Deheeger F. 2013. Metamodel-based importance sampling for structural reliability analysis. Probabilistic Engineering Mechanics, 33(1): 47-57.

Der Kiureghian A. and Stefano M. 1991. Efficient algorithm for second-order reliability analysis. Journal of Engineering Mechanics, 117(2):2904–2923.

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

Echard B., Gayton N. and Lemaire M. 2013. A combined importance sampling and Kriging reliability method for small failure probabilities with time-demanding numerical models. Reliability Engineering and System Safety, 111: 232–40.

Huang X., Chen J. and Zhu H. 2016. Assessing small failure probabilities by AK–SS: an active learning method combining Kriging and subset simulation. Structural Safety, 59: 86-95.

Hurtado M., Muravchik C.H. and Nehorai A. 2013. Enhanced sparse Bayesian learning via statistical thresholding for signals in structured noise. IEEE transactions on signal processing, 61(21): 5430-5443.

Hu Z., and Mahadevan S. 2016. Global sensitivity analysis-enhanced surrogate (GSAS) modeling for reliability analysis. Structural and Multidisciplinary Optimization, 53(3):1-21.

Marelli S. and Sudret, B. 2018 An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis. Structural Safety, 75: 67–74.

Marelli S, Schöbi R and Sudret B. 2015. Uqlab User Manual – Reliability Analysis, Technical Report, Report # UQLab-V0.9-107, Chair of Risk, Safety & Uncertainty Quantification, ETH Zurich.

Konakli K. and Sudret B. 2016. Reliability analysis of high-dimensional models using low-rank tensor approximations, Probabilistic Engineering Mechanics, 46:18–36

Pan Q. and Dias D. 2017. Sliced inverse regression-based sparse polynomial chaos expansions for reliability analysis in high dimensions. Reliability Engineering and System Safety, 167:484-493.

Rackwitz R. and Fiessler B. 1978. Structural reliability under combined random load sequences. Computer and Structure, 9:489–494.

Rajashekhar M.R. and Ellingwood B.R. 1993. A new look at the response surface approach for reliability analysis. Structural Safety,12(3):205-20.

Sacks J., Welch W.J., Mitchell T.J. and Wynn H.P. 1989. Design and analysis of computer experiments. Statistical science, 4: 409-423.

Schöbi R., Sudret B. and Wiart J. 2015. Polynomial-Chaos-based Kriging. International Journal for Uncertainty Quantification, 5 (2): 171–193.

Schöbi R., Sudret B. and Marelli S. 2016. Rare event estimation using polynomial-chaos kriging. ASCE-ASME Journal of Risk and Uncertainty in Engineering Systems, Part A: Civil Engineering. 3(2): D4016002.

Sun Z., Wang J., Li R. and Tong C. 2017. LIF: A new Kriging based learning function and its application to structural reliability analysis. Reliability Engineering and System Safety. 157:152-165.

Tipping M.E and Lawrence N.D. 2005. Variational inference for Student-t models: Robust Bayesian interpolation and generalised component analysis. Neurocomputing, 2005, 69(1/3):123-141.

Waterhouse S.R. MacKay D. and Robinson A. J. 1996. Bayesian methods for mixtures of experts. In *Advances in neural information processing systems*, pp. 351-357.

Wipf D.P. and Rao B.D. 2004. Sparse Bayesian learning for basis selection. IEEE Transactions on Signal processing, 52(8): 2153-2164.

Xiu D. and Karniadakis G.E. 2002. The Wiener-Askey Polynomial Chaos for Stochastic Differential Equations. SIAM journal on scientific computing, 24(2): 619-644.

Yun W., Lu Z.Z, Zhou Y. and Jiang X. 2018. An efficient reliability analysis method combining adaptive Kriging and modified importance sampling for small failure probability. Structural & Multidisciplinary Optimization, 58(4):1383-1393.

Zhang Z. and Rao B.D. 2011. Sparse signal recovery with temporally correlated source vectors using sparse Bayesian learning. IEEE Journal of Selected Topics in Signal Processing. 5(5): 912-926.