Structural reliability analysis by a Bayesian sparse polynomial chaos expansion

Biswarup Bhattacharyya

Version of Record: https://www.sciencedirect.com/science/article/pii/S0167473020301521
Manuscript_b158e1758dc20ca2218affbb3b15308c

Structural reliability analysis by a Bayesian sparse polynomial chaos expansion

Biswarup Bhattacharyya[a,*]

[a] *Univ Lyon, Université Claude Bernard Lyon 1, IFSTTAR, LBMC UMR_T9406, F69622, Lyon, France*

Abstract

Accurate computation of failure probability considering uncertain input parameters is very challenging within limited computational cost. An efficient surrogate model, referred to here as sparse variational Bayesian inference based polynomial chaos expansion (SVB-PCE), is formulated in this paper for reliability analysis. The sparsity in the polynomial basis terms is introduced by the automatic relevance determination (ARD) algorithm and the coefficients corresponding to the sparse polynomial bases are computed using the VB framework. The reliability analysis is performed on four typical numerical problems using the SVB-PCE model. The failure probability and the reliability index for all the examples are assessed accurately by the SVB-PCE model using fewer number of model evaluations as compared to the state-of-art methods. Further, the ARD enables to capture the most important terms in the polynomial bases which also reduces the computational cost in assessing the failure probability.

Keywords: Sparse Polynomial chaos expansion, variational Bayesian inference, Automatic relevance determination, Reliability analysis

1. Introduction

In most of the real life engineering problems, some inherent randomness are always present. The need of probabilistic analysis arises from these uncertainties. Reliability analysis aims at assessing the failure probability associated with the integral of the distribution over the failure region. The failure probability (P_f) is computed by integrating the joint probability distribution function (PDF) ($f_x(x)$) over all the random variables in the failure region of the limit state function ($g(x) \leq 0$) which is given by:

$$P_f = p\left(g\left(x\right) \leq 0\right) = \int_{g(x)\leq 0} f_x\left(x\right) dx \tag{1}$$

where $g(x)$ represents the limit state function (LSF) for the quantity of interest (QoI) of a problem. One of the important aspects in the assessment of P_f is the evaluation of the LSF in the failure

*Corresponding author
Email address: biswarupb6@gmail.com (Biswarup Bhattacharyya)

region. Monte Carlo simulation (MCS) is one of the useful methods for the computation of failure probability for any type of problem more accurately. However, MCS requires a large number of model evaluations for assessing an accurate P_f (especially for the low failure probability problems) which limits the method in applying for a large scale engineering problem. Other sampling based methods include subset simulation [1, 2], importance sampling [3, 4], line sampling [5], directional simulation [6], Latin hypercube sampling (LHS) [7]. On the other hand, there are some approximation techniques (based on gradient) available in literature such as first order reliability method (FORM) [8] and second order reliability method (SORM) [9]. These methods approximate the boundary of the failure region (i.e. $g(x) = 0$) and a search process is used to identify the most probable point (MPP). The failure probability of some linear and weakly nonlinear problems can be computed using FORM and SORM. However, in many complex situations, these methods are inaccurate due to highly nonlinear LSF and large number of uncertain parameters.

A distinct approach has been addressed in the literature to deal with the above-mentioned issues, it is based on surrogate models. Surrogate models approximate the LSF more efficiently as compared to the above-mentioned methods. Several surrogate models have been investigated in last two decades for the reliability analysis which are polynomial chaos expansion (PCE) [10, 11], Kriging [12], response surface method (RSM) [13, 14], support vector machine (SVM) [15], high dimensional model representation (HDMR) [16], radial basis function (RBF) [17], neural network [18] etc. PCE is one of the widely used surrogate models for reliability analysis. As a consequence, PCE has been improved in many ways according to the problems. The PCE method, which was developed initially only to account the Gaussian input random variables [19], was improved in [20] to account the other types of random variables. Further, it was improved in [10] for reliability analysis of high-dimensional problems. A different type of improvement has been proposed in [11] for reliability analysis using PCE. Later, the efficiency of both the PCE and the Kriging has been utilized in [21]. An active learning based PCE has also been used along with the PCE [22] recently.

The main issue with the PCE modelling is the curse of dimensionality. Different approaches have been adopted by the researchers to address the issue of curse of dimensionality as discussed previously. The important terms in the PCE model (the terms contributing to predict an accurate stochastic response) were selected using least angle regression (LARS) [23] in [24]. Bayesian approach has also been utilized for formulating an efficient PCE model. Bayesian compressed sensing [25, 26] has been used in [27] to formulate a sparse PCE model for high-dimensional problems. Further, a Kashyap information criterion based sparse Bayesian PCE model has been developed in [28]. A sparse PCE model has been developed in [29] using subspace pursuit approach [30]. Furthermore, D-optimal experimental design has been used to select the sample points adaptively. Two greedy algorithm has been proposed in [31] to formulate sparse PCE model for high-dimensional problems. An extensive review of the recently developed sparse PCE approaches

can be found in [32].

The issue of curse of dimensionality is handled in this paper using a Bayesian inference [33, 34, 35]. A fully Bayesian model is formulated for the PCE using the variational Bayesian (VB) inference [36, 37, 35]. Further, the sparsity in the PCE model terms is achieved through automatic relevance determination (ARD) [37]. The main aim of this formulation is the development of a fully sparse Bayesian PCE model for approximating the QoI in the failure region more accurately using limited number of model evaluations.

The rest of the paper is organized as follows. The general outline for reliability analysis using the PCE is explained in section 2. Then, the Bayesian formulation for the PCE is explained in section 3 and the sparse VB based PCE is formulated in section 4 for reliability analysis. The applicability of the proposed approach is illustrated through some numerical examples in section 5 and finally, the concluding remarks from the present study are discussed in section 6.

2. PCE for reliability analysis

PCE is investigated in a non-intrusive way in this paper. Usually, N number of model evaluations is required beforehand for the non-intrusive approach. Consider d-dimensional random variable $x = \{x_1, x_2, \ldots, x_d\}$ with N realizations in a matrix $\boldsymbol{X} = \{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_d\} \in \mathbb{R}^{N \times d}$ and the corresponding QoI are inscribed in a vector $\boldsymbol{Y} = \{Y_1, Y_2, \ldots, Y_N\}^T \in \mathbb{R}^{N \times 1}$. The QoI is expressed by the PCE in the following way:

$$\boldsymbol{Y}(\boldsymbol{X}) = \sum_{i=1}^{\infty} a_i \boldsymbol{\Psi}^{(i)}(\boldsymbol{X}) \tag{2}$$

where, $\boldsymbol{\Psi}^{(i)}(\boldsymbol{X})$ is the i-th multivariate orthogonal polynomial basis function which is the combinations of the univariate orthogonal polynomial bases. A multivariate orthogonal polynomial basis in terms of the univariate orthogonal polynomial bases is expressed as:

$$\boldsymbol{\Psi}^{(i)}(\boldsymbol{X}) = \prod_{j=1}^{d} \boldsymbol{\phi}^{(i,j)}(\boldsymbol{x}_j) \tag{3}$$

where $\boldsymbol{\phi}^{(i,j)}(\boldsymbol{x}_j)$ is the i-th orthogonal polynomial basis function for the j-th random variable $\boldsymbol{x}_j$. The univariate orthogonal polynomials for the different types of random variables can be constructed according to the Askey scheme [38]. The types of polynomial for a variety of input random variables are given in [20]. The multivariate orthogonal polynomials should satisfy the condition of inner product between two polynomials:

$$\left\langle \boldsymbol{\Psi}^{(i)} \boldsymbol{\Psi}^{(j)} \right\rangle = \int_{\mathbb{R}_X} \boldsymbol{\Psi}^{(i)}(\boldsymbol{X}) \boldsymbol{\Psi}^{(j)}(\boldsymbol{X}) f_{\boldsymbol{X}}(\boldsymbol{X}) d\boldsymbol{X} = h_n^2 \delta_{ij} \tag{4}$$

where, δ is the Kronecker delta which is 1 for $i = j$ and zeros for all the other cases. h_n is a constant which is 1 for the orthonormal polynomials. $f_{\boldsymbol{X}}(\boldsymbol{X})$ is the joint PDF for all the input

random variables, and considering all the variables independent, the joint PDF takes the form:

$$f_{\boldsymbol{X}}\left(\boldsymbol{X}\right) = \prod_{i=1}^{d} f_{\boldsymbol{X}}\left(\boldsymbol{x}_i\right) \tag{5}$$

The PCE as defined in Equation 2, must be truncated with some maximum degree P of the orthogonal polynomials for the practical implementation. Therefore, the truncated PCE is represented by:

$$\begin{aligned} \boldsymbol{Y}\left(\boldsymbol{X}\right) &= \sum_{i=1}^{n} a_i \boldsymbol{\Psi}^{(i)}\left(\boldsymbol{X}\right) + \boldsymbol{\varepsilon}_p \\ &= \boldsymbol{\Psi}\left(\boldsymbol{X}\right)\boldsymbol{a} + \boldsymbol{\varepsilon}_p \end{aligned} \tag{6}$$

where, $\boldsymbol{\varepsilon}_p$ is the residual vector of the truncated PCE and the elements of the vector are assumed to be Normally independent distributed with zero mean and precision ς^{-1}. i in the expansion represents the i-th polynomial basis in the truncated PCE. The total number of terms in the polynomial can be found from all the possible combinations of the input random variables along with the maximum degree of the polynomial P:

$$n = \binom{d+P}{P} = \frac{(d+P)!}{d!P!} \tag{7}$$

The multivariate orthogonal polynomial in Equation 6 is a matrix $\boldsymbol{\Psi}\left(\boldsymbol{X}\right) \in \mathbb{R}^{N\times n}$ and the coefficient is a vector $\boldsymbol{a} \in \mathbb{R}^{n\times 1}$. The QoI is calculated initially at some predefined experimental design (e.g. LHS, Sobol sequence) points which are $\boldsymbol{Y}\left(\boldsymbol{X}\right) \in \mathbb{R}^{N\times 1}$ in Equation 6. The only unknown in the PCE formulation is the coefficient vector $\boldsymbol{a}$.

The accuracy of the PCE modelling is greatly influenced by the computation of the PCE coefficients. The coefficients are calculated by minimizing the residual error $\boldsymbol{\varepsilon}_p$. Ordinary least square (OLS) [39, 40] is one of the most popularly used techniques for computing the PCE coefficients. The PCE coefficients are calculated using the OLS as follows:

$$\boldsymbol{a} = \left(\boldsymbol{\Psi}^T\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}^T\boldsymbol{Y} \tag{8}$$

In this paper, a distinct approach is proposed to compute the PCE coefficients by a VB inference [41, 37].

3. Bayesian approximation for PCE

Bayesian inference has already been used efficiently in the machine learning domain [42, 43, 44]. Besides, it has also been successfully applied for the nonlinear system identification problems [45, 46]. In this paper, the coefficients of the PCE model are computed for the reliability analysis using a VB inference [41, 37, 33].

3.1. Formulation of Bayesian inference for PCE

Recall the truncated form of the PCE as defined in Equation 6:

$$\boldsymbol{Y}(\boldsymbol{X}) = \boldsymbol{\Psi}(\boldsymbol{X})\,\boldsymbol{a} + \boldsymbol{\varepsilon}_p \tag{9}$$

In Equation 9, the coefficient vector $\boldsymbol{a}$ is calculated through the Bayesian formulation. For the Bayesian formulation in the present paper, all the notations are used without the functional terms i.e. $Y_i = Y(\boldsymbol{X}_i), \boldsymbol{\Psi}_i = \boldsymbol{\Psi}(\boldsymbol{X}_i); i = 1, 2, \ldots, N$. $\boldsymbol{X}_i$ and $\boldsymbol{\Psi}_i$ represent the i-th row of $\boldsymbol{X}$ and $\boldsymbol{\Psi}$ respectively. Therefore, having the QoI at N number of samples $\boldsymbol{Y} = \{Y_1, Y_2, \ldots, Y_N\}^T$ and the polynomial basis matrix $\boldsymbol{\Psi}$, the Bayesian model posterior is given by:

$$p(\Theta|\boldsymbol{Y}) = \frac{p(\boldsymbol{Y}|\boldsymbol{\Psi}, \Theta)\, p(\Theta)}{p(\boldsymbol{Y})} \tag{10}$$

In Equation 10, the Bayesian model parameter is represented by Θ and $p(\boldsymbol{Y}|\boldsymbol{\Psi}, \Theta)$ is the likelihood function which is computed based on the distribution of the Bayesian model parameter and the polynomial basis. $p(\Theta)$ is the prior distribution of the Bayesian model parameter. $p(\boldsymbol{Y})$ is the marginal likelihood which is given by:

$$p(\boldsymbol{Y}) = \int p(\boldsymbol{Y}|\boldsymbol{\Psi}, \Theta)\, p(\Theta)\, d\Theta \tag{11}$$

To infer an accurate Bayesian model parameter, the posterior must be estimated having the proper knowledge of the prior and the likelihood function. Given the orthogonal polynomials and the QoI at the initial experimental design points, the likelihood function is given by:

$$p(\boldsymbol{Y}|\boldsymbol{\Psi}, \boldsymbol{a}, \varsigma) = \prod_{i=1}^{N} p(Y_i|\boldsymbol{\Psi}_i, \boldsymbol{a}, \varsigma) \tag{12}$$

$$= \prod_{i=1}^{N} \mathcal{N}\left(Y_i|\boldsymbol{\Psi}_i\boldsymbol{a}, \varsigma^{-1}\right) \tag{13}$$

$$= \left(\frac{\varsigma}{2\pi}\right)^{\frac{N}{2}} \exp\left(-\frac{\varsigma}{2}\sum_{i=1}^{N}(Y_i - \boldsymbol{\Psi}_i\boldsymbol{a})^2\right) \tag{14}$$

In Equation 13, the likelihood function is estimated by the Gaussian distribution ($\mathcal{N}(\bullet)$ in Equation 13) with mean $\boldsymbol{\Psi}_i\boldsymbol{a}$ and variance ς^{-1}. To have the conjugacy and continuity in the formulation [35], the prior is inferred by the joint Gaussian-gamma distribution [47]:

$$p(\boldsymbol{a}, \varsigma|\boldsymbol{\alpha}) = p(\boldsymbol{a}|\varsigma, \boldsymbol{\alpha})\, p(\varsigma) \tag{15}$$

$$= \mathcal{N}\left(\boldsymbol{a}|0, (\varsigma\boldsymbol{\mathcal{A}})^{-1}\right) \mathrm{Gam}(\varsigma|A_0, B_0) \tag{16}$$

$$= (2\pi)^{-\frac{n}{2}} |\boldsymbol{\mathcal{A}}|^{\frac{1}{2}} \frac{B_0^{A_0}}{\Gamma(A_0)} \varsigma^{\frac{n}{2}+A_0-1} \exp\left(-\frac{\varsigma}{2}\left(\boldsymbol{a}^T\boldsymbol{\mathcal{A}}\boldsymbol{a} + 2B_0\right)\right) \tag{17}$$

where $\mathrm{Gam}(\bullet)$ defines the PDF of the gamma distribution with the distribution parameters A_0 and B_0. In Equation 17, the prior is expressed with a double exponential form which defines

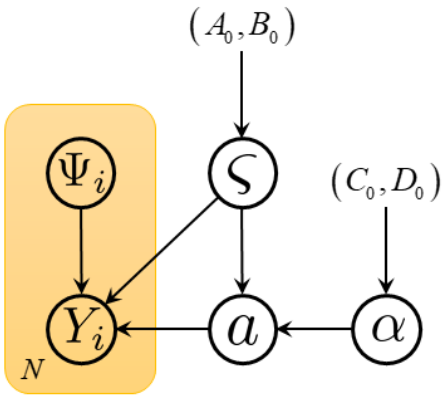


Figure 1: Graphical model representing the dependencies between the Bayesian model parameters

the absolute continuity of the prior [47, 37]. The prior is further parameterized by a hyper-prior $\boldsymbol{\alpha} = \{\alpha_1, \alpha_2, \ldots, \alpha_n\}^T = \text{diag}(\boldsymbol{\mathcal{A}})$. The determinant of $\boldsymbol{\mathcal{A}}$ is given by:

$$|\boldsymbol{\mathcal{A}}| = \prod_{i=1}^{n} \alpha_i \tag{18}$$

The exponential form of the Bayesian formulation is also maintained for the hyper-prior. Therefore, the hyper-prior is expressed with an independent gamma distribution which is given by:

$$p(\boldsymbol{\alpha}) = \prod_{j=1}^{n} \text{Gam}(\alpha_j | C_0, D_0) \tag{19}$$

$$= \prod_{j=1}^{n} \frac{D_0^{C_0}}{\Gamma(C_0)} \alpha_j^{C_0 - 1} \exp(-D_0 \alpha_j) \tag{20}$$

where $\Gamma(\bullet)$ denotes the gamma function and C_0, D_0 are the parameters of the gamma distribution for $\boldsymbol{\alpha}$. Incorporating all the model parameters, the posterior of the Bayesian formulation is given by:

$$p(\boldsymbol{a}, \varsigma, \boldsymbol{\alpha} | \boldsymbol{Y}) = \frac{p(\boldsymbol{Y} | \boldsymbol{\Psi}, \boldsymbol{a}, \varsigma)\, p(\boldsymbol{a} | \varsigma, \boldsymbol{\alpha})\, p(\varsigma)\, p(\boldsymbol{\alpha})}{p(\boldsymbol{Y})} \tag{21}$$

Having the polynomial basis matrix $\boldsymbol{\Psi} \in \mathbb{R}^{N \times n}$ and the QoI $\boldsymbol{Y} \in \mathbb{R}^{N \times 1}$ at the N experimental design points, the Bayesian model parameters $\Theta \in \{\boldsymbol{a}, \varsigma, \boldsymbol{\alpha}\}$ are inferred by the above-described Bayesian formulation. A graphical model is shown in Figure 1 which represents the dependencies between all the Bayesian model parameters. The main objective of this Bayesian formulation is to infer an appropriate PCE coefficient $\boldsymbol{a}$. However, the incorporation of the hyper-prior $\boldsymbol{\alpha}$ in the Bayesian formulation makes the problem intractable [35]. In this respect, the posterior is inferred by approximating the marginal likelihood which can be carried out by random sampling methods [48, 49]. However, to minimize the computational cost, a VB inference has been utilized in this paper to approximately infer the marginal likelihood.

3.2. Variational Bayesian inference

The posterior is approximated by inferring an approximate solution of the marginal likelihood $p(\boldsymbol{Y})$ through a simple optimization technique [37] by VB inference. For the simplicity in formulation, the marginal likelihood is considered here without the orthogonal polynomial basis matrix $\boldsymbol{\Psi}$. Recall the marginal likelihood from Equation 11:

$$p(\boldsymbol{Y}) = \int_{\Theta} p(\boldsymbol{Y}|\Theta)\, p(\Theta)\, d\Theta \tag{22}$$

An arbitrary variational distribution $q(\Theta)$ is considered for the approximation of the posterior $p(\Theta|\boldsymbol{Y})$. Further, a variational lower bound (VLB) can be constructed on the arbitrary variational distribution $q(\Theta)$ [35] and the VLB is represented by:

$$\mathscr{L}[q(\Theta)] = \int_{\Theta} q(\Theta) \ln \frac{p(\boldsymbol{Y}|\Theta)\, p(\Theta)}{q(\Theta)} d\Theta \tag{23}$$

$$= \int q(\Theta) \ln p(\boldsymbol{Y}, \Theta)\, d\Theta - \int q(\Theta) \ln q(\Theta)\, d\Theta \tag{24}$$

According to the proper probability distribution, the integral of the variational distribution with respect to the Bayesian model parameter is $\int_{\Theta} q(\Theta)\, \mathrm{d}\,\Theta = 1$. Consequently, the VLB is depicted by:

$$\mathscr{L}[q(\Theta)] = \int_{\Theta} q(\Theta) \ln \frac{p(\Theta|\boldsymbol{Y})}{q(\Theta)} d\Theta + \ln p(\boldsymbol{Y}) \tag{25}$$

and finally, the log-marginal likelihood is given by:

$$\ln p(\boldsymbol{Y}) = \mathscr{L}[q(\Theta)] - \int_{\Theta} q(\Theta) \ln \frac{p(\Theta|\boldsymbol{Y})}{q(\Theta)} d\Theta \tag{26}$$

$$= \mathscr{L}[q(\Theta)] + \mathrm{KL}(q(\Theta) \parallel p(\Theta|\boldsymbol{Y})) \tag{27}$$

In Equation 27, the log-marginal likelihood is represented by the VLB and the Kullback-Leibler (KL) divergence $\mathrm{KL}(\bullet)$ from q to p.

The variational distribution is incorporated in the formulation such that an approximate solution for the posterior is achieved by approximating the log-marginal likelihood through the variational distribution $q(\Theta)$. The variational distribution can be inferred either by minimizing the KL divergence [35, 50] or, by maximizing the VLB $\mathscr{L}$ with respect to $q(\Theta)$. One possible solution for the variational distribution $q(\Theta)$ is $q(\Theta) = p(\Theta|\boldsymbol{Y})$ when the KL divergence is equal to zero. A factorized distribution has been used to formulate the variational distribution [51, 52] in this paper. The variational distribution is described in context of the factorized distribution in the next section.

3.3. Factorized distribution based VB inference

To confine the distribution family for the variational distribution $q(\Theta)$, the variational distribution of the Bayesian model parameters are separated using the factorized distribution. Using the

factorized distribution, the parameters of the Bayesian model are constructed in a multiplicative form [35]:

$$q\left(\Theta\right) = \prod_{i=1}^{N_p} q\left(\Theta_i\right) \tag{28}$$

where N_p is the number of Bayesian model parameters in Θ. From the above-formulated Bayesian model, we have three components of the model parameter i.e. $\Theta \in \{\boldsymbol{a}, \varsigma, \boldsymbol{\alpha}\}$. Considering the joint probability distribution, $q\left(\Theta\right)$ is subdivided into two components and Equation 28 is given by:

$$q\left(\Theta\right) = q\left(\boldsymbol{a}, \varsigma\right) q\left(\boldsymbol{\alpha}\right) \tag{29}$$

The Bayesian model parameters are divided into $N_p = 2$ parts for the present Bayesian formulation. However, for the generalized formulation, the indicial notation as mentioned in Equation 28 will be used in the forthcoming derivation.

For the computation of the variational distribution, the lower bound $\mathscr{L}$ is maximized with respect to each component of the variational distribution independently, while the others are fixed. According to the factorized distribution (see Equation 28), the VLB in Equation 24 is formulated as:

$$\mathscr{L}\left[q\left(\Theta\right)\right] = \int \prod_{i=1}^{N_p} q\left(\Theta_i\right) \left(\ln p\left(\boldsymbol{Y}, \Theta\right) - \sum_{i=1}^{N_p} \ln q\left(\Theta_i\right) \right) d\Theta \tag{30}$$

$$= \int q\left(\Theta_j\right) \left(\int \ln p\left(\boldsymbol{Y}, \Theta\right) \prod_{i \neq j} q\left(\Theta_i\right) d\Theta_i \right) d\Theta_j$$

$$- \int q\left(\Theta_j\right) \ln q\left(\Theta_j\right) d\Theta_j + \text{constant} \tag{31}$$

$$= \int q\left(\Theta_j\right) \ln \tilde{p}\left(\boldsymbol{Y}, \Theta_j\right) d\Theta_j - \int q\left(\Theta_j\right) \ln q\left(\Theta_j\right) d\Theta_j + \text{constant} \tag{32}$$

$$= -\,\mathrm{KL}\left(q\left(\Theta_j\right) \parallel \tilde{p}\left(\boldsymbol{Y}, \Theta_j\right)\right) + \text{constant} \tag{33}$$

where $\sum_{i \neq j} \int q\left(\Theta_i\right) \ln q\left(\Theta_i\right) d\Theta_i$ are inserted in the constant term. It should be noted that a new probability distribution is defined in Equation 32 as $\tilde{p}\left(\boldsymbol{Y}, \Theta_j\right)$ [35] which is given by:

$$\ln \tilde{p}\left(\boldsymbol{Y}, \Theta_j\right) = \mathbb{E}_{i \neq j}\left[\ln p\left(\boldsymbol{Y}, \Theta\right)\right] \tag{34}$$

where $\mathbb{E}_{i \neq j}\left[\bullet\right]$ represents the expectation with respect to all the distribution $q\left(\Theta_i\right)$ for all $i \neq j$ which is a part in the first integral of Equation 31 i.e.

$$\mathbb{E}_{i \neq j}\left[\ln p\left(\boldsymbol{Y}, \Theta\right)\right] = \int \ln p\left(\boldsymbol{Y}, \Theta\right) \prod_{i \neq j} q\left(\Theta_i\right) d\Theta_i \tag{35}$$

The VLB in Equation 33 represents a negative KL divergence from $q\left(\Theta_j\right)$ to $\tilde{p}\left(\boldsymbol{Y}, \Theta_j\right)$. The variational distribution can also be inferred by minimizing the KL divergence which is ultimately

maximizing the VLB. The minimum of the KL divergence occurs when $q(\Theta_j) = \tilde{p}(\boldsymbol{Y}, \Theta_j)$. Therefore, the optimal solution for the j-th distribution (when the KL divergence is minimum for the j-th distribution), keeping the others fix at $q(\Theta_{i \neq j})$, is given by:

$$\ln q_k(\Theta_j) = \mathbb{E}_{i \neq j}[\ln p(\boldsymbol{Y}, \Theta)] \tag{36}$$

where k in the subscript represents the number of iterations for the maximization of the VLB. The optimization is performed using Equation 36 iteratively. The k-th VLB is computed using all the updated parameters in the $(k-1)$-th step. All the factors of $q(\Theta_i)$ are initialized to get an accurate estimation for the $q(\Theta_j)$. The convergence of the VLB is guaranteed [53] because the VLB is convex with respect to all the component of the variational distribution. The convergence of the VLB can be followed through Equation 24 which is written more explicitly as:

$$\mathscr{L}[q(\Theta)] = \mathbb{E}_{\Theta}[\ln p(\boldsymbol{Y}, \Theta)] - \mathbb{E}_{\Theta}[\ln q(\Theta)] \tag{37}$$

3.4. Factorized VB inference for PCE

In this section, the VB inference is formulated for the PCE using the factorized distribution. We need to formulate two components $q(\boldsymbol{a}, \varsigma)$ and $q(\boldsymbol{\alpha})$ separately for the PCE. Firstly, keeping the $q(\boldsymbol{\alpha})$ fix, the variational distribution $q(\boldsymbol{a}, \varsigma)$ is represented by Equation 36 as:

$$\ln q_k(\boldsymbol{a}, \varsigma) = \ln p(\boldsymbol{Y}|\boldsymbol{\Psi}, \boldsymbol{a}, \varsigma) + \mathbb{E}_{\boldsymbol{\alpha}}[\ln p(\boldsymbol{a}, \varsigma|\boldsymbol{\alpha})] \tag{38}$$

where $\ln p(\boldsymbol{Y}|\boldsymbol{\Psi}, \boldsymbol{a}, \varsigma)$ and $\ln p(\boldsymbol{a}, \varsigma|\boldsymbol{\alpha})$ can be directly substituted from Equation 14 and Equation 17 respectively. After substituting both the expressions in Equation 38 and taking all the terms independent of $\boldsymbol{a}$ and ς in constant, Equation 38 is re-formulated as:

$$\begin{aligned} \ln q_k(\boldsymbol{a}, \varsigma) = & \left(\frac{n}{2} + A_0 - 1 + \frac{N}{2}\right) \ln \varsigma \\ & - \frac{\varsigma}{2}\left(\boldsymbol{a}^T\left(\mathbb{E}_{\boldsymbol{\alpha}}[\boldsymbol{\mathcal{A}}] + \sum_{i=1}^{N} \boldsymbol{\Psi}_i^T \boldsymbol{\Psi}_i\right)\boldsymbol{a} + \sum_{i=1}^{N} Y_i^2 - 2\sum_{i=1}^{N} Y_i \boldsymbol{\Psi}_i \boldsymbol{a} + 2B_0\right) \\ & + \text{constant} \end{aligned} \tag{39}$$

The choice of the type of distribution for the VB inference is solely dependent on the type of distribution in the Bayesian formulation [35, 54]. Given the Gaussian distribution likelihood function (refer Equation 13) and the Gaussian-gamma prior (refer Equation 16), the factorized VB inference for $q_k(\boldsymbol{a}, \varsigma)$ is inferred using a conjugate Gaussian-gamma distribution which is given by:

$$q_k(\boldsymbol{a}, \varsigma) = \mathcal{N}\left(\boldsymbol{a}|\boldsymbol{a}_k, \varsigma^{-1}\boldsymbol{V}_k\right) \text{Gam}(\varsigma|A_k, B_k) \tag{40}$$

Here Equation 40 can also be expressed similar to Equation 17. After representing similarly with natural logarithm and by equating the coefficients of $-\frac{\varsigma}{2}\boldsymbol{a}^T\boldsymbol{a}$ between Equation 40 and Equation 39

(only expanding the Gaussian distribution), the inverse of $\boldsymbol{V}_k$ can be deduced as:

$$\boldsymbol{V}_k^{-1} = \sum_{i=1}^{N} \boldsymbol{\Psi}_i^T \boldsymbol{\Psi}_i + \mathbb{E}_{\boldsymbol{\alpha}} \left[\boldsymbol{\mathcal{A}}\right] \tag{41}$$

Similar to $\boldsymbol{V}_k$, the updated PCE coefficients can be found by equating the coefficient of $\boldsymbol{a}$, which is given by:

$$\boldsymbol{a}_k = \boldsymbol{V}_k \sum_{i=1}^{N} \boldsymbol{\Psi}_i^T Y_i \tag{42}$$

The PCE coefficients are updated in each iteration by Equation 42 during the optimization procedure. However, the parameters of the gamma distribution should also be updated in each iteration to have an updated variational distribution $q_k\left(\boldsymbol{a}, \varsigma\right)$. Taking natural logarithm, the part of gamma distribution in Equation 40 can be represented as:

$$\begin{aligned} \ln q_k\left(\boldsymbol{a}, \varsigma\right) = & \ln \mathcal{N}\left(\boldsymbol{a} | \boldsymbol{a}_k, \varsigma^{-1} \boldsymbol{V}_k\right) - \frac{\varsigma}{2}\left(\sum_{i=1}^{N} Y_i^2 + 2B_0 - a_k^T \boldsymbol{V}_k^{-1} \boldsymbol{a}_k\right) \\ & + \left(A_0 - 1 + \frac{N}{2}\right) \ln \varsigma \end{aligned} \tag{43}$$

According to the formulation, ς follows gamma distribution with parameter A_k and B_k. Equating the coefficients of ς of Equation 43 with the PDF of the gamma distribution, B_k is given by:

$$\begin{aligned} -\varsigma B_k &= -\frac{\varsigma}{2}\left(\sum_{i=1}^{N} Y_i^2 + 2B_0 - \boldsymbol{a}_k^T \boldsymbol{V}_k^{-1} \boldsymbol{a}_k\right) \\ B_k &= B_0 + \frac{1}{2}\left(\sum_{i=1}^{N} Y_i^2 - \boldsymbol{a}_k^T \boldsymbol{V}_k^{-1} \boldsymbol{a}_k\right) \end{aligned} \tag{44}$$

Similarly, another parameter of the Gamma distribution A_k is computed by comparing the coefficients of $\ln \varsigma$:

$$A_k = A_0 + \frac{N}{2} \tag{45}$$

The variational distribution $q\left(\boldsymbol{a}, \varsigma\right)$ is optimized by using Equation 41, 42, 44 and 45 in each iteration for the optimization of the variational distribution $q\left(\Theta\right)$.

For the optimization of the variational distribution $q\left(\Theta\right)$, another part of the factorized distribution $q\left(\boldsymbol{\alpha}\right)$ must be maximized. Following Equation 36, the log-variational distribution for α is given by:

$$\ln q_k\left(\boldsymbol{\alpha}\right) = \ln p\left(\boldsymbol{\alpha}\right) + \mathbb{E}_{\boldsymbol{a},\varsigma}\left[\ln p\left(\boldsymbol{a}, \varsigma | \boldsymbol{\alpha}\right)\right] \tag{46}$$

$$= \sum_{j=1}^{n} \left(C_0 - 1\right) \ln \alpha_j - D_0 \alpha_j + \frac{1}{2} \ln \alpha_j - \frac{\boldsymbol{\alpha}}{2} E_{\boldsymbol{a},\varsigma}\left[\varsigma a_j^2\right] + \text{constant} \tag{47}$$

where $p\left(\boldsymbol{\alpha}\right)$ and $p\left(\boldsymbol{a}, \varsigma | \boldsymbol{\alpha}\right)$ are substituted from Equation 20 and Equation 17 respectively. The terms independent of $\boldsymbol{\alpha}$ are considered in the constant. The hyper-prior has already been formulated in the Bayesian formulation by the gamma distribution. For the conjugacy in the VB

formulation, the variational distribution for $\boldsymbol{\alpha}$ is formulated by the gamma distribution which is given by:

$$\ln q_k(\boldsymbol{\alpha}) = \sum_{j=1}^{n} \ln \mathrm{Gam}\left(\alpha_j | C_k, D_{k_j}\right) \tag{48}$$

By expanding the above-equation and comparing Equation 47 and 48, the two parameters of the gamma distributions are found. Comparing the coefficients of $\ln \alpha_j$ and α_j between the two equations, the parameters of the gamma distribution are given by:

$$C_k = C_0 + \frac{1}{2} \tag{49}$$

$$D_{k_j} = D_0 + \frac{1}{2}\mathbb{E}_{\boldsymbol{a},\varsigma}\left[\varsigma a_j^2\right] \tag{50}$$

The expectation parameter in Equation 41 and the expectation in Equation 50 are calculated via standard moment of the corresponding distributions [35] which are given by:

$$\mathbb{E}_{\boldsymbol{\alpha}}[\boldsymbol{\mathcal{A}}] = \boldsymbol{\mathcal{A}}_k \tag{51}$$

$$\mathbb{E}_{\boldsymbol{a},\varsigma}\left[\varsigma a_j^2\right] = a_{k_j}^2 \frac{A_k}{B_k} + V_{k_{jj}} \tag{52}$$

where k_{jj} in the subscript denotes the diagonal elements of the matrix $\boldsymbol{V}$. The expression given in Equation 51 is a matrix having only the diagonal terms which correspond each of the orthogonal polynomial bases in PCE:

$$\mathbb{E}_{\boldsymbol{\alpha}}[\alpha_j] = \frac{C_k}{D_{k_j}} \tag{53}$$

All the diagonal elements of $\mathbb{E}_{\boldsymbol{\alpha}}[\boldsymbol{\mathcal{A}}]$ are calculated using Equation 53:

$$\mathbb{E}_{\boldsymbol{\alpha}}[\boldsymbol{\mathcal{A}}] = \mathrm{diag}\left(\mathbb{E}_{\boldsymbol{\alpha}}[\boldsymbol{\alpha}]\right) \tag{54}$$

In the above formulation of the variational distribution using the factorized distribution, the parameters A_k and C_k are constants during the optimization procedure. Whereas, B_k and D_k are dependent on the updated parameters of the previous iteration. For that reason, B_k and D_k are initialized at their initial values B_0 and D_0 respectively. $\boldsymbol{a}_k$ and $\boldsymbol{V}_k$ are initialized with the initial given input as:

$$\boldsymbol{a}_0 = \left(\boldsymbol{\Psi}^T\boldsymbol{\Psi}\right)^{-1}\boldsymbol{\Psi}^T\boldsymbol{Y} \tag{55}$$

$$\boldsymbol{V}_0 = \boldsymbol{\Psi}^T\boldsymbol{\Psi} \tag{56}$$

The main aim of the above-mentioned formulation is to estimate an accurate coefficients for the PCE model which is found at the end of the VB optimization process. The VB inference has been formulated in accordance with the factorized distribution which does not required the computation of the VLB. However, the convergence of the optimization is observed using the VLB

only. The procedure of computing the VLB is described in Appendix A. A convergence criterion is required to achieve the convergence of the VLB during the optimization process, which is given by:

$$\frac{\mathscr{L}\left[q\left(\Theta\right)\right]_k - \mathscr{L}\left[q\left(\Theta\right)\right]_{k-1}}{\mathscr{L}\left[q\left(\Theta\right)\right]_{k-1}} \times 100 \leq T_{\mathscr{L}} \tag{57}$$

where $T_{\mathscr{L}}$ is a threshold value defined in percentage. The threshold is required to define initially in percentage. The whole procedure of computing the PCE coefficients using the factorized VB inference is given in Algorithm 1.

Algorithm 1 Pseudo-code for the factorized VB inference

procedure VB($\boldsymbol{\Psi}, \boldsymbol{Y}, A_0, B_0, C_0, D_0$)
  $k = 0$
  $A_k = A_0 + \frac{N}{2}$ ▷ Refer Equation 45
  $B_k = B_0$
  $C_k = C_0 + \frac{1}{2}$ ▷ Refer Equation 49
  $D_k = D_0$
  for $i = 1 : n$ **do**
    $\mathbb{E}_{\boldsymbol{\alpha}}\left[\boldsymbol{\alpha}_j\right] = \frac{C_k}{D_{k_j}}$ ▷ Refer Equation 53
  end for
  $\mathbb{E}_{\boldsymbol{\alpha}}\left[\boldsymbol{\mathcal{A}}\right] = \mathrm{diag}\left(\mathbb{E}_{\boldsymbol{\alpha}}\left[\boldsymbol{\alpha}\right]\right)$ ▷ Refer Equation 54
  while $\frac{\mathscr{L}[q(\Theta)]_k - \mathscr{L}[q(\Theta)]_{k-1}}{\mathscr{L}[q(\Theta)]_{k-1}} \times 100 > T_{\mathscr{L}}$ **do** ▷ Refer Equation 57
    $k = k + 1$
    $\boldsymbol{V}_k^{-1} = \sum_{i=1}^{N} \boldsymbol{\Psi}_i^T \boldsymbol{\Psi}_i + \mathbb{E}_{\boldsymbol{\alpha}}\left[\boldsymbol{\mathcal{A}}\right]$ ▷ Refer Equation 41
    $\boldsymbol{a}_k = \boldsymbol{V}_k \sum_{i=1}^{N} \boldsymbol{\Psi}_i^T Y_i$ ▷ Refer Equation 42
    $B_k = B_0 + \frac{1}{2}\left(\sum_{i=1}^{N} Y_i^2 - \boldsymbol{a}_k^T \boldsymbol{V}_k^{-1} \boldsymbol{a}_k\right)$ ▷ Refer Equation 44
    for $j = 1 : n$ **do**
      $\mathbb{E}_{\boldsymbol{a},\varsigma}\left[\varsigma a_j^2\right] = a_{k_j}^2 \frac{A_k}{B_k} + V_{k_{jj}}$ ▷ Refer Equation 52
      $D_{k_j} = D_0 + \frac{1}{2}\mathbb{E}_{\boldsymbol{a},\varsigma}\left[\varsigma a_j^2\right]$ ▷ Refer Equation 50
      $\mathbb{E}_{\boldsymbol{\alpha}}\left[\alpha_j\right]_k = \frac{C_k}{D_{k_j}}$ ▷ Refer Equation 53
    end for
    $\mathbb{E}_{\boldsymbol{\alpha}}\left[\boldsymbol{\mathcal{A}}\right]_k = \mathrm{diag}\left(\mathbb{E}_{\boldsymbol{\alpha}}\left[\boldsymbol{\alpha}\right]\right)$ ▷ Refer Equation 54
    Update VLB $\mathscr{L}\left[q\left(\Theta\right)\right]_k$ ▷ Refer Equation 76
  end while
  return $\boldsymbol{a}, \mathscr{L}\left[q\left(\Theta\right)\right], \mathbb{E}_{\boldsymbol{\alpha}}\left[\boldsymbol{\mathcal{A}}\right]$
end procedure

3.5. Sparse VB inference based PCE

The main issue with the PCE is that the computational cost is increased with the increase of number of random variables, and the increase of the polynomial degree. The sparsity in the PCE polynomials is introduced here by Automatic Relevance Determination (ARD) [55, 56, 57] in accordance with the factorized VB inference. The usefulness of a polynomial basis is measured using the ARD, and the VLB $\mathscr{L}$ is utilized to measure the most useful model for a specific problem.

The introduction of the sparsity is solely related to the formulation of the factorized VB inference for the PCE model. The relation is established here with the hyper-prior $\boldsymbol{\alpha}$ as introduced in Equation 15. The standard moment of the hyper-priors $\mathbb{E}_{\boldsymbol{\alpha}}[\boldsymbol{\mathcal{A}}]$ has been calculated in the previous section, and this quantity is directly related to the relevance of each of the polynomials. $\mathbb{E}_{\boldsymbol{\alpha}}[\boldsymbol{\mathcal{A}}]$ is a diagonal matrix, and each of the diagonal terms corresponds to each of the orthogonal polynomials. Therefore, after convergence of the VB inference for a full PCE model, n number of hyper-priors for all the orthogonal polynomial are available and the inverse of that defines the ARD values for the full PCE model [58, 46]:

$$ARD^s = \text{diag}\left(\mathbb{E}_{\boldsymbol{\alpha}}[\boldsymbol{\mathcal{A}}]^{-1}\right) \tag{58}$$

where s in the superscript denotes the iteration number for the ARD value (see Algorithm 2). After performing the VB inference on the full PCE model, the polynomials which have an ARD value below a threshold are discarded. The threshold value is computed as:

$$\ln T^s_{ARD} = \min\left(\ln ARD^s\right) + \frac{\max\left(\ln ARD^s\right) - \min\left(\ln ARD^s\right)}{\rho} \tag{59}$$

The threshold can be changed by tuning the resolution of the threshold ρ. The pruning of orthogonal polynomials is performed until only one term remains in the polynomial basis (i.e. $n = 1$). At the end of the pruning procedure, several sets of polynomials along with the corresponding VLB values are available: the final most suitable sparse set corresponds to the highest VLB value. The final sparse polynomial basis is given by:

$$\boldsymbol{\Psi}_{sparse} = \boldsymbol{\Psi}^{s^*} \tag{60}$$

$$s^* = \text{ind}\left(\max_s \mathscr{L}\left(\Theta\right)^s\right) \tag{61}$$

where $\text{ind}\left(\bullet\right)$ is the index of the polynomial set having the maximum VLB value. The orthogonal polynomial set which estimates a highest VLB value would assess the most accurate result. The procedure of obtaining the most appropriate sparse PCE is shown in Algorithm 2.

Remark 1: The threshold for the construction of sparse PCE is obtained using natural logarithm of the ARD in Equation 59. The main reason behind this formulation is that the ARD value for a less relevant term may be very less as compared to the highly relevant term and the chance of pruning a less relevant term (but relevant for the PCE) would be high. To

appropriately detect the polynomial bases using the ARD value, natural logarithm is used in the present formulation.

Remark 2: The pruning of number of polynomial bases per cycle (s) is fully dependent on the tuning parameter ρ. Large number of polynomial bases are pruned with the low value of the tuning parameter ρ which may leads to an inaccurate set for the estimation using the sparse PCE. As a consequence, it is always safe to use a 'high' value for the tuning parameter ρ ($\rho = 1000$ is used in this paper).

Algorithm 2 Pseudo-code for the SVB framework

procedure SVB($\boldsymbol{\Psi}, \boldsymbol{Y}, \rho$)
  $s = 0$
  $\boldsymbol{\Psi}^0 = \boldsymbol{\Psi}$
  while $n > 1$ **do** ▷ Perform iteration until 1 basis remains
    $s = s + 1$
    $(\boldsymbol{a}^s, \mathscr{L}\left[q\left(\Theta\right)\right]^s, \mathbb{E}_{\boldsymbol{\alpha}}\left[\boldsymbol{\mathcal{A}}\right]^s) = \mathrm{VB}(\boldsymbol{\Psi}^{s-1}, \boldsymbol{Y}, A_0, B_0, C_0, D_0)$ ▷ Refer Algorithm 1
    $ARD^s = \mathrm{diag}\left(\mathbb{E}_{\boldsymbol{\alpha}}\left[\boldsymbol{\mathcal{A}}\right]^{-1}\right)$ ▷ Refer Equation 58
    Calculate $\ln T^s_{ARD}$ ▷ Refer Equation 59
    $\boldsymbol{\Psi}^- = \emptyset$
    for $j = 1 : n$ **do**
      if $\ln ARD^s \leq \ln T^s_{ARD}$ **then**
        $\Psi^- = \boldsymbol{\Psi}^- \cup \boldsymbol{\Psi}_j$
      end if
    end for
    $\boldsymbol{\Psi}^s = \boldsymbol{\Psi}^{s-1} \setminus \boldsymbol{\Psi}^-$
    $n = \mathrm{card}\left(\boldsymbol{\Psi}^s\right)$
  end while
  $s^* = \mathrm{ind}\left(\max \mathscr{L}\left(\Theta\right)^s\right)$
  $\boldsymbol{\Psi}_{sparse} = \boldsymbol{\Psi}^{s^*}$
  $\boldsymbol{a}_{sparse} = \boldsymbol{a}^{s^*}$
  $\mathrm{ind}_{sparse} = \mathrm{ind}\left(\mathrm{card}\left(\boldsymbol{\Psi}_{sparse}\right)\right)$ ▷ Get the index of final set
  return $\boldsymbol{a}_{sparse}, \boldsymbol{\Psi}_{sparse}, \mathrm{ind}_{sparse}$
end procedure

4. Sparse VB-PCE based reliability framework

Algorithm 1 and 2 are used here for reliability analysis. More specifically, the uncertain QoI of a physical system is estimated using the sparse VB based PCE (SVB-PCE). The PDF of a

QoI is computed here using the SVB-PCE surrogate model. Afterwards, the reliability analysis is conducted by assessing the tail of the PDF of QoI. The procedure of the reliability analysis using the SVB-PCE surrogate model is given in Algorithm 3.

The surrogate model is trained with some response QoI and the polynomial basis function at the initial sample points. LHS is used for all the numerical examples to generate the initial sample points.

As mentioned in section 3.4, some of the parameters in the factorized VB inference formulation must be initialized to start Algorithm 1. The initial parameters of the distributions are chosen as $A_0 = C_0 = 1 \times 10^{-2}$ and $B_0 = D_0 = 1 \times 10^{-4}$ to have an uninformative prior distribution [46]. To have a good converged results, the threshold of the VLB, $T_{\mathscr{L}}$ is chosen as 0.001%.

The two previously outlined algorithms are utilized for the reliability analysis using the above-mentioned initialized parameters. The proposed framework is shown in Algorithm 3. After getting the sparse set of the PCE basis matrix and the corresponding coefficients, a MCS is performed on the surrogate model with new samples using Equation 6, and the probability of failure (P_f) is computed just simply by calculating the number of samples exceeding the threshold value (τ) for the QoI which is given by:

$$P_f = \frac{\text{card}\left((\tau - Y_{new}) \leq 0\right)}{\text{card}\left(Y_{new}\right)} \tag{62}$$

where card $(\bullet)$ is the number of elements satisfying the condition for the corresponding vector. The reliability index is computed by:

$$\beta = -\Phi^{-1}\left(P_f\right) \tag{63}$$

where Φ^{-1} is the inverse of the standard normal variate. Further, the predictive distribution parameters are computed as described in Appendix B and 95% confidence interval (CI) of the predicted probability of failure is also computed from the distribution parameters.

5. Numerical examples

The applicability of the proposed SVB-PCE model in the reliability analysis is examined in this section through some numerical problems. More specifically, four problems are solved. Out of the four, first one is a highly nonlinear test function, second one is a low failure probability problem, and the last two examples are the structural engineering problems.

All the results assessed by the SVB-PCE model are compared with the sparse PCE [24] and the full PCE model [20]. The sparse PCE model [24] is designated in this paper by LARS-PCE due to the use of LARS algorithm [23] for selecting the important terms in the polynomial basis. In all the numerical examples, the MCS result is taken as the reference. Additionally, the results for the first and the last two examples are also compared with the results presented in [22]. The accuracy

Algorithm 3 Algorithm for performing reliability analysis by the SVB-PCE model

procedure SVBPCE_RELIABILITY(d, N, Distribution type)
  Initialize P, ρ, τ ▷ Degree of PCE, resolution of SVB and threshold for the QoI
  $\boldsymbol{X} = N$ samples for d dimension variables by LHS
  Compute $\boldsymbol{Y}(\boldsymbol{X})$ ▷ QoI at the N samples
  Compute $\boldsymbol{\Psi}(\boldsymbol{X})$ ▷ Orthogonal polynomial basis matrix
  $(\boldsymbol{a}_{sparse}, \boldsymbol{\Psi}_{sparse}, \text{ind}_{sparse})$=SVB$(\boldsymbol{\Psi}(\boldsymbol{X}), \boldsymbol{Y}(\boldsymbol{X}), \rho)$ ▷ Refer Algorithm 2
  Generate/load $\boldsymbol{X}_{new}$ ▷ New samples
  Compute $\boldsymbol{\Psi}(\boldsymbol{X}_{new}, \text{ind}_{sparse})$ ▷ Sparse polynomial basis matrix
  $\boldsymbol{Y}_{new} = \boldsymbol{\Psi}(\boldsymbol{X}_{new}, \text{ind}_{sparse})\, \boldsymbol{a}_{sparse}$ ▷ Estimate QoI at the new samples
  $P_f = \frac{\text{card}((\tau - \boldsymbol{Y}_{new}) \leq 0)}{\text{card}(\boldsymbol{Y}_{new})}$ ▷ Compute failure probability
  $\beta = -\Phi^{-1}(P_f)$ ▷ Compute reliability index
end procedure

and the efficiency of the proposed method are measured for all the examples. The efficiency is measured through the initial number of model evaluations and through the number of effective polynomial bases (n_{sparse}). A surrogate model having less n_{sparse} in the polynomial basis matrix is depicted as the most efficient with same number of model evaluations.

5.1. A four branch series system [22]

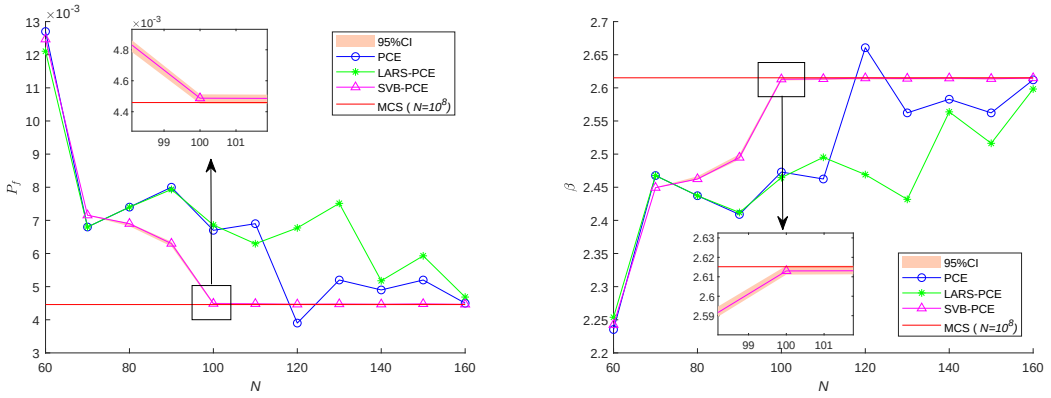
This is a common benchmark problem for the reliability estimation which was proposed in [59]. The functional form of the series system is given by:

$$g(x) = \min \left\{ \begin{array}{c} 3 + 0.1(x_1 - x_2)^2 - \frac{(x_1+x_2)}{\sqrt{2}} \\ 3 + 0.1(x_1 - x_2)^2 + \frac{(x_1+x_2)}{\sqrt{2}} \\ (x_1 - x_2) + \frac{6}{\sqrt{2}} \\ (x_2 - x_1) + \frac{6}{\sqrt{2}} \end{array} \right\} \tag{64}$$

where x_1 and x_2 are the two random variables which are normally distributed with mean zero and unit variance (i.e. $x_i \in \mathcal{N}[0, 1]$). Similar to [22], it is assumed that failure occurs when $g(x) \leq 0$.

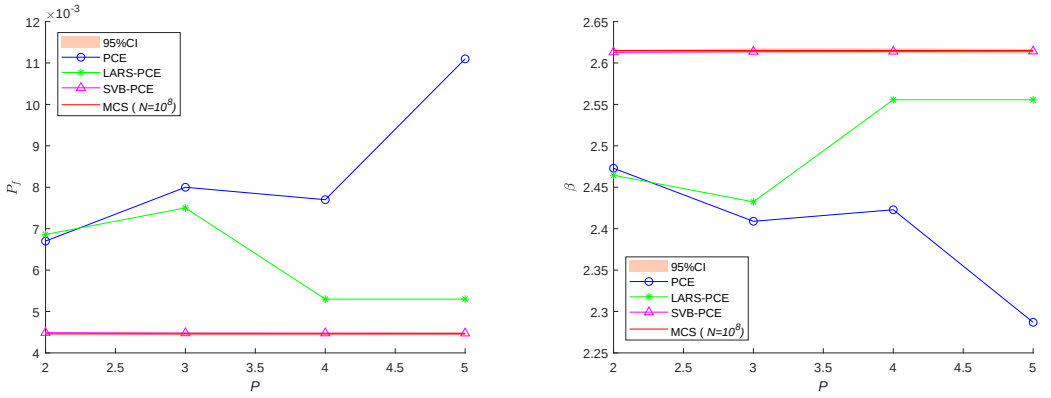
The failure probability and the reliability index are computed by the full-scale MCS using $N = 10^8$ number of model evaluations which is taken as the reference solution for this problem. The failure probability and the reliability index by the MCS are estimated as $P_f = 4.46 \times 10^{-3}$ and $\beta = 2.61$, respectively. Further, the failure probability is computed by the SVB-PCE, LARS-PCE and the full PCE model.

The evolution of the failure probability and the reliability index with different number of model evaluations for all the surrogate models are presented in Figure 2 along with 95% CI fixing $P = 2$. It is seen that the results by all the surrogates are converged to the reference result with the



(a) Failure probability (b) Reliability index

Figure 2: Evolution of failure probability and reliability index with the increase of the number of model evaluations by different methods for the series system



(a) Failure probability (b) Reliability index

Figure 3: Evolution of failure probability and reliability index with the increase of polynomial degree by different methods for the series system

increase of number of model evaluations. The convergence is noticeably rapid by the SVB-PCE model.

It is evident from Figure 2 that $N = 100$ is sufficient to obtain a good accuracy by the SVB-PCE model. Further, a study is performed by increasing the degree P keeping $N = 100$ for all the surrogate models. The evolution of P_f and β with P is shown in Figure 3. It is seen that the results predicted by the SVB-PCE model follow the MCS results, however the results by the PCE model are diverged with the increase of polynomial degree. Almost same results are predicted by the SVB-PCE model beyond $P = 2$ because the same terms are selected by the SVB approach beyond $P = 2$. For the LARS-PCE model, the results with $P = 4$ and $P = 5$ are almost the same because the same terms are selected for both degrees.

Table 1: Prediction of the failure probability and the reliability index for the series system by various methods

Method	N	P_f	95% CI (P_f)	β	95% CI (β)	P	n_{sparse}
MCS	10^8	4.46×10^{-3}	–	2.61	–	–	–
PCE	160	4.50×10^{-3}	–	2.61	–	2	–
LARS-PCE	160	4.68×10^{-3}	–	2.60	–	2	3
A-bPCE [22]	167	4.63×10^{-3}	$[4.5, 4.7]\times10^{-3}$	2.60	[2.59, 2.61]	5	12
SVB-PCE	100	4.48×10^{-3}	$[4.45, 4.51]\times10^{-3}$	2.61	[2.61, 2.62]	2	3

An accurate result is obtained by the SVB-PCE model using lower ($N = 100$) number of model evaluations as compared to the full PCE ($N = 160$) and LARS-PCE ($N = 160$) model. All the results are presented in Table 1 along with the results from [22]. It is seen clearly that the SVB-PCE model outperforms other surrogates in efficiency as it requires much less number of model evaluations. Furthermore, a narrow 95% CI suggests a stable prediction by the SVB-PCE model. Same number of sparse bases are noticed for the SVB-PCE and the LARS-PCE model, however, the accuracy of the results are higher for the SVB-PCE model. The number of important bases for A-bPCE model [22] is little higher due to the high degree polynomial.

5.2. A small failure probability problem [60]

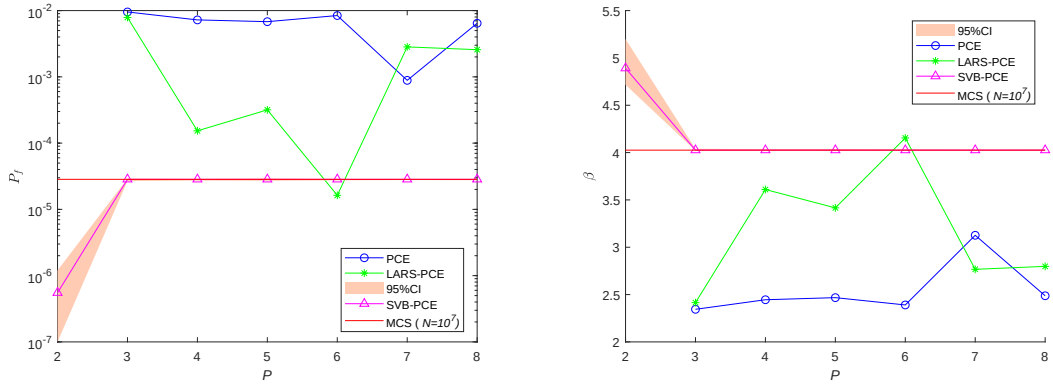
A small failure probability problem is considered in this example. The functional form is given by:

$$g(x) = 0.5(x_1 - 2)^2 - 1.5(x_2 - 5)^3 - 3 \tag{65}$$

where x_1 and x_2 follow standard normal distribution. The failure occurs when $g(x) \leq 0$. The reference solution is obtained by the MCS approach using 10^7 number of model evaluations. The reference failure probability and the reliability index are found as $P_f = 2.84 \times 10^{-5}$ and $\beta = 4.0257$, respectively.

The SVB-PCE, LARS-PCE and PCE models are used to predict the failure probability and the reliability index. The evolution of P_f and β with the increase of polynomial degree is shown in Figure 4 for $N = 15$. It is seen that the SVB-PCE model is predicted a good result only using $N = 15$ and $P = 3$. The results by the SVB-PCE model beyond $P = 3$ is almost same because the same terms are selected by the ARD algorithm beyond $P = 3$. Furthermore, the PCE model is unable to predict a good result due to the less number of model evaluations as compared to the number of terms. The LARS-PCE model is also predicted close to the MCS results using $P = 6$, however the results predicted by the SVB-PCE model is more accurate. The CI is quite high at the beginning, however it is very narrow when the results converge with the MCS results.

The results obtained by all surrogate models are listed in Table 2 along with the MCS results.



(a) Failure probability (b) Reliability index

Figure 4: Evolution of failure probability and reliability index with the increase of polynomial degree by different methods for the small failure probability problem

Table 2: Prediction of the failure probability and the reliability index for the small failure probability problem by various methods

Method	N	P_f	95% CI (P_f)	β	95% CI (β)	P	n_{sparse}
MCS	10^7	2.84×10^{-5}	–	4.0257	–	–	–
PCE	15	8.85×10^{-4}	–	3.1263	–	7	–
LARS-PCE	15	1.63×10^{-5}	–	4.1545	–	6	6
SVB-PCE	15	2.84×10^{-5}	$[2.67, 2.91]\times 10^{-5}$	4.0257	[4.02, 4.04]	3	6

For the PCE and the LARS-PCE models, the best obtained results are listed in the table. It is noticed that the PCE model is predicted the worst results. The LARS-PCE model predicted results are quite close to the MCS results, however, a more accurate result is predicted by the SVB-PCE model only with $P = 3$. The 95% CI are also computed from the predictive distribution parameters (Appendix B) for the SVB-PCE model. The CI is small, therefore the uncertainty in the prediction is quite low for the SVB-PCE model.

5.3. A 23 bars truss problem [24]

To check the applicability of the SVB-PCE on structural engineering problems, a 2D truss is taken in this example. The truss structure is shown in Figure 5.

The truss shown in Figure 5, is a simply supported two dimensional truss having 6 point loads $(P_1, P_2, \ldots, P_6)$ acting on it. All the loads are considered uncertain for this problem and follow the Gumbel distribution. The Young's modulus and the cross-sectional areas of all the members are also considered uncertain and follow a Lognormal distribution. All the uncertain parameters of the truss are given in Table 3 along with their mean and standard deviation. A_1 and E_1 correspond

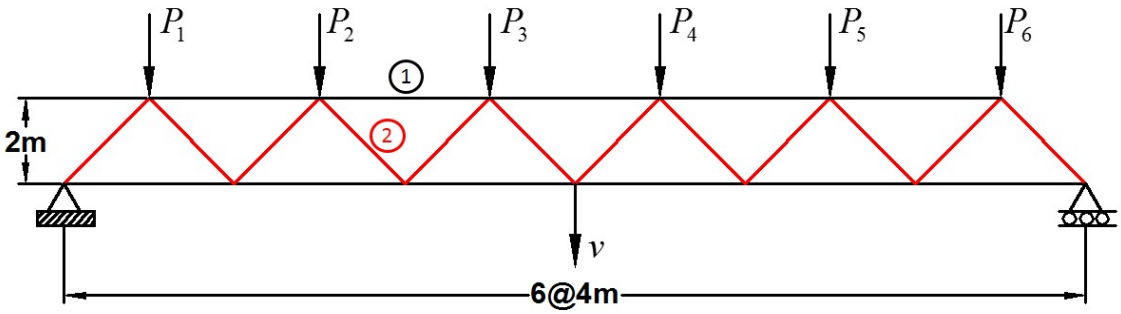


Figure 5: Geometrical view of the 2D truss with all the loading conditions

Table 3: Description of random variables for the truss structure

Random variable	Distribution	Mean	Standard deviation	Unit
E_1, E_2	Lognormal	2.1×10^{11}	2.1×10^{10}	Pa
A_1	Lognormal	2.0×10^{-3}	2.0×10^{-4}	m^2
A_2	Lognormal	1.0×10^{-3}	1.0×10^{-4}	m^2
$P_1, P_2, \ldots, P_6$	Gumbel	5.0×10^4	7.5×10^3	N

to the black members, and A_2 and E_2 correspond to the red members in Figure 5.

The QoI for this problem is considered as the vertical deflection at the mid span of the truss (v). For the reliability analysis, the threshold value of the vertical deflection is considered as $\tau_v = 0.12$m. The LSF for this problem is given by:

$$g(X) = \tau_v - v(X) \tag{66}$$

where $g(X) \leq 0$ denotes the failure region.

The failure probability and the reliability index for this problem is computed by the full-scale MCS using 10^6 number of model evaluations. Each of the model evaluations is performed in a MATLAB based finite element framework.

For the computation of P_f and β by the full PCE, the LARS-PCE and the SVB-PCE, the degree of the polynomial is taken as $P = 2$. As a consequence, the full PCE model consists of $n = 66$ terms in the polynomial basis matrix.

The failure probability and the reliability index computed by all the methods are listed in Table 4 along with the results reported in [22]. An accurate result is achieved by the SVB-PCE model using much less number of model evaluations as compared to the other surrogate models. As a result, a low computational cost is required for assessing the failure probability by the SVB-PCE model. Along with this, only 30.30% terms in the polynomial basis are required for the assessment of QoI by the SVB-PCE. On the other hand, LARS-PCE requires 63.64% terms in the

Table 4: Prediction of the failure probability and the reliability index for the 2D truss by various methods

Method	N	P_f	95% CI (P_f)	β	95% CI (β)	P	n_{sparse}
MCS	10^6	1.50×10^{-3}	–	2.97	–	–	–
PCE	180	1.47×10^{-3}	–	2.97	–	2	–
LARS-PCE	150	1.41×10^{-3}	–	2.98	–	2	42
A-bPCE [22]	129	1.48×10^{-3}	$[1.43, 1.54]\times10^{-3}$	2.97	[2.96, 2.98]	3	43
SVB-PCE	90	1.49×10^{-3}	$[1.45, 1.54]\times10^{-3}$	2.97	[2.96, 2.98]	2	20

Table 5: Details of the elements of the frame structure

Element	Modulus of elasticity	Moment of inertia	Cross-sectional area
B_1	E_4	I_{10}	A_{18}
B_2	E_4	I_{11}	A_{19}
B_3	E_4	I_{12}	A_{20}
B_4	E_4	I_{13}	A_{21}
C_1	E_5	I_6	A_{14}
C_2	E_5	I_7	A_{15}
C_3	E_5	I_8	A_{16}
C_4	E_5	I_9	A_{17}

polynomial basis to assess P_f and β. A similar number of terms were also required by the A-bPCE model [22]. Furthermore, the 95% CI for the SVB-PCE and A-bPCE models are very close to each other. The level of uncertainty in the failure probability prediction is very low, however, the SVB-PCE model is much efficient considering a lower number of model evaluations.

5.4. A multi-storied frame structure [61]

The fourth illustration is a two dimensional multi-storied frame structure. This example has been investigated previously in several research articles [61, 10, 22]. The frame structure along with all the loadings and dimensions are shown in Figure 6. The frame consists of four different types of beams $(B_1, \ldots, B_4)$ and four different types of columns $(C_1, \ldots, C_4)$. All the beams and columns are marked in Figure 6 and the properties of all the elements are given in Table 5. For the present frame structure, all the loads $(P_1, \ldots, P_3)$, modulus of elasticities (E_4, E_5), moment of inertias $(I_6, \ldots, I_{13})$ and cross-sectional areas $(A_{14}, \ldots, A_{21})$ are considered uncertain. A total 21 random variables are present in the system which can be represented as $\boldsymbol{X} = \{P_1, \ldots, P_3, E_4, E_5, I_6, \ldots, I_{13}, A_{14}, \ldots, A_{21}\}$.

Along with these uncertainties, the random variables are correlated with each other. The

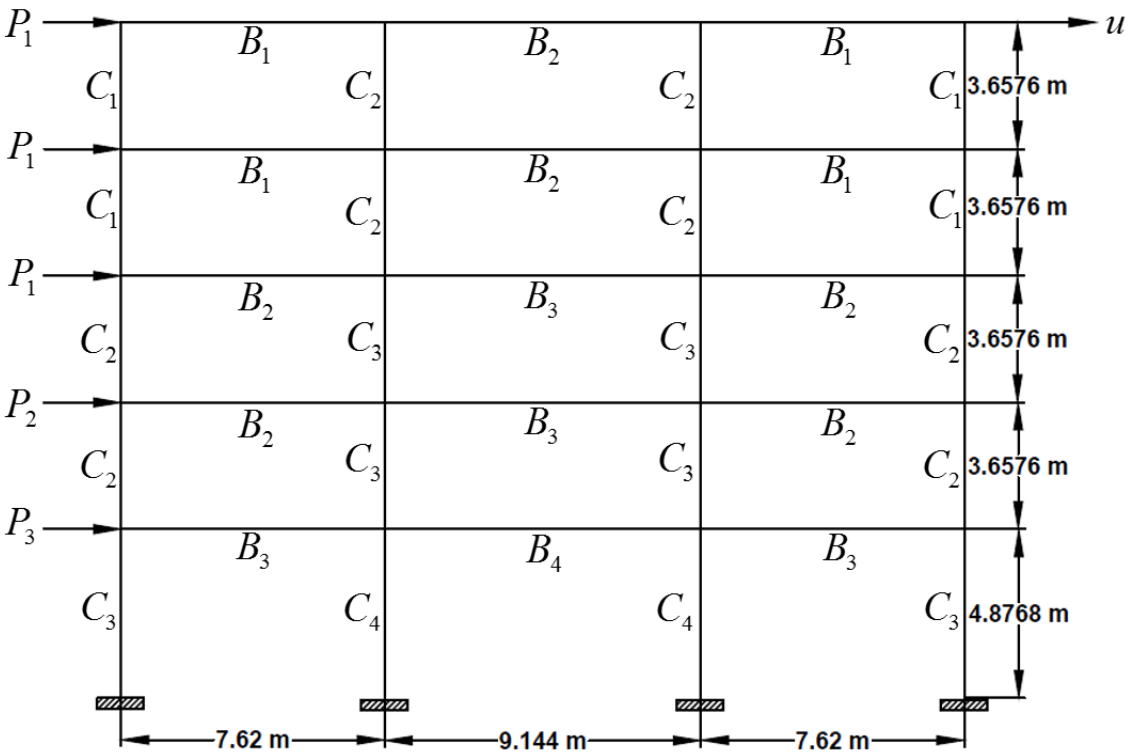


Figure 6: Geometrical view of the multi-storied frame

correlation coefficients between the random variables are given below:

- The modulus of elasticities are highly correlated i.e. $r_{E_4,E_5} = 0.9$.
- The moment of inertia and the cross-sectional of the each element in Table 5 are also highly correlated i.e. $r_{I_i,A_i} = 0.95$.
- The correlation coefficients between the other material properties are very low i.e. $r_{I_i,A_j} = r_{I_i,I_j} = r_{A_i,A_j} = 0.13$.
- The correlation coefficients between all the other variables are zero.

The correlation coefficients for this example are more explicitly given in [62]. In the formulation of the surrogate models, the correlated variables are modelled with the Gaussian copula [63] and the random variables are casted as the independent random variables with respect to their marginal probability distribution. The mean and standard deviation for all the random variables are listed in Table 6.

[1]The Truncated Gaussian distributed variables are truncated in $[0, \infty]$, however, the means and the standard deviations are given for the full Gaussian distributions

Table 6: Description of the random variables for the frame structure

Random variable	Distribution [1]	Mean	Standard deviation	Unit
P_1	Lognormal	133.454	40.04	kN
P_2	Lognormal	88.97	35.59	kN
P_3	Lognormal	71.175	28.47	kN
E_4	Truncated Gaussian	2.1738×10^7	1.9152×10^6	$\mathrm{kN\,m^{-2}}$
E_5	Truncated Gaussian	2.3796×10^7	1.9152×10^6	$\mathrm{kN\,m^{-2}}$
I_6	Truncated Gaussian	8.1344×10^{-3}	1.0834×10^{-3}	$\mathrm{m^4}$
I_7	Truncated Gaussian	1.1509×10^{-2}	1.2980×10^{-3}	$\mathrm{m^4}$
I_8	Truncated Gaussian	2.1375×10^{-2}	2.5961×10^{-3}	$\mathrm{m^4}$
I_9	Truncated Gaussian	2.5961×10^{-2}	3.0288×10^{-3}	$\mathrm{m^4}$
I_{10}	Truncated Gaussian	1.0812×10^{-2}	2.5961×10^{-3}	$\mathrm{m^4}$
I_{11}	Truncated Gaussian	1.4105×10^{-2}	3.4615×10^{-3}	$\mathrm{m^4}$
I_{12}	Truncated Gaussian	2.3279×10^{-2}	5.6249×10^{-3}	$\mathrm{m^4}$
I_{13}	Truncated Gaussian	2.5961×10^{-2}	6.4902×10^{-3}	$\mathrm{m^4}$
A_{14}	Truncated Gaussian	3.1256×10^{-1}	5.5815×10^{-2}	$\mathrm{m^2}$
A_{15}	Truncated Gaussian	3.7210×10^{-1}	7.4420×10^{-2}	$\mathrm{m^2}$
A_{16}	Truncated Gaussian	5.0606×10^{-1}	9.3025×10^{-2}	$\mathrm{m^2}$
A_{17}	Truncated Gaussian	5.5815×10^{-1}	1.1163×10^{-1}	$\mathrm{m^2}$
A_{18}	Truncated Gaussian	2.5302×10^{-1}	9.3025×10^{-2}	$\mathrm{m^2}$
A_{19}	Truncated Gaussian	2.9117×10^{-1}	1.0232×10^{-1}	$\mathrm{m^2}$
A_{20}	Truncated Gaussian	3.7303×10^{-1}	1.2093×10^{-1}	$\mathrm{m^2}$
A_{21}	Truncated Gaussian	4.1860×10^{-1}	1.9537×10^{-1}	$\mathrm{m^2}$

Table 7: Prediction of the failure probability and the reliability index for the frame by various methods

Method	N	P_f	95% CI (P_f)	β	95% CI (β)	P	n_{sparse}
MCS	10^5	1.51×10^{-3}	–	2.96	–	–	–
PCE	300	1.50×10^{-3}	–	2.97	–	2	–
LARS-PCE	300	1.54×10^{-3}	–	2.96	–	2	41
A-bPCE [22]	235	1.49×10^{-3}	$[1.42, 1.62]\times10^{-3}$	2.97	[2.94, 2.98]	2	30
SVB-PCE	200	1.52×10^{-3}	$[1.39, 1.67]\times10^{-3}$	2.96	[2.93, 2.99]	2	33

The QoI for this problem is the horizontal displacement at the top right corner of the frame (u) (see Figure 6) and the threshold value for the displacement is given by $\tau_u = 0.05$m. Consequently, the LSF for the frame is given by:

$$g(X) = \tau_u - u(X) \tag{67}$$

The failure probability and the reliability index are computed by MCS using 10^5 model evaluations. The model evaluations are performed with a MATLAB based FEM code. Similar to the previous examples, the results are assessed also by the full PCE, LARS-PCE and the SVB-PCE surrogate model.

The maximum degree of the full PCE model is considered as $P = 2$. As a consequence, a total $n = 253$ terms are found in the polynomial basis matrix for the full PCE model. The results obtained by all surrogate models are given in Table 7 along with the results reported in [22]. The failure probability and the reliability index are computed by the SVB-PCE model using $N = 200$ model evaluations. Whereas, the LARS-PCE and PCE require $N = 300$ model evaluations to assess an accurate result. A little higher number of model evaluations were used in [22] by the A-bPCE model as compared to the SVB-PCE model. Therefore, a fair reduction in the number of model evaluations is noticed. At the same time, n_{sparse} for the SVB-PCE model is also lower than that for the LARS-PCE model, and it is little higher than the number of terms selected by the A-bPCE model [22]. The efficiency is achieved in the two steps using the SVB-PCE model over the other surrogates. The 95% CI is also obtained from the predictive distribution parameters. It is observed that the CI for the SVB-PCE model is little higher than the A-bPCE model. Nevertheless, the required number of model evaluations is much less for the SVB-PCE model as compared to all other approaches reported in Table 7.

6. Conclusion

The computation of an accurate PDF for a QoI in the failure region is the most important aspect in the reliability analysis of a system. Many attempts have been made in the past to

estimate an accurate failure probability by different surrogate models [10, 12, 13, 15]. In this paper, a surrogate model has been proposed by combining the PCE with the sparse variational Bayesian (SVB) inference, which is called as SVB-PCE. The SVB framework is a combination of VB inference and ARD algorithm. Few drawbacks of the PCE model have been addressed in the SVB-PCE model. Mainly, the curse of dimensionality has been addressed using the ARD algorithm. Along with this, the SVB-PCE requires much less number of model evaluations for assessing an accurate failure probability.

The applicability of the SVB-PCE model has been illustrated through the reliability analysis of four typical numerical examples. The failure probability and the reliability index have been computed for all the examples using much less number of model evaluations by the SVB-PCE model as compared to the LARS-PCE and the full PCE model. For three examples, the results have also been compared with the results reported in [22] by the A-bPCE model. The required number of model evaluations for the A-bPCE model was higher than the SVB-PCE model for the three examples. Along with this, the SVB-PCE requires very few polynomial bases to assess an accurate result.

Appendix A Computation of VLB

The VLB must be computed for monitoring the convergence of the variational distribution $q(\Theta)$. The VLB is increased during the optimization procedure in each iteration [64]. Here, the VLB is computed using the factorized distribution. The VLB as introduced in Equation 23, is expressed by the factorized distribution:

$$\mathscr{L}\left[q\left(\boldsymbol{a},\varsigma,\boldsymbol{\alpha}\right)\right]=\iiint q\left(\boldsymbol{a},\varsigma,\boldsymbol{\alpha}\right)\ln\frac{p\left(\boldsymbol{Y}|\boldsymbol{\Psi},\boldsymbol{a},\varsigma,\boldsymbol{\alpha}\right)p\left(\boldsymbol{a},\varsigma|\boldsymbol{\alpha}\right)p\left(\boldsymbol{\alpha}\right)}{q\left(\boldsymbol{a},\varsigma,\boldsymbol{\alpha}\right)}d\boldsymbol{a}d\varsigma d\boldsymbol{\alpha} \tag{68}$$

However, the VLB is estimated using Equation 37 by expanding the terms of the equation which is given by:

$$\begin{aligned}\mathscr{L}\left[q\left(\Theta\right)\right]&=\mathbb{E}_{\Theta}\left[\ln p\left(\boldsymbol{Y},\Theta\right)\right]-\mathbb{E}_{\Theta}\left[\ln q\left(\Theta\right)\right] &(69)\\ &=\mathbb{E}_{\boldsymbol{a},\varsigma}\left[\ln p\left(\boldsymbol{Y}|\boldsymbol{\Psi},\boldsymbol{a},\varsigma\right)\right]+\mathbb{E}_{\boldsymbol{a},\varsigma,\boldsymbol{\alpha}}\left[\ln p\left(\boldsymbol{a},\varsigma|\boldsymbol{\alpha}\right)\right]+\mathbb{E}_{\boldsymbol{\alpha}}\left[\ln p\left(\boldsymbol{\alpha}\right)\right]\\ &\quad-\mathbb{E}_{\boldsymbol{a},\varsigma}\left[\ln q\left(\boldsymbol{a},\varsigma\right)\right]-\mathbb{E}_{\boldsymbol{\alpha}}\left[\ln q\left(\boldsymbol{\alpha}\right)\right] &(70)\end{aligned}$$

The terms as shown in Equation 70 can be found by taking moments with respect to the respective parameters. All the terms of Equation 70 can be found by taking the expectations over the previous derivations. In this way, the first term in Equation 70 is given by taking expectation of Equation 14 [35]:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{a},\varsigma}\left[\ln p\left(\boldsymbol{Y}|\boldsymbol{\Psi},\boldsymbol{a},\varsigma\right)\right]=&\frac{N}{2}\left(\psi\left(A_k\right)-\ln B_k-\ln 2\pi\right)\\ &-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{A_k}{B_k}\left(Y_i-\boldsymbol{\Psi}_i\boldsymbol{a}\right)^2+\boldsymbol{\Psi}_i\boldsymbol{V}_k\boldsymbol{\Psi}_i^T\right)\end{aligned} \tag{71}$$

where $\psi(\bullet)$ is the Digamma function. Similarly, the second term is computed by taking the expectation of Equation 17 [35] which is given by:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{a},\varsigma,\boldsymbol{\alpha}}\left[\ln p\left(\boldsymbol{a},\varsigma|\boldsymbol{\alpha}\right)\right] =& \frac{n}{2}\left(\psi\left(A_k\right)-\ln B_k+\psi\left(C_k\right)-\ln 2\pi\right)-B_0\frac{A_k}{B_k}\\ &-\frac{1}{2}\sum_{j=1}^{n}\left(\ln D_{k_j}+\frac{C_k}{D_{k_j}}\left(\frac{A_k}{B_k}a_{k_j}^2+V_{k_{jj}}\right)\right)\\ &-\ln\Gamma\left(A_0\right)+A_0\ln B_0+\left(A_0-1\right)\left(\psi\left(A_k\right)-\ln B_k\right)\end{aligned} \tag{72}$$

The rest of terms of Equation 70 are computed similarly by taking the expectations of Equation 20, 40 and 43 [35] and are given by:

$$\begin{aligned}\mathbb{E}_{\boldsymbol{\alpha}}\left[\ln p\left(\boldsymbol{\alpha}\right)\right] =& -n\left(\ln\Gamma\left(C_0\right)+C_0\ln D_0\right)\\ &+\sum_{j=1}^{n}\left(\left(C_0-1\right)\left(\psi\left(C_k\right)-\ln D_{k_j}\right)-D_0\frac{C_k}{D_{k_j}}\right)\end{aligned} \tag{73}$$

$$\begin{aligned}\mathbb{E}_{\boldsymbol{a},\varsigma}\left[\ln q_k\left(\boldsymbol{a},\varsigma\right)\right] =& \frac{n}{2}\left(\psi\left(A_k\right)-\ln B_k-\ln 2\pi-1\right)-\frac{1}{2}\ln\left|\boldsymbol{V}_k\right|-\ln\Gamma\left(A_k\right)\\ &+A_k\ln B_k+\left(A_k-1\right)\left(\psi\left(A_k\right)-\ln B_k\right)-A_k\end{aligned} \tag{74}$$

$$\mathbb{E}_{\boldsymbol{\alpha}}\left[\ln q_k\left(\boldsymbol{\alpha}\right)\right]=\sum_{j=1}^{n}\left(\left(C_k-1\right)\psi\left(C_k\right)+\ln D_{k_j}\right)-n\left(\ln\Gamma\left(C_k\right)+C_k\right) \tag{75}$$

By substituting from Equation 71 to 75 in Equation 70, the VLB is given by:

$$\begin{aligned}\mathscr{L}\left[q\left(\Theta\right)\right] =& -\frac{N}{2}\ln 2\pi+\frac{1}{2}\ln\left|\boldsymbol{V}_k\right|-B_0\frac{A_k}{B_k}-\frac{1}{2}\sum_{i=1}^{N}\left(\frac{A_k}{B_k}(Y_i-\boldsymbol{\Psi}_i\boldsymbol{a}_k)^2+\boldsymbol{\Psi}_i\boldsymbol{V}_k\boldsymbol{\Psi}_i^T\right)\\ &+\ln\Gamma\left(A_k\right)-A_k\ln B_k+A_k-\ln\Gamma\left(A_0\right)+A_0\ln B_0\\ &-\sum_{j=1}^{n}\left(C_k\ln D_{k_j}\right)+n\left(\frac{1}{2}-\ln\Gamma\left(C_0\right)+C_0\ln D_0+\ln\Gamma\left(C_k\right)\right)\end{aligned} \tag{76}$$

Appendix B Prediction distribution

The predictive distribution at the new samples $\boldsymbol{X}_{new}$ can be computed having the available informations $\mathcal{D}\in\{\boldsymbol{X},\boldsymbol{Y}\}$ [35]. Marginalizing over the parameters, the predictive distribution is given by:

$$p\left(\boldsymbol{Y}_{new}|\boldsymbol{\Psi}new,\mathcal{D}\right)= \iiint p\left(\boldsymbol{Y}_{new}|\boldsymbol{\Psi}_{new},\boldsymbol{a},\varsigma\right)p\left(\boldsymbol{a},\varsigma,\omega|\mathcal{D}\right)d\boldsymbol{a}d\varsigma d\omega \tag{77}$$

$$\approx \iiint p\left(\boldsymbol{Y}_{new}|\boldsymbol{\Psi}_{new},\boldsymbol{a},\varsigma\right)q\left(\boldsymbol{a},\varsigma\right)q\left(\omega\right)d\boldsymbol{a}d\varsigma d\omega \tag{78}$$

$$= \iint \mathcal{N}\left(\boldsymbol{Y}_{new}|\boldsymbol{\Psi}_{new}\boldsymbol{a}_k,\varsigma^{-1}\right)\mathcal{N}\left(\boldsymbol{a}|\boldsymbol{a}_k,\varsigma^{-1}h_k\right)\mathrm{Gam}\left(\varsigma|A_k,B_k\right)dad\varsigma \tag{79}$$

$$= \int \mathcal{N}\left(\boldsymbol{Y}_{new}|\boldsymbol{\Psi}_{new}\boldsymbol{a}_k,\varsigma^{-1}\left(1+\boldsymbol{\Psi}_{new}h_k\boldsymbol{\Psi}_{new}^T\right)\right)\mathrm{Gam}\left(\varsigma|A_k,B_k\right)d\varsigma \tag{80}$$

$$= \mathrm{St}\left(\boldsymbol{Y}_{new}|\boldsymbol{\mu},\boldsymbol{\lambda},\nu\right) \tag{81}$$

where the predictive distribution is Student's t-distribution (denoted by St) with the parameters $\boldsymbol{\mu}$, $\boldsymbol{\lambda}$ and ν. After constructing the SVB-PCE model, the distribution parameters are computed as:

$$\boldsymbol{\mu} = \boldsymbol{\Psi}_{new}\boldsymbol{a} \tag{82}$$

$$\boldsymbol{\lambda} = \frac{A_k}{B_k}\left(1 + \boldsymbol{\Psi}_{new} h_k \boldsymbol{\Psi}_{new}^T\right)^{-1} \tag{83}$$

$$\nu = 2A_k \tag{84}$$

Furthermore, the standard deviation of the predictive distribution is computed as:

$$\boldsymbol{\sigma} = \sqrt{(1 + \boldsymbol{\Psi}_{new} h_k \boldsymbol{\Psi}_{new}^T)\frac{B_k}{A_k - 1}} \tag{85}$$

Acknowledgements

Some fruitful suggestions from Prof. Eric Jacquelin and Dr. Denis Brizard are greatly acknowledged during the writing of this paper.

References

[1] S. K. Au, J. L. Beck, Estimation of small failure probabilities in high dimensions by subset simulation, Probabilistic Engineering Mechanics 16 (4) (2001) 263–277.

[2] S. Au, J. Ching, J. Beck, Application of subset simulation methods to reliability benchmark problems, Structural Safety 29 (3) (2007) 183–193.

[3] S. Engelund, R. Rackwitz, A benchmark study on importance sampling techniques in structural reliability, Structural Safety 12 (4) (1993) 255–276.

[4] S. K. Au, J. L. Beck, A new adaptive importance sampling scheme for reliability calculations, Structural Safety 21 (2) (1999) 135–158.

[5] H. J. Pradlwarter, G. I. Schuëller, P. S. Koutsourelakis, D. C. Charmpis, Application of line sampling simulation method to reliability benchmark problems, Structural Safety 29 (3) (2007) 208–221.

[6] R. Melchers, Structural system reliability assessment using directional simulation, Structural Safety 16 (1-2) (1994) 23–37.

[7] A. Olsson, G. Sandberg, O. Dahlblom, On Latin hypercube sampling for structural reliability analysis, Structural Safety 25 (1) (2003) 47–68.

[8] A. M. Hasofer, N. C. Lind, Exact and Invariant Second Moment Code Format, Journal of the Engineering Mechanics Division 100 (EMI) (1974) 111–121.

[9] R. Rackwitz, B. Fiessler, Structural reliability under combined random load sequences, Computers & Structures 9 (5) (1978) 489–494.

[10] G. Blatman, B. Sudret, An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis, Probabilistic Engineering Mechanics 25 (2) (2010) 183–197.

[11] C. Hu, B. D. Youn, Adaptive-sparse polynomial chaos expansion for reliability analysis and design of complex engineering systems, Structural and Multidisciplinary Optimization 43 (3) (2010) 1–24.

[12] I. Kaymaz, Application of kriging method to structural reliability problems, Structural Safety 27 (2) (2005) 133–151.

[13] C. G. Bucher, U. Bourgund, A fast and efficient response surface approach for structural reliability problems, Structural Safety 7 (1) (1990) 57–66.

[14] N. Gayton, J. Bourinet, M. Lemaire, CQ2RS: a new statistical approach to the response surface method for reliability analysis, Structural Safety 25 (1) (2003) 99–121.

[15] J.-M. Bourinet, F. Deheeger, M. Lemaire, Assessing small failure probabilities by combined subset simulation and Support Vector Machines, Structural Safety 33 (6) (2011) 343–353.

[16] R. Chowdhury, B. N. Rao, Assessment of high dimensional model representation techniques for reliability analysis, Probabilistic Engineering Mechanics 24 (1) (2009) 100–115.

[17] J. Deng, Structural reliability analysis for implicit performance function using radial basis function network, International Journal of Solids and Structures 43 (11-12) (2006) 3255–3291.

[18] J. B. Cardoso, J. R. de Almeida, J. M. Dias, P. G. Coelho, Structural reliability analysis using Monte Carlo simulation and neural networks, Advances in Engineering Software 39 (6) (2008) 505–513.

[19] N. Wiener, The homogeneous chaos, American Journal of Mathematics 60 (4) (1938) 897–936.

[20] D. Xiu, G. E. Karniadakis, The Wiener-Askey polynomial chaos for stochastic differential equation, SIAM Journal on Scientific Computing Scientific Computing 24 (2) (2002) 619–644.

[21] P. Kersaudy, B. Sudret, N. Varsier, O. Picon, J. Wiart, A new surrogate modeling technique combining Kriging and polynomial chaos expansions - Application to uncertainty analysis in computational dosimetry, Journal of Computational Physics 286 (2015) 103–117.

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

[23] B. Efron, T. Hastie, I. Johnstone, R. Tibshirani, Least angle regression, The Annals of Statistics 32 (2) (2004) 407–499.

[24] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, Journal of Computational Physics 230 (6) (2011) 2345–2367.

[25] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, IEEE Transactions on Signal Processing 56 (6) (2008) 2346–2356.

[26] S. D. Babacan, R. Molina, A. K. Katsaggelos, Bayesian compressive sensing using laplace priors, IEEE Transactions on Image Processing 19 (1) (2010) 53–63.

[27] K. Sargsyan, C. Safta, H. N. Najm, B. J. Debusschere, D. Ricciuto, P. Thornton, Dimensionality reduction for complex models via Bayesian compressive sensing, International Journal for Uncertainty Quantification 4 (1) (2014) 63–93.

[28] Q. Shao, A. Younes, M. Fahs, T. A. Mara, Bayesian sparse polynomial chaos expansion for global sensitivity analysis, Computer Methods in Applied Mechanics and Engineering 318 (2017) 474–496.

[29] P. Diaz, A. Doostan, J. Hampton, Sparse polynomial chaos expansions via compressed sensing and D-optimal design, Computer Methods in Applied Mechanics and Engineering 336 (2018) 640–666.

[30] W. Dai, O. Milenkovic, Subspace pursuit for compressive sensing signal reconstruction, IEEE Transactions on Information Theory 55 (5) (2009) 2230–2249.

[31] R. Baptista, V. Stolbunov, P. B. Nair, Some greedy algorithms for sparse polynomial chaos expansions, Journal of Computational Physics 387 (2019) 303–325.

[32] N. Lüthen, S. Marelli, B. Sudret, Sparse Polynomial Chaos Expansions: Literature Survey and Benchmark, Tech. rep. (2020). `arXiv:2002.01290v1`.

[33] H. Attias, A variational Bayesian framework for graphical models, in: 12th International Conference on Neural Information Processing Systems, MIT Press, 1999, pp. 209–215.

[34] Y. W. Teh, D. Newman, M. Welling, A Collapsed Variational Bayesian Inference Algorithm for Latent Dirichlet Allocation, in: 19th International Conference on Neural Information Processing Systems, 2006, pp. 1353–1360.

[35] C. M. Bishop, Pattern recognition and machine learning, Springer, 2006.

[36] M. I. Jordan, Z. Ghahramani, T. S. Jaakkola, L. K. Saul, An Introduction to Variational Methods for Graphical Models, Machine Learning 37 (2) (1999) 183–233.

[37] Z. Ghahramani, M. J. Beal, Propagation Algorithms for Variational Bayesian Learning, in: Advances in Neural Information Processing Systems, 2001, pp. 507–513.

[38] R. Koekoek, R. F. Swarttouw, The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue, Tech. Rep. 98-17, Delft University of Technology, Faculty of Information Technology and Systems, Department of Technical Mathematics and Informatics (1998).

[39] B. Bhattacharyya, A Critical Appraisal of Design of Experiments for Uncertainty Quantification, Archives of Computational Methods in Engineering 25 (3) (2018) 727–751.

[40] S. Salehi, M. Raisee, M. J. Cervantes, A. Nourbakhsh, An efficient multifidelity 1-minimization method for sparse polynomial chaos, Computer Methods in Applied Mechanics and Engineering 334 (2018) 183–207.

[41] C. W. Fox, S. J. Roberts, A tutorial on variational Bayesian inference, Artificial Intelligence Review 38 (2) (2012) 85–95.

[42] C. M. Bishop, M. Tipping, Variational Relevance Vector Machines, in: Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI2000), 2000, pp. 46–53.

[43] M. E. Tipping, Sparse Bayesian Learning and the Relevance Vector Machine, Journal of Machine Learning Research 1 (2001) 211–244.

[44] Q. Wan, H. Duan, J. Fang, H. Li, Z. Xing, Robust Bayesian compressed sensing with outliers, Signal Processing 140 (2017) 104–109.

[45] P. Green, Bayesian system identification of a nonlinear dynamical system using a novel variant of Simulated Annealing, Mechanical Systems and Signal Processing 52-53 (2015) 133–146.

[46] W. R. Jacobs, T. Baldacchino, T. J. Dodd, S. R. Anderson, Sparse Bayesian Nonlinear System Identification using Variational Inference, IEEE Transactions on Automatic Control 63 (12) (2018) 4172–4187.

[47] J. E. Griffin, P. J. Brown, Inference with normal-gamma prior distributions in regression problems, Bayesian Analysis 5 (1) (2010) 171–188.

[48] W. R. Gilks, S. Richardson, D. J. Spiegelhalter, Markov chain Monte Carlo in practice, Chapman & Hall, 1996.

[49] S. Sun, A review of deterministic approximate inference techniques for Bayesian machine learning, Neural Computing and Applications 23 (7-8) (2013) 2039–2050.

[50] I. M. Franck, P. Koutsourelakis, Sparse Variational Bayesian approximations for nonlinear inverse problems: Applications in nonlinear elastography, Computer Methods in Applied Mechanics and Engineering 299 (2016) 215–244.

[51] G. Parisi, Statistical field theory, Addison-Wesley, 1988.

[52] R. Peierls, On a Minimum Property of the Free Energy, Physical Review 54 (11) (1938) 918–919.

[53] S. Boyd, L. Vandenberghe, Convex Optimization, Cambridge University Press, 2004.

[54] M. J. Beal, Variational algorithms for approximate Bayesian inference, Ph.D. thesis, University College London (2003).

[55] F. R. Burden, M. G. Ford, D. C. Whitley, D. A. Winkler, Use of Automatic Relevance Determination in QSAR Studies Using Bayesian Neural Networks, Journal of Chemical Information and Modelling 40 (6) (2000) 1423–1430.

[56] Y. Li, C. Campbell, M. Tipping, Bayesian automatic relevance determination algorithms for classifying gene expression data, Bioinformatics 18 (10) (2002) 1332–1339.

[57] V. Y. F. Tan, C. Fevotte, Automatic Relevance Determination in Nonnegative Matrix Factorization with the /spl beta/-Divergence, IEEE Transactions on Pattern Analysis and Machine Intelligence 35 (7) (2013) 1592–1605.

[58] D. Wipf, S. Nagarajan, A New View of Automatic Relevance Determination, in: Advances in neural information processing systems, 2008, pp. 1625–1632.

[59] P. H. Waarts, Structural reliability using finite element methods: an appraisal of DARS: directional adaptive response surface sampling, Ph.D. thesis, Technical University of Delft (2000).

[60] K. Song, Y. Zhang, X. Zhuang, X. Yu, B. Song, An adaptive failure boundary approximation method for reliability analysis and its applications, Engineering with Computers (2020) 1–16.

[61] P.-L. Liu, A. Der Kiureghian, Optimization algorithms for structural reliability, Structural Safety 9 (3) (1991) 161–177.

[62] D. Wei, S. Rahman, Structural reliability analysis by univariate decomposition and numerical integration, Probabilistic Engineering Mechanics 22 (1) (2007) 27–38.

[63] R. Lebrun, A. Dutfoy, Do Rosenblatt and Nataf isoprobabilistic transformations really differ?, Probabilistic Engineering Mechanics 24 (4) (2009) 577–584.

[64] M. Svensén, C. M. Bishop, Robust Bayesian mixture modelling, Neurocomputing 64 (2005) 235–252.