Available online at www.sciencedirect.com
ScienceDirect
Comput. Methods Appl. Mech. Engrg. 386 (2021) 114130
Computer methods in applied mechanics and engineering
www.elsevier.com/locate/cma

# A hybrid sequential sampling strategy for sparse polynomial chaos expansion based on compressive sampling and Bayesian experimental design

Bei-Yang Zhang, Yi-Qing Ni*

*Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region*

*Hong Kong Branch of the National Engineering Research Centre on Rail Transit Electrification and Automation, Hung Hom, Kowloon, Hong Kong Special Administrative Region*

Received 28 November 2020; received in revised form 18 August 2021; accepted 18 August 2021
Available online 7 September 2021

**Abstract**

Sparse representation of Polynomial Chaos Expansion (PCE) has been widely used in the field of Uncertainty Quantification (UQ) due to its simple model structure and low computational cost. The sample quality is a crucial issue that affects the precision of sparse PCE model. In this paper, a hybrid sequential sampling strategy is proposed to collect samples with high quality and in relatively small quantities for training PCE model. To achieve fast convergence rate and modelling stability, the proposed strategy takes into account both input information and target model feature by combining compressive sampling and Bayesian experimental design. First, a sequential sampling framework is established to collect samples that approximately match the coherence-optimal distribution, which is derived from the compressive sampling theory, during the iteration process. Then, by resorting to the Bayesian Compressive Sensing (BCS) method and information theory, favourable sampling points in each iteration are determined according to the modelling results, substituting for randomly selecting sampling points. The performance of the proposed sampling strategy is evaluated on several analytical functions through comparison with three input-dependent only sampling methods and two output-dependent only sampling methods. Results show that the proposed strategy outperforms the input-dependent only methods and has no worse performance than the output-dependent only methods in convergence rate and computational stability in most circumstances. The proposed strategy is further applied to two engineering cases for global sensitivity analysis of structural static and dynamic properties. It is illustrated that with automatically collected samples and observations, the PCE models can be obtained with desired accuracy, and the sensitivity analysis can be pursued with low computational cost and high precision.


*Keywords:* Polynomial chaos expansion (PCE); Bayesian compressive sensing; Sequential sampling; Coherence-optimal sampling; Bayesian experimental design

## 1. Introduction

Simulation models generated by computer to model physical phenomena or engineering systems have been made available for decades with the development of computer technology. More and more complex simulation models

* Corresponding author at: Department of Civil and Environmental Engineering, The Hong Kong Polytechnic University, Hung Hom, Kowloon, Hong Kong Special Administrative Region.

*E-mail address:* ceyqni@polyu.edu.hk (Y.Q. Ni).

https://doi.org/10.1016/j.cma.2021.114130

are built to attain analysis results with high accuracy, and these models can be used as reliable references to help solve practical problems. However, the simulation modelling result will inevitably have discrepancy to the real system observations. The uncertainty existent in the input parameters of a simulation model, which may induce estimation difficulty or inaccurate estimation of these parameters, is one of the causes, and this may produce erroneous judgement to the problems that we focus on [1]. An efficient strategy in dealing with such problems is to perform Uncertainty Quantification (UQ) on input parameters to explore the uncertainty propagation from the input parameters to the model output [2]. With the aid of UQ, further analysis on the input parameters or target system could be conducted, such as parameter sensitivity analysis and system reliability analysis [3–5]. In the past decades, numerous techniques to address the UQ problem have been developed, such as Monte Carlo Simulation (MCS) [6], perturbation method [7,8], first-order and second-order reliability methods (FORM/SORM) [9,10], and surrogate modelling approaches [2,11]. However, with the increasing demand on accuracy, the target simulation model becomes much more complicated, which is expensive and time-consuming to execute. Some of the methods, such as MCS method, are extremely costly and intractable when a huge number of samples and model observations are needed in exploring the UQ problem. In such situations, the surrogate modelling techniques have attracted considerable attention. By use of a simple mathematical model to map the relation between inputs and outputs, the surrogate modelling techniques can build a surrogate model as an alternative to a complex analytical or computational model, which is in general cheap to evaluate. If an adequate surrogate model is formulated, the model observations of interest can be obtained with low cost.

Polynomial Chaos Expansion (PCE) is one of surrogate modelling techniques, which has wide applications in the field of UQ due to its low training cost in modelling highly complex systems [12]. The basic idea behind PCE is to expand the model responses onto a series of polynomials. By calculating the PCE coefficients of each polynomial component, it could be trained to represent the model responses of interest as a polynomial function of the inputs. A significant property of these polynomials is that they are orthogonal to each other with respect to the joint probability density function of the inputs, which makes the computation simpler. Several methods were proposed to estimate the PCE coefficients, in which the non-intrusive method is more adaptive to most problems since no information about the governing equations is needed compared with the intrusive method [13]. This means that the target problem could be regarded as a 'black box' problem where observations can be collected using different samples of input variables. There are two main categories in the non-intrusive method, i.e., projection method [14] and regression method [15–17]; and the latter is more popular because of its efficiency in dealing with multivariate problems [18]. When applying the PCE technique to problems with high input dimension or high polynomial degree, the full expansion of the PCE model will involve massive unknown coefficients, thereby requiring more samples and observations in modelling [19,20]. It would be highly expensive for complex problems in performing the regression calculation. The recently emerging compressive sensing and sparse representation theory can help alleviate the "curse of dimensionality" issue that is a major drawback in applying the PCE technique. Owing to the "sparsity of effects" principle that most phenomenon-describing models are dominated by main effects and interactions of low order [21], real-world problems are generally considered compressible on polynomial chaos [22]. The sparse representation methods feature an appealing merit in garnering the PCE coefficients with less amount of model evaluations. In the compressive sensing framework, the sparsity is realised through modifying the regression equation by adding a regularisation term, such as $l_0$-norm or $l_1$-norm term [23,24]. A variety of algorithms have been developed to solve the sparse PCE, such as Least Angle Regression (LAR), Orthogonal Matching Pursuit (OMP), and Basis Pursuit Denoising (BPDN) [22,24–27]. These algorithms enable to recover a sparse approximation of PCE model with desired accuracy, while the number of samples and observations required for model training can be less than that of the unknown coefficients.

Clearly, the sample quality affects the solution quality when the sample quantity is restricted. In order to obtain a reliable sparse PCE model, it is crucial to choose a set of well-designed samples. The sample set is known as the Experimental Design (ED) [28–30]. In the context of PCE, many sampling strategies have been proposed, such as Monte Carlo [31], Latin Hypercube Sampling (LHS) [28,32], D-optimal design [33,34], and others [35–37]. A recent literature review of the sampling methods is available [38]. Typically, samples will have good quality (e.g., the corresponding observations could discover more comprehensive information about the target model) when they spread over the input domain in a space filling way [39,40]. MCS is a traditional way to generate random samples from a given distribution. However, when only a few samples are collected, the sample distribution may have a large discrepancy from the target distribution, that is, the samples will not be able to evenly spread across the probability space. LHS can overcome this drawback [41]. Even when the sample number is small, the LHS method can collect samples from the target distribution in a space filling way. To make the samples more contributive to the training of PCE model, D-optimal sampling strategy was proposed, which was originally designed for Ordinary Least Square (OLS) method [42]. To attain the smallest estimation uncertainty of coefficients in using the OLS method, samples are collected at those places which minimise the determinant of the information matrix. Thus, the training of PCE model with these samples will circumvent ill-conditioned regression matrix. Diaz et al. [34] further extended this method to $l_1$-minimisation problem by employing QR factorisation. Based on the compressive sensing theory, Hampton and Doostan [43] proposed a coherence-optimal sampling strategy, in which a lower bound of sample quantity which can address the $l_1$-minimisation problem with desired accuracy in a high likelihood was deduced. In their work, the input distribution was modified by multiplying a weight function to form a coherence-optimal distribution; thus the regression matrix would have the lowest coherence when sampling from this distribution, and the samples needed for best recovery of sparse representation would be the minimum.

Although the methods mentioned above were originated from different theories, they were designed to collect all samples at once. However, the best size (less but enough) of ED generally remains unknown when the target model information is not known *a priori*. It is hard to determine the size of ED in advance. Sequential sampling strategy, or called active learning, has been proposed to cope with this problem in the field of PCE through refinement of the non-sequential sampling methods [25,28,30]. The sequential sampling strategy begins with a small ED size, and then gradually adds new samples and the corresponding model evaluations to the current selection set until a predefined stop criterion is satisfied (e.g., the precision of the trained PCE model). This kind of strategy is flexible to help determine the ED size because the previously selected samples and the corresponding observations will be kept in the subsequent iterations, and the final sample size could be determined automatically. In the previous studies, LHS was coupled with sequential sampling to form a nested LHS method for the purpose of space filling [28]. D-optimal design was refined with sequential sampling to achieve a robust evaluation of PCE model [34].

The above sampling strategies, either sequential or non-sequential, were developed only based on the input distribution and polynomials, which we refer to as input-dependent only approaches. The target model feature is not considered. It results in the same sampling strategy for problems when their PCE models have identical input dimension, input distribution and polynomial truncation degree. However, it is common sense that a simple target model shall need fewer samples to train a PCE model with desired accuracy than a complex model. In view of this, ED method based on Bayesian regression has been developed to collect samples in the light of information about the experiment outcome [44], which is in general called Bayesian Experimental Design (BED) [45,46]. To apply BED in a sequential sampling framework, the PCE model is trained with samples and observations at each iteration by Bayesian modelling approaches. The information in the modelling results could be exploited by assessing the prediction uncertainty to instruct the next sampling position. Thus, the target model information in each sample and corresponding observation can be fully utilised to instruct the next sampling position under the sequential sampling framework. Zhou et al. [27] employed Sparse Bayesian Learning (SBL) to train the PCE model. In their study, each unknown coefficient was assigned with a Gaussian prior distribution with a unique hyperparameter, which induces an equivalent sparsity effect to the iteratively reweighted least-squares method [47]. At each iteration step of the sampling process, the trained PCE model could make predictions to unobserved points. Prediction variance and cross-validation error were obtained at each point, and the point which owns both large prediction variance and large cross-validation error was sampled in the next iteration. This strategy is designed to optimise the sampling process for both global exploration and local exploitation.

Nevertheless, BED is largely contingent on the modelling quality, and it will lose its effectiveness when the modelling results are highly inaccurate, i.e., at the beginning of sequential sampling process when samples are scarce. Under this circumstance, instructing sampling process based on the modelling results is likely to be incredible. To overcome this drawback and to make full use of both input and output information, a hybrid sequential sampling strategy is proposed in this paper by combining BED and compressive sampling theory. In the proposed method, a sequential sampling framework is first designed to collect samples from the coherence-optimal distribution. This framework provides restriction on the available sampling region in the input space in each iteration by employing the so-called Progressive Latin Hypercube Sampling (PLHS) method so that the collected samples overall will approximately match the coherence-optimal distribution with appealing space filling property. Making use of Bayesian compressive sensing (BCS) as a sparse representation approach to elicit the coefficients, BED is then pursued to determine the specific point to be collected, instead of random sampling, in the restricted sampling region in each iteration. Benefitting from the Bayesian modelling technique, the obtained PCE model in each iteration could make predictions to anywhere in the restricted sampling region, and the differential entropy is used to interpret the modelling results and quantify the uncertainty of the predicted results. The next sample position is determined at a point which has large prediction uncertainty. If the predefined stop criterion is satisfied, the sequential process terminates and samples at the current stage will be the best result. In the proposed hybrid sequential sampling strategy, the collected samples which overall approximately match the coherence-optimal distribution in each iteration could help pursue a small sample quantity for accurate recovery of $l_1$ minimisation, and the sample collected in each iteration contributes more to PCE modelling in the subsequent training stage. It can also be deemed that the sampling process in accordance with BED is restrained to make the distribution of overall samples in each iteration close to the coherence-optimal distribution. Even though the modelling results at the beginning of the sampling process might not be able to afford effective information to instruct the next sample, the whole sample set is always controlled to hold a favourable space filling property, and the regression equation can be formulated in good quality. The proposed sampling strategy will be compared with three input-dependent only methods and two output-dependent only methods by validation on analytical functions with different degrees of complexity. Additionally, two engineering cases will be presented to illustrate the applicability of the proposed method for global sensitivity analysis.

The rest of this paper is organised as follows. In Section 2, the basic knowledge about PCE is briefed for completeness of description, and the sparse regression method used to build a sparse PCE model is introduced. In Section 3, two popular sampling strategies related to this study are introduced. Then, the new hybrid sequential sampling strategy is proposed in Section 4, which is followed by comparison and validation with several case studies given in Section 5. Finally, conclusions are drawn in Section 6.

## 2. Polynomial chaos expansion and sparse representation

For completeness of description and unity of symbols, the principle of PCE is briefed here. PCE is a surrogate model technique that represents the scalar model output $y$ as an expansion of a set of orthogonal polynomials of input variables $\boldsymbol{\xi} = (\xi_1, \xi_2, \ldots \xi_d)$ with $d$ dimensions. Here the orthogonal polynomials defined in a probabilistic space can be expressed as:

$$\int \phi_i(\boldsymbol{\zeta})\phi_j(\boldsymbol{\zeta})p(\boldsymbol{\zeta})d\boldsymbol{\zeta} = w_i\delta_{ij} \quad (1)$$

where $\phi_i(\boldsymbol{\zeta})$, $\phi_j(\boldsymbol{\zeta})$ are two orthogonal polynomials of the random variables $\boldsymbol{\xi}$ with different degrees and $p(\boldsymbol{\zeta})$ is the probability density function of $\boldsymbol{\xi}$; $w_i$ is a constant; $\delta_{ij}$ is the Dirac function, which is equal to one when $i = j$ and otherwise zero. For computational convenience, the orthogonal polynomials are commonly normalised with respect to the probability density function so that $w_i$ is equal to one when $i = j$, namely orthonormal polynomials [48]. Thus, the PCE can be expressed as:

$$y = \sum_{\boldsymbol{\alpha}\in\mathbb{N}^d} c_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) \quad (2)$$

where $c_{\boldsymbol{\alpha}}$ are unknown coefficients; $\psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi})$ are multivariate orthonormal polynomials, which can be written as a tensor product of univariate polynomials when the input variables are assumed to be independent from each other:

$$\psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) = \prod_{i=1}^{d} \phi_{\alpha_i}(\xi_i) \quad (3)$$

and $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \ldots \alpha_d)$ represent the indices of the polynomial bases, in which $\alpha_i (i \in [1, d])$ denotes the degree of each independent variable $\xi_i$ in a polynomial term. Therefore, for different distributions, there are different types of orthogonal polynomials with respect to the probability density functions. For example, the Hermite polynomials are associated with the Gaussian distribution and the Legendre polynomials are orthogonal with respect to the uniform distribution. Some commonly used polynomial types are summarised in Table 1 [48]:

In practice, to facilitate calculation, the PCE representation in Eq. (2) will be truncated such that only a finite number of coefficients are kept in training. A general truncation strategy is to keep the total degree of PCE not exceeding a given degree $p$:

$$\mathcal{A}^{p,d} \equiv \left\{\boldsymbol{\alpha} \in \mathbb{N}^d : \|\boldsymbol{\alpha}\|_1 \le p\right\} \quad (4)$$

**Table 1**
Type of univariate orthogonal polynomials with different continuous variables.

| Random variable | Polynomial type | Support |
|---|---|---|
| Uniform | Legendre | $[a, b]$ |
| Gaussian | Hermite | $(-\infty, +\infty)$ |
| Beta | Jacobi | $[a, b]$ |
| Gamma | Laguerre | $[0, +\infty)$ |

and thus, the cardinality in the truncated PCE is:

$$P = \binom{d+p}{d} = \frac{(d+p)!}{d!p!} \tag{5}$$

It is obvious that the cardinality of PCE will suffer from the "curse of dimensionality" issue with the growing dimension or total degree. The model output $y$ is then expressed as a sum of the truncated PCE and a truncation error $\varepsilon$,

$$y = \sum_{\boldsymbol{\alpha} \in \mathcal{A}^{p,d}} c_{\boldsymbol{\alpha}} \, \psi_{\boldsymbol{\alpha}}(\boldsymbol{\xi}) + \varepsilon = \boldsymbol{\Psi}(\boldsymbol{\xi})\, \boldsymbol{c} + \varepsilon \tag{6}$$

where $\boldsymbol{\Psi}(\boldsymbol{\xi}) = [\psi_{\boldsymbol{\alpha}_1}(\boldsymbol{\xi}), \psi_{\boldsymbol{\alpha}_2}(\boldsymbol{\xi}), \ldots, \psi_{\boldsymbol{\alpha}_P}(\boldsymbol{\xi})]$ and $\boldsymbol{c} = [c_{\boldsymbol{\alpha}_1}, c_{\boldsymbol{\alpha}_2}, \ldots, c_{\boldsymbol{\alpha}_P}]^T$. By resorting to the regression methods (e.g., the OLS method), if an ED with $N$ samples from the random variables $\{\boldsymbol{\xi}^{(i)}\}_{i=1}^N = \{\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \ldots, \boldsymbol{\xi}^{(N)}\}$ is made and the corresponding model observations $\boldsymbol{Y} = \{\boldsymbol{y}^{(i)}\}_{i=1}^N$ at these sampled points are obtained, we can reformulate the above equation as:

$$\boldsymbol{Y} = \boldsymbol{\Psi}\boldsymbol{c} + \varepsilon \tag{7}$$

$$[\boldsymbol{y}^{(1)}, \boldsymbol{y}^{(2)}, \ldots, \boldsymbol{y}^{(N)}]^T = [\boldsymbol{\Psi}(\boldsymbol{\xi}^{(1)}), \boldsymbol{\Psi}(\boldsymbol{\xi}^{(2)}), \ldots, \boldsymbol{\Psi}(\boldsymbol{\xi}^{(N)})]^{\boldsymbol{T}} \boldsymbol{c} + \varepsilon \tag{8}$$

The OLS method seeks to obtain the solution of $\boldsymbol{c}$ by solving:

$$\hat{\boldsymbol{c}} = arg \min_{\boldsymbol{c}} \|\boldsymbol{\Psi}\boldsymbol{c} - \boldsymbol{Y}\|_2 \tag{9}$$

To get a unique solution, the number of samples $N$ should be larger or at least equal to the number of unknown coefficients, and an oversampling rate of 2∼3 times is recommended for obtaining reliable and robust results [49]. Due to the "curse of dimensionality" issue, the required model observations will grow exponentially with the increase of input dimension or PCE degree, which would be unaffordable for complex target models. Thus, the sparse representation methods that can solve Eq. (9) with sampling number far less than the number of unknown coefficients tend to be more promising. Some of them belong to $l_1$-norm regularisation methods and some have equivalent effect of $l_0$-norm regularisation. By placing penalty on the unknown coefficients, the regression problem in Eq. (9) can be redefined as the following optimisation problem:

$$\hat{\boldsymbol{c}} = arg \min_{\boldsymbol{c}} \left\{ \|\boldsymbol{\Psi}\boldsymbol{c} - \boldsymbol{Y}\|_2^2 + \lambda \|\boldsymbol{c}\| \right\} \tag{10}$$

where $\lambda$ is a regularisation coefficient which controls the penalty weight to the coefficients, and the norm in $\|\boldsymbol{c}\|$ can be either $l_0$, $l_1$ or $l_p (0 < p < 1)$ norm to induce different sparsity. Among them, $l_1$ minimisation is preferable since it can drive a sparse solution and meanwhile is more tractable than $l_0$- and $l_p$-norm regularisation [50]. In the present study, the BCS method proposed in [51], which has been demonstrated to endow a similar effect to $l_1$ minimisation through introducing a hierarchical form of Laplace priors to the coefficients, is recalled as the sparse regression procedure to solve Eq. (10). The principle of this method is briefly introduced in the following according to [51].

By assuming the truncation error follows a zero mean Gaussian distribution with variance equal to $\beta^{-1}$, the likelihood function can be formulated according to Eq. (7) as:

$$p(\boldsymbol{y}|\boldsymbol{c}, \beta) = N\left(\boldsymbol{y}|\boldsymbol{\Psi}\mathbf{c}, \beta^{-1}\right) \tag{11}$$

A conjugate prior, Gamma distribution $p(\beta)$, is placed on $\beta$ to facilitate the calculation. The $l_1$ regularisation formulation in Eq. (10) is equivalent to applying a Laplace prior on the coefficients $\boldsymbol{c}$. However, this setting is not a conjugate prior to the likelihood function for tractable Bayesian inference. Hence, hierarchical priors are applied to alleviate this problem [51,52]:

$$p(\boldsymbol{c}|\boldsymbol{\lambda}) = \prod_{i=1}^{P} N(c_i|0, \lambda_i) \tag{12}$$

$$p(\lambda_i|\kappa) = \Gamma(\lambda_i|1, \kappa/2) = \frac{\kappa}{2}\exp(-\frac{\kappa\lambda_i}{2}) \tag{13}$$

Likewise, $\kappa$ is modelled with a Gamma hyperprior $p(\kappa)$.

Now, the Bayesian inference can be deduced to calculate the posterior distribution of all the parameters:

$$p(\boldsymbol{c}, \boldsymbol{\lambda}, \kappa, \beta|\boldsymbol{y}) = \frac{p(\boldsymbol{y}|\boldsymbol{c}, \beta)\, p(\boldsymbol{c}|\boldsymbol{\lambda})\, p(\boldsymbol{\lambda}|\kappa)\, p(\beta)p(\kappa)}{p(\boldsymbol{y})} \tag{14}$$

Nevertheless, the calculation of the marginal likelihood $p(\boldsymbol{y})$ is analytically intractable, so an asymptotic approximation is employed to generate an iteration process to gradually approach the optimal parameter values [51,53]. To this end, the posterior distribution is decomposed as:

$$p(\boldsymbol{c}, \boldsymbol{\lambda}, \kappa, \beta|\boldsymbol{y}) = p(\boldsymbol{c}|\boldsymbol{\lambda}, \kappa, \beta, \boldsymbol{y})\, p(\boldsymbol{\lambda}, \kappa, \beta|\boldsymbol{y}) \tag{15}$$

The posterior distribution of the target coefficients $\boldsymbol{c}$ conditional on all other parameters and observations can be obtained from $p(\boldsymbol{c}|\boldsymbol{\lambda}, \kappa, \beta, \boldsymbol{y})$, which is a Gaussian distribution $N(\boldsymbol{c}|\mu_c, \Sigma_c)$ with:

$$\mu_c = \Sigma_c \beta \boldsymbol{\Psi}^T \boldsymbol{y} \tag{16}$$

$$\Sigma_c = \left(\beta \boldsymbol{\Psi}^T \boldsymbol{\Psi} + \mathrm{diag}(1/\lambda_i)\right)^{-1} \tag{17}$$

Consequently, the unknown parameters in Eqs. (16) and (17) can be estimated through maximising $p(\boldsymbol{\lambda}, \kappa, \beta|\boldsymbol{y})$, which can be alternatively pursued by maximising $p(\boldsymbol{y}, \boldsymbol{\lambda}, \kappa, \beta)$. To reduce the computational cost and speed up the iteration process in the asymptotic approximation, a fast Laplace algorithm was proposed [51]. The central idea is to update only a single $\lambda_i$ instead of updating the whole $\boldsymbol{\lambda}$ in each iteration so that the updates of $\mu_c$ and $\Sigma_c$ become efficient. The details of the algorithm can be found in [51].

At the beginning of the algorithm, the model will be set as empty, namely $\boldsymbol{\lambda} = \boldsymbol{0}$ [51]. So only $\beta$ needs to be assigned with a proper value to launch the algorithm. Even though a fixed value, $\beta^{-1} = 0.01\,\|\boldsymbol{y}\|_2^2$, was suggested in [51] for initialisation, it is found that different values of $\beta$ will affect the solution. Cross-Validation (CV) has been proven a good tool to provide proper choice of the hyperparameters in PCE, such as estimating the polynomial degree and the error tolerance [24,32], when no validation dataset is available. Here we choose CV technique to help determine a proper value for $\beta$ as initialisation, which was also adopted in [22]. The principle of the so-called $K$-fold CV is to randomly partition the training data into $K$ parts with equal size; one part is regarded as test data while the remaining $K-1$ parts are used as training data. Thus, the average predicted error on the test data could be evaluated. By taking each part as test data in turn, a total of $K$ average predicted errors can be obtained, and the average of these $K$ values is defined as the CV error. When applying the $K$-fold CV to choose a proper value of the model parameter, a bunch of values of the target parameter should be chosen in advance. For each value with given training data, a CV error will be estimated. The parameter value with the smallest CV error will be viewed as the best choice. In this study, a 10-fold CV method is used to select the initial value of the parameter $\beta$.

BCS is a sparse representation method that can provide high degree sparsity to solutions [51]. As such, the PCE model resulting from BCS can be expressed with a very simple structure. An appealing benefit of BCS compared with other regularisation algorithms is that, based on the modelling result, this method can make predictions to unknown points with a distribution instead of a deterministic way, which provides a reference of our confidence on the estimated value at a point. This property will be utilised in BED as described in the next section.

## 3. Coherence-optimal sampling and BED

In this section, two sampling schemes to be utilised in our method are outlined. One is the coherence-optimal sampling strategy originated from the compressive sampling theory, and the other is BED. It has been proven that the former can achieve the minimum sample quantity while preserving satisfactory recovery performance in solving $l_1$ minimisation problem [43]. This sampling method was compared with several input-dependent only sampling methods by verifying on several benchmark function tests in [22] and was shown promising to collect high-quality samples. The latter capitalises on the information provided by model observations through Bayesian modelling techniques, which has wide applicability to various problems [46].

### 3.1. Coherence-optimal sampling

The coherence parameter in the context of PCE is defined as [43]:

$$\mu(\boldsymbol{\xi}, \mathcal{A}^{p,d}) = \sup \max_{\boldsymbol{\alpha} \in \mathcal{A}^{p,d}} |\psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})|^2 \tag{18}$$

which has been testified to be a vital parameter of bounding the number of samples needed for accurate recovery of $l_1$ minimisation problem. A lower coherence parameter value represents a smaller bound of the sample number. Eq. (10) can be rewritten in a weighted form:

$$\hat{\boldsymbol{c}} = arg \min_{\boldsymbol{c}} \left\{ \|\boldsymbol{W\Psi c} - \boldsymbol{WY}\|_2^2 + \lambda \|\boldsymbol{c}\|_1 \right\} \tag{19}$$

Here we focus on $l_1$ minimisation problem. The weight matrix $\boldsymbol{W}$ is an identity matrix in Eq. (10) so that $\boldsymbol{W\Psi} = \boldsymbol{\Psi}$. The regression matrix $\boldsymbol{W\Psi}$ is controlled by the truncated polynomials at sampled points. To achieve the lowest sample bound, the concept of isotropy of regression matrix is introduced. The regression matrix in the context of PCE is isotropy if the ED is sampled from the input distribution, and the lowest sampling number could be achieved in a large probability for accurate reconstruction [43,54]. With the input variables $\boldsymbol{\xi}$ and the corresponding truncated polynomials, a coherence parameter value can be determined to bound the lowest sample number for accurate recovery of $l_1$ minimisation problem in a high likelihood. To attain this bound, the samples should be collected from the input distribution. Therefore, in target to recover an $l_1$ minimisation problem with satisfactory performance and as small number of samples as possible, one should reduce the coherence parameter value and sample from the corresponding distribution to make the regression matrix isotropic.

Now we consider a new polynomial basis which is modified from the original standard polynomials by multiplying a weight coefficient, where the coherence parameter of the new polynomial basis would be the minimum as defined below:

$$\mu_{min}(\boldsymbol{\gamma}, \mathcal{A}^{p,d}) = \sup \max_{\boldsymbol{\alpha} \in \mathcal{A}^{p,d}} |w(\boldsymbol{\zeta})\psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})|^2 \tag{20}$$

in which the weight coefficient is determined by:

$$w(\boldsymbol{\zeta}) = \frac{1}{cB(\boldsymbol{\zeta})} \tag{21}$$

where $B(\boldsymbol{\zeta}) = \max_{\boldsymbol{\alpha} \in \mathcal{A}^{p,d}} |\psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})|$. Then, new variables $\boldsymbol{\gamma}$ with distribution in compliance with the new orthogonal polynomial basis $\psi_{\boldsymbol{\alpha}}^{new}(\boldsymbol{\gamma}) = w(\boldsymbol{\zeta})\, \psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})$ are obtained. Thus, the weight matrix can be calculated from the weight coefficient:

$$\boldsymbol{W}(i, i) = w\left(\gamma^{(i)}\right) \tag{22}$$

where $\gamma^{(i)}$ is a sample from the distribution of $\boldsymbol{\gamma}$. The detailed statements of the coherence-optimal sampling and its convergence theorems can be found in [43].

In principle, the coherence-optimal sampling strategy amounts to discovering a weighted orthogonal polynomial system $w(\boldsymbol{\zeta})\psi_{\boldsymbol{\alpha}}(\boldsymbol{\zeta})$ that achieves the lowest coherence parameter value; then sampling from the corresponding new distribution will make the new regression matrix $\boldsymbol{W\Psi}$ isotropic [43]. One difficulty in this method is that the new distribution is generally not a standard or known distribution, and therefore direct sampling from this distribution is undoubtedly hard. In [43], a Monte Carlo Markov Chain (MCMC) approach was utilised to generate samples without calculating the specific expression of the distribution. This sampling method was originally proposed for Hermite and Legendre polynomials. For problems with input distribution that is neither uniform distribution nor Gaussian distribution but has known probability density function, the isoprobabilistic transform can be adopted to convert the original input distribution to uniform or Gaussian distribution [30]. Then the coherence-optimal sampling method can be performed on the transformed distribution, and the corresponding PCE model can be built and trained.

### 3.2. BED

It is worth mentioning that, in the coherence-optimal sampling method, only information from the input distribution and the truncated orthogonal polynomials is used; but the model observations, which represent the target model characteristic, are not utilised. Most of the commonly used sampling methods such as LHS, D-optimal, etc., are all input-dependent only. For various problems which can be modelled by the PCE technique with the same input dimension, input distribution and total polynomial degree, the formulated coherence-optimal distributions for sampling in the coherence-optimal sampling method are identical to each other. However, for problems with different complexity, the target model characteristic will influence the sampling result. BED takes advantage of this kind of information to instruct the sampling process [55–58]. A noticeable metric used in BED to quantify the information inherent in variables is the Shannon entropy in the context of information theory. The core idea in BED with information theory is that samples will be collected from a predefined candidate pool to maximise the information gain about the target model [58]. When applying BED for sequential sampling, the unobserved positions in the candidate pool are assumed as random variables, and the trained PCE model by use of Bayesian modelling techniques can make predictions to the unobserved positions with distributions instead of deterministic evaluations. The unreliability of the trained model can be reflected in the predicted results by evaluating the prediction uncertainty at each unobserved point. Then, we intend to get the next sample at a location which owns a large prediction uncertainty, because the observation collected at this position will provide more information to train the PCE model in the next iteration [56]. The PCE model trained using such collected samples and observations would reduce its uncertainty to a large extent. The target of interest is usually assumed as a continuous variable, so the differential entropy which extends from the Shannon entropy is [56,59]:

$$H_D(X) = -\int f(x) \log(f(x))\, dx \tag{23}$$

where $f(x)$ denotes the probability density function of a continuous variable $X$. The unit of the differential entropy depends on the base of the logarithm. The commonly used bases are 2, Euler's number and 10, which generate entropy units of bits, nats and bans, respectively. A large entropy value implies that less information is known about $x$ and more uncertainty exists in this variable.

Obviously, BED will make better use of the modelling results to make the trained PCE model more and more accurate. In another aspect, however, it has no robust guarantee on the calculation stability. Moreover, the Bayesian modelling approaches require enough data for training. The uncertainty will keep in a high level at almost all predicted points when scarce data is used. Under this circumstance, the predicted results are not accurate and will provide invalid instruction in the sampling process [60]. The initial sample size for a sequential sampling process is usually arbitrary and in general insufficient for modelling. Purely relying on BED will get samples with poor quality at the beginning of sampling process. In the next section, a new sequential sampling strategy in connection with BED will be proposed, where the samples will be constrained simultaneously to approximately follow the coherence-optimal distribution and to ensure the quality of samples.

## 4. Sequential experimental design

In this section, a method to sequentially sample from the coherence-optimal distribution is first introduced. Then BED in conjunction with the differential entropy is pursued where the BCS method is utilised to build the sparse PCE. Finally, these two methods are fused to develop a hybrid sequential sampling strategy which is expected to achieve high convergence rate and stable modelling performance.

### 4.1. Sequential sampling strategy from coherence-optimal distribution

To sequentially sample from a given distribution, a general framework is to first generate an initial sample set of small size from this distribution, and then gradually add samples until the quantity meets the requirement. There are two challenges in forming this framework.

The first one is about the generation of a small size of initial samples from a given distribution. In principle, the LHS method is amenable to generating small number of samples from a known distribution with good space filling property. It generates LHS samples in a multi-dimensional space. In each dimension, the definition domain

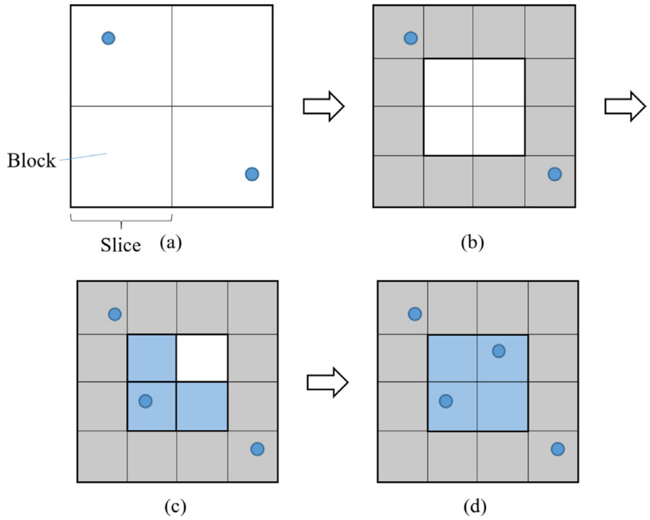


**Fig. 1.** Simple 2-D example of PLHS algorithm. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)

of the variable is uniformly divided into $N$ equal probability slices according to the marginal distribution, where $N$ is the target number of samples. By randomly generating one sample in each slice, a total of $N$ scalar samples are generated in this dimension. Then, the scalar samples from different dimensions are randomly matched together; the generated samples in this multi-dimensional space are referred to as LHS samples, and this distribution property is called Latin hypercube property. However, there are two issues in applying the LHS method to collect samples from coherence-optimal distribution. First, LHS was originally designed for problems with independent variables, while the coherence-optimal distribution may have non-negligible dependence among variables in a multivariable problem. To apply LHS, the dependence among input variables is ignored in this study. As a result, the collected LHS samples will follow a quasi-coherence-optimal distribution instead of the coherence-optimal distribution, but these samples can distribute over the input space in a more space filling way. Second, the generation of LHS samples from the quasi-coherence-optimal distribution is still difficult since each univariate distribution is mostly not a standard distribution. Equally dividing the input space by probability distribution is intractable. In view of this, an approximate operation is proposed in this study. First, a large number of samples are generated from the target distribution by use of MCMC method or other procedures. Then, according to the initial size of ED, the input space in each dimension is partitioned into slices so that each slice contains the same quantity of samples. As the sample quantity approaches infinity, these slices will have the same probability. Therefore, with enough samples (e.g., $10^5$) being generated from the target distribution, these divided slices in each dimension could be thought to have approximately equal probability, and random sampling from these slices will achieve a near LHS sample set. We refer to this method as Near-LHS (NLHS) algorithm, which will be implemented to generate LHS samples from a non-standard distribution. This near LHS sample set conforms to the quasi-coherence-optimal distribution in a space filling way.

The second challenge is how to ensure that the samples always possess a favourable space filling property during the sampling process, since each new sample will destroy the Latin hypercube property of the overall samples. Here, a 'doubling procedure' adapted from the so-called Progressive-LHS (PLHS) algorithm [40] is employed to maintain the distribution property of the collected samples during the sequential sampling process. When a near LHS sample set of size $n_1$ has been generated by the NLHS algorithm, the input space is uniformly sliced into $n_1$ equal probability slices by the marginal distribution in each dimension. To add samples, the 'doubling procedure' is to perform extra $n_2 = n_1$ slicing operations in the input space of each dimension. A total of $2n_1$ equal probability slices in each dimension are generated. In other words, each slice from the last operation is equally divided into two parts in each input dimension. Slice from different dimensions will form an input subspace, named a block. Those blocks, where slices in all dimensions have no samples located, are defined as active blocks. Subsequently, extra $n_1$ samples are randomly sampled from these newly generated active blocks, and the resulting $2n_1$ samples in total are still LHS samples. For easy interpretation, a simple 2-D example is illustrated in Fig. 1.

Assume that the initial sample set has two sample points, so the marginal probability distributions of two variables in this 2-D plane are uniformly divided into two slices, that is, four blocks, as shown in Fig. 1(a). Two samples are generated so that each slice has one sample projection in this dimension. Then, to seek for new samples, the two distributions are doubly sliced as depicted in Fig. 1(b), where the new samples together with the original ones comply with the target distribution. The newly generated active blocks are marked in white and inactive blocks are in grey. By randomly choosing one active block and sampling once from it, three of the original four active blocks become inactive according to the definition of active block, and they are marked in blue in Fig. 1(c). So, the last sample can only be collected in the final active block, as shown in Fig. 1(d), and all the blocks become inactive. When no more active block exists, the ‘doubling procedure’ will be executed again if more samples are needed. It is apparent that this algorithm imposes a constraint to each sampling operation, and finally the samples will be the near LHS samples. One drawback of this algorithm is that it is less flexible in controlling the sample size. It is known that the samples are near LHS samples only when no active block exists. After each ‘doubling procedure’, the number of samples in need to construct the LHS samples will be doubled (e.g. 2, 4, 8, 16, and so on in the previous case). But this is not a problem since generating LHS samples in a strict way during the iteration process is not imperative in this study. By use of the NLHS algorithm with the ‘doubling procedure’ (NLHS-DP), a sequential sampling framework can be formulated to sample from the quasi-coherence-optimal distribution. The detailed description of this algorithm is shown in Algorithm 1.

**Algorithm 1**. NLHS-DP algorithm

**Input:** Coherence-optimal samples $\mathcal{X}_N$; Initial sampling number $N_0$; Total sampling number $N_s$.

**Initialisation:**

1. Divide the coherence-optimal samples $\mathcal{X}_N$ into $N_0$ slices uniformly according to the input dimension $d$; Totally $d$ sets of slices will be obtained, and in each set, there are $N_0$ slices;
2. In each dimension, randomly collect one sample from each slice; $d$ sets of $N_0$ samples corresponding to the input dimensions will be obtained, and they will be randomly combined in dimension to get the initial $N_0$ samples.

**Iteration:**

1. Check the slices in each dimension; If no block is active, get into step 2; otherwise go to step 3;
2. Under the current sample number $N_k$, re-divide input space in each dimension into $2N_k$ slices so that each slice contains $\frac{N}{2N_k}$ coherence-optimal samples. New $N_k^d$ active blocks are generated;
3. Randomly select one active block, and randomly collect one sample from this active block;
4. With the newly added sample, some active blocks become inactive; re-evaluate the blocks to find the remaining active blocks;
5. If the total sampling number is reached, quit iteration.

### 4.2. BCS-based BED

Under the framework of NLHS-DP, samples are still randomly generated since at each iteration, at least one block is active. One should first randomly choose an active block (if there is than one active block), and then randomly sample once in this block. So, the NLHS-DP algorithm is thought to be less robust and slow to convergence. In view of this, BED is introduced in this study. By use of the BCS method to calculate the PCE coefficients, the trained PCE model can be used to make predictions to those unknown places with Gaussian distributions. By considering a set of points $\mathcal{X}_{N_t} = \left\{\xi^{(1)}, \xi^{(2)}, \dots, \xi^{(N_t)}\right\}$ to be predicted, the polynomial basis will form a new matrix $\boldsymbol{\Psi}_{new}$ with $N_t$ rows. According to the posterior distribution of the coefficients, $N(\boldsymbol{c}|\mu_c, \Sigma_c)$, the predictions to these points can be obtained as [61]:

$$Y_{pre}(\mathcal{X}_{N_t}) \sim N\left(\boldsymbol{\Psi}_{new}\mu_c, \boldsymbol{\Psi}_{new}\Sigma_c{\boldsymbol{\Psi}_{new}}^T\right) \tag{24}$$

Then the prediction $Y_{pre}\left(\xi^{(i)}\right) \sim N(\mu_i, {\sigma_i}^2)$ at each point $\xi^{(i)}$ can be obtained. Here $Y_{pre}^{(i)}$ denotes the variable with its prediction at point $\xi^{(i)}$. $f(Y^{(i)})$ represents the probability density function of $Y_{pre}^{(i)}$. Thus, the differential entropy at each predicted point can be calculated by:

$$H_D(Y_{pre}^{(i)}) = -\int f(Y^{(i)}) \ln\left(f(Y^{(i)})\right) dY^{(i)} \tag{25}$$

Here we use Euler's number as base of logarithm. A point with high entropy value denotes that the obtained PCE model is less certain at this point, and the predicted result at this point has a high uncertainty. Thus, this point has a higher priority to be selected than other points since observation at this point will provide more information about the target model. Training a PCE model with this observation will reduce the prediction uncertainty to a large extent. For Gaussian distribution, the differential entropy can be elicited as:

$$H_D\left(Y_{pre}^{(i)}\right) = \frac{1}{2}\ln\left|{\sigma_i}^2\right| + \frac{1}{2}\ln(2\pi) + \frac{1}{2} \tag{26}$$

which is related only to the variance value and easy to be calculated.

### 4.3. Hybrid sequential sampling strategy

In this section, a hybrid sequential sampling strategy is proposed by fusing the sequential sampling framework and BED. Assume that a set $\mathcal{X}_{N_t} = \{\boldsymbol{\xi}^{(1)}, \boldsymbol{\xi}^{(2)}, \ldots, \boldsymbol{\xi}^{(N_t)}\}$ with $N_t$ samples is generated in the input space beforehand. In each iteration, the model observations corresponding to the selected samples are collected to train the PCE model. Taking advantage of the BCS method, the obtained PCE model can make predictions to the points which belong to the active blocks in the candidate set. Then the differential entropy values of these points can be obtained from the prediction results, and the point with the maximum entropy value is selected as the next sample. By doing so, the random sampling operation in sequential sampling framework is replaced by a different sampling criterion which employs the differential entropy to instruct the sampling process. This new sampling strategy is termed coherence-entropy (Coh-entro) algorithm. The algorithmic procedure is detailed in Algorithm 2.

In the proposed method, the collected samples are controlled to conform to the quasi-coherence-optimal distribution in each iteration. Even though the samples do not strictly obey the coherence-optimal distribution, they still have a good space filling property to cover the input space, and it is speculated that such distribution property could still help bound the number of samples. The isotropic property of the regression matrix cannot be achieved,

**Algorithm 2.** Coherence-entropy algorithm

**Input:** Problem input dimension $d$; total degree $p$; initial sampling number $N_0$; total sampling number $N_{ts}$ or desired accuracy $\varepsilon$; candidate set $\mathcal{X}_{N_t}$; selection set $S$ and observation set $Y$.

**Initialisation:**
1. Generate $N$ coherence-optimal samples $\mathcal{X}_N$ by employing MCMC sampling method according to the problem dimension and total degree;
2. Generate $N_0$ initial samples which yield the quasi-coherence-optimal distribution and add them into the selection set $S = \mathcal{X}_{N_0}$;
3. Get model observations at sampled points in the selection set $Y = \boldsymbol{Y}(\mathcal{X}_{N_0})$.

**In the $i$th iteration:**
1. Build PCE model according to the selection set $S$ and the corresponding observations $Y$ by employing the BCS method;
2. If the total sampling number $N_{ts}$ is reached or the desired accuracy $\varepsilon$ is achieved, quit iteration; otherwise, get into step 3;
3. Check all the blocks; If no more active block exists, get into step 4; otherwise get into step 5;
4. Under the current sample number $N_S$, re-divide the input space into $2N_S$ slices in each dimension so that each slice contains $\frac{N}{2N_S}$ coherence-optimal samples. $N_S^d$ new active blocks are generated;
5. Abandon the candidate points which do not belong to active blocks temporarily;
6. Make prediction to the remaining candidate points and calculate the corresponding differential entropy;
7. Add the candidate point $\boldsymbol{\xi}^{(i)}$ which owns the largest differential entropy value to the selection set $S = S \cup \boldsymbol{\xi}^{(i)}$, and get model observation at this point $Y = Y \cup \boldsymbol{Y}_{\boldsymbol{\xi}^{(i)}}$;
8. According to the current sample set, re-evaluate the active blocks;
9. Update the candidate set by deleting the selected point $\mathcal{X}_{N_t} = \mathcal{X}_{N_t} \backslash \boldsymbol{\xi}^{(i)}$.

but it is expected that the regression matrix can be in good condition for regression calculation when using $l_1$ sparse representation method. BED based on the BCS method and differential entropy is employed to substitute the random selection operation in sequential sampling framework. It can be deemed that the BED is constrained so that the samples can have a good distribution property while enriching the target model information received from the corresponding observations. Thus, the PCE model can be trained to be more and more accurate with the gradually increased target model information. In addition, even if the collected samples and the corresponding observations at the beginning of the iteration process are not enough for training an accurate PCE model, which gives rise to a worse performance of the BED, the generation of samples from the quasi-coherence-optimal distribution still provides a foundation for further sampling and PCE modelling. The proposed method not only makes full use of the input information, but also earns much information from the observations and modelling results. It is expected to have both better convergence rate and computational stability.

### 4.4. Termination criterion

Several criteria have been proposed to terminate the sampling process through evaluating the precision of the obtained PCE model, such as Kullback–Leibler Divergence (KLD) and Leave-One-Out (LOO) error [15,35]. Inspired by the idea of comparing the responses obtained from the PCE model in successive iterations for modelling accuracy evaluation, a simple criterion capitalising on the changes of the PCE model mean and standard deviation (std) in successive iterations, is employed in this study to assess convergence of the obtained PCE model. These values can be easily obtained from the PCE coefficients, and they are good statistical measures to represent the PCE model. When the PCE model mean and std values in successive iterations keep stable or the change values of these two statistical measures are smaller than a given threshold, the PCE model is considered to have converged with satisfactory modelling accuracy.

## 5. Numerical examples

In this section, two analytical benchmark functions with different input dimensions and degrees are addressed to validate the proposed method. The PCE models are also truncated on different degrees to test the algorithm performance. Three state-of-art input-dependent only sampling methods (Coh-Opt, D-Coh-Opt, and Seq-D-Coh-Opt) proposed in [34] and [43], which are based on coherence-optimal sampling and D-optimal design, are compared with the proposed method (Coh-entro). Among them, the Coh-Opt and D-Coh-Opt methods are non-sequential sampling strategies while the Seq-D-Coh-Opt method is a sequential sampling strategy. In addition, two output-dependent sequential sampling methods are also compared in this study. One is the component of the proposed method, BED with differential entropy. The other is the CV-ELF criterion which was proposed in [27]. The CV-ELF criterion is based on the PCE modelling results inferred by SBL to instruct the next sample, which is an output-oriented method. For the convenience of comparison, the termination criterion used for the benchmark tests is set as sample upper limit instead of convergence evaluation. The size of the candidate sample set in the Coh-entro algorithm is set as $N_t = 10^4$. In order to assess the real precision of the obtained PCE models, a validation set with $N_{val} = 10^4$ random samples and the exact model values are used. The relative root mean square error (RMSE) $\varepsilon_{RMSE}$ [34] is calculated by:

$$\varepsilon_{RMSE} = \sqrt{\frac{\sum_{i=1}^{N_{val}} (y_{true}^{(i)} - y_{PCE}^{(i)})^2}{\sum_{i=1}^{N_{val}} {y_{true}^{(i)}}^2}} \tag{27}$$

After the benchmark study, the proposed method will be applied to two engineering problems to build accurate PCE models for parameter sensitivity analysis. The convergence is assessed by the termination criterion which automatically determines the ED size with desired accuracy.

The Matlab codes of the Coh-Opt, D-Coh-Opt, and Seq-D-Coh-Seq methods are available online [34]. An in-house code of the CV-ELF criterion has been developed according to [27], but the sparse solver there has been changed in this study to BCS with Laplace prior for fair comparison. The code of the BCS method is also available [51]. For the sake of brevity, the methods of Coh-Opt, D-Coh-Opt, Seq-D-Coh-Seq, and BED with differential entropy are denoted as Coh, D-coh, Coh-seq, and Entro hereafter.

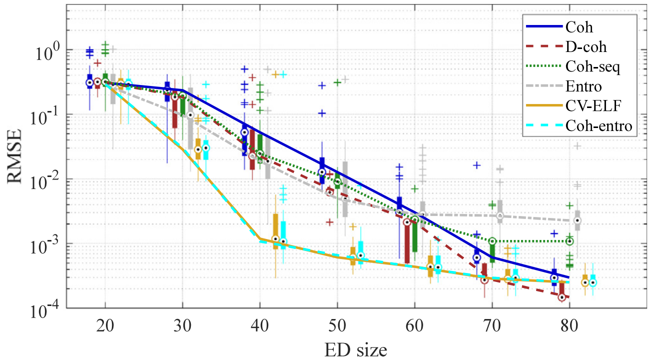


**Fig. 2.** RMSE with different ED size under degree 9.

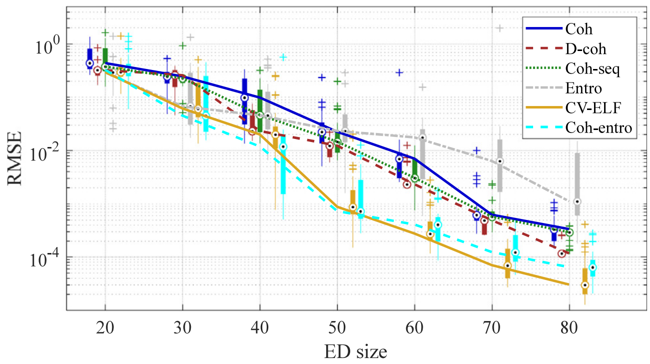


**Fig. 3.** RMSE with different ED size under degree 11.

### 5.1. Two-dimensional function

The first test function is a low-dimensional analytical function as given in Eq. (28), which has been studied as a benchmark test function for PCE modelling [35,62]. The random input variables are uniformly distributed with a mean of 2.0 and a probability density function height of 0.7222. The analytical results of the mean and std of this function are 0.079 and 1.124, respectively.

$$y = \ln\left(1 + {x_1}^2\right) * \sin\left(5x_2\right) \tag{28}$$

Two total degrees, 9 and 11, are chosen as truncation degree of the PCE model, which will result in the full PCE expansions of $P = 55$ and $P = 78$ multivariate Legendre polynomials. The initial sampling number is set as 10. Each method is calculated 30 times to ensure statistical stability. The RMSE results with respect to the increase of ED size are shown by box plots in Figs. 2 and 3.

In the box plots, bold vertical lines represent the range between upper and lower quartiles of the 30 results, and fine vertical lines represent the 1.5 times interquartile range which constrains the normal value limitation. Values out of them are regarded as outliers, which are depicted as plus symbols in the plots. The dot inside the white circle denotes the median of RMSE, and the lines that vary with respect to the ED size represent the variation of the median of RMSE obtained by different methods. It can be observed that the CV-ELF criterion and the proposed Coh-entro strategy have lower validation error than other three input-dependent only sampling methods and the Entro sampling method under almost all circumstances. In the final converged results, except for the D-coh strategy with PCE degree 9, the PCE models with samples generated by the Coh-entro strategy and the CV-ELF criterion both have smaller validation error than the other four methods. In the circumstances of ED size much larger than the number of unknown coefficients, samples from the D-coh strategy can construct a better regression matrix by

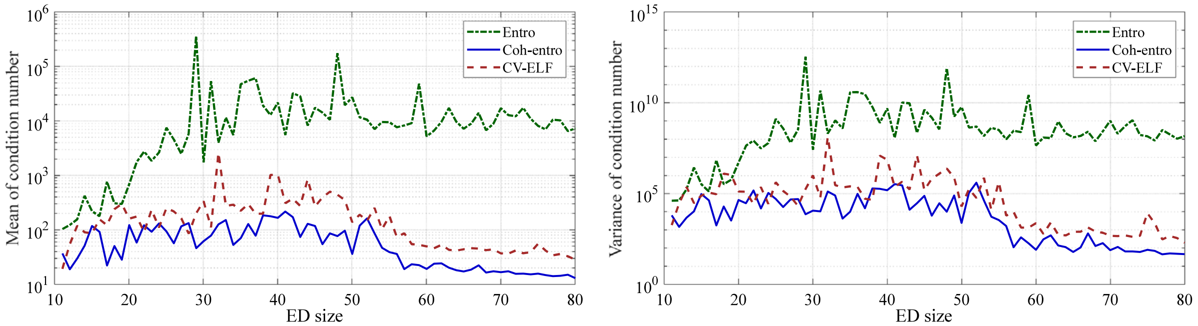


**Fig. 4.** Mean (left) and variance (right) of the condition number with increasing ED size under degree 9.

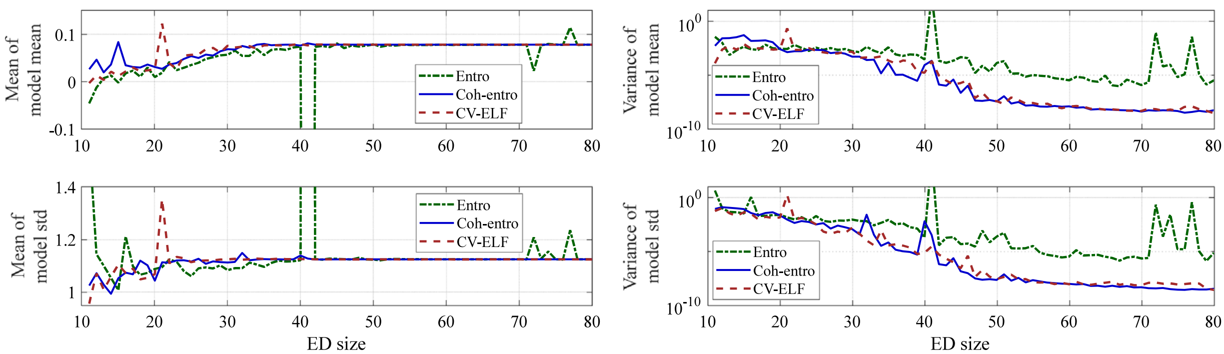


**Fig. 5.** Mean (left) and variance (right) of the model means and standard deviations with increasing ED size under degree 9.

sampling at once to form well-conditioned equations which make the solutions stable. For the PCE models with degree 11, the final converged results from the Coh-entro strategy have a bit higher validation error than the CV-ELF criterion, while the Coh-entro strategy has a better performance than the three input-dependent only methods and the Entro method.

To illustrate the solution stability of the regression calculation during the sampling process, the condition number of the regression matrix and the mean and std values of the obtained PCE model calculated during the iteration process are depicted. The condition number is defined as [35,63]:

$$\textit{condition number} = \|\boldsymbol{\Psi}_{\boldsymbol{c}}\| \cdot \left\|\boldsymbol{\Psi}_c^{\dagger}\right\| \tag{29}$$

where $\boldsymbol{\Psi}_{\boldsymbol{c}}$ represents a submatrix of $\boldsymbol{\Psi}$ in Eq. (7) with columns corresponding to non-zero coefficients; † denotes Moore–Penrose inverse. The norm can be of any form, such as 1-norm, 2-norm, $\infty$-norm, etc. Here 2-norm is used. A large condition number represents poor performance of the regression matrix and implies that the regression solution becomes more sensitive to changes in the input values and observations, i.e., the correct solution is hard to find. Hence, a large condition number appearing in the sequential sampling process means that the corresponding regression solution may be incorrect. Three strategies, the proposed Coh-entro strategy, the Entro method and the CV-ELF criterion, which are all sequential sampling methods and output-oriented, are compared here. Figs. 4 and 6 show the mean and variance of the condition number during the 30 repeated tests with respect to the increase of ED size. It can be seen that under the PCE degree 9, the condition number of the Entro method is always the largest among the three strategies. The condition number of Coh-entro has its mean values mostly lower than CV-ELF during the whole iteration process and ultimately becomes stable at a very low value. The variance values of the condition number have a similar variation trend with the mean values, which demonstrates that the condition number is stable in the repeated tests with the increased samples. Under the PCE degree 11, the results of the three methods all show large fluctuation, where the condition number of the Entro method is again the largest during nearly the whole iteration process. In the first half of iteration process before 50 samples, the Coh-entro strategy gives rise to lower condition number than the CV-ELF criterion; but they have similar performance in the second

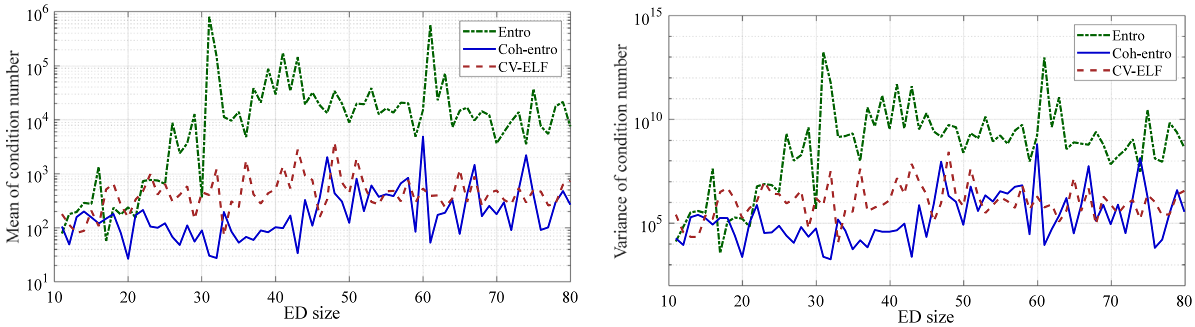


**Fig. 6.** Mean (left) and variance (right) of the condition number with increasing ED size under degree 11.

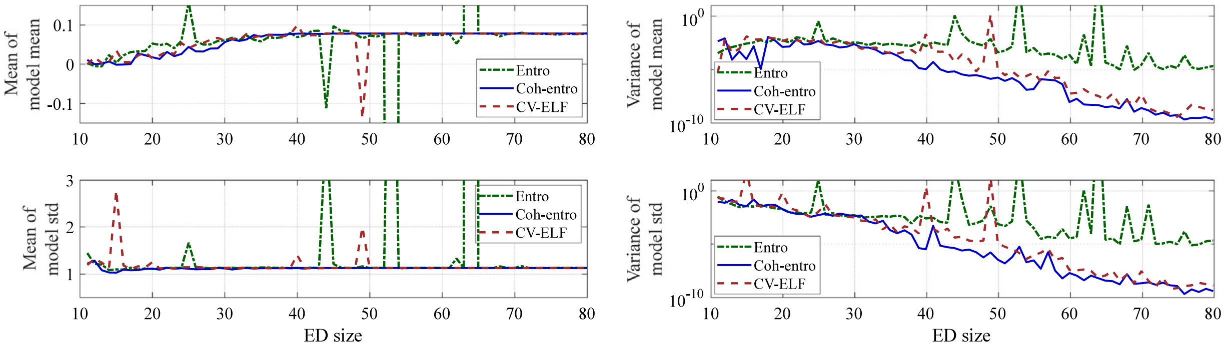


**Fig. 7.** Mean (left) and variance (right) of the model means and standard deviations with increasing ED size under degree 11.

half of iteration process. The iteration stability of the obtained PCE models is represented by the model mean and std values. Figs. 5 and 7 show the statistical properties (mean and variance) of the obtained PCE model mean and std values during the 30 repeated tests against the increase of ED size. The PCE models with samples generated by Coh-entro and CV-ELF have better convergence performance than the Entro method. The PCE model obtained by the Coh-entro strategy is more stable than that resulting from the CV-ELF criterion under the PCE degree 11 as illustrated in Fig. 7 since the variance values from Coh-entro are a bit lower than CV-ELF. Several large outliers emerge during the iteration process with the CV-ELF criterion.

The final samples from the three methods in one test under degree 11 are provided in Fig. 8 for comparison. It is clear that the samples generated by the Entro method almost concentrate on the corners and edges of the input space, and the samples generated by CV-ELF concentrate more on the edges of the input space than Coh-entro. In contrast, the proposed Coh-entro strategy places samples in a more space filling way.

In summary, the proposed Coh-entro strategy and the CV-ELF criterion outperform the input-dependent only methods and the Entro method in terms of convergence rate, and the Entro method suffers from stability problem during the iteration process. Through constraining the sample distribution to approximately match the coherence-optimal distribution, the proposed strategy overcomes this drawback and improves the solution stability. It not only generates samples in a more space filling way but also results in a low condition number in regression calculation. One more benchmark test on a low-dimensional function is provided in the Appendix, where the Ishigami function is studied.

### *5.2. High-dimensional function*

The second benchmark test function is a high-dimensional function [22,35]. Its expression is as follows:

$$y = 3 - \frac{5}{d}\sum_{i=1}^{d} i x_i + \frac{1}{d}\sum_{i=1}^{d} i x_i^3 + ln\left[\frac{1}{3d}\sum_{i=1}^{d} i(x_i^2 + x_i^4)\right] \tag{30}$$

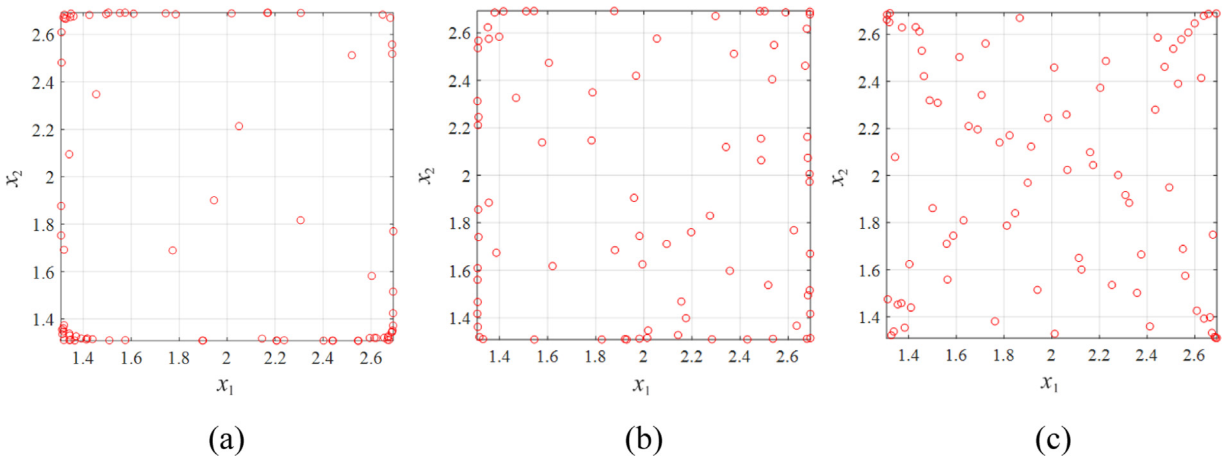


**Fig. 8.** Final samples from (a) Entro; (b) CV-ELF; (c) Coh-entro.

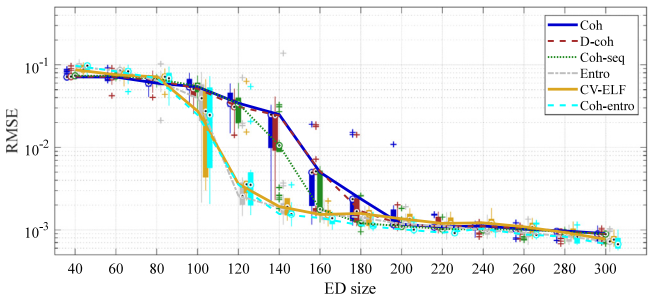


**Fig. 9.** RMSE with different ED size under input dimension 15.

where $d$ is the input dimension chosen by user. All the input variables conform to uniform distributions defined on the interval $[1, 2]$. The input dimension is selected as $d = 15$, and the total degree of the PCE model is chosen as 3, which results in polynomial items with $P = 816$. The initial sampling number is set as 20. The operation is repeated for 30 times to ensure statistical stability.

The RMSE results with respect to the increase of ED size are shown in Fig. 9. The output-oriented methods exhibit similar convergence trends and perform better than the input-dependent only methods in terms of convergence rate. All six methods converge to similar validation errors, and the proposed Coh-entro strategy has the lowest validation error at the end of the sampling process.

The condition number of the regression matrix and the obtained PCE model mean and std values of the three output-oriented sampling methods are depicted in Figs. 10 and 11. It is apparent that the condition number from the Entro method and the CV-ELF criterion fluctuates more than the Coh-entro strategy during the whole iteration process, especially when the model has 100 to 140 samples. By contrast, the Coh-entro strategy has less extremely large values, which all concentrate at the beginning of the iteration process where ED size is less than 60. When the samples are more than 60, the condition number keeps stable at a very low value. The variance of the condition number shows that the proposed strategy is robust in the repeated tests.

The solution stability of the regression calculation during the iteration process is demonstrated through the model mean and std given in Fig. 11. After convergence, the PCE model from the Coh-entro strategy has no obvious outlier while the Entro method and the CV-ELF criterion show more fluctuations. Moreover, it can be found from the variances of the model mean and std that the Coh-entro strategy converges to a lower value of variance than the Entro method and the CV-ELF criterion, which indicates that the PCE model from the Coh-entro strategy performs more stable than those from the Entro method and the CV-ELF criterion.

In addition, another input dimension $d = 20$ is tested with polynomial terms $P = 1771$ when the total truncation degree is chosen as 3. Due to the heavy computational burden, the number of repeated tests in this case is set as 10.

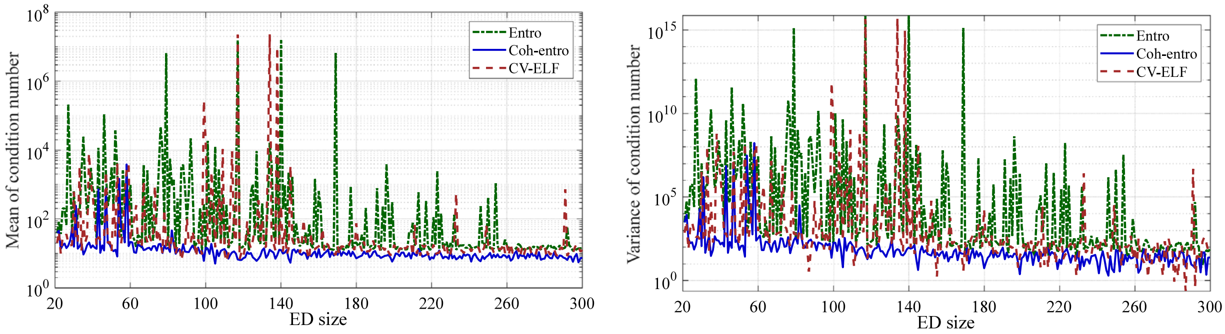


**Fig. 10.** Mean (left) and variance (right) of the condition number with increasing ED size under input dimension 15.

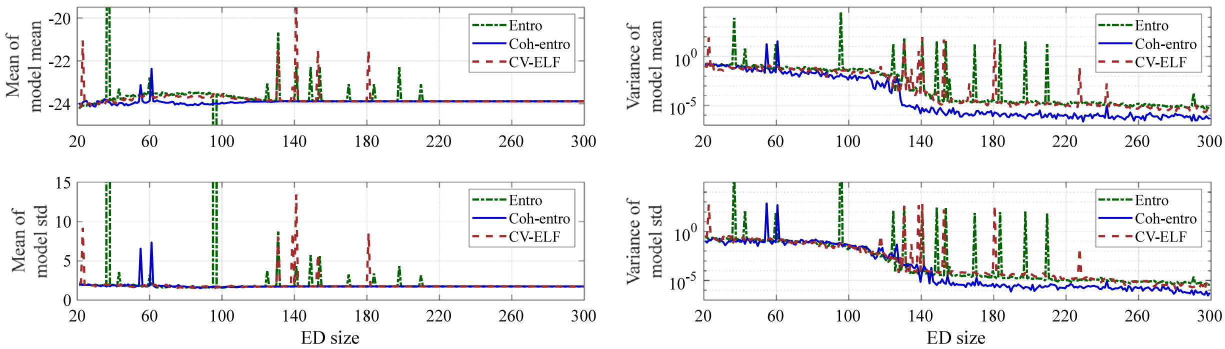


**Fig. 11.** Mean (left) and variance (right) of the model means and standard deviations with increasing ED size under input dimension 15.

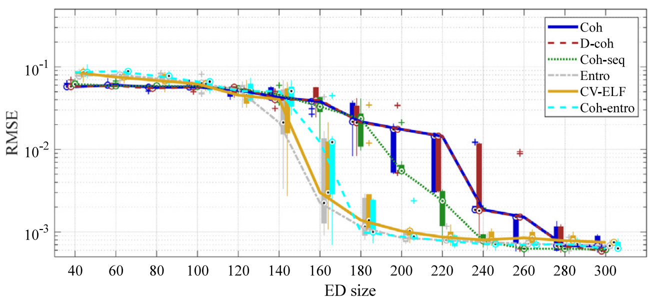


**Fig. 12.** RMSE with different ED size under input dimension 20.

The initial sampling number is set as 20 which is the same as the previous case. The RMSE results with respect to the increase of ED size are illustrated in Fig. 12. The three output-oriented methods outperform the input-dependent only methods, and both have similar convergence rate.

The condition number of the regression matrix and the obtained PCE model mean and std values from the output-oriented methods are plotted in Figs. 13 and 14. It can be seen from the condition number that the three methods have similar performance. All of them give rise to large fluctuations during the iteration process. The reason might be because the number of samples is too small compared with the polynomial terms, making the calculation unstable with sparse representation. However, even with large fluctuation on the condition number, the convergence performance of Coh-entro is still a bit better than the Entro method and the CV-ELF criterion, as illustrated in Fig. 14. Moreover, the variances of the model mean and std obtained by the Coh-entro strategy finally converge to lower values than those from the Entro method and the CV-ELF criterion, demonstrating that the proposed method can help obtain more stable PCE models during the sampling process.

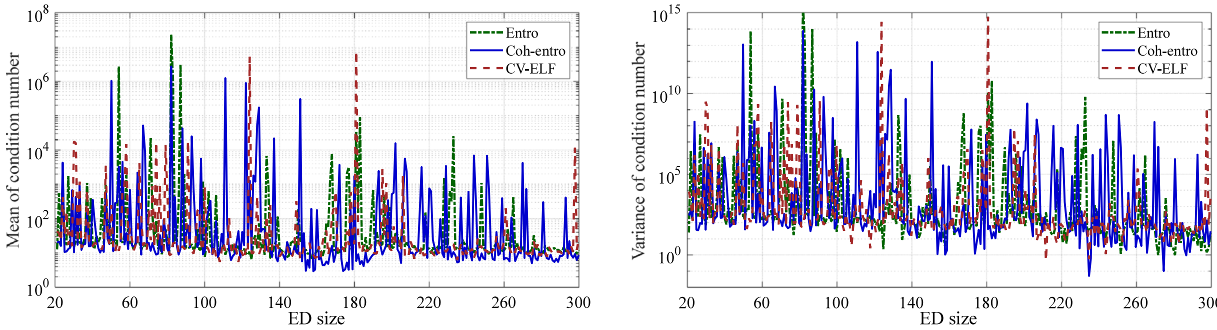


**Fig. 13.** Mean (left) and variance (right) of the condition number with increasing ED size under input dimension 20.

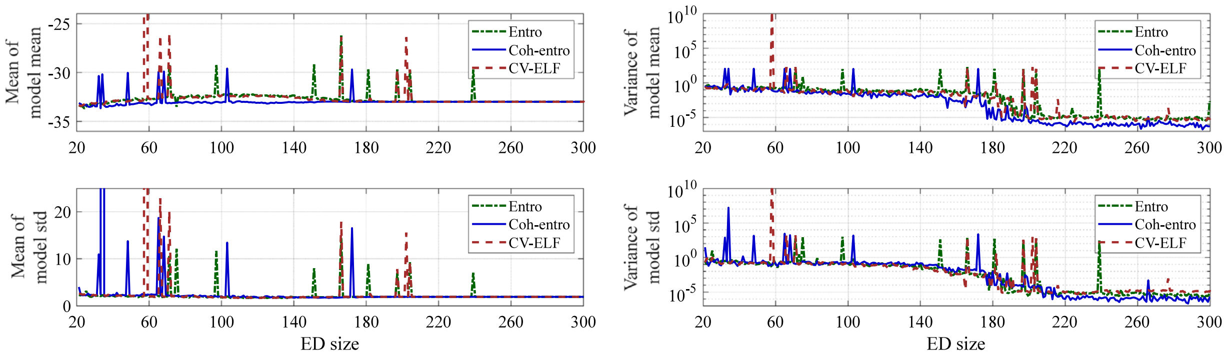


**Fig. 14.** Mean (left) and variance (right) of the model means and standard deviations with increasing ED size under input dimension 20.

### 5.3. A 2-D truss structure

In order to demonstrate the application of the proposed method, two engineering cases with different target responses are devoted to parameter sensitivity analysis. Here, the Analysis Of Variance (ANOVA, or called Sobol' indices) method is used to analyse the parameter sensitivity [64–66]. Firstly, a brief introduction to the ANOVA method is given. The ANOVA method is a variance-based global sensitivity analysis method, which works on the entire input domain and analyses the influence of each independent input variable on the model output. Suppose a model can be decomposed as the following summands [67]:

$$f(\xi_1, \xi_2, \ldots, \xi_d) = f_0 + \sum_{i=1}^{d} f_i(\xi_i) + \sum_{1 \le i < j \le d} f_{ij}(\xi_i, \xi_j) + \cdots + f_{12\cdots d}(\xi_1, \xi_2, \ldots, \xi_d) \tag{31}$$

and

$$\int_D f_{i_1 i_2 \cdots i_k}\left(\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_k}\right) p(\xi_{i_r}) d\xi_{i_r} = 0 \quad r \in [1, k] \quad k \le d \tag{32}$$

where $D$ is the definition domain of input variables; $p(\xi_{i_r})$ is the marginal probabilistic density function of $\boldsymbol{\xi}_{i_r}$. The total variance of the model $f(\xi_1, \xi_2, \ldots, \xi_d)$ is decomposed into the summation of each summand variance which can be simply calculated by integration. The total variance is obtained by:

$$V_{total} = \sum_{1 \le i \le d} V_i + \sum_{1 \le i < j \le d} V_{i,j} + \cdots + V_{1,2,\ldots,d} \tag{33}$$

where

$$V_{i_1 i_2 \cdots i_k} = \int_D \cdots \int_D f_{i_1 i_2 \cdots i_k}(\xi_{i_1}, \xi_{i_2}, \ldots, \xi_{i_k})^2 p(\xi_{i_1} \cdots \xi_{i_k}) d\xi_{i_1} \cdots d\xi_{i_k} \quad k \le d \tag{34}$$

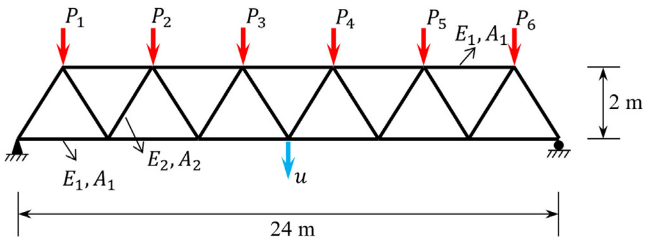


**Fig. 15.** Layout of the truss structure.

**Table 2**
Input distributions of the truss structure.

| Variable | Distribution | Mean | Standard deviation |
|---|---|---|---|
| Elastic moduli $E_1$, $E_2$ (Pa) | Lognormal | 2.10e11 | 2.10e10 |
| Cross-section area $A_1$ (m$^2$) | Lognormal | 2.0e−3 | 2.0e−4 |
| Cross-section area $A_2$ (m$^2$) | Lognormal | 1.0e−3 | 1.0e−4 |
| Vertical forces $P_1 \sim P_6$ (N) | Gumbel | 5.0e4 | 7.5e3 |

Subsequently, the variance of each summand can be normalised by the total variance $V_{total}$, and the variance contribution of each summand can be obtained by

$$S_{i_1 i_2 \cdots i_k} = \frac{V_{i_1 i_2 \cdots i_k}}{V_{total}} \tag{35}$$

To quantify the variance contribution of each variable, the first-order index is defined as the $S$ value that only contains variance value from one variable in numerator, which represents the independent variance contribution from this variable to the model output. The total index is the sum of all $S$ values that involve the target variable in numerator, including interaction terms with other variables. By comparing the first-order index and the total order index, the interaction effect of input variables on the output can be identified. When applying ANOVA based on the PCE model, the calculation of indices becomes convenient since the multivariate orthogonal polynomials satisfy the condition in Eq. (32). The ANOVA indices can be easily calculated from the PCE coefficients.

A 2-D truss structure is first introduced to validate the applicability of the proposed sampling method to build PCE model for sensitivity analysis, which has been widely studied as a benchmark test in the past [15,28,68]. As depicted in Fig. 15, this structure comprises 23 bars and 13 nodes, and the target is the maximum deflection in the midspan under 6 vertical forces.

A total of ten input variables are considered in this case, which include six vertical forces ($P_1$, $P_2$, $P_3$, $P_4$, $P_5$, $P_6$), the elastic moduli ($E_1$, $E_2$) and the cross-section areas ($A_1$, $A_2$) of the horizontal bars and diagonal bars. The distributions of these variables are reported in Table 2 [28] and they are assumed to be mutually independent. A PCE model using Hermite polynomials is built to investigate the relation between the vertical deflection $u$ and the ten variables. The input variables are first transformed into standard Gaussian variables to ease computational burden. Second-order PCE has been used to characterise engineering structures [12]. Here, the total degree to truncate the PCE model is chosen as 3 to compromise between model accuracy and computational burden. Thus, the total number of polynomial terms is $P = 286$. The initial sampling number is set as 20. To ensure computational stability, the variation values of the model mean and std in two successive iterations smaller than a given threshold is considered as the convergence criterion. The threshold value is selected as $10^{-5}$ in this case. To test the robustness of the proposed sampling method when applied to this engineering problem, the modelling process is repeated 10 times, and only the CV-ELF criterion is compared herein. A validation dataset with $10^4$ finite element runs is used to calculate RMSE of the obtained PCE model. The RMSE values and the numbers of ED after convergence in 10 repeated tests are shown in Fig. 16. It can be observed that the RMSE values after convergence from the Coh-entro strategy are much lower than those from the CV-ELF criterion, and the obtained PCE model with CV-ELF is not stable since the validation errors have a large discrepancy in the repeated tests. Finally, samples from the Coh-entro strategy are around 210 and range between 180 to 250, which are a bit more than samples used in the CV-ELF criterion, but the former guarantees the obtained PCE model in high accuracy and stability.

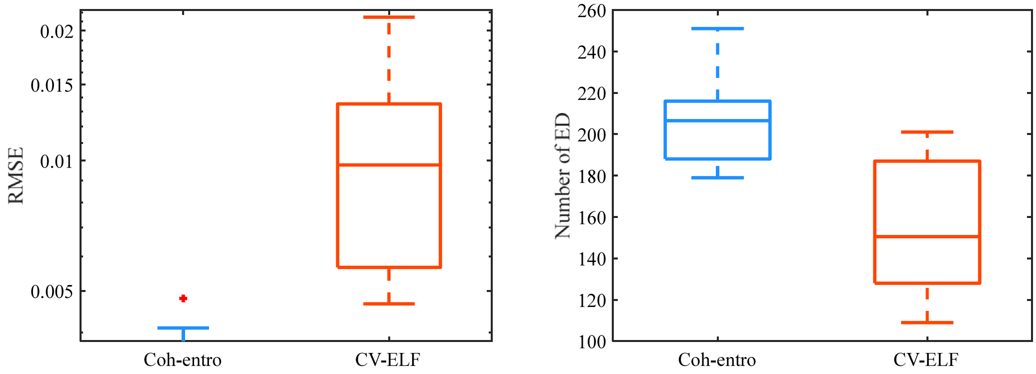


**Fig. 16.** Convergence results for truss structure.

**Table 3**
Sensitivity indices of input variables for the truss structure.

| Variables | First-order indices | | | | | Total indices | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | MC | CV-ELF | | Coh-entro | | MC | CV-ELF | | Coh-entro | |
| | | Mean | Std | Mean | Std | | Mean | Std | Mean | Std |
| $E_1$ | 0.370 | 0.360 | 6.46e−3 | 0.367 | 4.02e−3 | 0.372 | 0.366 | 6.70e−3 | 0.372 | 3.75e−3 |
| $E_2$ | 0.011 | 0.013 | 1.53e−3 | 0.012 | 6.29e−4 | 0.013 | 0.013 | 1.53e−3 | 0.012 | 5.87e−4 |
| $A_1$ | 0.362 | 0.371 | 1.65e−2 | 0.365 | 4.80e−3 | 0.373 | 0.377 | 1.60e−2 | 0.370 | 5.04e−3 |
| $A_2$ | 0.012 | 0.013 | 1.08e−3 | 0.012 | 5.82e−4 | 0.013 | 0.014 | 1.11e−3 | 0.013 | 5.53e−4 |
| $P_1$ | 0.004 | 0.004 | 1.47e−3 | 0.004 | 5.19e−4 | 0.005 | 0.004 | 1.52e−3 | 0.005 | 5.16e−4 |
| $P_2$ | 0.037 | 0.036 | 3.12e−3 | 0.039 | 2.17e−3 | 0.038 | 0.036 | 3.15e−3 | 0.039 | 1.98e−3 |
| $P_3$ | 0.077 | 0.075 | 1.16e−2 | 0.076 | 1.77e−3 | 0.078 | 0.076 | 1.18e−2 | 0.077 | 1.74e−3 |
| $P_4$ | 0.077 | 0.076 | 4.43e−3 | 0.076 | 1.77e−3 | 0.078 | 0.077 | 4.37e−3 | 0.077 | 1.90e−3 |
| $P_5$ | 0.037 | 0.038 | 2.63e−3 | 0.037 | 1.29e−3 | 0.038 | 0.039 | 2.58e−3 | 0.038 | 1.50e−3 |
| $P_6$ | 0.004 | 0.004 | 8.40e−4 | 0.004 | 2.28e−4 | 0.005 | 0.005 | 8.59e−4 | 0.004 | 2.74e−4 |

With the PCE models obtained by the CV-ELF criterion and the Coh-entro strategy, the mean and std values of Sobol' indices of the ten input variables are given in Table 3. The Sobol' indices obtained by Monte Carlo (MC) simulations are also shown in Table 3 as a reference. The sample size of MC simulations is $10^5$. It can be observed from the table that, for given distributions of the input variables, the elastic moduli and the cross-section areas of horizontal bars have the largest impact on the deflection $u$, and the material properties of diagonal bars show low impact. Among the six vertical forces, the deflection is more sensitive to the forces which are closer to the midspan. It can also be observed that the first-order indices are approximately equal to the total indices, indicating that the input variables have few correlated influences on the target response. Through comparison between two methods, it is found that the results obtained by Coh-entro are closer to the reference values. Moreover, the Coh-entro strategy can get smaller std values than the CV-ELF criterion, which indicates that the results obtained by Coh-entro are more stable than CV-ELF in 10 repeated tests.

### *5.4. An extradosed cable-stayed bridge*

The second engineering application is the modelling of the dynamic characteristics of a bridge structure with respect to structural material properties and the analysis of parameter sensitivity. Only the proposed sampling method is used. The target structure is an extradosed cable-stayed bridge with three spans of 460 metres long in total. The girders are designed as single box with three rooms of 33.5 metres wide. Two towers of 40 metres high each are rigidly consolidated with the girders. The girders are continuously supported on piers. A finite element model of the superstructure is built as the target model without modelling the piers, which is displayed in Fig. 17. The girders and towers use C55 concrete, and displacements at the second support counted from the left end are constrained.

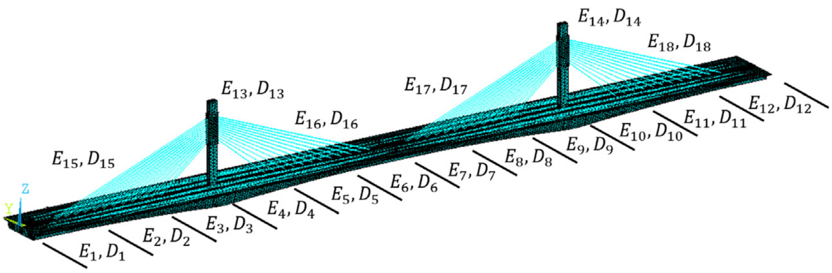


**Fig. 17.** Finite element model and partitioned substructures.

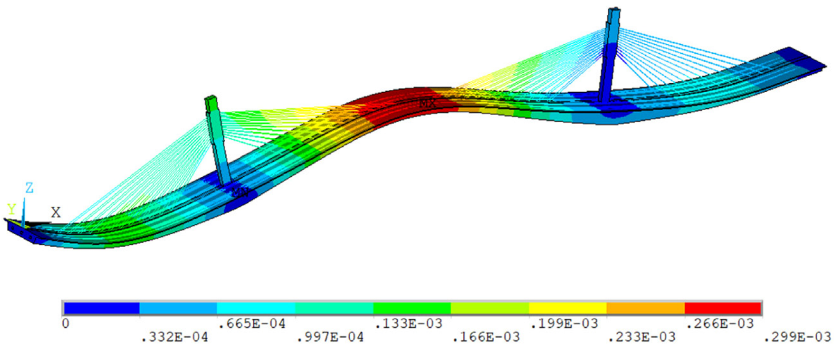


**Fig. 18.** First vertical mode.

**Table 4**
Input distributions of the extradosed cable-stayed bridge.

| Variable | Distribution | Mean | Standard deviation |
|---|---|---|---|
| Elastic modulus $E_1 \sim E_{12}$ (girder) (Pa) | Lognormal | 3.55e10 | 3.55e9 |
| Elastic modulus $E_{13}$, $E_{14}$ (tower) (Pa) | Lognormal | 3.55e10 | 3.55e9 |
| Elastic modulus $E_{15} \sim E_{18}$ (cable) (Pa) | Lognormal | 1.95e11 | 1.95e10 |
| Density $D_1 \sim D_{12}$ (girder) (kg/m$^3$) | Weibull | 2549 | 254.9 |
| Density $D_{13}$, $D_{14}$ (tower) (kg/m$^3$) | Weibull | 2549 | 254.9 |
| Density $D_{15} \sim D_{18}$ (cable) (kg/m$^3$) | Weibull | 8005 | 800.5 |

Our target on the bridge dynamic characteristics is the first vertical natural frequency. The corresponding mode shape is shown in Fig. 18. To identify the impact of the structure material properties on the target characteristic, each side span of the girder is divided into three equal-length regions and the middle span is divided into six equal-length regions. Moreover, the cables on the same side of each tower are grouped into one group, as shown in Fig. 17. Together with the two towers, eighteen portions (substructures) are considered. The elastic moduli and densities of these eighteen substructures, 36 input variables in total, are taken as input variables in this study. Distributions of these variables are reported in Table 4, and they are assumed to be mutually independent. The mean values of these variables are their nominal values, and the variance is set as 0.1 coefficient of variation, which is defined as the ratio of the std to the mean [12]. The input variable distributions are chosen subjectively to reflect the parameter uncertainty, but this setting does not affect the assessment of the proposed method. The input variables are transformed into standard Gaussian variables to build a Hermite PCE model. The total degree to truncate the PCE model is chosen as 3 to ensure the calculation accuracy, and this will generate a PCE model with polynomial terms $P = 9139$. The initial sampling number is set as 20. The threshold to terminate the sampling process is set as $10^{-5}$. Finally, a total of 192 samples are obtained to attain the precision. The obtained PCE model is then used to quantify the parameter sensitivity by calculating the ANOVA indices. Moreover, due to the heavy computational burden in generating massive MC simulations in this case, the reference solution of the ANOVA indices is obtained by building a sparse PCE model trained with 5000 LHS samples for comparison. The first-order indices and the

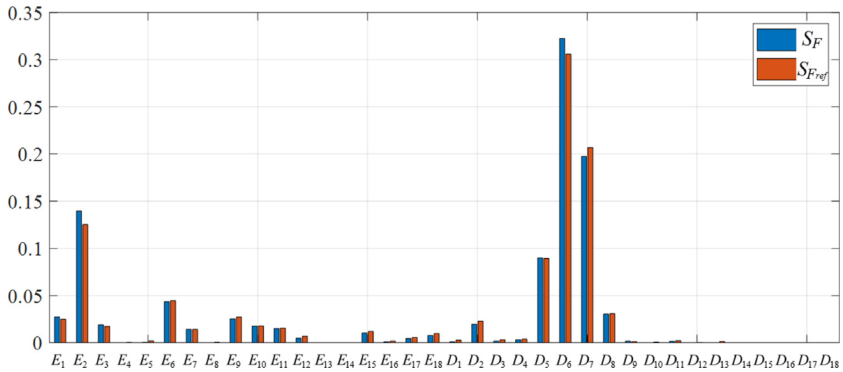


**Fig. 19.** First-order indices of input variables of the extradosed cable-stayed bridge.

total indices of the 36 variables are calculated. Since these two indices of each parameter have almost the same value, only the first-order indices are plotted in Fig. 19, in which $S_F$ represents the ANOVA indices calculated with samples collected by the proposed Coh-entro strategy, and $S_{F_{ref}}$ denotes the reference value.

From the results of sensitivity analysis, it can be concluded that the ANOVA indices obtained by the proposed Coh-entro strategy have very tiny discrepancy compared with the reference solutions, demonstrating that the PCE model trained with 192 samples collected by Coh-entro is reliable for sensitivity analysis. In 36 input variables, the material properties of the two towers, $E_{13}$, $E_{14}$ and $D_{13}$, $D_{14}$, are insensitive to the first vertical nature frequency, which is in line with our knowledge that the towers have less influence on the girder vibration. Moreover, the densities of the stay cables, $D_{15}{\sim}D_{18}$, are insensitive to the first vertical nature frequency; and the elastic moduli of the cables, $E_{15}{\sim}E_{18}$, are also in low sensitivity. $E_2$ and $D_5{\sim}D_7$ are the most sensitive parameters, and $E_1$, $E_3$, $E_6$, $E_7$, $E_9$, $E_{10}$, $E_{11}$ and $D_2$, $D_8$ have the second most influence on the frequency. The remaining parameters have extremely small sensitivity values, which means that these parameters are less influential. Moreover, it is worth noting that for each of the parameters the first-order indices and the total indices have nearly the same values, which implies that there are slight interaction effects among these parameters on the first vertical natural frequency.

## 6. Conclusions

This paper proposes a new sequential sampling method termed coherence-entropy strategy for polynomial chaos expansion (PCE) modelling, which comprises two popular sampling strategies, coherence-optimal sampling and Bayesian experimental design (BED), to take advantage of input and output information simultaneously. The Bayesian compressive sensing (BCS) is employed as a sparse regression procedure to calculate the unknown coefficients associated with a simple representation of PCE model, which also provides the foundation for BED. The input information is first utilised to form a coherence-optimal distribution in line with compressive sampling theory. Sampling in this distribution ensures a lower bound on sample quantity for accurate recovery of the PCE coefficients. In order to build a sequential sampling framework, the Latin hypercube sampling (LHS) method is instead employed to collect samples from the quasi-coherence-optimal distribution, which ignores the dependence among input variables in the coherence-optimal distribution. BED is encompassed in the sequential sampling framework by use of the differential entropy to expedite convergence. It also benefits sampling at the beginning of iteration process in that the collected samples in the quasi-coherence-optimal distribution will always have a favourable space filling property.

To validate the proposed method, three analytical functions (one shown in the Appendix) with different complexity were studied, and the results from the proposed method were compared with those from three input-dependent only methods and two output-oriented methods, in which the Coh-Opt method and the Entro method are the components of the proposed method. It is shown that the proposed approach and the output-oriented methods generally outperform the input-dependent only methods in convergence rate, and meanwhile the modelling results have good accuracy after convergence. The proposed method generally has the fastest convergence rate. Among the output-oriented methods, the proposed strategy gets samples in a more space filling way than the Entro method and the CV-ELF criterion, and the obtained PCE models are in more stable convergence performance as shown in the repeated tests. For the problems with relatively less unknown PCE coefficients, the proposed strategy gives rise to the condition number of the regression matrix in a low value, thus performing better than the Entro method and the CV-ELF criterion. For the problems with the number of unknown coefficients much higher than the required ED size, the condition number is hardly kept at a low value, but the proposed sampling strategy still ensures a fast convergence rate and a high PCE modelling accuracy after convergence. Also, it is observed that the proposed method has a faster convergence rate than the Coh-Opt method and gets more stable modelling results than the Entro method, which demonstrates that the combination of BED and coherence-optimal sampling could help improve both methods. Furthermore, two engineering cases on parameter sensitivity analysis for static and dynamic structural properties were investigated. It is demonstrated that, by imposing an accuracy threshold to the PCE model, the best sample number can be determined automatically and reliable sensitivity analysis results can be achieved.

Generation of samples to obey the coherence-optimal distribution with good space filling property is in general difficult under the sequential sampling framework, so the lower bound of sample number for accurate recovery of $l_1$ minimisation problem under the compressive sensing theory could not be achieved in the proposed approach. However, combining both input and output information to instruct sampling process was demonstrated effective. In the future work, some other input-dependent sampling procedures, which may be easy to be implemented under the sequential sampling framework, will be considered to combine with BED to achieve better outcomes.

## Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

## Acknowledgements

The work described in this paper is supported in part by a grant from the Research Grants Council of the Hong Kong Special Administrative Region (SAR), China (Project No. PolyU 152024/17E). The authors also appreciate the funding support by the Innovation and Technology Commission of Hong Kong SAR Government to the Hong Kong Branch of National Engineering Research Centre on Rail Transit Electrification and Automation (Grant No. K-BBY1).

## Appendix. Benchmark test on the Ishigami function

The Ishigami function is a highly nonlinear and nonmonotonic function with three input variables which has been extensively studied in the past [22,27,35]. All the variables conform to uniform distributions on the interval $[-\pi, \pi]$. The expression of this function is as follows:

$$y = \sin x_1 + a\,(\sin x_2)^2 + b{x_3}^4 \sin x_1 \tag{A.1}$$

where $a$ and $b$ are two parameters which are commonly set as 7 and 0.1, respectively. The analytical mean is 3.5 and the std is 3.7208. Three total degrees, 9, 12 and 14, are considered in this study, which generate different

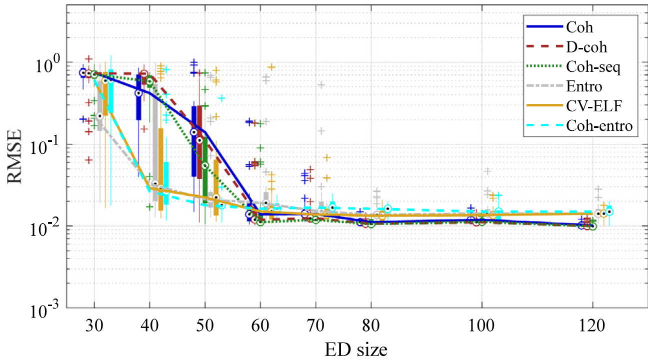


**Fig. A.1.** RMSE with different ED size under degree 9.

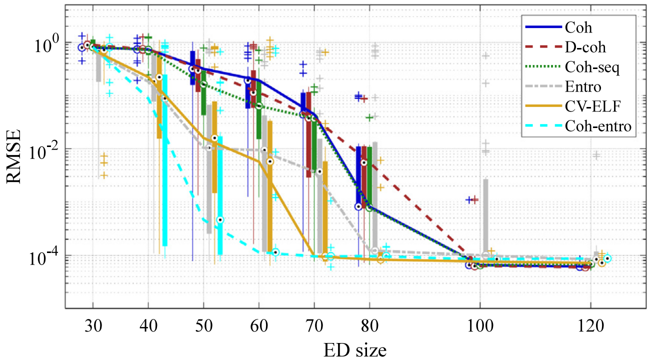


**Fig. A.2.** RMSE with different ED size under degree 12.

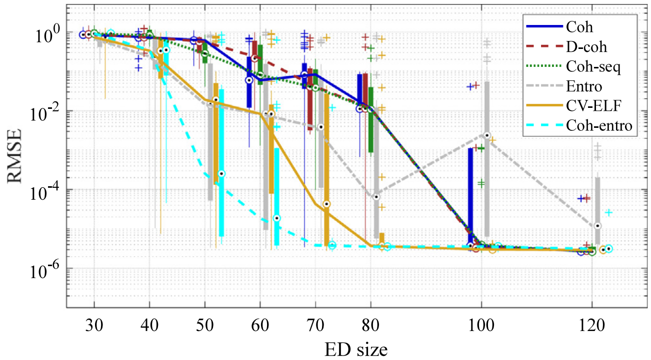


**Fig. A.3.** RMSE with different ED size under degree 14.

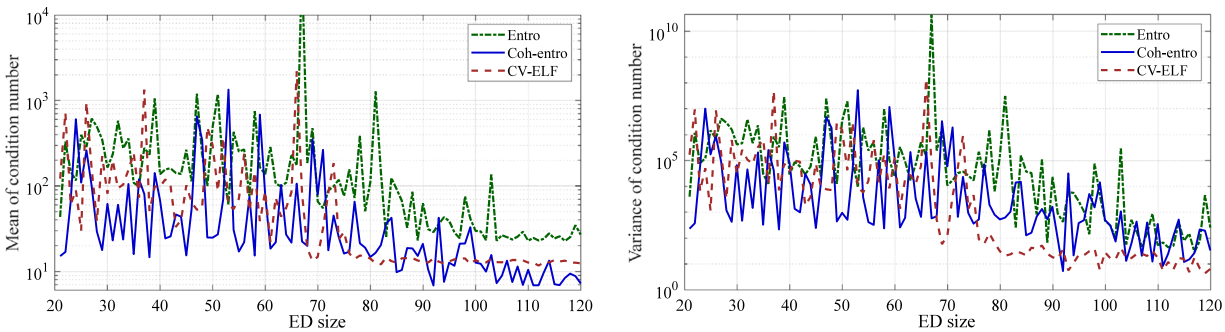


**Fig. A.4.** Mean (left) and variance (right) of the condition number with increasing ED size under degree 9.

polynomial items of $P = 220$, $P = 455$, and $P = 680$. The initial sampling number is set as 20. Each method is calculated for 30 times to ensure statistical stability.

The RMSE results with respect to the increase of ED size are shown by box plots in Figs. A.1 to A.3. In this case, the CV-ELF criterion and the proposed Coh-entro strategy have much faster convergence rate than the input-dependent only methods in all the cases with different degrees. The Entro method performs worse than CV-ELF and Coh-entro except under degree 9. Only around 60, 80 and 80 samples are needed for convergence of CV-ELF and Coh-entro with PCE degrees of 9, 12 and 14, respectively, and the input-dependent only methods need around 70, 100 and 120 samples respectively for convergence. The Entro method needs around 70 and 120 samples for convergence with PCE degrees of 9 and 12, respectively. In degree 14, the validation error from the Entro method decreases fast but it cannot converge to a stable result. This demonstrates that by use of the model observation

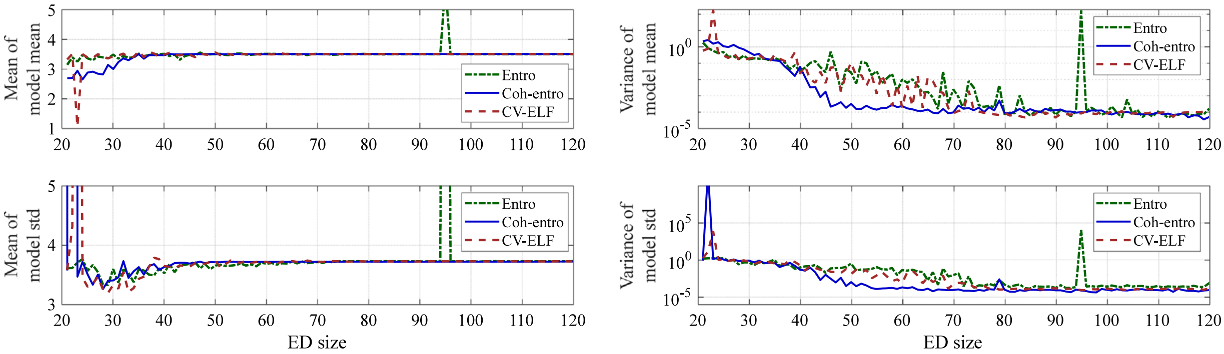


**Fig. A.5.** Mean (left) and variance (right) of the model means and standard deviations with increasing ED size under degree 9.

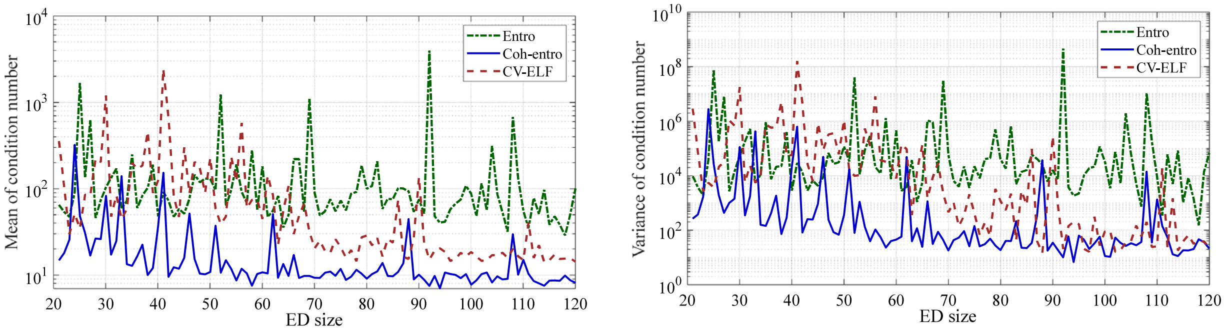


**Fig. A.6.** Mean (left) and variance (right) of the condition number with increasing ED size under degree 12.

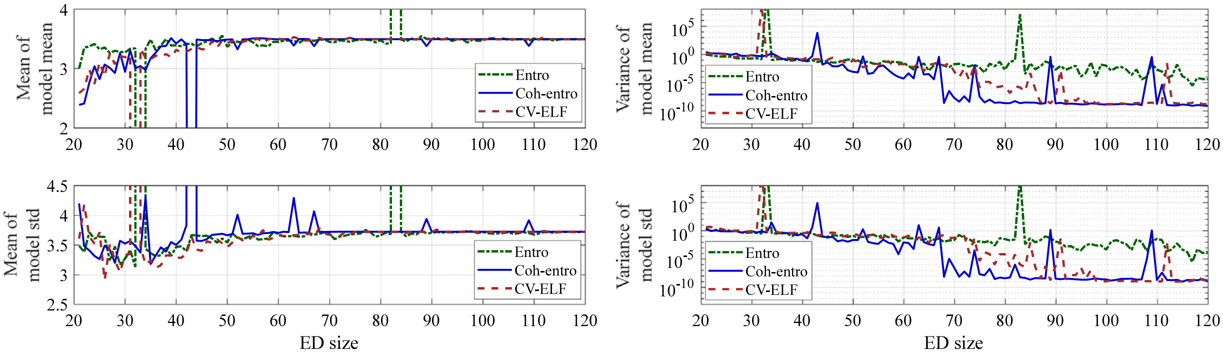


**Fig. A.7.** Mean (left) and variance (right) of the model means and standard deviations with increasing ED size under degree 12.

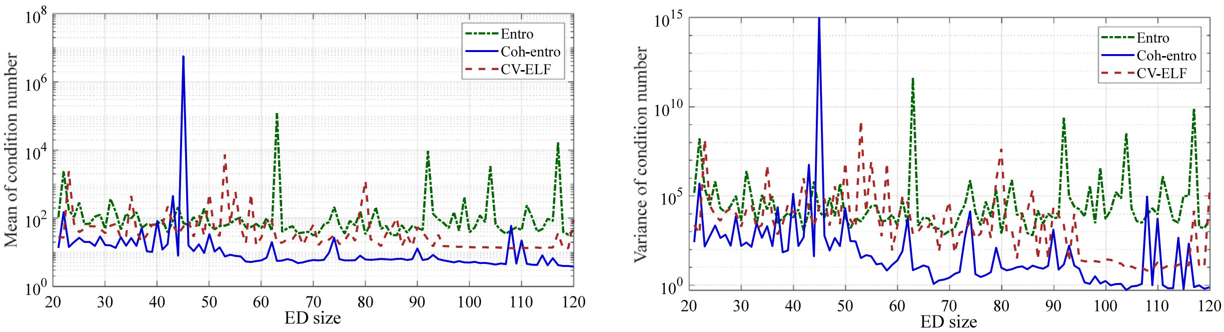


**Fig. A.8.** Mean (left) and variance (right) of the condition number with increasing ED size under degree 14.

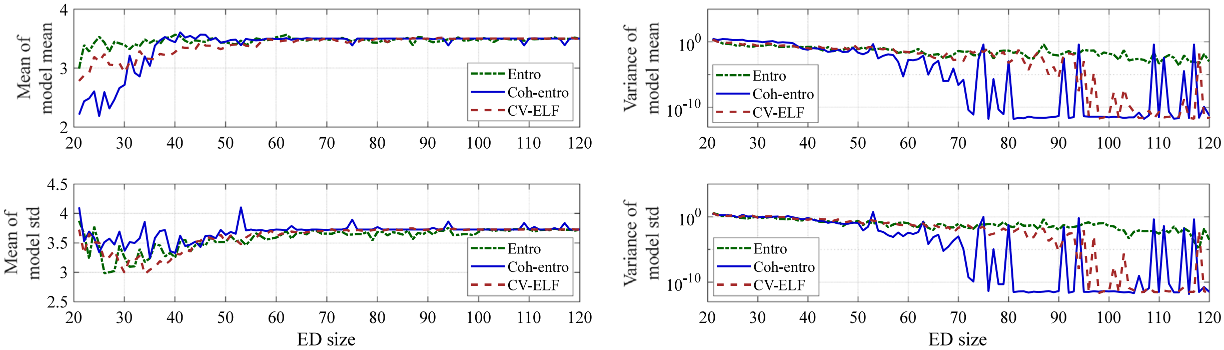


**Fig. A.9.** Mean (left) and variance (right) of the model means and standard deviations with increasing ED size under degree 14.

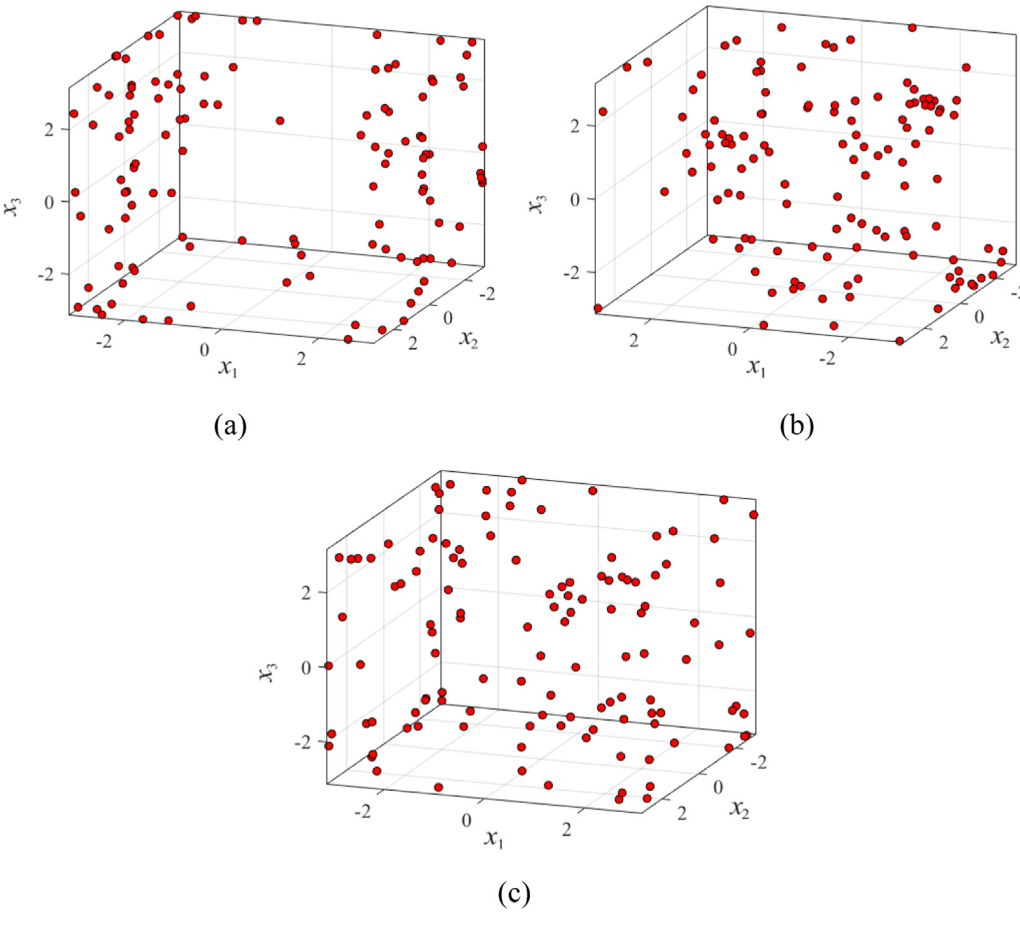


**Fig. A.10.** Final samples from (a) Entro; (b) CV-ELF; (c) Coh-entro.

information under the Bayesian framework, the sampling process will have a quite fast convergence rate; but using only BED may not be able to achieve stable solutions. Moreover, the proposed Coh-entro strategy has an apparently better convergence performance than the CV-ELF criterion, particularly for the PCE models with degrees $p = 12$ and $p = 14$.

To provide an intuitive insight into the computational stability of the Entro method, the Coh-entro criterion and the CV-ELF method, the condition number of the regression matrix and the statistical properties (mean and variance) of the model mean and std values under different degrees are depicted in Figs. A.4 to A.9. As illustrated in Fig. A.4 under the PCE degree 9, the condition numbers from the CV-ELF criterion and the Coh-entro method exhibit similar trends with increasing ED, while the condition numbers from the Entro method show a bit larger values than these two methods. For the PCE models of degrees 12 and 14, the condition numbers from the Coh-entro strategy mostly keep lower than those from the Entro method and the CV-ELF criterion, except for a few emerged peaks. The variance of the condition numbers in these two cases shows that the Coh-entro strategy generates more stable and lower condition numbers than the Entro method and the CV-ELF criterion in most circumstances, especially after convergence. In regard to the model mean and std under the PCE degree 9 in Fig. A.5, all the three methods render the PCE models converge to the target one. When the PCE degree is altered to 12, Coh-entro has a quick convergence rate compared with CV-ELF, and the variances of model mean and std from Coh-entro mostly keep at low values after convergence, as shown in Fig. A.7 (right); whereas the Entro method cannot converge well. The same phenomenon can be observed in Fig. A.9 for the PCE degree 14.

The final samples of these three methods in one test under degree 14 are shown in Fig. A.10 for comparison. It is seen that the samples from the Entro method still concentrate more on the edges of the input space, and the samples from the Coh-entro strategy are more space filling than those from the CV-ELF criterion.

## References


[1] G. Blatman, Adaptive Sparse Polynomial Chaos Expansions for Uncertainty Propagation and Sensitivity Analysis (Ph.D. thesis), Université Blaise Pascal, Clermont-Ferrand, 2009.

[2] K. Sepahvand, S. Marburg, H.-J. Hardtke, Uncertainty quantification in stochastic systems using polynomial chaos expansion, Int. J. Appl. Mech. 02 (02) (2010) 305–353, http://dx.doi.org/10.1142/S1758825110000524.

[3] P. Ni, Y. Xia, J. Li, H. Hao, Using polynomial chaos expansion for uncertainty and sensitivity analysis of bridge structures, Mech. Syst. Signal Process. 119 (2019) 293–311, http://dx.doi.org/10.1016/j.ymssp.2018.09.029.

[4] X. Sun, Y.Y. Choi, J.-I. Choi, Global sensitivity analysis for multivariate outputs using polynomial chaos-based surrogate models, Appl. Math. Model. 82 (2020) 867–887, http://dx.doi.org/10.1016/j.apm.2020.02.005.

[5] K. Cheng, Z. Lu, Active learning polynomial chaos expansion for reliability analysis by maximizing expected indicator function prediction error, Internat. J. Numer. Methods Engrg. 121 (14) (2020) 3159–3177, http://dx.doi.org/10.1002/nme.6351.

[6] A. Kareem, Effect of parametric uncertainties on wind excited structural response, J. Wind Eng. Ind. Aerodyn. 30 (1988) 233–241, http://dx.doi.org/10.1016/0167-6105(88)90088-8.

[7] W.K. Liu, T. Belytschko, A. Mani, Probabilistic finite elements for nonlinear structural dynamics, Comput. Methods Appl. Mech. Engrg. 56 (1986) 61–81, http://dx.doi.org/10.1016/0045-7825(86)90136-2.

[8] H. Leng, Z. He, Computing eigenvalue bounds of structures with uncertain-but-non-random parameters by a method based on perturbation theory, Commun. Numer. Methods. Eng. 23 (11) (2007) 973–982, http://dx.doi.org/10.1002/cnm.936.

[9] A. Vishwanathan, G.A. Vio, Efficient quantification of material uncertainties in reliability-based topology optimization using random matrices, Comput. Methods Appl. Mech. Engrg. 351 (2019) 548–570, http://dx.doi.org/10.1016/j.cma.2019.03.047.

[10] J. Zhang, X. Du, A second-order reliability method with first-order efficiency, J. Mech. Des. 132 (2010) 101006, http://dx.doi.org/10.1115/1.4002459.

[11] K. Zhou, J. Tang, Uncertainty quantification of mode shape variation utilizing multi-level multi-response Gaussian process, J. Vib. Acoust. 143 (1) (2021) 011003, http://dx.doi.org/10.1115/1.4047700.

[12] H.P. Wan, W.X. Ren, M.D. Todd, Arbitrary polynomial chaos expansion method for uncertainty quantification and global sensitivity analysis in structural dynamics, Mech. Syst. Signal Process. 142 (2020) 106732, http://dx.doi.org/10.1016/j.ymssp.2020.106732.

[13] R. Pulch, Stochastic collocation and stochastic Galerkin methods for linear differential algebraic equations, J. Comput. Appl. Math. 262 (2014) 281–291, http://dx.doi.org/10.1016/j.cam.2013.10.046.

[14] O.P. Le Maître, O.M. Knio, Spectral Methods for Uncertainty Quantification, Springer Netherlands, Dordrecht, 2010, http://dx.doi.org/10.1007/978-90-481-3520-2.

[15] G. Blatman, B. Sudret, Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach, C. R. Méc. 336 (6) (2008) 518–523, http://dx.doi.org/10.1016/j.crme.2008.02.013.

[16] P.S. Palar, T. Tsuchiya, G.T. Parks, Multi-fidelity non-intrusive polynomial chaos based on regression, Comput. Methods Appl. Mech. Engrg. 305 (2016) 579–606, http://dx.doi.org/10.1016/j.cma.2016.03.022.

[17] M. Raisee, D. Kumar, C. Lacor, A non-intrusive model reduction approach for polynomial chaos expansion using proper orthogonal decomposition, Internat. J. Numer. Methods Engrg. 103 (4) (2015) 293–312, http://dx.doi.org/10.1002/nme.4900.

[18] Y. Zhou, Z. Lu, K. Cheng, Y. Shi, An expanded sparse Bayesian learning method for polynomial chaos expansion, Mech. Syst. Signal Process. 128 (2019) 153–171, http://dx.doi.org/10.1016/j.ymssp.2019.03.032.

[19] S. Salehi, M. Raisee, M.J. Cervantes, A. Nourbakhsh, An efficient multifidelity $\ell 1$-minimization method for sparse polynomial chaos, Comput. Methods Appl. Mech. Engrg. 334 (2018) 183–207, http://dx.doi.org/10.1016/j.cma.2018.01.055.

[20] H. Zhao, Z. Gao, F. Xu, Y. Zhang, J. Huang, An efficient adaptive forward–backward selection method for sparse polynomial chaos expansion, Comput. Methods Appl. Mech. Engrg. 355 (2019) 456–491, http://dx.doi.org/10.1016/j.cma.2019.06.034.

[21] D.C. Montgomery, Design and Analysis of Experiments, eighth ed., John Wiley & Sons, Inc., 2012.

[22] N. Lüthen, S. Marelli, B. Sudret, Sparse polynomial chaos expansions: Literature survey and benchmark, 2020, arXiv preprint arXiv:2002.01290.

[23] J. Peng, J. Hampton, A. Doostan, A weighted $\ell 1$-minimization approach for sparse polynomial chaos expansions, J. Comput. Phys. 267 (2014) 92–111, http://dx.doi.org/10.1016/j.jcp.2014.02.024.

[24] X. Huan, C. Safta, K. Sargsyan, Z.P. Vane, G. Lacaze, J.C. Oefelein, H.N. Najm, Compressive sensing with cross-validation and stop-sampling for sparse polynomial chaos expansions, SIAM/ASA J. Uncertain. Quantif. 6 (2) (2018) 907–936, http://dx.doi.org/10.1137/17M1141096.

[25] G. Blatman, B. Sudret, Adaptive sparse polynomial chaos expansion based on least angle regression, J. Comput. Phys. 230 (6) (2011) 2345–2367, http://dx.doi.org/10.1016/j.jcp.2010.12.021.

[26] X. Sun, Q. Tu, J. Chen, C. Zhang, X. Duan, Probabilistic load flow calculation based on sparse polynomial chaos expansion, IET Gener. Transm. Distrib. 12 (2018) 2735–2744, http://dx.doi.org/10.1049/iet-gtd.2017.0859.

[27] Y. Zhou, Z. Lu, K. Cheng, C. Ling, An efficient and robust adaptive sampling method for polynomial chaos expansion in sparse Bayesian learning framework, Comput. Methods Appl. Mech. Engrg. 352 (2019) 654–674, http://dx.doi.org/10.1016/j.cma.2019.04.046.

[28] G. Blatman, B. Sudret, An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis, Probab. Eng. Mech. 25 (2) (2010) 183–197, http://dx.doi.org/10.1016/j.probengmech.2009.10.003.

[29] F. Ni, P.H. Nguyen, J.F.G. Cobben, Basis-adaptive sparse polynomial chaos expansion for probabilistic power flow, IEEE Trans. Power Syst. 32 (2017) 694–704, http://dx.doi.org/10.1109/TPWRS.2016.2558622.

[30] N. Fajraoui, S. Marelli, B. Sudret, Sequential design of experiment for sparse polynomial chaos expansions, SIAM/ASA J. Uncertain. Quantif. 5 (2017) 1061–1085, http://dx.doi.org/10.1137/16M1103488.

[31] G. Migliorati, F. Nobile, E. von Schwerin, R. Tempone, Analysis of discrete $L^2$ projection on polynomial spaces with random evaluations, Found. Comput. Math. 14 (2014) 419–456, http://dx.doi.org/10.1007/s10208-013-9186-4.

[32] J.D. Jakeman, M.S. Eldred, K. Sargsyan, Enhancing $\ell 1$-minimization estimates of polynomial chaos expansions using basis selection, J. Comput. Phys. 289 (2015) 18–34, http://dx.doi.org/10.1016/j.jcp.2015.02.025.

[33] E. Burnaev, I. Panin, B. Sudret, Efficient design of experiments for sensitivity analysis based on polynomial chaos expansions, Ann. Math. Artif. Intell. 81 (2017) 187–207, http://dx.doi.org/10.1007/s10472-017-9542-1.

[34] P. Diaz, A. Doostan, J. Hampton, Sparse polynomial chaos expansions via compressed sensing and D-optimal design, Comput. Methods Appl. Mech. Engrg. 336 (2018) 640–666, http://dx.doi.org/10.1016/j.cma.2018.03.020.

[35] M. Thapa, S.B. Mulani, R.W. Walters, Adaptive weighted least-squares polynomial chaos expansion with basis adaptivity and sequential adaptive sampling, Comput. Methods Appl. Mech. Engrg. 360 (2020) 112759, http://dx.doi.org/10.1016/j.cma.2019.112759.

[36] Y. Shin, D. Xiu, Nonadaptive quasi-optimal points selection for least squares linear regression, SIAM J. Sci. Comput. 38 (1) (2016) A385–A411, http://dx.doi.org/10.1137/15M1015868.

[37] N. Alemazkoor, H. Meidani, A near-optimal sampling strategy for sparse recovery of polynomial chaos expansions, J. Comput. Phys. 371 (2018) 137–151, http://dx.doi.org/10.1016/j.jcp.2018.05.025.

[38] M. Hadigol, A. Doostan, Least squares polynomial chaos expansion: A review of sampling strategies, Comput. Methods Appl. Mech. Engrg. 332 (2018) 382–407, http://dx.doi.org/10.1016/j.cma.2017.12.019.

[39] K. Crombecq, E. Laermans, T. Dhaene, Efficient space-filling and non-collapsing sequential design strategies for simulation-based modeling, European J. Oper. Res. 214 (2011) 683–696, http://dx.doi.org/10.1016/j.ejor.2011.05.032.

[40] R. Sheikholeslami, S. Razavi, Progressive Latin hypercube sampling: An efficient approach for robust sampling-based analysis of environmental models, Environ. Model. Softw. 93 (2017) 109–126, http://dx.doi.org/10.1016/j.envsoft.2017.03.010.

[41] M. Stein, Large sample properties of simulations using Latin hypercube sampling, Technometrics 29 (1987) 143–151, http://dx.doi.org/10.2307/1269769.

[42] S. Zein, B. Colson, F. Glineur, An efficient sampling method for regression-based polynomial chaos expansion, Commun. Comput. Phys. 13 (4) (2013) 1173–1188, http://dx.doi.org/10.4208/cicp.020911.200412a.

[43] J. Hampton, A. Doostan, Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies, J. Comput. Phys. 280 (2015) 363–386, http://dx.doi.org/10.1016/j.jcp.2014.09.019.

[44] Q. Pan, X. Qu, L. Liu, D. Dias, A sequential sparse polynomial chaos expansion using Bayesian regression for geotechnical reliability estimations, Int. J. Numer. Anal. Methods Geomech. 44 (2020) 874–889, http://dx.doi.org/10.1002/nag.3044.

[45] P. Sebastiani, H.P. Wynn, Maximum entropy sampling and optimal Bayesian experimental design, J. R. Stat. Soc. Ser. B Stat. Methodol. 62 (2000) 145–157, http://dx.doi.org/10.1111/1467-9868.00225.

[46] K. Chaloner, I. Verdinelli, Bayesian experimental design: A review, Stat. Sci. 10 (1995) 273–304, http://dx.doi.org/10.1214/ss/1177009939.

[47] R. Hou, Y. Xia, X. Zhou, Y. Huang, Sparse Bayesian learning for structural damage detection using expectation–maximization technique, Struct. Control Health Monit. 26 (2019) e2343, http://dx.doi.org/10.1002/stc.2343.

[48] D. Xiu, G.E. Karniadakis, The Wiener-Askey polynomial chaos for stochastic differential equations, SIAM J. Sci. Comput. 24 (2002) 619–644, http://dx.doi.org/10.1137/S1064827501387826.

[49] S. Hosder, R. Walters, M. Balch, Efficient sampling for non-intrusive polynomial chaos applications with multiple uncertain input variables, in: 48th AIAA/ASME/ASCE/AHS/ASC Struct. Struct. Dyn. Mater. Conf., American Institute of Aeronautics and Astronautics, Honolulu, Hawaii, USA, 2007, http://dx.doi.org/10.2514/6.2007-1939.

[50] A.M. Bruckstein, D.L. Donoho, M. Elad, From sparse solutions of systems of equations to sparse modeling of signals and images, SIAM Rev. 51 (2009) 34–81, http://dx.doi.org/10.1137/060657704.

[51] S.D. Babacan, R. Molina, A.K. Katsaggelos, Bayesian compressive sensing using Laplace priors, IEEE Trans. Image Process. 19 (1) (2010) 53–63, http://dx.doi.org/10.1109/TIP.2009.2032894.

[52] T. Jiang, X.W. Zhang, Y. Li, Bayesian compressive sensing using reweighted Laplace priors, AEU - Int. J. Electron. Commun. 97 (2018) 178–184, http://dx.doi.org/10.1016/j.aeue.2018.10.005.

[53] J.L. Beck, L.S. Katafygiotis, Updating models and their uncertainties. I: Bayesian statistical framework, J. Eng. Mech. 124 (1998) 455–461, http://dx.doi.org/10.1061/(ASCE)0733-9399(1998)124:4(455).

[54] F. Krahmer, R. Ward, Stable and robust sampling strategies for compressive imaging, IEEE Trans. Image Process. 23 (2) (2014) 612–622, http://dx.doi.org/10.1109/TIP.2013.2288004.

[55] D.J.C. MacKay, Information-based objective functions for active data selection, Neural Comput. 4 (1992) 590–604, http://dx.doi.org/10.1162/neco.1992.4.4.590.

[56] S. Ji, Y. Xue, L. Carin, Bayesian compressive sensing, IEEE Trans. Signal Process. 56 (2008) 2346–2356, http://dx.doi.org/10.1109/TSP.2007.914345.

[57] M.W. Seeger, H. Nickisch, Compressed sensing and Bayesian experimental design, in: Proc. 25th Int. Conf. Mach. Learn. - ICML '08, ACM Press, New York, USA, 2008, pp. 912–919, http://dx.doi.org/10.1145/1390156.1390271.

[58] C. Papadimitriou, J.L. Beck, S.K. Au, Entropy-based optimal sensor location for structural model updating, J. Vib. Control 6 (5) (2000) 781–800, http://dx.doi.org/10.1177/107754630000600508.

[59] F. Nielsen, On a generalization of the Jensen–Shannon divergence and the JS-symmetrization of distances relying on abstract means, Entropy 21 (5) (2019) 485, http://dx.doi.org/10.3390/e21050485.

[60] B.Y. Zhang, Y.Q. Ni, An inversed greedy method for information-based optimal sensor placement on bridges, in: Structural Health Monitoring 2019, San Francisco, USA, 2019. http://dx.doi.org/10.12783/shm2019/32433.

[61] M.E. Tipping, Sparse Bayesian learning and the relevant vector machine, J. Mach. Learn. Res. (2001) 211–244, http://dx.doi.org/10.1162/15324430152748236.

[62] M. Thapa, S.B. Mulani, R.W. Walters, A new non-intrusive polynomial chaos using higher order sensitivities, Comput. Methods Appl. Mech. Engrg. 328 (2018) 594–611, http://dx.doi.org/10.1016/j.cma.2017.09.024.

[63] G. Zielke, Some remarks on matrix norms, condition numbers, and error estimates for linear equations, Linear Algebra Appl. 110 (1988) 29–41, http://dx.doi.org/10.1016/0024-3795(83)90130-1.

[64] I.M. Sobol', Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Math. Comput. Simulation 55 (2001) 271–280, http://dx.doi.org/10.1016/S0378-4754(00)00270-6.

[65] Ö.F. Alış, H. Rabitz, Efficient implementation of high dimensional model representations, J. Math. Chem. 29 (2001) 127–142, http://dx.doi.org/10.1023/A:1010979129659.

[66] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliab. Eng. Syst. Saf. 93 (7) (2008) 964–979, http://dx.doi.org/10.1016/j.ress.2007.04.002.

[67] T. Crestaux, O. Le Maıtre, J.-M. Martinez, Polynomial chaos expansion for sensitivity analysis, Reliab. Eng. Syst. Saf. 94 (2009) 1161–1172, http://dx.doi.org/10.1016/j.ress.2008.10.008.

[68] S. Marelli, B. Sudret, An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis, Struct. Saf. 75 (2018) 67–74, http://dx.doi.org/10.1016/j.strusafe.2018.06.003.