



HAL
open science

Global sensitivity analysis: A Bayesian learning based polynomial chaos approach

Biswarup Bhattacharyya

► **To cite this version:**

Biswarup Bhattacharyya. Global sensitivity analysis: A Bayesian learning based polynomial chaos approach. *Journal of Computational Physics*, 2020, 415, 47p. 10.1016/j.jcp.2020.109539 . hal-02971331

HAL Id: hal-02971331

<https://hal.science/hal-02971331v1>

Submitted on 19 Oct 2020

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Global sensitivity analysis: A Bayesian learning based polynomial chaos approach

Biswarup Bhattacharyya^{a,*}

^aUniv Lyon, Université Claude Bernard Lyon 1, IFSTTAR, LBMC UMR_T9406, F69622, Lyon, France

Abstract

A novel sparse polynomial chaos expansion (PCE) is proposed in this paper for global sensitivity analysis (GSA). The proposed model combines variational Bayesian (VB) inference and automatic relevance determination (ARD) with the PCE model. The VB inference is utilized to compute the PCE coefficients. The PCE coefficients are obtained through a simple optimization procedure in the VB framework. On the other hand, the curse of dimensionality issue of PCE model is tackled using the ARD which reduces the number of polynomial bases significantly. The applicability of the proposed approach is illustrated by performing GSA on five numerical examples. The results show that the proposed approach outperforms a similar state-of-art surrogate model in obtaining an accurate sensitivity indices using limited number of model evaluations. For all the examples, the PCE models are highly sparse, which require very few polynomial bases to assess an accurate sensitivity indices.

Keywords: Sparse polynomial chaos expansion, Global sensitivity analysis, Bayesian approach, Automatic relevance determination, Sobol' indices

1. Introduction

Sensitivity analysis (SA) measures the effect of the input parameters on the output parameters for a system [1]. SA is globally divided in two distinct

*Corresponding author

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

categories: local SA [1] and global SA (GSA) [2]. The former measures the local effect of an input parameter on the response through the derivative of the output with respect to the input parameter [3]. On the other hand, GSA measures the effect of the whole input space on the output quantity considering uncertainties [1, 2]. GSA is the point of study for this paper. Various approaches are available in the literature for GSA such as regression based approach [3], variance based approach [4, 5], moment independent approach [6, 7] and moment based approach [8]. Out of all, the variance based approach has attained more popularity than the others. The variance based approach for GSA was proposed in [4, 5] for the first time. However, the main idea came from Pearson's study [9]. The main idea of the variance based approach is the computation of the contribution of variance in the output quantity for a single or, a combine effect of inputs [10]. According to Sobol' [5], the variance based approach is performed by decomposing the output variance into the single and the combine effect of the inputs. Often this method is known as the ANalysis Of VAriance (ANOVA) [11]. The approach as proposed by Sobol' [5, 2] is quite simple to use and is widely acknowledged to compute the sensitivity indices. To perform GSA for an uncertain system, multidimensional integrals over the input space can be computed by Monte Carlo simulation (MCS) [12] technique. However, the overburden computational cost of the MCS limits the technique in applying for the high-dimensional complex systems. For that reason, meta-model or, surrogate based approach has been investigated in the last decade for GSA which includes polynomial chaos expansion (PCE) [13, 14, 15], radial basis function [16, 17], Kriging [18], quasi random sampling-high dimensional model representation [19].

PCE has been used extensively in the last decade for GSA in different types of problems [13, 20, 21]. PCE was utilized by Sudret in [13] first time for GSA. The orthogonal property of PCE has reduced the computational burden substantially to compute the Sobol' sensitivity indices [2]. However, for the high-dimensional complex problems, PCE requires high degree polynomials which increases the number of polynomial bases exponentially. As a result, the

number of model evaluations for selecting a good enough PCE model increases drastically. Often all the terms in the polynomial bases do not contribute in the response quantity. In this context, several researchers have investigated PCE to reduce the number of polynomial bases [22, 23, 15, 24]. Therefore, a forward-backward regression based technique has been used in [14] to build up a sparse PCE for GSA. Further, the ℓ_1 minimization technique (least angle regression) has been utilized in [23] to efficiently capture the most important terms in the polynomial bases. A weighted ℓ_1 minimization approach has been proposed in [25] to compute the PCE coefficients sparsely based on some prior information about the PCE coefficients. Further, a gradient enhanced ℓ_1 minimization approach has been studied in [26] for computing the PCE coefficients and it was efficient as compared to the standard ℓ_1 minimization approach. Recently, the Kashyap information criterion based Bayesian sparse PCE is proposed for GSA [15]. An expanded sparse Bayesian learning approach has been investigated in [27] by allocating the priors only for the important polynomial bases. In [24], a support vector regression based sparse PCE has been proposed based on the variance contribution in the output and a partial least square based sparse PCE has been proposed in [28].

A sparse PCE model is developed in this paper by utilizing the variational Bayesian (VB) inference [29, 30] and automatic relevance determination (ARD) [31, 32]. More specifically, ARD is used to select the important polynomial bases and the coefficients of the PCE model are computed using the VB inference. The usefulness of the proposed approach is that ARD uses the results obtained from the VB inference to select the important polynomial basis functions. Hence, VB inference and ARD approach are fully connected to each other. The proposed approach is utilized in this paper for GSA. Explicitly, the Sobol' sensitivity indices [2] are computed for different types of problems using the proposed approach.

This paper is organized in the following way. The Sobol' sensitivity indices are explained in the next section and the PCE formulation is described in section 3. The procedure of computing the PCE coefficients using Bayesian

inference and the formulation of sparse PCE are explained in section 4. The procedure of computing the Sobol' sensitivity indices using the proposed approach is explained in section 5. Afterwards, the applicability of the proposed approach is illustrated through some numerical examples in section 6 and the conclusion of the present study is discussed in section 7.

2. Sobol' sensitivity indices

For GSA, Sobol' [2] proposed two different types of sensitivity indices: partial sensitivity indices (PSI) and total sensitivity indices (TSI). The former takes care of only the variance contribution due to a particular term in the Sobol' decomposition whereas the later considers the total sensitivity of a random variable including the interaction terms for the corresponding variable.

Consider d -dimensional independent input random variables in a vector $\Xi = \{\xi_1, \xi_2, \dots, \xi_d\}$ and a square integrable model response $Y = g(\Xi)$. For the independent random variables, the joint probability density function (PDF) may be represented as:

$$\rho(\Xi) = \prod_{i=1}^d \rho_{\xi_i}(\xi_i) \quad (1)$$

where ρ_{ξ_i} is the marginal PDF for the random variable ξ_i .

The model response is decomposed into the main effect and the composition effect of the input variables as [5, 2]:

$$\begin{aligned} Y = g(\Xi) = & g_0 + \sum_{i=1}^d g_i(\xi_i) + \sum_{1 \leq i < j \leq d} g_{i,j}(\xi_i, \xi_j) + \dots \\ & + \sum_{1 \leq i_1 < \dots < i_s \leq d} g_{i_1, \dots, i_s}(\xi_{i_1}, \dots, \xi_{i_s}) + \dots + g_{1, \dots, d}(\xi_1, \dots, \xi_d) \quad (2) \end{aligned}$$

where g_0 is the constant of the expansion which is also the mean of Y . One of the important properties of the decomposition is that the summand of the decomposition must satisfy:

$$\begin{aligned} \int_{\Omega_{\xi_k}} g_{i_1, \dots, i_s}(\xi_{i_1}, \dots, \xi_{i_s}) \rho_{\xi_k}(\xi_k) d\xi_k = 0 \\ 1 \leq i_1 < \dots < i_s \leq d; k \in \{i_1, \dots, i_s\} \quad (3) \end{aligned}$$

where Ω_{ξ_k} is the support of the random variable ξ_k . Along with this, the orthogonality condition is also required to satisfy for all the terms except the constant which is given by:

$$\int_{\Omega} g_{i_1, \dots, i_s}(\xi_{i_1}, \dots, \xi_{i_s}) g_{j_1, \dots, j_t}(\xi_{j_1}, \dots, \xi_{j_t}) d\Xi = 0 \quad (4)$$

$$\forall \{i_1, \dots, i_s\} \neq \{j_1, \dots, j_t\}$$

Each of the terms in Equation 2 can be found by:

$$\begin{aligned} g_0 &= \mathbb{E}(Y) \\ g_i(\xi_i) &= \mathbb{E}_{\Xi_{\sim i}}(Y|\xi_i) - \mathbb{E}(Y) \\ g_{i,j}(\xi_i, \xi_j) &= \mathbb{E}_{\Xi_{\sim i,j}}(Y|\xi_i, \xi_j) - \mathbb{E}(Y) \\ &\vdots \end{aligned} \quad (5)$$

where $\mathbb{E}(\bullet)$ is the expectation of the corresponding term. $\mathbb{E}_{\Xi_{\sim i}}$ is the expectation with respect to all the variables except ξ_i , and similarly, $\mathbb{E}_{\Xi_{\sim i,j}}$ is the expectation with respect to all the variables except ξ_i and ξ_j . Due to the orthogonal property between the model terms in Equation 2, the corresponding variance can be represented as:

$$V(Y) = \sum_{i=1}^d V_i + \sum_{1 \leq i < j \leq d} V_{i,j} + \dots + V_{1, \dots, d} \quad (6)$$

where $V(Y)$ is the total variance of the response quantity Y . V_i is the partial variance of the response due to the i -th variable and so on. Therefore, after dividing both sides of Equation 6 by the total variance, the partial sensitivity indices (PSI) are given by:

$$\sum_{i=1}^d S_i + \sum_{1 \leq i < j \leq d} S_{i,j} + \dots + S_{1, \dots, d} = 1 \quad (7)$$

In Equation 7, S_i is the PSI for the i -th variable and $S_{1, \dots, d}$ is the combined PSI for the interaction of the corresponding variables. Often the variance contribution of a single variable is computed (S_i) which is also known as the first order sensitivity index. More explicitly, the PSI can be written as [11]:

$$S_i = \frac{V_{\xi_i}(\mathbb{E}_{\Xi_{\sim i}}(Y|\xi_i))}{V(Y)} \in [0, 1] \quad (8)$$

where $V_{\xi_i} (\mathbb{E}_{\Xi_{\sim i}} (Y|\xi_i))$ measures the partial contribution of the variance on the output due to the variable ξ_i . On the other hand, TSI measures the whole effect of a variable on the output quantity [33]. TSI considers the PSI along with the interaction terms for a particular variable. TSI for a variable is given by:

$$S_{Ti} = \frac{\mathbb{E}_{\Xi_{\sim i}} (V_{\xi_i} (Y|\Xi_{\sim i}))}{V(Y)} = 1 - \frac{V_{\Xi_{\sim i}} (\mathbb{E}_{\xi_i} (Y|\Xi_{\sim i}))}{V(Y)} \quad (9)$$

where $V_{\Xi_{\sim i}} (\mathbb{E}_{\xi_i} (Y|\Xi_{\sim i}))$ is the variance contribution of all the terms other than ξ_i . PSI and TSI are investigated in the present work. For the MCS computation, the procedure proposed in [11] are utilized in this paper (refer Appendix A). Several other approaches can be found in [5, 33, 34, 35].

3. PCE representation

PCE is a surrogate based approach for approximating an unknown function. Having the d -dimensional random variable $\Xi = \{\xi_1, \xi_2, \dots, \xi_d\} \in \mathbb{R}^{N \times d}$ and the corresponding response quantities $Y = \{Y_1, Y_2, \dots, Y_N\}^T \in \mathbb{R}^{N \times 1}$ at N sample points, PCE is represented by [36]:

$$Y(\Xi) = \sum_{i=1}^{\infty} a_i \Phi^{(i)}(\Xi) \quad (10)$$

where a_i are the coefficients of the expansion and i represents the i -th term in the expansion. $\Phi^{(i)}$ are the multivariate orthogonal polynomial bases. The multivariate orthogonal polynomial bases are constructed by the product of the univariate orthogonal polynomial bases as:

$$\Phi^{(i)}(\Xi) = \prod_{j=1}^d \phi^{(i,j)}(\xi_j) \quad (11)$$

where $\phi^{(i,j)}$ is the i -th polynomial basis function for the j -th variable ξ_j . The univariate orthogonal polynomial bases are computed via the Askey scheme [37]. Some important polynomials for the commonly used random variables are listed in [36, 38]. The orthogonality condition for the multivariate polynomials is given by:

$$\mathbb{E} [\Phi^{(i)}(\Xi), \Phi^{(j)}(\Xi)] = h_n^2 \delta_{ij} \quad (12)$$

where h_n is a constant and $h_n = 1$ for the orthonormal polynomial. δ_{ij} is the Kronecker delta which is 1 for $i = j$ and zeros for the other cases.

Remark 1. This study uses the usual way of obtaining the orthogonal polynomial bases based on the known PDF of the input random variables. However, it may happen in practical cases that the PDFs of the random variables are unknown. In that case, the polynomial basis functions can be constructed using the Gram-Schmidt orthogonalization, however, these are not always ‘exact orthonormal’ [39]. For that reason, a near optimal orthonormal polynomial has been developed in [39] and this can be used in conjunction with the proposed approach in this paper.

Often PCE as defined in Equation 10, is terminated with finite number of terms in the expansion. Therefore, the degree of the orthogonal polynomial function is terminated at some finite degree p . The truncated PCE is given by:

$$Y(\Xi) = \sum_{i=1}^n a_i \Phi^{(i)}(\Xi) + \varepsilon_p \quad (13)$$

where ε_p is the residual of the truncated PCE, which is assumed as Gaussian white noise with zero mean and variance κ^{-1} . n is the finite number of terms in the expansion, which is computed as:

$$n = \frac{(d+p)!}{d!p!} \quad (14)$$

Therefore, the multivariate orthogonal polynomial of the truncated PCE and the coefficients are of dimensions $\Phi(\Xi) \in \mathbb{R}^{N \times n}$ and $a \in \mathbb{R}^{n \times 1}$, respectively.

Remark 2. It is evident from the previous study [40] that the orthogonal polynomials (e.g. Hermite polynomial) are often not stable and does not satisfy the condition of orthogonality (Equation 12) for high degree polynomials. For that reason, the orthonormal polynomials are used in this paper.

The coefficients in Equation 13 are often computed using the ordinary least square (OLS) approach [38] for the truncated PCE model. However, it has been found in the literature that all the terms in the truncated PCE model

underestimate or, overestimate the fitting [14]. Therefore, the most influenced terms in the polynomial bases are used in the later studies [14, 24, 28] to fit the PCE model accurately. In this regards, a fully Bayesian approach is utilized in this paper for computing the PCE coefficients. Along with this, a combination of Bayesian inference and ARD is used for selecting the most influenced terms in the polynomial bases.

4. Bayesian sparse PCE

Sparse PCE has already been proposed in many researches [14, 15, 24, 41] to substantially increase the computational efficiency and the accuracy. Bayesian inference has already gained interest in the machine learning domain [42, 43, 32, 29]. However, the application of Bayesian inference in the PCE formulation is very limited in the literature [15].

4.1. Bayesian inference in PCE

A generalized Bayesian inference is formulated in this section for the truncated PCE as defined in Equation 13 which can be written as:

$$Y(\Xi) = \Phi(\Xi) a + \varepsilon_p \quad (15)$$

where $\Phi(\Xi) \in \mathbb{R}^{N \times n}$ is the orthonormal polynomial basis matrix and $a \in \mathbb{R}^{n \times 1}$ is the PCE coefficient vector. The main aim of the Bayesian formulation is to infer the coefficient vector having the known response vector $Y(\Xi) \in \mathbb{R}^{N \times 1}$ and the polynomial basis matrix $\Phi(\Xi)$. The posterior of the Bayesian model parameter is inferred as:

$$p(\mathbf{H}|Y) = \frac{p(Y|\Phi, \mathbf{H}) p(\mathbf{H})}{p(Y)} \quad (16)$$

where \mathbf{H} is the Bayesian model parameter and $p(Y|\Phi, \mathbf{H})$ is the likelihood function of the Bayesian model. $p(\mathbf{H})$ is the prior of the Bayesian model and $p(Y)$ is the marginal likelihood which is expressed as:

$$p(Y) = \int p(Y|\Phi, \mathbf{H}) p(\mathbf{H}) d\mathbf{H} \quad (17)$$

For the estimation of the posterior in the Bayesian model, the likelihood function is inferred by the Gaussian distribution which is given by:

$$p(Y|\Phi, a, \kappa) = \prod_{i=1}^N p(Y_i|\Phi_i, a, \kappa) \quad (18)$$

$$= \prod_{i=1}^N \mathcal{N}(Y_i|\Phi_i a, \kappa^{-1}) \quad (19)$$

$$= \left(\frac{\kappa}{2\pi}\right)^{\frac{N}{2}} \exp\left(-\frac{\kappa}{2} \sum_{i=1}^N (Y_i - \Phi_i a)^2\right) \quad (20)$$

In Equation 19, the response is modelled by the Gaussian distribution with mean $\Phi_i a$ and variance κ^{-1} . i in the subscript denotes the i -th row of the corresponding matrix/vector. Each of the sample points is modelled as the product of the Gaussian distribution. As a result, the likelihood function is given by an exponential function. For that reason, the prior should be expressed by the exponential function to maintain conjugacy [30] in the formulation. A joint Gaussian-Gamma distribution is used to express the prior [44] in this formulation which is given by:

$$p(a, \kappa|\theta) = p(a|\kappa, \theta) p(\kappa) \quad (21)$$

$$= \mathcal{N}\left(a|0, (\kappa\Theta)^{-1}\right) \text{Gam}(\kappa|A_0, B_0) \quad (22)$$

$$= (2\pi)^{-\frac{n}{2}} |\Theta|^{\frac{1}{2}} \frac{B_0^{A_0}}{\Gamma(A_0)} \kappa^{\frac{n}{2} + A_0 - 1} \exp\left(-\frac{\kappa}{2} (a^T \Theta a + 2B_0)\right) \quad (23)$$

An absolute continuity in the formulation is noticed in Equation 23 with the likelihood function. A_0 and B_0 are the parameters of the Gamma distribution. The prior is further parameterized with a hyper-prior $\theta = \{\theta_1, \theta_2, \dots, \theta_n\}^T = \text{diag}(\Theta)$. $\Theta \in \mathbb{R}^{n \times n}$ is a sparse matrix having only the diagonal terms. The determinant of Θ is given by:

$$|\Theta| = \prod_{i=1}^n \theta_i \quad (24)$$

The hyper-prior in Equation 21 is formulated using the independent Gamma distribution to maintain the conjugacy in the formulation. Therefore, the hyper-

prior is given by:

$$p(\theta) = \prod_{j=1}^n \text{Gam}(\theta_j | C_0, D_0) \quad (25)$$

$$= \prod_{j=1}^n \frac{D_0^{C_0}}{\Gamma(C_0)} \theta_j^{C_0-1} \exp(-D_0 \theta_j) \quad (26)$$

where C_0 and D_0 are the parameters of the Gamma distribution, and $\Gamma(\bullet)$ represents the Gamma function. Therefore, it is clear from Equation 26 that the number of hyper-priors (n) is same with the number of terms in the truncated polynomial basis function. Consequently, a polynomial basis is directly dependent on the corresponding hyper-prior in the Bayesian formulation. Further, considering the joint probability distribution, the posterior of the Bayesian inference can be inferred as:

$$p(a, \kappa, \theta | Y) = \frac{p(Y | \Phi, a, \kappa) p(a | \kappa, \theta) p(\kappa) p(\theta)}{p(Y)} \quad (27)$$

The computation of the Bayesian model parameters would compute the PCE coefficients also. Therefore, the main objective of the Bayesian formulation is the computation of the Bayesian model parameter $\mathbf{H} \in \{a, \kappa, \theta\}$. However, the hyper-prior in the Bayesian formulation makes the inference intractable. For that reason, the Bayesian formulation is solved by approximating the marginal likelihood of the Bayesian inference. In general, the random sampling schemes [45, 46] can be used for such approximations. However, due to the large number of sample points, a variation Bayesian (VB) inference is used in the current formulation. The VB inference based PCE is formulated in the next section.

4.2. VB inference

4.2.1. Formulation of VB inference for Bayesian model parameter

The VB inference is utilized for the Bayesian formulation such that the posterior is inferred by approximating the marginal likelihood (Equation 17). The approximation is made by constructing a variational lower bound (VLB) on a variational distribution $q(\mathbf{H})$. Therefore, the VLB is given by [30]:

$$\mathcal{L}[q(\mathbf{H})] = \int_{\mathbf{H}} q(\mathbf{H}) \ln \frac{p(Y|\mathbf{H})p(\mathbf{H})}{q(\mathbf{H})} d\mathbf{H} \quad (28)$$

After manipulating and imposing the condition of the proper probability distribution with respect to the Bayesian model parameter ($\int_{\mathbf{H}} q(\mathbf{H}) d\mathbf{H} = 1$), the log-marginal likelihood function is expressed in term of the VLB as:

$$\ln p(Y) = \mathcal{L}[q(\mathbf{H})] - \int_{\mathbf{H}} q(\mathbf{H}) \ln \frac{p(\mathbf{H}|Y)}{q(\mathbf{H})} d\mathbf{H} \quad (29)$$

$$= \mathcal{L}[q(\mathbf{H})] + \text{KL}(q(\mathbf{H}) \parallel p(\mathbf{H}|Y)) \quad (30)$$

where $\text{KL}(\bullet)$ is the Kullback-Leibler (KL) divergence from q to p which is represented more explicitly as:

$$\text{KL}(q(\mathbf{H}) \parallel p(\mathbf{H}|Y)) = - \int_{\mathbf{H}} q(\mathbf{H}) \ln \frac{p(\mathbf{H}|Y)}{q(\mathbf{H})} d\mathbf{H} \quad (31)$$

The log-marginal likelihood is represented by the VLB and KL divergence in Equation 30. The approximate solution of the log-marginal likelihood is achieved either by minimizing the KL divergence [30, 47] or, by maximizing the VLB \mathcal{L} with respect to $q(\mathbf{H})$. To have an appropriate solution, the family of variational distribution is required to be restricted. For that reason, the variational distribution is partitioned using a factorized distribution [48, 49]. The Bayesian model parameters are represented by the factorized distribution as follows [30]:

$$q(\mathbf{H}) = \prod_{i=1}^{N_b} q(\mathbf{H}_i) \quad (32)$$

In Equation 32, N_b indicates the number of Bayesian model parameters. Considering the joint probability distribution, the variational distribution is depicted by the multiplication of the two distributions as follows:

$$q(\mathbf{H}) = q(a, \kappa) q(\theta) \quad (33)$$

The above-mentioned factorized distribution depicts $N_b = 2$. Therefore, the VLB is maximized with respect to each of the factorized distribution parameters for assessing an appropriate solution of the variational distribution. The factorized distribution based VLB is formulated in Appendix B: it suggests that the VLB is maximized when each of the Bayesian model parameters is approximated by Equation 74. The factorized VB inference based PCE is formulated in the next section.

4.2.2. Factorized VB inference based PCE

The PCE coefficients of the truncated PCE model are computed here by the factorized VB (FVB) formulation. The Bayesian model parameters $\mathbf{H} \in \{a, \kappa, \theta\}$ are computed using the FVB inference. For that reason, two components of the variational distribution (Equation 33) are formulated separately. The first component of the variational distribution $q(a, \kappa)$ is represented by Equation 74 as follows:

$$\ln q_r(a, \kappa) = \ln p(Y|\Phi, a, \kappa) + \mathbb{E}_\theta [\ln p(a, \kappa|\theta)] \quad (34)$$

Therefore, after substituting $\ln p(Y|\Phi, a, \kappa)$ and $\ln p(a, \kappa|\theta)$ from Equation 20 and Equation 23 respectively, Equation 34 is represented by:

$$\begin{aligned} \ln q_r(a, \kappa) &= \left(\frac{n}{2} + A_0 - 1 + \frac{N}{2} \right) \ln \kappa \\ &\quad - \frac{\kappa}{2} \left(a^T \left(\mathbb{E}_\theta [\Theta] + \sum_{i=1}^N \Phi_i^T \Phi_i \right) a + \sum_{i=1}^N Y_i^2 - 2 \sum_{i=1}^N Y_i \Phi_i a + 2B_0 \right) \\ &\quad + \text{const} \end{aligned} \quad (35)$$

The terms independent of a and κ are considered in the constant term. In a similar way to the formulation of Bayesian inference, the prior is formulated in the FVB inference using the conjugate Gaussian-Gamma distribution:

$$q_r(a, \kappa) = \mathcal{N}(a|a_r, \kappa^{-1}\chi_r) \text{Gam}(\kappa|A_r, B_r) \quad (36)$$

Now, Equation 36 can be compared with Equation 35 taking the natural logarithm on the both sides. Therefore, after expanding Equation 36 in a similar way to Equation 20 and comparing the coefficients of $-\frac{\kappa}{2}a^T a$ (between Equation 36 and Equation 35), the unknown term χ_r is given by:

$$\chi_r^{-1} = \sum_{i=1}^N \Phi_i^T \Phi_i + \mathbb{E}_\theta [\Theta] \quad (37)$$

The most important unknown term in this formulation a_r can be computed by comparing the coefficients of a between Equation 36 and Equation 35 as follows:

$$a_r = \chi_r \sum_{i=1}^N \Phi_i^T Y_i \quad (38)$$

In Equation 36, the part of Gaussian distribution can be computed having the updated a_r and χ_r . However, to update the variational distribution $q_r(a, \kappa)$, the updated parameters of the Gamma distribution are also required to be computed. For that reason, only expanding the part of Gamma distribution with natural logarithm, Equation 36 is given by:

$$\begin{aligned} \ln q_r(a, \kappa) &= \ln \mathcal{N}(a|a_r, \kappa^{-1}\chi_r) - \frac{\kappa}{2} \left(\sum_{i=1}^N Y_i^2 + 2B_0 - a_r^T \chi_r^{-1} a_r \right) \\ &\quad + \left(A_0 - 1 + \frac{N}{2} \right) \ln \kappa \end{aligned} \quad (39)$$

In Equation 36, κ follows Gamma distribution with parameters A_r and B_r . After equating the coefficients of κ between Equation 39 and the PDF of Gamma distribution, the parameter B_r is given by:

$$\begin{aligned} -\kappa B_r &= -\frac{\kappa}{2} \left(\sum_{i=1}^N Y_i^2 + 2B_0 - a_r^T \chi_r^{-1} a_r \right) \\ B_r &= B_0 + \frac{1}{2} \left(\sum_{i=1}^N Y_i^2 - a_r^T \chi_r^{-1} a_r \right) \end{aligned} \quad (40)$$

The other parameter A_r is computed by comparing the coefficients of $\ln \kappa$:

$$A_r = A_0 + \frac{N}{2} \quad (41)$$

Having the updated parameters of the respective distributions, the part of the variational distribution $q(a, \kappa)$ can be updated during the optimization procedure.

The second part of the variational distribution i.e. the hyper-prior $q(\theta)$ is also required to optimize for the variational distribution $q(\mathbf{H})$. Therefore, in a similar way to the first part, the second part is also given by Equation 74 as follows:

$$\ln q_r(\theta) = \ln p(\theta) + \mathbb{E}_{a, \kappa} [\ln p(a, \kappa|\theta)] \quad (42)$$

$$= \sum_{j=1}^n (C_0 - 1) \ln \theta_j - D_0 \theta_j + \frac{1}{2} \ln \theta_j - \frac{\theta}{2} \mathbb{E}_{a, \kappa} [\kappa a_j^2] + \text{const} \quad (43)$$

In Equation 42, $p(\theta)$ and $p(a, \kappa|\theta)$ are replaced by Equation 26 and Equation 23 respectively. In this case also, we follow the conjugacy in the formulation.

Therefore, the variational distribution for the hyper-prior is given by the Gamma distribution:

$$\ln q_r(\theta) = \sum_{j=1}^n \ln \text{Gam}(\theta_j | C_r, D_{r_j}) \quad (44)$$

Equation 44 can be compared with Equation 43. Therefore, the parameters of the Gamma distribution C_r and D_r are computed by comparing the coefficients of $\ln \theta_j$ and θ_j respectively. The parameters are given by:

$$C_r = C_0 + \frac{1}{2} \quad (45)$$

$$D_{r_j} = D_0 + \frac{1}{2} \mathbb{E}_{a,\kappa}[\kappa a_j^2] \quad (46)$$

The factorized distribution is often called the expectation-maximization (EM) formulation [50]. The expectation parts in Equation 37 and Equation 46 are computed through the standard moments [30] and are given by:

$$\mathbb{E}_\theta[\Theta] = \Theta_r \quad (47)$$

$$\mathbb{E}_{a,\kappa}[\kappa a_j^2] = a_{r_j}^2 \frac{A_r}{B_r} + \chi_{r_{jj}} \quad j = 1, \dots, n \quad (48)$$

where each term in the diagonal matrix Θ can be computed directly by using the updated Gamma distribution parameters as follows:

$$\mathbb{E}_\theta[\theta_j] = \frac{C_r}{D_{r_j}} \quad j = 1, \dots, n \quad (49)$$

and Equation 47 is expressed more explicitly as:

$$\mathbb{E}_\theta[\Theta] = \text{diag}(\mathbb{E}_\theta[\theta]) \quad (50)$$

The above-mentioned FVB formulation is used to compute the coefficients of the truncated PCE model. The FVB inference is ultimately optimizing the VLB. Therefore, a truncation criterion is required to be imposed for converging the optimization process. The error value for the VLB must satisfy the threshold value, which is defined as:

$$\varepsilon_{\mathcal{L}} = \frac{\mathcal{L}[q(\mathbf{H})]_r - \mathcal{L}[q(\mathbf{H})]_{r-1}}{\mathcal{L}[q(\mathbf{H})]_{r-1}} \times 100 \leq T_{\mathcal{L}} \quad (51)$$

where $T_{\mathcal{L}}$ is the threshold value for the error $\varepsilon_{\mathcal{L}}$. The threshold value for the VLB is decided as $T_{\mathcal{L}} = 0.001\%$ to have a good convergence. As a result, the VLB value is required to compute in each iteration during the optimization process. An explicit expression for computing the VLB is given in Appendix C.

4.3. ARD based FVB-PCE

The sparse PCE is formulated here using the Automatic Relevance Determination (ARD) [31, 51]. Therefore, the most important terms contributing in the stochastic response are identified here by the ARD. This identification leads the PCE more accurate as compared to the truncated PCE. For the sparse PCE formulation, the hyper-prior of the FVB formulation is utilized. It should be noted that the hyper-priors are computed corresponding to each of the terms in the PCE polynomial bases. Therefore, the diagonal terms in Θ are taken into account for the sparse PCE formulation.

Having Θ for the multivariate orthonormal polynomial basis matrix Φ , the one to one correspondence between the polynomial basis and the hyper-prior is achieved. The ARD values [52, 53] for all the multivariate orthonormal polynomial bases are computed as follows:

$$\lambda_{\beta} = \text{diag} (\Theta^{-1}) \quad (52)$$

where λ is the vector containing the ARD values for all the orthonormal polynomial bases. β in the subscript denotes the iteration number during the construction of the sparse PCE (refer Figure 1). Therefore, on converging the VLB \mathcal{L} , the ARD value is computed for the n number of orthonormal polynomial bases and a few orthonormal bases are pruned at this step which are below some threshold ARD value. The threshold ARD value is defined as:

$$\ln T_{\beta} = \min (\ln \lambda_{\beta}) + \frac{\max (\ln \lambda_{\beta}) - \min (\ln \lambda_{\beta})}{\epsilon} \quad (53)$$

where ϵ is an adjustable parameter for the threshold value. The threshold value (T) is computed taking the natural logarithm of the ARD values. Therefore, after the convergence of the VLB, the orthonormal polynomial bases having less

ARD values than the threshold ($\ln \lambda < \ln T$) are discarded. In the next iteration ($\beta + 1$), $n_{\beta+1} < n_\beta$ number of orthonormal polynomial bases are considered for the PCE and the corresponding coefficients are computed by the FVB formulation. This process continues until $n = 1$ orthonormal polynomial basis remains in Φ . Therefore, the selection of the most important sparse polynomial set is made by the VLB value. The one sparse set of orthonormal bases are selected which yields highest VLB value. The sparse orthonormal basis matrix and the corresponding PCE coefficients are given by:

$$\hat{\Phi} = \Phi_{\beta_m} \quad (54)$$

$$\hat{a} = a_{\beta_m} \quad (55)$$

$$\beta_m = \text{ind} \left(\max_{\beta} \mathcal{L}(\mathbf{H})_{\beta} \right) \quad (56)$$

where $\hat{\Phi}$ is the sparse orthonormal polynomial basis matrix and β_m is the index of β having the highest VLB value. Therefore, the final sparse PCE is constructed using the sparse polynomial basis matrix, which is regarded as AFVB-PCE (ARD based FVB-PCE) model.

Remark 3. To reduce the chances of pruning the minor important terms (but important for the PCE model), the natural logarithm of the ARD value is utilized in Equation 53.

Remark 4. The number of pruning terms in each iteration (β) is highly dependent on the adjustable parameter ϵ in Equation 53. A large number of terms may be pruned in each iteration with a low value of ϵ which may lead the AFVB-PCE model erroneous. Therefore, it is useful to set a high value for the adjustable parameter. $\epsilon = 1000$ is used in the present study.

5. AFVB-PCE based global sensitivity analysis

The AFVB-PCE model is used here for the computation of the Sobol' indices. The final AFVB-PCE model can be defined as:

$$\hat{Y} = \sum_{j=1}^{n_f} \hat{a}_j \hat{\Phi}_j \quad (57)$$

where n_f is the number of terms in the AFVB-PCE model which must be less than n (i.e. $n_f < n$). \hat{a}_j are the computed coefficients corresponding to the sparse polynomial bases $\hat{\Phi}_j$ and \hat{Y} is the approximate response. Therefore, with the analogy of Equation 2, the AFVB-PCE can be expressed more explicitly as:

$$\hat{Y} = \sum_{\substack{j=1 \\ 1 \leq i_1 \leq \dots \leq i_s \leq d}}^{n_f} \hat{a}_j \hat{\Phi}_j(\xi_{i_1}, \dots, \xi_{i_s}) \quad (58)$$

The multivariate polynomial bases $\hat{\Phi}_j(\xi_{i_1}, \dots, \xi_{i_s})$ are orthonormal. Therefore, from the orthonormal property, the total variance is given by:

$$V(\hat{Y}) = \sum_{j=2}^{n_f} \hat{a}_j^2 \quad (59)$$

and the partial contributing variance is given by:

$$V_{\xi_i} = \sum_{\substack{j=1 \\ p_{\xi_{i,j}} \neq 0 \& p_{\xi_{\sim i,j}} = 0}}^{n_f} \hat{a}_j^2 \quad (60)$$

where $p_{\xi_{i,j}} \neq 0 \& p_{\xi_{\sim i,j}} = 0$ defines the j -th multivariate polynomial corresponding to the j -th coefficient should not have the zero degree univariate polynomial for the i -th variable and should have the zero degree univariate polynomials for the other variables. Therefore, the PSI for each of the random variables can be computed by directly substituting Equation 59 and 60 in Equation 8 (replacing $V(Y)$ with $V(\hat{Y})$). Similarly, the TSI for each random variable is given by:

$$S_{T_i} = \frac{1}{V(\hat{Y})} \sum_{\substack{j=1 \\ p_{\xi_{i,j}} \neq 0}}^{n_f} \hat{a}_j^2 \quad (61)$$

The Sobol' sensitivity indices are computed by post-processing the coefficients from the AFVB-PCE model which requires very less computational cost. Therefore, an accurate prediction of the PSI and the TSI depends on the accuracy of the coefficients and the important terms in the AFVB-PCE model. A flowchart for computing the AFVB-PCE model and for computing the sensitivity indices is shown in Figure 1.

Remark 5. This study has been conducted considering the input random variables as independent. GSA for the correlated random variables is performed using ANalysis Of COVariance (ANCOVA) [54]. The computation of the covariance based sensitivity indices by PCE approach has been described in [55] for the correlated random variables.

6. Numerical examples

In this section, GSA is performed on five numerical examples using the AFVB-PCE model. Out of the five examples, the first two examples are non-linear test functions and the last three examples are engineering problems. All the results are compared with the MCS (see Appendix A) results which are considered as the benchmark solutions. Along with this, GSA is also performed using the sparse PCE [23, 56] model for all the examples and the results are compared with the proposed model. As the sparse PCE [23] model uses LARS algorithm, it is denoted by L-PCE in this paper. For the construction of the surrogate models, Sobol sequence [57] is utilized to generate the initial samples.

6.1. Ishigami function

Ishigami function [13] is one of the commonly used test functions for GSA. The functional form of this function is given by:

$$Y = \sin \xi_1 + a \sin^2 \xi_2 + b \xi_3^4 \sin \xi_1 \quad (62)$$

where ξ_i are the independent random variables uniformly distributed in $[-\pi, \pi]$. a and b are the constants which are considered as 7 and 0.1 respectively. The PSI and the TSI for the three random variables are computed here by the AFVB-PCE and L-PCE. The analytical expressions for computing Sobol' indices are also available for this problem [58]. The Sobol' indices are also computed by the MCS using 5×10^6 Sobol sequence samples (refer Appendix A). All the results are shown in Table 1.

First of all, the robustness of the distribution parameters used in the Bayesian formulation is investigated. For that reason, the PSI and the TSI are computed

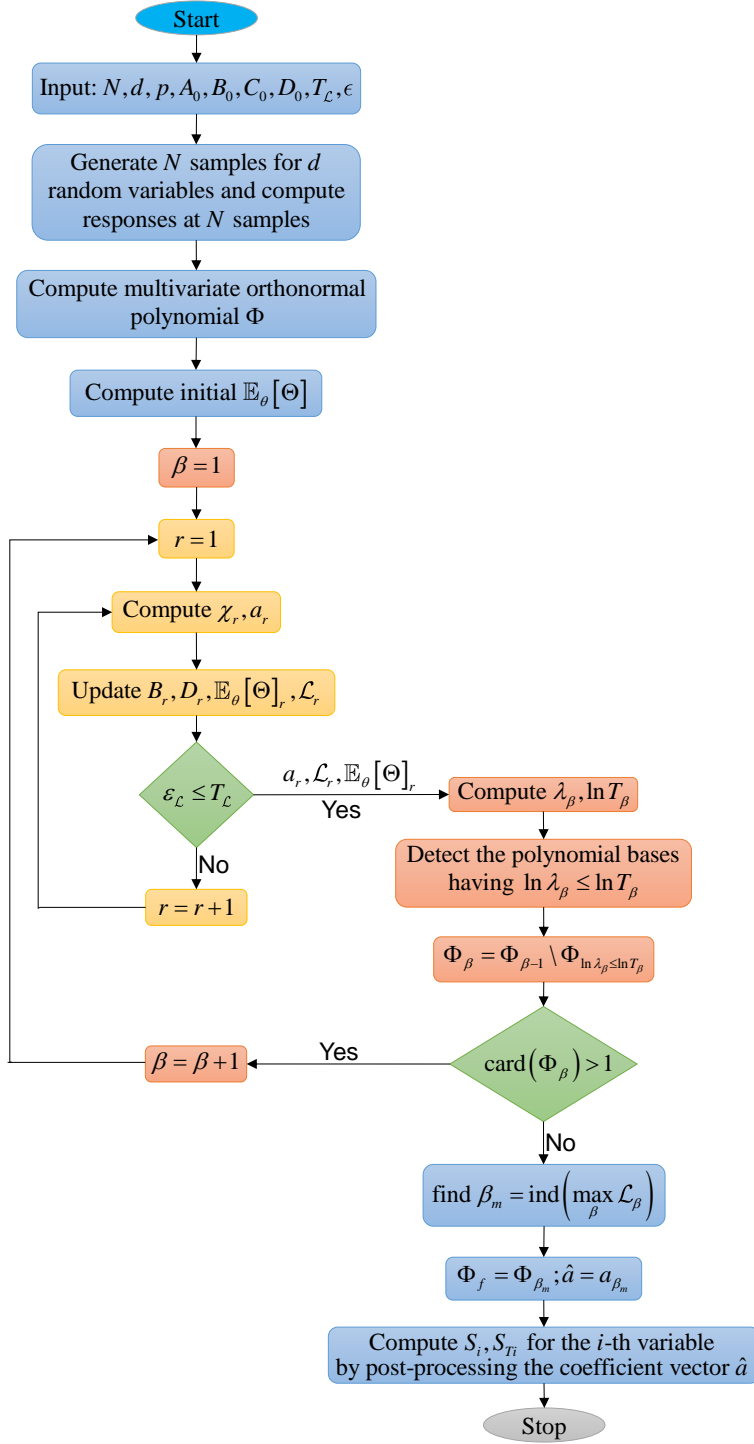


Figure 1: Flowchart for the construction of the AFVB-PCE model

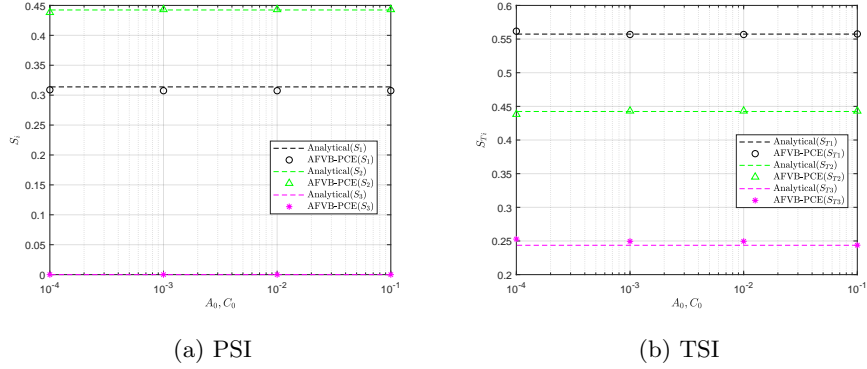


Figure 2: Predicted sensitivity indices for the Ishigami function with $B_0 = D_0 = 1 \times 10^{-4}$ and varying A_0, C_0

by the AFVB-PCE model using $N = 35$ and $p = 8$. The Legendre polynomial is used for computing the PCE bases. The distribution parameters are varied from 1×10^{-4} to 1×10^{-1} . The predicted sensitivity indices by the AFVB-PCE model along with the analytical results for all the variables are presented in Figure 2 and Figure 3. In Figure 2, $B_0 = D_0 = 1 \times 10^{-4}$ are constant, while A_0, C_0 are varied altogether. It is seen that there is no significant effect of A_0, C_0 on the predicted PSI and TSI: the sensitivity indices are predicted quite well by the AFVB-PCE model. On contrary, Figure 3 depicts an influence of B_0, D_0 on the predicted PSI and TSI for the variables ξ_1 and ξ_2 is noticed when $A_0 = C_0 = 1 \times 10^{-1}$ are constant. Indeed some discrepancy is noticed in the results predicted using $B_0 = D_0 = 1 \times 10^{-1}$ and $A_0 = C_0 = 1 \times 10^{-1}$. Although $A_0 = C_0 = 1 \times 10^{-2}$ and $B_0 = D_0 = 1 \times 10^{-4}$ has been suggested in [53] for system identification problem, the present study depicts $A_0 = C_0 = 1 \times 10^{-1}$ and $B_0 = D_0 = 1 \times 10^{-4}$ are reliable for the PCE model formulation and therefore, these values of the distribution parameters are used for all the examples in this paper.

A convergence study is then performed for both the surrogate models varying the number of samples with $p = 8$. The variation of the predicted PSI and TSI with the increase of sample points is shown in Figure 4. It is seen the PSI and

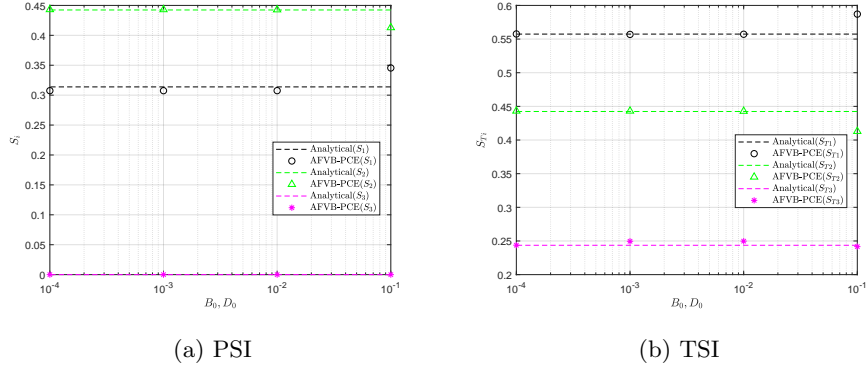


Figure 3: Predicted sensitivity indices for the Ishigami function with $A_0 = C_0 = 1 \times 10^{-1}$ and varying B_0, D_0

the TSI for all the variables are converged with the analytical results using less samples by the AFVB-PCE model ($N = 35$) than the L-PCE model ($N = 40$). It is noticeable that the predicted TSI for all the random variables are much accurate by the AFVB-PCE model with the increase of samples as compared to the L-PCE predicted results.

As the good results have been achieved by the AFVB-PCE model using $N = 35$ samples, the sensitivity indices computed by the surrogate models using $N = 35$ are listed in Table 1. The sensitivity indices are computed by the AFVB-PCE model using only $n_f = 13$ terms in the polynomial bases, while the full PCE model have $n = 165$ terms in the polynomial basis. Therefore, the PSI and the TSI for the Ishigami function are computed using only 13 PCE coefficients. On contrary, L-PCE requires $n_f = 14$ terms in the polynomial bases. It is seen clearly from the table that the PSI and the TSI for all the random variables are assessed using the AFVB-PCE more accurately as compared to the L-PCE using same number of model evaluations. The CPU time for all the approaches are also reported in Table 1: it is quite low for the MCS because of a mathematical function. Noticeably the L-PCE model requires less time as compared to the AFVB-PCE model, however, the CPU time for the AFVB-PCE model is not so high.

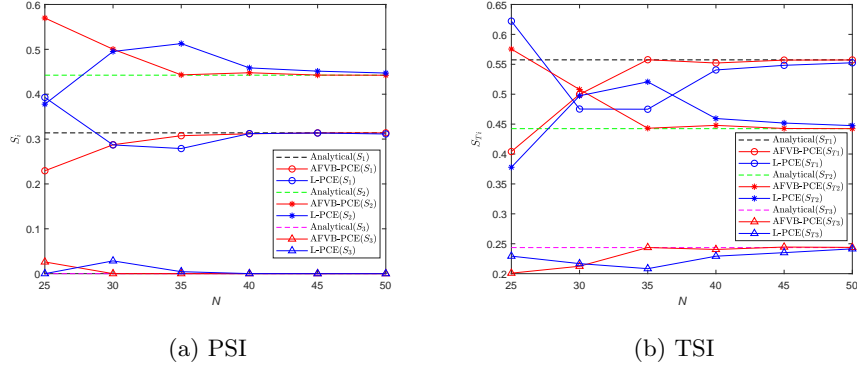


Figure 4: Convergence study for the predicted sensitivity indices by the surrogate models for the Ishigami function

Table 1: GSA results for the Ishigami function

| Sensitivity indices | Analytical results [58] | MCS | L-PCE [23] | AFVB-PCE |
|---------------------|-------------------------|-----------------|------------|---------------|
| S_1 | 0.3138 | 0.3139 | 0.2789 | 0.3075 |
| S_2 | 0.4424 | 0.4424 | 0.5128 | 0.4431 |
| S_3 | 0.0000 | 0.0000 | 0.0044 | 0.0000 |
| S_{T1} | 0.5574 | 0.5576 | 0.4749 | 0.5576 |
| S_{T2} | 0.4424 | 0.4424 | 0.5207 | 0.4424 |
| S_{T3} | 0.2436 | 0.2437 | 0.2084 | 0.2437 |
| N | — | 5×10^6 | 35 | 35 |
| p | — | — | 8 | 8 |
| n_f | — | — | 14 | 13 |
| CPU time (s) | — | 0.65 | 0.66 | 1.33 |

6.2. Sobol' function

The Sobol' function [59] is considered for the illustration as the second example. The function is given by:

$$Y = \prod_{i=1}^d \frac{|4\xi_i - 2| + a_i}{1 + a_i} \quad (63)$$

where ξ_i are the uniformly distributed random variables over $[0, 1]$ and a_i are the constants. For the present illustration, $d = 8$ is chosen. The constants are given by $\mathbf{a} = [1, 2, 5, 10, 20, 50, 100, 500]$ [13]. For this function, the analytical PSI and TSI are computed by the expressions given in [59, 13]. The MCS is also performed using 10×10^6 number of model evaluations to compute the sensitivity indices.

In a similar way to the previous example, a convergence study is performed for this problem also. The convergence of sensitivity indices for the first four random variables are shown in Figure 5. The sample points are varied from $N = 80$ to $N = 120$ at a step of 5 samples. The degree of the Legendre polynomial is selected as $p = 3$. Therefore, $n = 165$ terms are found in the PCE polynomial basis matrix. It is noticeable that a good results have been achieved by the AFVB-PCE model using $N = 90$ samples and a little divergence is noticed for S_2 after $N = 110$ sample points. On contrary, the discrepancies are much higher for the results predicted by the L-PCE model even with the increase of sample points.

The PSI and the TSI for all the random variables are listed in Table 2 using $N = 90$ samples. It is noticeable that a more accurate results are computed by the AFVB-PCE model as compared to the L-PCE model. Although L-PCE is capable to compute a desirable result for the second random variable (with $n_f = 9$), the results for the rest of the random variables are very poor. On contrary, AFVB-PCE model uses very less number of PCE coefficients ($n_f = 8$) and predicts much better results for all the random variables. This result is suggesting that an accurate PCE bases are selected using the ARD algorithm and the corresponding coefficients are computed more accurately using the VB

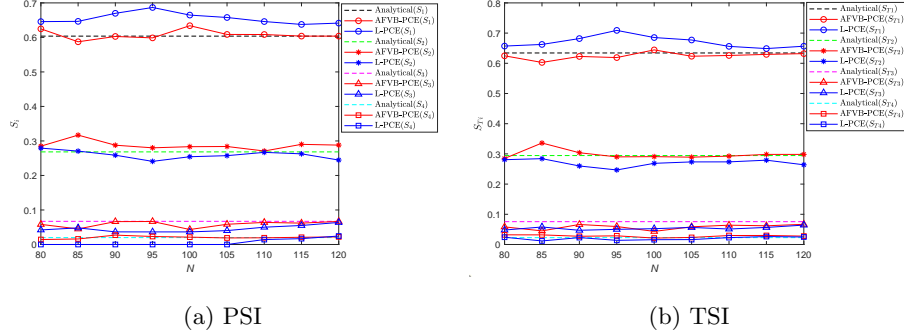


Figure 5: Convergence study for the predicted sensitivity indices of first four random variables by the surrogate models for the Sobol' function

inference. The CPU time for all the approaches are also obtained, which is quite low for both the surrogate models.

6.3. A roof truss

In this example, a roof truss [60] is considered for GSA. The roof truss is shown in Figure 6. The length of the truss is L and the height is $L/6$. A uniformly distributed load $qN m^{-1}$ is acting on the truss. The truss members have two different cross sectional areas A_c and A_s . The corresponding elastic modulus are E_c and E_s . All the parameters are considered normally distributed for the roof truss. All the random variables along with the parameters of the distribution are given in Table 3. The vertical deflection at point C is considered the response quantity for the roof truss. The response can be computed from the structural mechanics [60] which is given by:

$$Y = \Delta_C = \frac{qL^2}{2} \left(\frac{3.81}{A_c E_c} + \frac{1.13}{A_s E_s} \right) \quad (64)$$

For this problem, no analytical solution is available. Therefore, the reference solution is computed by the MCS using 8×10^6 model evaluations. A convergence study is also performed for this problem. The degree of the Hermite polynomial is chosen as $p = 3$. As a result, $n = 28$ terms are found in the full PCE model. The evaluation of the sensitivity indices with the increase of model evaluation

Table 2: GSA results for the Sobol' function

| Sensitivity indices | Analytical results [59] | MCS | L-PCE [23] | AFVB-PCE |
|---------------------|-------------------------|------------------|------------|---------------|
| S_1 | 0.6037 | 0.6038 | 0.6698 | 0.6027 |
| S_2 | 0.2683 | 0.2683 | 0.2587 | 0.2877 |
| S_3 | 0.0671 | 0.0671 | 0.0366 | 0.0663 |
| S_4 | 0.0200 | 0.0200 | 0.0000 | 0.0272 |
| S_5 | 0.0055 | 0.0055 | 0.0000 | 0.0028 |
| S_6 | 0.0009 | 0.0009 | 0.0000 | 0.0000 |
| S_7 | 0.0002 | 0.0002 | 0.0000 | 0.0000 |
| S_8 | 0.0000 | 0.0000 | 0.0000 | 0.0000 |
| S_{T1} | 0.6342 | 0.6342 | 0.6819 | 0.6227 |
| S_{T2} | 0.2945 | 0.2945 | 0.2600 | 0.3038 |
| S_{T3} | 0.0756 | 0.0756 | 0.0471 | 0.0663 |
| S_{T4} | 0.0227 | 0.0226 | 0.0229 | 0.0272 |
| S_{T5} | 0.0062 | 0.0062 | 0.0127 | 0.0103 |
| S_{T6} | 0.0011 | 0.0011 | 0.0172 | 0.0008 |
| S_{T7} | 0.0003 | 0.0002 | 0.0164 | 0.0058 |
| S_{T8} | 0.0000 | 0.0000 | 0.0054 | 0.0000 |
| N | – | 10×10^6 | 90 | 90 |
| p | – | – | 3 | 3 |
| n_f | – | – | 9 | 8 |
| CPU time (s) | – | 1.35 | 1.06 | 1.95 |

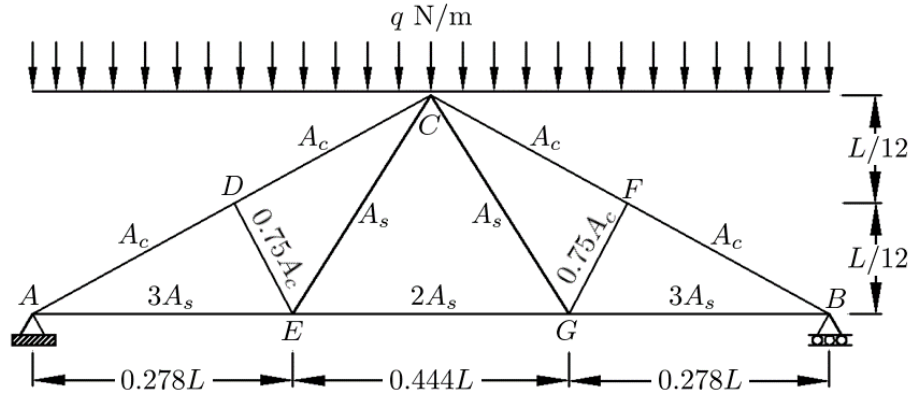


Figure 6: Geometrical view of the roof truss

Table 3: Description of the random variables for the roof truss

| Variable number | Random variable | Distribution | Mean | Coefficient of variation | Unit |
|-----------------|-----------------|--------------|-----------------------|--------------------------|-------------------|
| 1 | q | Normal | 2×10^4 | 0.08 | N m^{-1} |
| 2 | L | Normal | 12 | 0.02 | m |
| 3 | A_s | Normal | 9.82×10^{-4} | 0.06 | m^2 |
| 4 | A_c | Normal | 0.04 | 0.20 | m^2 |
| 5 | E_s | Normal | 2×10^{11} | 0.07 | N m^{-2} |
| 6 | E_c | Normal | 3×10^{10} | 0.08 | N m^{-2} |

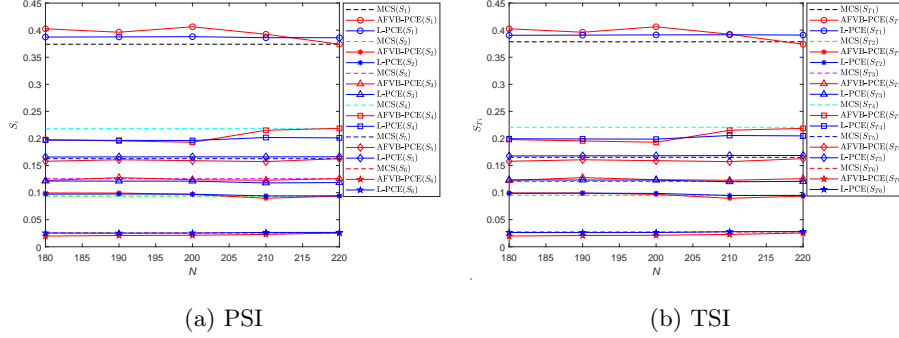


Figure 7: Convergence study of the predicted sensitivity indices by the surrogate models for the roof truss

is shown in Figure 7. It is noticeable that the sensitivity indices are almost constant with the increase of sample points by the L-PCE model. A good accuracy is achieved by the AFVB-PCE model for all the variables with the increase of model evaluations, however, the accuracies for the sensitivity indices of q and A_c have not increased by the L-PCE model.

As the AFVB-PCE model obtained a good result using $N = 220$ model evaluations, the sensitivity indices assessed by the L-PCE and the AFVB-PCE model are listed in Table 4 using $N = 220$. It is seen from the table that a higher accuracy is achieved by the AFVB-PCE model as compared to the L-PCE model for all the random variables. Along with this, ARD algorithm selects very small number of polynomial bases ($n_f = 8$) for assessing an accurate sensitivity indices. Therefore, the coefficients corresponding to the important polynomials are computed much accurately using the VB inference in the AFVB-PCE model. The CPU time for this example is significantly less by the AFVB-PCE model as compared to the other approaches.

From the GSA results it is clear that the PSI and the TSI for each of the random variables are very close to each other which suggest the interaction between the random variables are much less for this example. It is seen that q and A_c are highly sensitive to the response Δ_C . The sensitivities of $\{L, A_s, E_s\}$ are not so high, but are sensitive too. A low sensitivity is noticed for E_c .

Table 4: GSA results for the roof truss

| Sensitivity indices | MCS | L-PCE [23] | AFVB-PCE |
|---------------------|-----------------|------------|---------------|
| S_1 | 0.3747 | 0.3860 | 0.3740 |
| S_2 | 0.0935 | 0.0939 | 0.0935 |
| S_3 | 0.1189 | 0.1186 | 0.1258 |
| S_4 | 0.2178 | 0.2010 | 0.2186 |
| S_5 | 0.1632 | 0.1662 | 0.1627 |
| S_6 | 0.0257 | 0.0265 | 0.0254 |
| S_{T1} | 0.3785 | 0.3909 | 0.3747 |
| S_{T2} | 0.0950 | 0.0948 | 0.0935 |
| S_{T3} | 0.1205 | 0.1207 | 0.1215 |
| S_{T4} | 0.2206 | 0.2048 | 0.2190 |
| S_{T5} | 0.1651 | 0.1685 | 0.1630 |
| S_{T6} | 0.0274 | 0.0280 | 0.0254 |
| N | 8×10^6 | 220 | 220 |
| p | – | 2 | 2 |
| n_f | – | 20 | 8 |
| CPU time (s) | 6.25 | 0.65 | 0.12 |

Table 5: Description of the random variables for the HYMOD model

| Variable number | Random variable | Description | Distribution | Bounds |
|-----------------|-----------------|---------------------------------------|--------------|---------|
| 1 | S_M | Maximum soil moisture capacity (mm) | Uniform | [0,400] |
| 2 | β | Exponent in soil moisture routine (-) | Uniform | [0,2] |
| 3 | α | Partition coefficient (-) | Uniform | [0,1] |
| 4 | R_s | Slow reservoir coefficient (-) | Uniform | [0,0.1] |
| 5 | R_f | Fast reservoir coefficient (-) | Uniform | [0.1,1] |

6.4. A hydrological model

The hydrological models are often used to predict the run-off, climate change scenario and flood. The hydrological model considered in this example is called HYMOD model [61] which is used for the run-off simulation accounting several modules such as snow module, soil-moisture accounting module and routing module. The study area considered in the present study is the Leaf river catchment (1950 km²) [62] which is situated in the north of Collins, Mississippi, USA.

The HYMOD model has five uncertain parameters which affects the model performance such as flow. All the uncertain parameters are uniformly distributed and the distribution parameters of all the random variables are given in Table 5. This example aims at investigating the sensitivities of the random variables on the Nash-Sutcliffe efficiency. This example has also been studied in [24] for GSA. The HYMOD model is obtained by the SAFE toolbox [63] in this paper.

In a similar way to the previous examples, a convergence study is performed for the sensitivity indices. The reference results are obtained by the MCS using 7×10^5 model evaluations. The polynomial degree for the surrogate models is considered as $p = 4$. Hence, $n = 126$ terms are found in the full PCE model. The variation of the sensitivity indices with the model evaluation is shown in Figure 8. It is seen clearly that the accuracies of the predicted sensitivity indices by the AFVB-PCE model are increased with the increase of model evaluations. On contrary, the converging nature is not noticed by the L-PCE model.

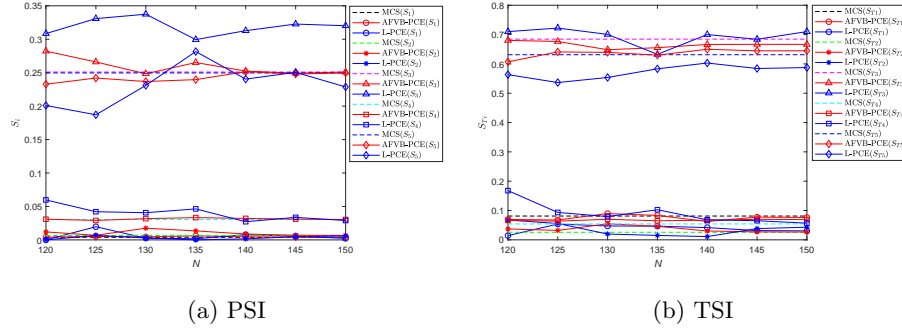


Figure 8: Convergence study of the predicted sensitivity indices by the surrogate models for the HYMOD model

All the GSA results for the HYMOD model by the surrogate models using $N = 150$ model evaluations are listed in Table 6. The PSI and the TSI are predicted quite well by the AFVB-PCE model using less number of significant terms in the polynomial basis as compared to the L-PCE model. On the other hand, L-PCE has predicted only S_2, S_4, S_5, S_{T4} quite well. The PSI and the TSI for the Partition coefficients and the fast reservoir coefficient are quite high as compared to the other random variables. It is also noticeable that the TSI for these two random variables are quite higher as compared to the PSI which suggest there is a strong significance of the interaction terms in the predicted TSI. The independent effect of S_M and β can be neglected due to much low PSI, however, the total effect of these two variable can not be neglected. Hence, the interaction terms have a strong impact for these two variables also. The CPU time for this problem is quite high by the MCS, whereas the CPU times by the surrogate models are significantly less.

6.5. A multi-storied frame structure

In the last example, a multi-storied frame structure is investigated for GSA. The frame is shown in Figure 9. Eight horizontal point loads are acting on the frame at the eight floors. All the loads follow gumbel distribution. The cross-sectional area of the each member is considered square and the sides of

Table 6: GSA results for the HYMOD model

| Sensitivity indices | MCS | L-PCE [23] | AFVB-PCE |
|---------------------|-----------------|------------|---------------|
| S_1 | 0.0061 | 0.0023 | 0.0039 |
| S_2 | 0.0072 | 0.0060 | 0.0067 |
| S_3 | 0.2698 | 0.3202 | 0.2510 |
| S_4 | 0.0298 | 0.0290 | 0.0305 |
| S_5 | 0.2399 | 0.2287 | 0.2496 |
| S_{T1} | 0.0811 | 0.0306 | 0.0763 |
| S_{T2} | 0.0257 | 0.0432 | 0.0267 |
| S_{T3} | 0.6844 | 0.7102 | 0.6665 |
| S_{T4} | 0.0542 | 0.0569 | 0.0690 |
| S_{T5} | 0.6313 | 0.5875 | 0.6447 |
| N | 7×10^5 | 150 | 150 |
| p | – | 4 | 4 |
| n_f | – | 45 | 43 |
| CPU time (s) | 505.49 | 1.23 | 0.97 |

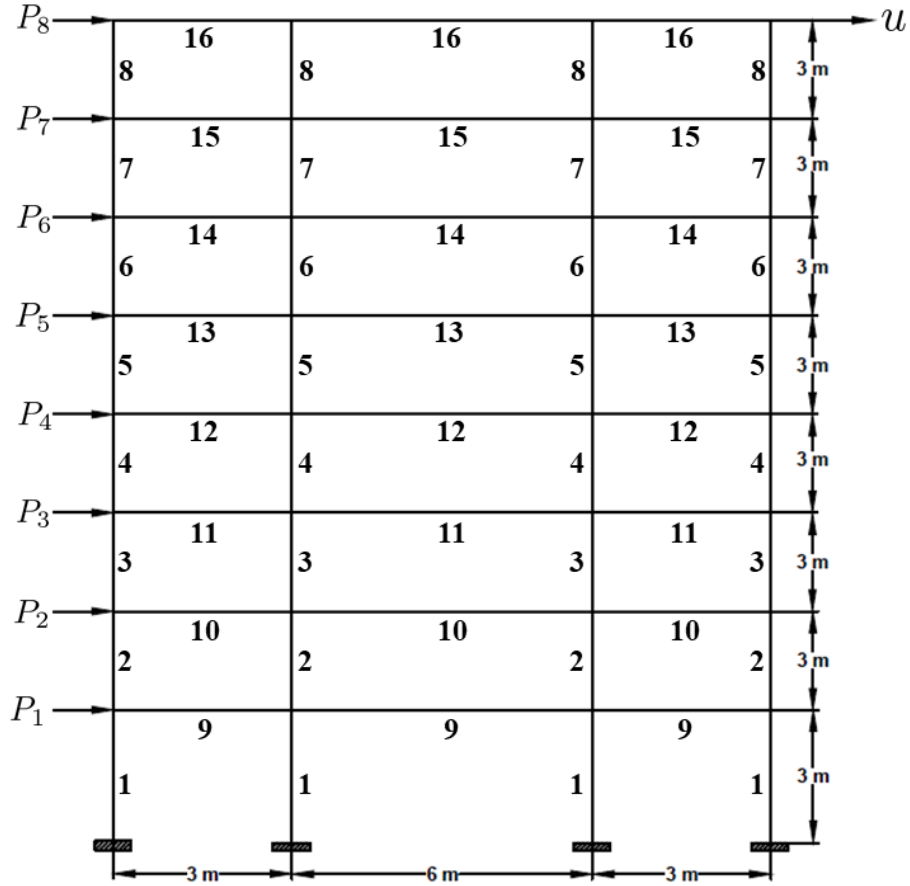


Figure 9: Geometrical view of the multi-storied frame structure

squares are considered lognormally distributed ($b_i; i = 1, 2, \dots, 16$ in Figure 9). The corresponding elastic modulus ($E_i; i = 1, 2, \dots, 16$) are also lognormally distributed. As a result, $d = 40$ random variables are present in this frame. All the random variables along with their distribution parameters are listed in Table 7.

For GSA, the response is considered as the horizontal deflection at the top right corner of the frame (u in Figure 9). The reference PSI and TSI are com-

¹The subscript in each random variable corresponds to the element number in Figure 9

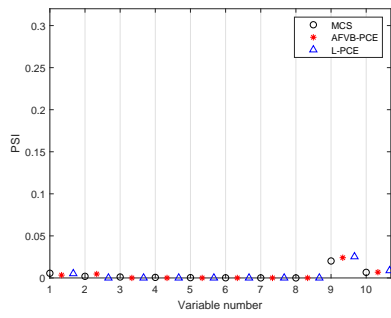
Table 7: Description of the random variables for the multi-storied frame structure

| Variable number | Random variable ¹ | Distribution | Mean | Standard deviation | Unit |
|-----------------|------------------------------|--------------|--------------------|--------------------|-------------------|
| 1, ..., 8 | E_1, \dots, E_8 | Lognormal | 3×10^{10} | 2×10^9 | N m^{-2} |
| 9, ..., 16 | b_1, \dots, b_8 | Lognormal | 0.4 | 0.015 | m |
| 17, ..., 24 | E_9, \dots, E_{16} | Lognormal | 3×10^{10} | 2×10^9 | N m^{-2} |
| 25, ..., 32 | b_9, \dots, b_{16} | Lognormal | 0.2 | 0.01 | m |
| 33 | P_1 | Gumbel | 1×10^4 | 3×10^3 | N |
| 34 | P_2 | Gumbel | 1.5×10^4 | 3.5×10^3 | N |
| 35, ..., 40 | P_3, \dots, P_8 | Gumbel | 2×10^4 | 4×10^3 | N |

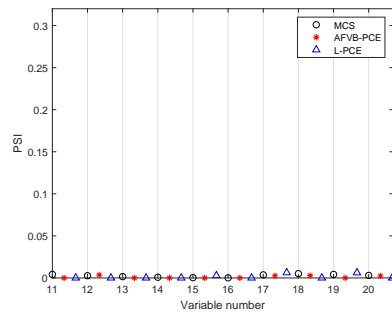
puted by the MCS using 42×10^6 model evaluations. Each of the models is evaluated using finite element method in a MATLAB framework.

For the surrogate models, the PSI and the TSI are computed using $N = 250$ model evaluations. The maximum degree of the polynomial is considered as $p = 2$. As a result, $n = 861$ polynomial bases are found in the truncated PCE model. The PSI and the TSI for all the random variables are plotted in Figure 10 and Figure 11 respectively. It is seen that the elastic modulus of all the members ($E_1 - E_{16}$) and the cross-sectional areas of the columns ($b_2 - b_8$) are less sensitive. Only b_2 is little sensitive for the lateral displacement. On the other hand, PSI and TSI for the cross-sectional areas of beams i.e. $\{b_9, b_{10}, b_{11}, b_{12}, b_{13}, b_{14}\}$ are sensitive to the lateral displacement, however, b_{15}, b_{16} are not so sensitive. The sensitivities of the loads increase with the height of the building. It is seen from the figures that the loads P_1 and P_2 are less sensitive, however, the other loads have higher sensitivity on the lateral displacement. It is also noticeable that the PSI and the TSI for each of the random variables are very close to each other. Hence, the dependency on the other random variables is much less for a TSI.

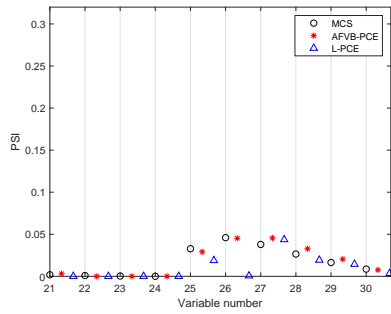
For the most of the random variables, the PSI and the TSI are computed more accurately by the AFVB-PCE model as compared to the L-PCE model. One important aspect is that the AFVB-PCE requires only $n_f = 30$ polynomial bases to compute the sensitivity indices. In contrast, L-PCE assesses the



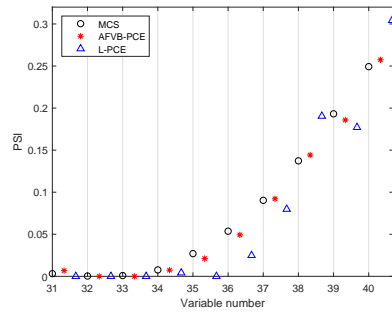
(a) Variable: 1-10



(b) Variable: 11-20

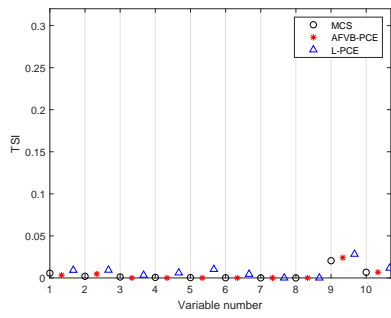


(c) Variable: 21-30

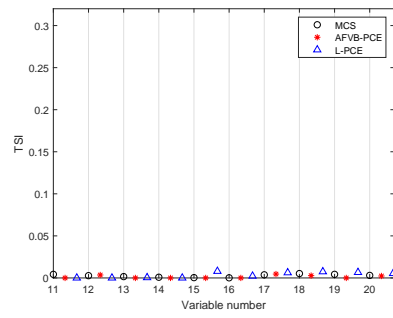


(d) Variable: 31-40

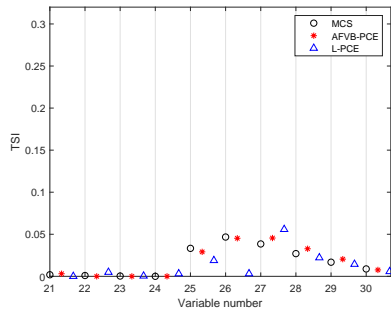
Figure 10: PSI for all random variables using different approaches



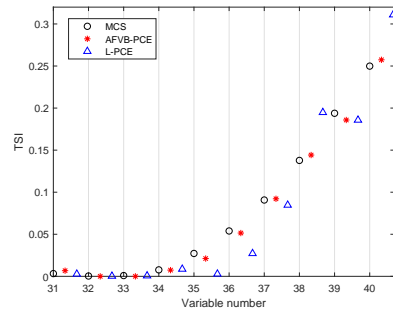
(a) Variable: 1-10



(b) Variable: 11-20



(c) Variable: 21-30



(d) Variable: 31-40

Figure 11: TSI for all random variables using different approaches

sensitivity indices using $n_f = 45$ polynomial bases and an erroneous result is noticed for the sensitivity indices of $b_{10}, P_3, P_4, P_5, P_7, P_8$. Therefore, the accuracy level in computing the sensitivity indices for this high-dimensional problem is maintained using a very few model evaluations and using much less polynomial bases by the AFVB-PCE model. The CPU time for this example by MCS is quite high (45 002.71 s). On contrary, the CPU time for the AFVB-PCE and the L-PCE are 2.67 s and 1.79 s, respectively.

7. Conclusion

A novel sparse methodology is developed for PCE to perform GSA in this paper. The main contributions of this paper are the computation of the PCE coefficients using VB inference and the selection of the important polynomial bases using ARD algorithm. The proposed approach is called AFVB-PCE. GSA is performed on five numerical examples to check the adequacy of the AFVB-PCE model. Out of all the examples, the first two are mathematical functions and the last three are the real life engineering problems. A promising accuracy has been achieved for all the numerical examples in assessing the PSI and the TSI using the AFVB-PCE model. Moreover, the sensitivity indices are predicted quite well using low order polynomials by the AFVB-PCE model for the real engineering problems. Along with this, the ARD algorithm selects very few polynomial bases for all the examples. For a high-dimensional system (having 40 random variables), AFVB-PCE requires only 3.48% terms in the polynomial bases to compute the PSI and the TSI for all the random variables. Therefore, the AFVB-PCE model selects the most useful polynomial bases for a system accurately.

Appendix A Computation of Sobol' indices by MCS

The MCS computation of the Sobol' indices is performed by the approach suggested in [11]. For the MCS, the approach proposed in [11] requires a single loop simulation for each variable to compute both the sensitivity indices (PSI

and TSI). However, a two set of independent random variable are required with N_{MCS} samples. Therefore, the two set of random variables can be denoted by $\Xi 1 \in \mathbb{R}^{N_{MCS} \times d}$ and $\Xi 2 \in \mathbb{R}^{N_{MCS} \times d}$. According to [11], $V_{\xi_i}(\mathbb{E}_{\Xi \sim i}(Y|\xi_i))$ in Equation 8 is given by:

$$V_{\xi_i}(\mathbb{E}_{\Xi \sim i}(Y|\xi_i)) = \frac{1}{N_{MCS}} \sum_{j=1}^{N_{MCS}} g(\Xi 2)_j \left[g(\Xi 1_{\Xi 2}^i)_j - g(\Xi 1)_j \right] \quad (65)$$

where j in the subscript represents the j -th simulation. $\Xi 1_{\Xi 2}^i$ denotes that all the variables are taken from the matrix $\Xi 1$ but the i -th variable is considered from the matrix $\Xi 2$. For computing the TSI, $\mathbb{E}_{\Xi \sim i}(V_{\xi_i}(Y|\Xi \sim i))$ is given by [11]:

$$\mathbb{E}_{\Xi \sim i}(V_{\xi_i}(Y|\Xi \sim i)) = \frac{1}{2N_{MCS}} \sum_{j=1}^{N_{MCS}} \left[g(\Xi 1)_j - g(\Xi 1_{\Xi 2}^i)_j \right]^2 \quad (66)$$

Therefore, substituting the expressions from Equation 65 and Equation 66 in Equation 8 and Equation 9 respectively, PSI and TSI can be computed with a single loop simulation process. The single loop simulation process requires $N_{MCS}(d+2)$ number of model evaluations for obtaining both the sensitivity indices.

Appendix B Formulation of VLB by factorized distribution

The VLB in Equation 28 is formulated using the factorized distribution as follows:

$$\mathcal{L}[q(\mathbf{H})] = \int_{\mathbf{H}} q(\mathbf{H}) (\ln p(Y, \mathbf{H}) - \ln q(\mathbf{H})) d\mathbf{H} \quad (67)$$

$$= \int \prod_{i=1}^{N_b} q(\mathbf{H}_i) \left(\ln p(Y, \mathbf{H}) - \sum_{i=1}^{N_b} \ln q(\mathbf{H}_i) \right) d\mathbf{H} \quad (68)$$

$$= \int q(\mathbf{H}_j) \left(\int \ln p(Y, \mathbf{H}) \prod_{i \neq j} q(\mathbf{H}_i) d\mathbf{H}_i \right) d\mathbf{H}_j - \int q(\mathbf{H}_j) \ln q(\mathbf{H}_j) d\mathbf{H}_j + \text{const} \quad (69)$$

$$= \int q(\mathbf{H}_j) \ln \tilde{p}(Y, \mathbf{H}_j) d\mathbf{H}_j - \int q(\mathbf{H}_j) \ln q(\mathbf{H}_j) d\mathbf{H}_j + \text{const} \quad (70)$$

$$= -\text{KL}(q(\mathbf{H}_j) \parallel \tilde{p}(Y, \mathbf{H}_j)) \quad (71)$$

In Equation 70, a new probability distribution is introduced ($\tilde{p}(Y, \Theta_j)$) which is also represented as [30]:

$$\ln \tilde{p}(Y, \mathbf{H}_j) = \mathbb{E}_{i \neq j} [\ln p(Y, \mathbf{H})] \quad (72)$$

where $\mathbb{E}_{i \neq j} [\bullet]$ defines the expectation with respect to all the distribution $q(\mathbf{H}_{i \neq j})$ which is a part in the first integral of Equation 69 i.e.

$$\mathbb{E}_{i \neq j} [\ln p(Y, \mathbf{H})] = \int \ln p(Y, \mathbf{H}) \prod_{i \neq j} q(\mathbf{H}_i) d\mathbf{H}_i \quad (73)$$

We need to maximize the VLB to have a good approximation for the variational distribution $q(\mathbf{H})$. The VLB is represented in Equation 71 by the negative KL divergence. Therefore, minimizing the KL divergence would maximize the VLB and the minimum of KL divergence occurs at $q(\Theta_j) = \tilde{p}(Y, \Theta_j)$. The optimum solution for the j -th distribution occurs at the minimum KL divergence for the j -th distribution, while the others $q(\mathbf{H}_{i \neq j})$ are fixed, which is given by:

$$\ln q_r(\mathbf{H}_j) = \mathbb{E}_{i \neq j} [\ln p(Y, \mathbf{H})] \quad (74)$$

The j -th Bayesian model parameter is approximated using Equation 74. The subscript r denotes the iteration number in the optimization procedure.

Appendix C Computation of VLB

The VLB is computed in each iteration during the optimization procedure to follow the convergence of the variational distribution $q(\mathbf{H})$. The VB inference for the PCE model is formulated using the factorized distribution in section 4.2.2. Therefore, it is important to formulate the VLB using the factorized distribution as well. The VLB in Equation 67 is formulated using Equation 74:

$$\mathcal{L}[q(\mathbf{H})] = \mathbb{E}_{\mathbf{H}}[\ln p(Y, \mathbf{H})] - \mathbb{E}_{\mathbf{H}}[\ln q(\mathbf{H})] \quad (75)$$

$$\begin{aligned} &= \mathbb{E}_{a, \kappa}[\ln p(Y|\Phi, a, \kappa)] + \mathbb{E}_{a, \kappa, \theta}[\ln p(a, \kappa|\theta)] + \mathbb{E}_{\theta}[\ln p(\theta)] \\ &\quad - \mathbb{E}_{a, \kappa}[\ln q(a, \kappa)] - \mathbb{E}_{\theta}[\ln q(\theta)] \end{aligned} \quad (76)$$

All the terms in Equation 76 are found by taking the moments from the previous expressions. The first term in Equation 76 is found by taking the moment of Equation 20 [30]:

$$\begin{aligned} \mathbb{E}_{a, \kappa}[\ln p(Y|\Phi, a, \kappa)] &= \frac{N}{2} (\psi(A_r) - \ln B_r - \ln 2\pi) \\ &\quad - \frac{1}{2} \sum_{i=1}^N \left(\frac{A_r}{B_r} (Y_i - \Phi_i a)^2 + \Phi_i \chi_r \Phi_i^T \right) \end{aligned} \quad (77)$$

where $\psi(\bullet)$ is the Digamma function. Similarly, the second component of Equation 76 is computed by taking the moment of Equation 23 [30]:

$$\begin{aligned} \mathbb{E}_{a, \kappa, \theta}[\ln p(a, \kappa|\theta)] &= \frac{n}{2} (\psi(A_r) - \ln B_r + \psi(C_r) - \ln 2\pi) - B_0 \frac{A_r}{B_r} \\ &\quad - \frac{1}{2} \sum_{j=1}^n \left(\ln D_{r_j} + \frac{C_r}{D_{r_j}} \left(\frac{A_r}{B_r} a_{r_j}^2 + \chi_{r_j j} \right) \right) \\ &\quad - \ln \Gamma(A_0) + A_0 \ln B_0 + (A_0 - 1) (\psi(A_r) - \ln B_r) \end{aligned} \quad (78)$$

The rest of the terms of Equation 76 are computed similarly by taking the moments of Equation 26, 36 and 39 [30] and are given by:

$$\begin{aligned} \mathbb{E}_{\theta}[\ln p(\theta)] &= -n (\ln \Gamma(C_0) + C_0 \ln D_0) \\ &\quad + \sum_{j=1}^n \left((C_0 - 1) (\psi(C_r) - \ln D_{r_j}) - D_0 \frac{C_r}{D_{r_j}} \right) \end{aligned} \quad (79)$$

$$\begin{aligned}\mathbb{E}_{a,\kappa} [\ln q_r (a, \kappa)] &= \frac{n}{2} (\psi (A_r) - \ln B_r - \ln 2\pi - 1) - \frac{1}{2} \ln |\chi_r| - \ln \Gamma (A_r) \\ &\quad + A_r \ln B_r + (A_r - 1) (\psi (A_r) - \ln B_r) - A_r\end{aligned}\quad (80)$$

$$\mathbb{E}_\theta [\ln q_r (\theta)] = \sum_{j=1}^n ((C_r - 1) \psi (C_r) + \ln D_{r_j}) - n (\ln \Gamma (C_r) + C_r) \quad (81)$$

The final expression for computing the VLB is given by substituting Equation 77 to 81 in Equation 76:

$$\begin{aligned}\mathcal{L} [q (\mathbf{H})] &= -\frac{N}{2} \ln 2\pi + \frac{1}{2} \ln |\chi_r| - \frac{1}{2} \sum_{i=1}^N \left(\frac{A_r}{B_r} (Y_i - \Phi_i a_r)^2 + \Phi_i \chi_r \Phi_i^T \right) \\ &\quad + \ln \Gamma (A_r) + A_r \left(1 - \ln B_r - \frac{B_0}{B_r} \right) - \ln \Gamma (A_0) + A_0 \ln B_0 \\ &\quad - \sum_{j=1}^n (C_r \ln D_{r_j}) + n \left(\frac{1}{2} - \ln \Gamma (C_0) + C_0 \ln D_0 + \ln \Gamma (C_r) \right)\end{aligned}\quad (82)$$

The VLB is computed in each iteration using Equation 82 during the optimization procedure. The convergence of the optimization procedure is observed through a threshold value of the VLB as defined in Equation 51.

References

- [1] E. Borgonovo, E. Plischke, Sensitivity analysis: A review of recent advances, *European Journal of Operational Research* 248 (3) (2016) 869–887.
- [2] I. M. Sobol, Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, *Mathematics and Computers in Simulation* 55 (13) (2001) 271–280.
- [3] J. C. Helton, Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal, *Reliability Engineering & System Safety* 42 (2-3) (1993) 327–367.
- [4] R. L. Iman, S. C. Hora, A Robust Measure of Uncertainty Importance for Use in Fault Tree System Analysis, *Risk Analysis* 10 (3) (1990) 401–406.

- [5] I. M. Sobol, Sensitivity analysis for nonlinear mathematical models, *Mathematical Modelling Computational Experiments* 1 (4) (1993) 407–414.
- [6] E. Borgonovo, A new uncertainty importance measure, *Reliability Engineering & System Safety* 92 (6) (2007) 771–784.
- [7] S. Chakraborty, R. Chowdhury, A hybrid approach for global sensitivity analysis, *Reliability Engineering & System Safety* 158 (2017) 50–57.
- [8] A. Dell’Oca, M. Riva, A. Guadagnini, Moment-based metrics for global sensitivity analysis of hydrological systems, *Hydrology and Earth System Sciences* 21 (12) (2017) 6219–6234.
- [9] K. Pearson, *On the general theory of skew correlation and non-linear regression*, Dulau and co., 1905.
- [10] B. Efron, C. Stein, The Jackknife Estimate of Variance, *The Annals of Statistics* 9 (3) (1981) 586–596.
- [11] A. Saltelli, P. Annoni, I. Azzini, F. Campolongo, M. Ratto, S. Tarantola, Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index, *Computer Physics Communications* 181 (2) (2010) 259–270.
- [12] I. M. Sobol, Quasi-Monte Carlo methods, *Progress in Nuclear Energy* 24 (1-3) (1990) 55–61.
- [13] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, *Reliability Engineering & System Safety* 93 (7) (2008) 964–979.
- [14] G. Blatman, B. Sudret, Efficient computation of global sensitivity indices using sparse polynomial chaos expansions, *Reliability Engineering & System Safety* 95 (11) (2010) 1216–1229.
- [15] 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.

- [16] J. Park, I. W. Sandberg, Universal Approximation Using Radial-Basis-Function Networks, *Neural Computation* 3 (2) (1991) 246–257.
- [17] Z. Wu, W. Wang, D. Wang, K. Zhao, W. Zhang, Global sensitivity analysis using orthogonal augmented radial basis function, *Reliability Engineering & System Safety* 185 (2019) 291–302.
- [18] Q. Ge, B. Ciuffo, M. Menendez, Combining screening and metamodel-based methods: An efficient sequential approach for the sensitivity analysis of model outputs, *Reliability Engineering & System Safety* 134 (2015) 334–344.
- [19] M. M. Zuniga, S. Kucherenko, N. Shah, Metamodelling with independent and dependent inputs, *Computer Physics Communications* 184 (6) (2013) 1570–1580.
- [20] K. Tang, P. M. Congedo, R. Abgrall, Adaptive surrogate modeling by ANOVA and sparse polynomial dimensional decomposition for global sensitivity analysis in fluid simulation, *Journal of Computational Physics* 314 (2016) 557–589.
- [21] S. Abraham, M. Raisee, G. Ghorbaniasl, F. Contino, C. Lacor, A robust and efficient stepwise regression method for building sparse polynomial chaos expansions, *Journal of Computational Physics* 332 (2017) 461–474.
- [22] G. Blatman, B. Sudret, Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach, *Comptes Rendus Mécanique* 336 (6) (2008) 518–523.
- [23] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, *Journal of Computational Physics* 230 (6) (2011) 2345–2367.
- [24] K. Cheng, Z. Lu, Adaptive sparse polynomial chaos expansions for global sensitivity analysis based on support vector regression, *Computers & Structures* 194 (2018) 86–96.

- [25] J. Peng, J. Hampton, A. Doostan, A weighted 1-minimization approach for sparse polynomial chaos expansions 267 (2014) 92–111.
- [26] J. Peng, J. Hampton, A. Doostan, On polynomial chaos expansion via gradient-enhanced 1-minimization, *Journal of Computational Physics* 310 (2016) 440–458.
- [27] Y. Zhou, Z. Lu, K. Cheng, Y. Shi, An expanded sparse Bayesian learning method for polynomial chaos expansion, *Mechanical Systems and Signal Processing* 128 (2019) 153–171.
- [28] Y. Zhou, Z. Lu, K. Cheng, Sparse polynomial chaos expansions for global sensitivity analysis with partial least squares and distance correlation, *Structural and Multidisciplinary Optimization* 59 (1) (2019) 229–247.
- [29] H. Attias, A variational Bayesian framework for graphical models, in: *12th International Conference on Neural Information Processing Systems*, MIT Press, 1999, pp. 209–215.
- [30] C. M. Bishop, *Pattern recognition and machine learning*, Springer, 2006.
- [31] 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.
- [32] Z. Ghahramani, M. J. Beal, Propagation Algorithms for Variational Bayesian Learning, in: *Advances in Neural Information Processing Systems*, 2001, pp. 507–513.
- [33] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, *Reliability Engineering & System Safety* 52 (1) (1996) 1–17.
- [34] M. J. Jansen, Analysis of variance designs for model output, *Computer Physics Communications* 117 (1-2) (1999) 35–43.

- [35] I. M. Sobol, Global sensitivity indices for the investigation of nonlinear mathematical models, *Matematicheskoe Modelirovanie* 19 (11) (2007) 23–24.
- [36] 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.
- [37] R. Koekoek, R. F. Swarttouw, The Askey-scheme of hypergeometric orthogonal polynomials and its q-analogue (1996).
- [38] B. Bhattacharyya, A Critical Appraisal of Design of Experiments for Uncertainty Quantification, *Archives of Computational Methods in Engineering* 25 (3) (2018) 727–751.
- [39] H. Lei, J. Li, P. Gao, P. Stinis, N. A. Baker, A data-driven framework for sparsity-enhanced surrogates with arbitrary mutually dependent randomness, *Computer Methods in Applied Mechanics and Engineering* 350 (2019) 199–227.
- [40] S. Rahman, WienerHermite polynomial expansion for multivariate Gaussian probability measures, *Journal of Mathematical Analysis and Applications* 454 (1) (2017) 303–334.
- [41] W. Zhao, L. Bu, Global sensitivity analysis with a hierarchical sparse meta-modeling method, *Mechanical Systems and Signal Processing* 115 (2019) 769–781.
- [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] J. E. Griffin, P. J. Brown, Inference with normal-gamma prior distributions in regression problems, *Bayesian Analysis* 5 (1) (2010) 171–188.
- [45] W. R. Gilks, S. Richardson, D. J. Spiegelhalter, *Markov chain Monte Carlo in practice*, Chapman & Hall, 1996.
- [46] S. Sun, A review of deterministic approximate inference techniques for Bayesian machine learning, *Neural Computing and Applications* 23 (7-8) (2013) 2039–2050.
- [47] 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.
- [48] G. Parisi, *Statistical field theory*, Addison-Wesley, 1988.
- [49] R. Peierls, On a Minimum Property of the Free Energy, *Physical Review* 54 (11) (1938) 918–919.
- [50] M. J. Beal, *Variational algorithms for approximate Bayesian inference*, Ph.D. thesis, University College London (2003).
- [51] V. Y. F. Tan, C. Fevotte, Automatic Relevance Determination in Nonnegative Matrix Factorization with the β -Divergence, *IEEE Transactions on Pattern Analysis and Machine Intelligence* 35 (7) (2013) 1592–1605.
- [52] D. Wipf, S. Nagarajan, A New View of Automatic Relevance Determination, in: *Advances in neural information processing systems*, 2008, pp. 1625–1632.
- [53] 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.

- [54] G. Li, H. Rabitz, P. E. Yelvington, O. O. Oluwole, F. Bacon, C. E. Kolb, J. Schoendorf, Global Sensitivity Analysis for Systems with Independent and/or Correlated Inputs, *The Journal of Physical Chemistry A* 114 (19) (2010) 6022–6032.
- [55] B. Sudret, Y. Caniou, Analysis of covariance (ANCOVA) using polynomial chaos expansions, in: *11th International Conference on Structural Safety and Reliability (ICOSSAR 2013)*, New York, USA, 2013, pp. 3275 – 3281.
- [56] S. Marelli, B. Sudret, UQLab: A Framework for Uncertainty Quantification in Matlab, in: *Second International Conference on Vulnerability and Risk Analysis and Management (ICVRAM)*, American Society of Civil Engineers, Liverpool, UK, 2014, pp. 2554–2563.
- [57] S. Galanti, A. R. Jung, Low-Discrepancy Sequences: Monte Carlo Simulation of Option Prices, *Journal of Derivatives* 5 (1997) 63–83.
- [58] T. Ishigami, T. Homma, An importance quantification technique in uncertainty analysis for computer models, in: *First international symposium on uncertainty modeling and analysis*, 1990, pp. 398–403.
- [59] I. Sobol', Theorems and examples on high dimensional model representation, *Reliability Engineering & System Safety* 79 (2) (2003) 187–193.
- [60] S. Kucherenko, S. Song, L. Wang, Quantile based global sensitivity measures, *Reliability Engineering & System Safety* 185 (2019) 35–48.
- [61] T. Wagener, D. P. Boyle, M. J. Lees, H. S. Wheater, H. V. Gupta, S. Sorooshian, A framework for development and application of hydrological models, *Hydrology and Earth System Sciences* 5 (1) (2001) 13–26.
- [62] S. Sorooshian, V. K. Gupta, J. L. Fulton, Evaluation of Maximum Likelihood Parameter estimation techniques for conceptual rainfall-runoff models: Influence of calibration data variability and length on model credibility, *Water Resources Research* 19 (1) (1983) 251–259.

- [63] F. Pianosi, F. Sarrazin, T. Wagener, A Matlab toolbox for Global Sensitivity Analysis, *Environmental Modelling and Software* 70 (2015) 80–85.