

RESEARCH

Open Access



Optimized sparse polynomial chaos expansion with entropy regularization

Sijie Zeng¹, Xiaojun Duan¹, Jiangtao Chen² and Liang Yan^{1*}

*Correspondence:

yanliang@nudt.edu.cn

¹College of Liberal Arts and Sciences, National University of Defense Technology, Changsha, China

Full list of author information is available at the end of the article

Abstract

Sparse Polynomial Chaos Expansion (PCE) is widely used in various engineering fields to quantitatively analyse the influence of uncertainty, while alleviating the problem of dimensionality curse. However, current sparse PCE techniques focus on choosing features with the largest coefficients, which may ignore uncertainties propagated with high order features. Hence, this paper proposes the idea of selecting polynomial chaos basis based on information entropy, which aims to retain the advantages of existing sparse techniques while considering entropy change as output uncertainty. A novel entropy-based optimization method is proposed to update the state-of-the-art sparse PCE models. This work further develops an entropy-based synthetic sparse model, which has higher computational efficiency. Two benchmark functions and a computational fluid dynamics (CFD) experiment are used to compare the accuracy and efficiency between the proposed method and classical methods. The results show that entropy-based methods can better capture the features of uncertainty propagation, improving accuracy and reducing sparsity while avoiding over-fitting problems.

Keywords: Uncertainty quantification, Sparse polynomial chaos expansion, Entropy, Regularization

1 Introduction

Due to the variety of uncertainties frequently involved in engineering applications, which may cause fluctuations in the performance of a system, it is necessary to comprehensively consider the impact of uncertain factors [1, 2]. To deal with the problem of limited experiment resources, various surrogate models [3, 4] have been proposed, aiming to construct a mathematical model that accurately mimics the behavior of the original problem with an affordable experimental design [5]. Popular surrogate techniques include the Kriging method [6–9], artificial neural network [10–12], polynomial chaos expansion (PCE) method [13–16] etc. In this paper, we focus on the PCE model, which has been widely used in engineering to quantify uncertainties efficiently.

The basic idea of PCE is to expand an exact solution of a stochastic process space by polynomial expansion [17–19], and for the model generated by which can generally be solved by ordinary least squares (OLS). Many successful applications of uncertainty quantification based on the PCE method have been achieved [20–22]. However, the

cost of constructing a PCE model increases exponentially with the dimension of input parameters, i.e. the curse of dimensionality, thus severely restricts practical applications of the model at the industrial level [23–25]. In order to solve this problem, lots of sparse algorithms have been developed in recent years. Some typical examples of well-established methods are the sparse regression method [26, 27] and the compressive sampling method [28, 29].

A problem with some OLS-solution methods is the possibility of over-fitting but can be avoided by regularization, which is another practical solution to seek a *sparse* representation [30]. Regularization is a typical model selection method that introduces a regularizer to the empirical risk. From the perspective of Bayesian estimation, the regularizer corresponds to the prior probability of the model [31].

There are many regularization methods, but it is ℓ_0 regularization that is required to obtain a true sparse model. However, solving model with ℓ_0 regularization is an NP-Hard problem [32, 33], so researchers usually adopt the ℓ_1 regularization when dealing with sparse problems [34], which is an optimal convex approximation of the ℓ_0 paradigm, and can be easily solved optimally to obtain a sparse model.

Although scholars have conducted a considerable amount of research on sparse polynomial chaos expansions, most of the existing sparse methods are derived from the perspective of regression [35]. In these methods, researchers usually pay attention to one or two numerical characteristics in the distribution, which are important statistics to measure uncertainty. However, the distribution is very complicated in common situations, especially when there is a noise in the original problem. Furthermore, the aforementioned methods focus on statistics, such as the relative mean squared error (MSE) or the cross-validation error, to optimize parameters, which ignores the overall uncertainties propagated from the input variables.

To depict the *full* uncertainty of the output, we introduce information entropy as a measure of the evolution of the output distribution to further optimize the sparse representation of a PCE model. Information entropy was first introduced to measure microscopic uncertainty and multiplicity by Shannon [36]. The idea was then extended to learning tasks to measure the information changes with different statistical models. Jaynes [37] proposed the maximum entropy principle (MEP) to provide an optimization criterion. B. Grechuk et al. [38] have studied the relationships among error, likelihood, and entropy in regression analysis, and found that a normal distribution can be recovered from the maximum entropy principle. Wang [39] introduced a maximum entropy penalty to model and incorporate the entropy-controlled framework with other conventional learning algorithms. Liang [40] proposed a sparse subspace clustering with entropy-norm by using information entropy as the regularizer of the objective function. Researchers further made variable selection through the evaluation of entropy, and extended the use of entropy as an important criterion for model selection [41, 42], which can be embedded in existing methods but does not form a framework.

In this work, we aim to use entropy to propagate uncertainty of input-output while retaining the advantages of existing sparse PCE techniques. Hence, we propose a novel method where entropy is an auxiliary penalty term. Firstly, a general re-optimization model is proposed which chooses the features with the largest information entropy-based on the existing optimized model. Although such a strategy is easy to implement, it may ignore some key features in the first stage. A hybrid entropy-based synthetic method

embedded in several commonly used classical sparse criteria is thus proposed. The novel regularization structure takes the value and volatility of each feature into account simultaneously. The entropy term can be regarded as a trade-off between the predictive mean value and uncertainty along with the selected features. The above two algorithms can be easily computed with existing sparse algorithms, which improves their usage scope and availability. Experiments are performed to compare the proposed two algorithms and classical sparse methods, i.e. the Orthogonal Matching Pursuit (OMP), the Least Angle Regression (LARS), the Subspace Pursuit (SP), and the Bayesian Compressive Sensing (BCS). Results show that the general re-optimization method is of simple operation and universality because it can optimize any sparse PCE model, while the hybrid entropy-based synthetic method has strong pertinence, and is superior to the former in high-dimensional complex applications.

The remainder of the paper is organized as follows. Section 2 introduces the fundamental theory of sparse PCE models. Section 3 proposes the theories and algorithms of the two optimization methods mentioned in the previous paragraph. Section 4 shows the performance of the proposed methods, illustrated by numerical examples and comparisons. Finally, the conclusions are summarized in Section 5.

2 The fundamental theory

2.1 Sparse polynomial chaos expansion

The basics of PCE are elaborated briefly as follows. Let \mathcal{M} be the original model and $Y = \mathcal{M}(\xi)$ be the output. The finite-order PCE model expanded in the full polynomial space can be expressed as:

$$Y \approx Y_{PC} = \sum_{i=0}^{P-1} c_i \Psi_i(\xi), \quad (1)$$

where ξ is d -dimensional random variables with a probability density function (PDF) of $f(\xi)$, and $\Psi_i(\xi)$ is a set of polynomial basis functions truncated at the p -th order with P denoting the degree of the PCE. Generally, P varies with different truncation schemes. For example, when the *total degree space* is chosen,

$$P = \frac{(p+d)!}{p!d!}. \quad (2)$$

with p as the maximum degree for each dimension. Given a design of experiment (DoE) $\{\xi, Y\}$, where $\xi = (\xi_1, \xi_2, \dots, \xi_n) \in \mathbb{R}^{d \times n}$ is a specific sample and $Y(\xi) \in \mathbb{R}^n$ are corresponding responses. The main effort of constructing a PCE model is solving the following generalized linear equation system:

$$\begin{bmatrix} Y(\xi_0) \\ Y(\xi_1) \\ \vdots \\ Y(\xi_n) \end{bmatrix} = \begin{bmatrix} \Psi_0(\xi_0) & \Psi_1(\xi_0) & \cdots & \Psi_{P-1}(\xi_0) \\ \Psi_0(\xi_1) & \Psi_1(\xi_1) & \cdots & \Psi_{P-1}(\xi_1) \\ \vdots & \vdots & \ddots & \vdots \\ \Psi_0(\xi_n) & \Psi_1(\xi_n) & \cdots & \Psi_{P-1}(\xi_n) \end{bmatrix} \begin{bmatrix} c_0 \\ c_1 \\ \vdots \\ c_{P-1} \end{bmatrix} \quad (3)$$

which can be rewritten in matrix form:

$$Y = \Psi c. \quad (4)$$

Although we are able to obtain a least squares estimate $\hat{c} = (\Psi^T \Psi)^{-1} \Psi^T Y$ easily, the degree P increases dramatically with p and d , hence we have to solve an underdetermined

system with limited DoE resources, and the least squares solution can be inaccurate and unstable. The sparse PCE method is proposed to solve this problem, and it restores the complete model response almost accurately by selecting a small number of basis that dominates the system output. ℓ_0 -norm is the most widely used regularization criterion, i.e. limit the degree of non-zero coefficients by solving the following optimization problem:

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathbb{R}^P} \|\mathbf{c}\|_0 \text{ subject to } \mathbf{Y} = \Psi \mathbf{c}. \tag{5}$$

However, Eq. 5 is a non-convex optimization problem, which can be difficult to solve in practice due to its NP-hardness. A typical practical technique is to replace the ℓ_0 norm term with the ℓ_1 norm term [43], and the new objective function becomes

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathbb{R}^P} \|\mathbf{c}\|_1 \text{ subject to } \mathbf{Y} = \Psi \mathbf{c}. \tag{6}$$

2.2 Error estimation and model evaluation

The optimization problem defined in Eq. 6 is usually solved by relaxing the constraint, i.e.

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathbb{R}^P} \|\mathbf{c}\|_1 \text{ subject to } \|\Psi \mathbf{c} - \mathbf{Y}\|_2 \leq \varepsilon, \tag{7}$$

where ε is recorded as the truncation error, determined by measurement noise, which is a more natural parameterization choice. If ε is too large, the reconstructed PC is not accurate enough; however, if ε is too small, the reconstructed PC may be over-fitting. The leave-one-out (LOO) error is used to measure the degree of over-fitting in practice, which can be expressed as follows:

$$E_{LOO} = \frac{1}{n} \sum_{i=1}^n \left(\frac{y_i - y_{PC}(\xi_i)}{1 - h_i} \right)^2, \tag{8}$$

where h_i denotes the i -th diagonal element of the matrix $\Psi (\Psi^T \Psi)^{-1} \Psi^T$. The E_{LOO} can be calculated easily when the least squares solution *w.r.t* full design are available (see [27]). We choose the PCE model with the smallest LOO error during the iteration of finding the best sparse solution.

When comparing the performance of different surrogate models, our interest lies in the precision and uncertainty. The relative MSE is a widely used measure to quantify the precision of a surrogate, which is defined by the following equation:

$$\text{Relative MSE} = \frac{\sum_{i=1}^n (y_i - y_{PC}(\xi_i))^2}{\sum_{i=1}^n (y_i - \bar{y})^2}, \tag{9}$$

where $\bar{y} = \frac{1}{n} \sum_{i=1}^n y_i$. As for the uncertainty, we pay special attention to the probability density function (PDF) or cumulative distribution function (CDF).

2.3 The concept of information entropy

The type of entropy addressed in this work is information entropy that was first proposed by Shannon in 1948 [36]. The most general explanation of entropy is the measure of uncertainty, which refers to the information contained in the system [44]. Specifically,

the more information the system contains, the smaller the uncertainty and corresponding entropy.

For any random variable $X = \{x_1, \dots, x_n\}$, the definition of information entropy is expressed in terms of a discrete set of probabilities $P(x_i)$, as follows:

$$H(X) = \sum_{i=1}^n P(x_i) \ln(P(x_i)). \quad (10)$$

Similarly, for a continuous random variable, differential entropy can be obtained:

$$H(x) = - \int p(x) \ln p(x) dx. \quad (11)$$

Based on information entropy, when predicting a probability model of random variables, the main idea of the MEP is that among all candidate distributions, the distribution that maximizes entropy should be selected. In other words, the best probability distribution has maximal entropy, which comes up with an optimization problem:

$$\text{maximize } H = - \int p(y) \ln p(y) dy \text{ subject to } \|\Psi \mathbf{c} - \mathbf{Y}\|_2 \leq \varepsilon, \quad (12)$$

where $p(y)$ is the probability density function of element y in \mathbf{Y} .

Hence, by combining MEP and the classical sparse PCE techniques, information entropy can be used as a sparse optimization criterion for the sparse PCE model for it can retain the largest amount of uncertainty while making it sparse.

3 Methodology

In this section, two adaptive methods are proposed. The first is the re-optimization of results achieved with the alternative model based on entropy, which can be thought of as a simple external plug-in. The second is hybrid optimization using entropy as a regularizer, which is the main method of this paper and it is more advantageous in high-dimensional problems.

3.1 Entropy-based re-optimization sparse PCE

As mentioned at the end of Section 2.3, the aim of this paper is to learn a sparse model and preserve the maximum uncertainty of the results. Thus, we can directly perform entropy-based secondary optimization from Eq. 12 based on existing sparse algorithms (i.e., entropy-based re-optimization sparse PCE). Typically, solving problems with the sparse PCE surrogate model can be converted to:

$$\mathbf{Y} = \sum_{i \in \mathcal{I}} \Psi_i \mathbf{c}, \quad (13)$$

where \mathcal{I} is the sparse index set.

With the classical sparse PCE method, we get the model response \mathbf{Y}_S . Additional model responses are produced through the repeated sampling process, which can then be re-optimized based on entropy sorting. Our goal is to select W items with the largest entropy values. The objective function during re-optimization is defined as follows:

$$\hat{\mathbf{c}} = \arg \min_{\mathcal{J} \subset \mathcal{I}} \|\Psi_{\mathcal{J}} \mathbf{c} - \mathbf{Y}_S\|_2, \quad (14)$$

where Y_S is the model response obtained by the classical sparse method, and index set \mathcal{J} satisfies:

$$\mathcal{J} = \arg \max_{\mathcal{J} \subset \mathcal{I}} H(Y_S) \text{ subject to } \text{card}(\mathcal{J}) \leq W, \tag{15}$$

with $H(Y_S) \triangleq \sum_{j=1}^W H(y_j)$, where $y_j = \Psi_j c_j$.

Since calculating the differential entropy of Y_S is more complicated, it can be approximated by the Shannon entropy of Monte Carlo. For example, if we repeat sampling m times as the verification set, the Shannon information entropy of y_j can be obtained:

$$H(y_j) = - \sum_{i=1}^m y_{ij}^2 \ln y_{ij}^2. \tag{16}$$

The pseudo code of the algorithm for entropy-based re-optimization sparse PCE is as follows:

Algorithm 1 Entropy-based re-optimization sparse PCE (Ent-PCE)

Input: Ψ : a set of candidate basis functions; y : observation vector. W : desired sparsity of Ent-PCE;

- 1: Establish a surrogate model based on a classical sparse PCE algorithm.
 - 2: Repeat sampling m times as the verification set and get the response Y_S .
 - 3: **for** each $j = 1, \dots, p$ **do**
 - 4: $y_j = \Psi_j c_j$
 - 5: Calculate Shannon entropy of y_j : $H_j = - \sum_{i=1}^m y_{ij}^2 \ln y_{ij}^2$
 - 6: **end for**
 - 7: Sort the indicators according to H ; \mathcal{J} is W indices corresponding to the largest value of entropy.
 - 8: Least-squares solution based on index set \mathcal{J} : $\hat{c} = \arg \min_{\mathcal{J} \in \mathcal{I}} \|\Psi_{\mathcal{J}} c - Y_S\|_2$
-

The main advantage of this method is that it can be easily implemented and extended to various sparse methods without the need to modify their internal codes. In other words, it is universal, maneuverable, and can be considered as an external plug-in. However, it requires longer computation time when dealing with high-dimensional cases because of the OLS algorithm used in the re-optimization process.

3.2 Hybrid entropy-based comprehensive sparse PCE

In Section 3.1, a universal entropy-based re-optimization sparse PCE method is proposed, which is essentially a subsequent optimization of the results trained by classical sparse PCE methods. However, its re-optimization is performed with the features and coefficients retained by a sparse algorithm, hence we cannot retrieve features that have been discarded, which may be of great influence for the overall uncertainty. As the complexity of the model increases, the discarded features by the classical sparse method will also increase, so the optimization effectiveness of Ent-PCE will be limited. Therefore, in different sparse methods, we can consider the role of information entropy when selecting features. In fact, there are several commonly used sparse algorithms in the literature, some of which can be found in open-source toolboxes. In this section, for these specific

sparse PCE algorithms, a method involving hybrid entropy-based comprehensive sparse PCE will be proposed.

In the PCE model, the unknown coefficients correspond to polynomial chaos basis. When we reduce the dimension of the coefficients, it can be seen as a feature selection for the polynomial chaos basis. From the perspective of regression, we usually select the feature basis that has a greater impact on the residual, such as the OMP method. However, the polynomial chaos basis stores the probabilistic information of the predicted response. From the MEP perspective, the feature basis with larger entropy contains more uncertainty information and is more important to the response.

To maximize the input uncertainty distribution in the response, we introduce the entropy of the basis function to regularize the classical model. Hence, the objective function is re-defined as follows:

$$\hat{\mathbf{c}} = \arg \min_{\mathbf{c} \in \mathbb{R}^P} \|\Psi \mathbf{c} - \mathbf{Y}\|_2 + \lambda (\|\mathbf{c}\|_1 + \gamma H(\Psi_{\mathcal{I}})), \quad (17)$$

An example of a hybrid entropy-based comprehensive sparse PCE algorithm based on the classical OMP method is Algorithm 2. The suffix indicates the classical sparse PCE algorithm being optimized.

Algorithm 2 Hybrid entropy-based orthogonal matching pursuit (HEnt-OMP)

Input: Ψ : a set of candidate basis functions; \mathbf{y} : observation vector.

- 1: Initialize: feature index set: $\mathcal{I} = \emptyset$; residual vector: $\mathbf{r} = \mathbf{y}$
 - 2: **for** each $i = 1, \dots, n$ **do**
 - 3: Find $\Psi_i \in \Psi$ which is closely related to the residual and has a larger entropy value by solving:
 - 4: $\Psi_i = \arg \max_{\Psi_i \in \Psi \setminus \Psi_{\mathcal{I}}} \left(\frac{|\langle \Psi_i, \mathbf{r} \rangle|}{\|\Psi_i\|_2} - \gamma \sum_{j=1}^P \Psi_{ij}^2 \ln(\Psi_{ij}^2) \right)$
 - 5: Update the index set: $\mathcal{I} = \mathcal{I} \cup \{i\}$
 - 6: Compute the coefficients \mathbf{c}^* using only the active indices by least-squares:
 - 7: $\mathbf{c}^* = \arg \min_{\mathcal{I} \in \mathbb{R}} \|\mathbf{y} - \Psi_{\mathcal{I}} \mathbf{c}^*\|_2$
 - 8: Update the residual: $\mathbf{r} = \mathbf{y} - \Psi_{\mathcal{I}} \mathbf{c}^*$
 - 9: **end for**
-

In each iteration, OMP selects the regressor that is most closely correlated with the current residual, while HEnt-OMP introduces entropy as a penalty term on this basis. The regressor must not only be related to the residual but also ensure that its entropy value is large. From this, we can find that HEnt-PCE only needs to calculate the entropy of the basis function once at the beginning of feature selection, so the extra computational effort is almost negligible.

It is worth noting that in the methodology, the prefixes ‘Ent-’ and ‘HEnt-’ indicate optimization routine and the suffixes ‘PCE’ represent universal sparse PCE algorithms. While in the algorithm as well as in the experiments, the suffix is replaced by the specified classical sparse PCE algorithm. Compared with Algorithm 1, this method is more targeted and has faster computing speeds.

However, due to the different classical methods, the selection of regularization parameters in different methods is also worthy of attention.

3.3 Parameter optimization and selection

Before conducting experiments, we need to set some parameters for these algorithms. For example, in Algorithm 1, we are concerned with the effects of changing parameter W on the final sparse result. Similarly, when using hybrid entropy-based algorithms, γ is a parameter worthy of attention. For each of the various methods used in entropy-based algorithms, the choice of parameters will be different.

The selection of parameter W is natural because we can set it directly according to the desired size of the active basis. While regarding the selection of γ , we can use a small size of experiment design in preliminary experiments before conducting the formal experiment, so that γ can get an ideal value. The algorithm is designed as Algorithm 3, the idea of which comes from DIRECT algorithm [45, 46]. The algorithm equates the search interval into k subintervals, and calculates the value of function at the boundary point of the subinterval, then takes the point with the lowest value as the centroid, the length of the subinterval as the length of the interval, thus establishing a new space to repeat the previous process until the value of function is below the set value ε .

Algorithm 3 Parameter selection

Input: \mathcal{S} : surrogate model with unknown γ ; γ_0 : the upper bound of γ

- 1: **while** The variance of Relative MSE is less than ε **do**
 - 2: **for** each $i = 0, \dots, k$ **do**
 - 3: $l = \frac{\gamma_k - \gamma_0}{k}, \gamma_i = \gamma_0 + i \times l$
 - 4: Run the model \mathcal{S} with γ_i and calculate Relative MSE: e_i
 - 5: **end for**
 - 6: Find the index i Find the index that minimizes the Relative MSE: $i = \arg \min_i e_i$
 - 7: Update the boundary: $\gamma_0 = \gamma_i - \frac{1}{2}l, \gamma_k = \gamma_i + \frac{1}{2}l$
 - 8: **end while**
-

4 Experiment analysis

In the experiments, we analyze and compare the effectiveness of four classical sparse PCE algorithms (OMP [47], LARS [27], SP [48], and BCS [49]) and two entropy-based optimized algorithms on the basis of them.

To implement the benchmark, we consider an Ishigami function ($d = 3$), an OAKLEY & O'HAGAN function ($d = 15$) and a high-dimensional function ($d = 54$) to show the performance of the methods in low-dimensional and high-dimensional cases respectively. Both benchmark functions have been inserted small random interference items to better fit the actual. The LHS sequence is used to generate sample points in the experiment. Each experiment was repeated 50 times, and the mean value is given to compare performance of the used methods. To further prove the effective performance of the proposed method, the approach is applied to a more realistic CFD application: RAE2822 airfoil ($d = 5$). For all experiments, we use the general-purpose uncertainty quantification software UQLab [35].

4.1 Toy model

Before starting the formal experiment, we first consider a toy model to explain the problem of this paper. The classical sparse method only considers the value of coefficients,

Table 1 Comparison of relative MSE and KL (Kullback-Leibler) divergence for the toy model

Results	OMP	Ent-OMP	HEnt-OMP
Relative MSE	0.237	0.217	0.249
KL Divergence	912.14	706.64	365.34

rather than the impact of feature changes on uncertainty. So we constructed a function that is a combination of smooth function and high-order features.

$$f(x, y) = 6\sin(y) + 4x(\cos(x) + 1) + 5x^6 + 10^{-10}9^x \cdot x^{22} \tag{18}$$

If classical methods such as OMP are used directly, these high-level features will be discarded, regardless of whether they contain important information or not. As shown in Table 1 and Fig. 1(a) and (b), using the two methods proposed in this paper to optimize can make the probability distribution closer to the original distribution, while keeping the relative MSR at a small value. Meanwhile, Fig. 1(c) shows the retained basis functions, and the meaning of the label of it is the method (number of retained terms): index among all basis functions. It can be found that Ent-PCE is sparse on the result of OMP, so the index set of selected features are subsets of the index set of the original result, while HEnt-PCE takes the role of information entropy into account at the time of feature selection, so some higher-order features that were originally discarded by OMP are retained.

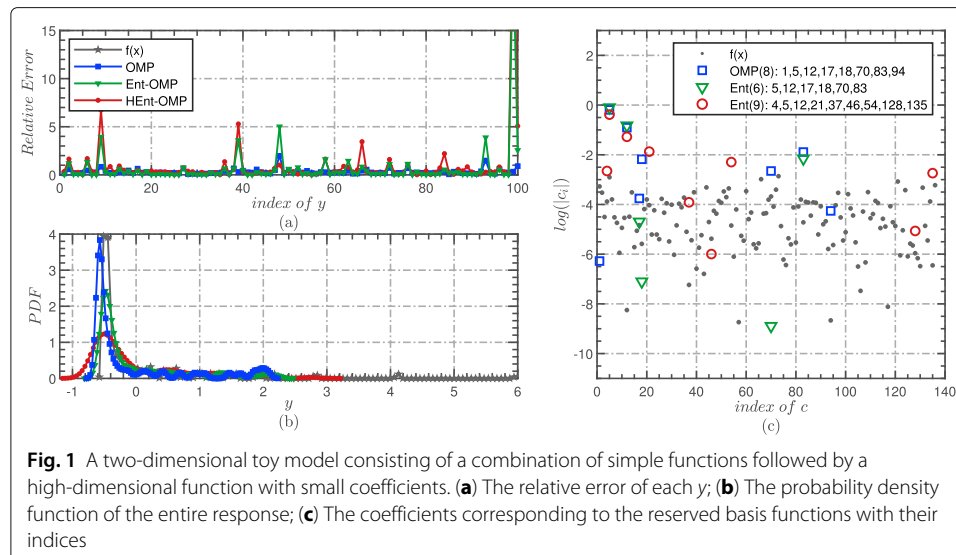
4.2 Case 1: the Ishigami function

The Ishigami model is a non-linear, non-monotonic smooth three-dimensional function:

$$Y(x_1, x_2, x_3) = \sin x_1 + a \sin^2 x_2 + bx_3^4 \sin x_1, \tag{19}$$

where $x_1, x_2,$ and x_3 are three independent input random variables uniformly distributed on $[-\pi, \pi]$, and in typical $a = 7, b = 0.1$.

We use the entropy-based re-optimization method and hybrid entropy-based comprehensive method on the basis of other sparse methods for optimization, and the results are shown in Figs. 2, 3, 4, 5.



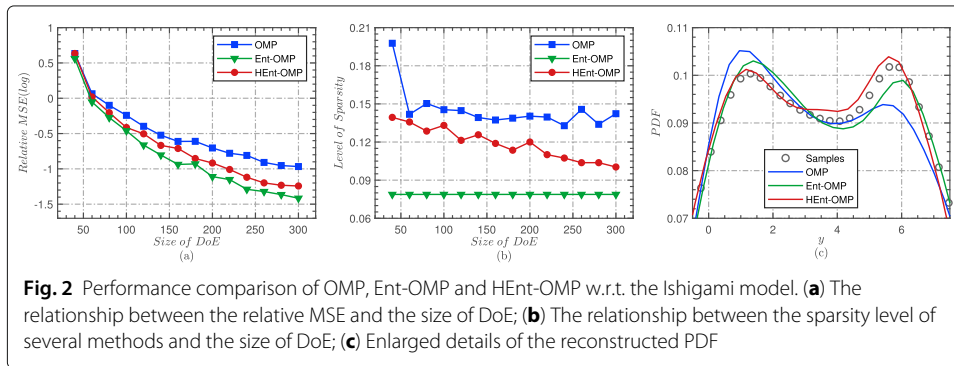


Fig. 2 Performance comparison of OMP, Ent-OMP and HEnt-OMP w.r.t. the Ishigami model. **(a)** The relationship between the relative MSE and the size of DoE; **(b)** The relationship between the sparsity level of several methods and the size of DoE; **(c)** Enlarged details of the reconstructed PDF

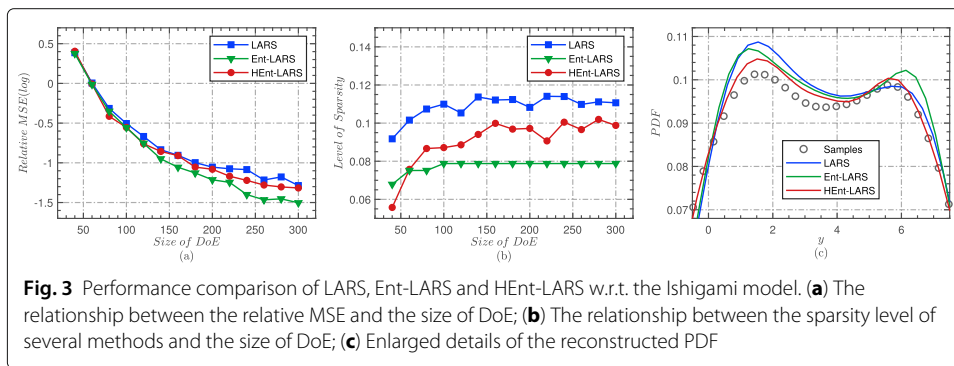


Fig. 3 Performance comparison of LARS, Ent-LARS and HEnt-LARS w.r.t. the Ishigami model. **(a)** The relationship between the relative MSE and the size of DoE; **(b)** The relationship between the sparsity level of several methods and the size of DoE; **(c)** Enlarged details of the reconstructed PDF

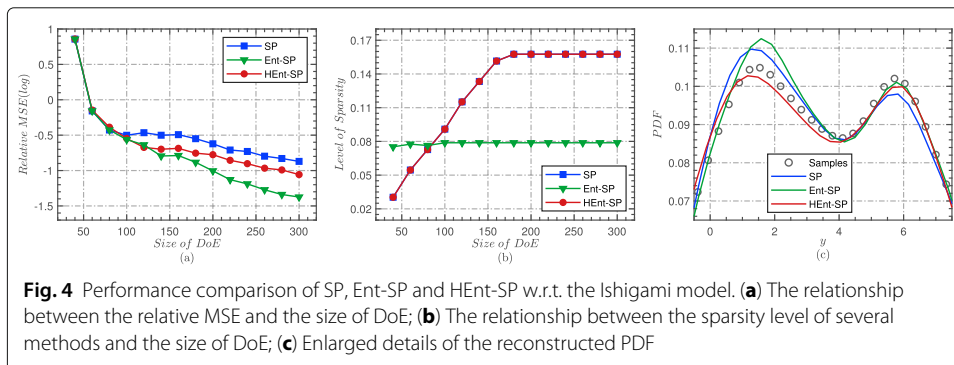


Fig. 4 Performance comparison of SP, Ent-SP and HEnt-SP w.r.t. the Ishigami model. **(a)** The relationship between the relative MSE and the size of DoE; **(b)** The relationship between the sparsity level of several methods and the size of DoE; **(c)** Enlarged details of the reconstructed PDF

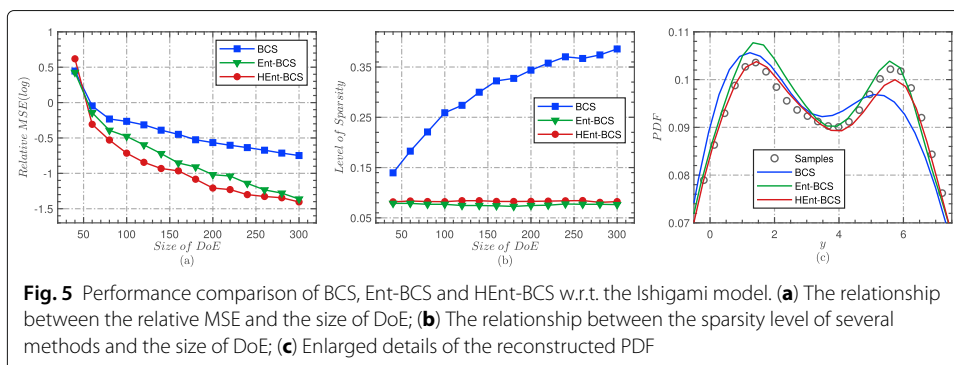
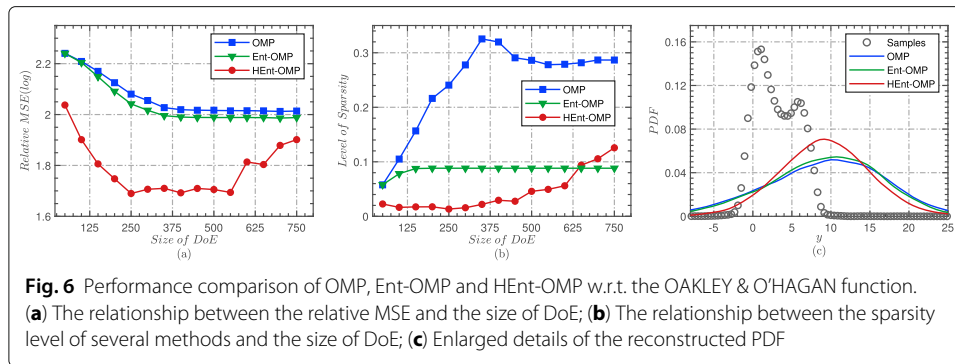


Fig. 5 Performance comparison of BCS, Ent-BCS and HEnt-BCS w.r.t. the Ishigami model. **(a)** The relationship between the relative MSE and the size of DoE; **(b)** The relationship between the sparsity level of several methods and the size of DoE; **(c)** Enlarged details of the reconstructed PDF



In (a) and (b), the x-axis is the size of DoE and the y-axis is relative MSE and level of sparsity (number of items retained/number of items in complete PCE) respectively. It can be noticed that both optimization methods can achieve a good degree of optimization in low dimensions. Specifically, relative MSE and level of sparsity are both reduced compared to the classical method when using the same size of DoE. In other words, entropy-based methods require fewer sample points when a certain level of precision or sparsity is required. Furthermore, (c) show the PDF using 200 DoE. It can be observed that the optimized PDF is closer to the sample point, which is consistent with the previous theory.

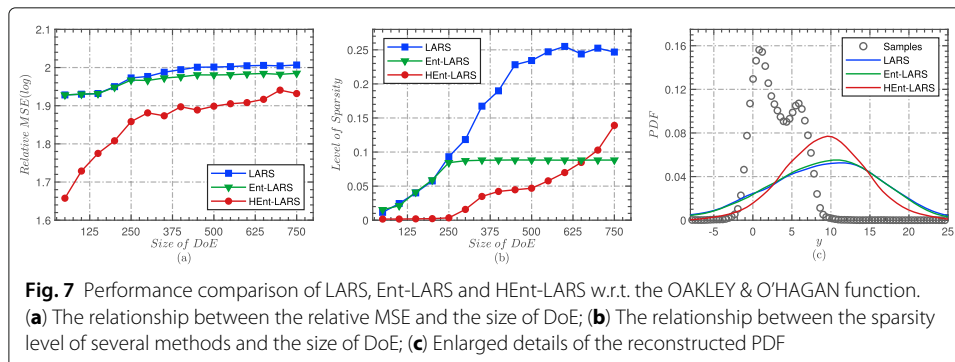
4.3 Case 2: the OAKLEY & O'HAGAN function

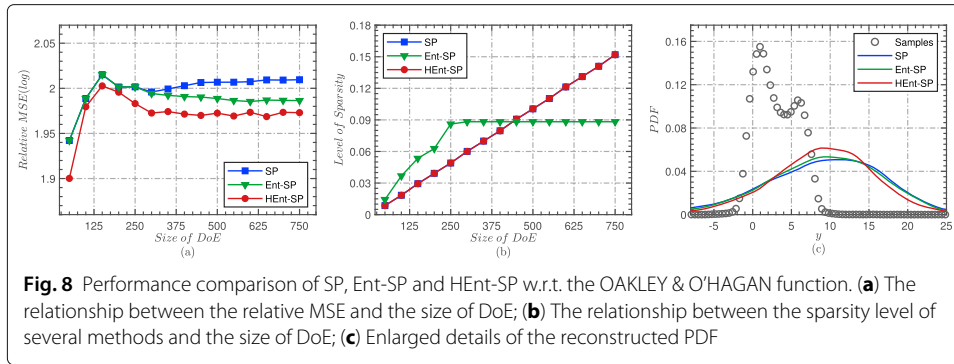
The OAKLEY & O'HAGAN function was proposed in 2010 [50]. It is a high-dimensional function and commonly used in uncertainty quantification and sensitivity analysis, which is expressed as follows:

$$f(\mathbf{x}) = \mathbf{a}_1^T \mathbf{x} + \mathbf{a}_2^T \sin(\mathbf{x}) + \mathbf{a}_3^T \cos(\mathbf{x}) + \mathbf{x}^T \mathbf{M} \mathbf{x}. \tag{20}$$

The independent distributions of the input random variables are $x_i \sim \mathcal{N}(0, 1), i = 1, \dots, 15$. $\mathbf{a}_1, \mathbf{a}_2, \mathbf{a}_3$ are 1×15 vectors, and \mathbf{M} is a square matrix, and the data are taken directly from [50].

From the results of Relative MSE shown in (a) of Figs. 6, 7, 8, 9, in the case of high dimensionality, both optimization methods can improve on the original classical methods. However, Ent-PCE can only be slightly optimized based on the existing results, while HEnt-PCE can achieve better optimization results. This is because Ent-PCE is a re-optimization method based on the classical sparse PCE. In the low-dimensional case, the





low-dimensional features selected by the classical sparse algorithm may be sufficient to represent the overall features of the original model. And when the model dimension is high, the classical sparse method may discard those high-dimensional features that affect the overall uncertainty, and we cannot retrieve them in the next re-optimization process. Therefore, HEnt-PCE is much better than Ent-PCE in the high-dimensional case for optimization. Compared to HEnt-PCE, Ent-PCE can control sparsity better, as can be seen in (b). However, in the higher-dimensional cases, Ent-PCE requires a large-scale OLS calculation, which can lead to an increase in computation time, and therefore HEnt-PCE is preferred for the higher dimensional cases where sparsity is not required.

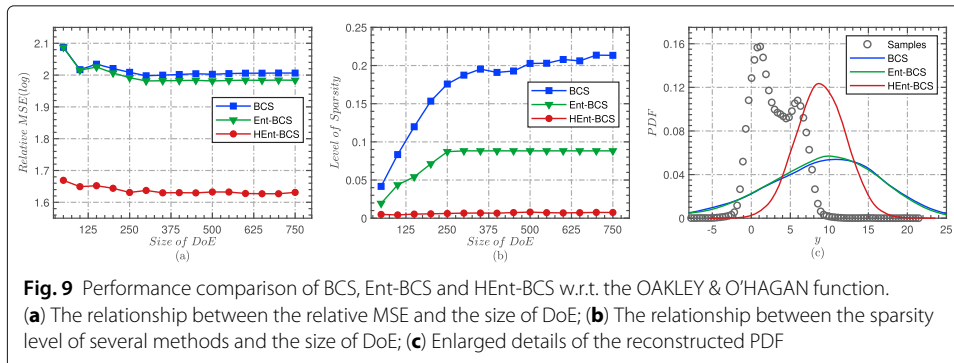
4.4 Case 3: the high-dimensional function

To demonstrate the performance of HEnt-PCE in higher dimensional conditions, consider a high-dimensional function from UQLab, which is an analytical model of the form:

$$f(\mathbf{X}) = 3 - \frac{5}{54} \sum_{i=1}^{54} iX_i + \frac{1}{54} \sum_{i=1}^{54} iX_i^3 + \ln \left(\frac{1}{162} \sum_{i=1}^{54} i(X_i^2 + X_i^4) \right) + X_1X_2^2 + X_2X_4 - X_3X_5 + X_{51} + X_{50}X_{54}^2 \quad (21)$$

where $X_i \sim \mathcal{U}([1, 2])$, $i \neq 20$, and $X_{20} \sim \mathcal{U}([1, 3])$.

As methods we proposed are further optimized based on existing sparse PCE model, the performance would be limited for high dimensional cases. For the Ent-PCE models, since the high order features are dumped by the sparse PCE model, there exists little room for further optimization. As a result, we only compare the results of HEnt-PCE and classical PCE, as can be seen in Figs. 10, 11, 12, 13. It can be seen that the HEnt-PCE still outperforms the classical method in low-order settings.



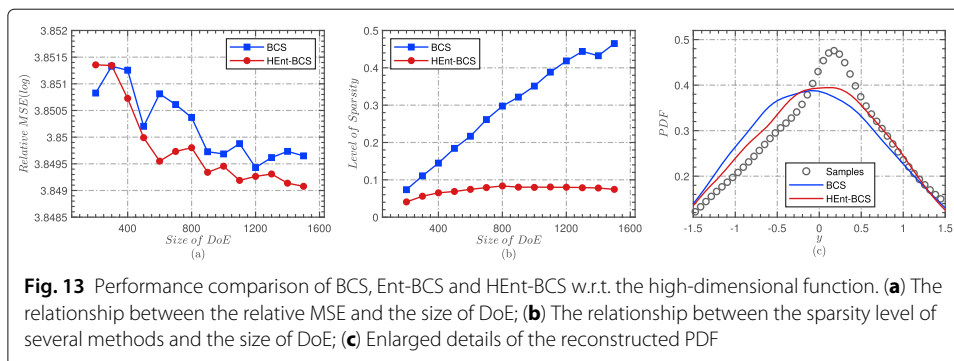
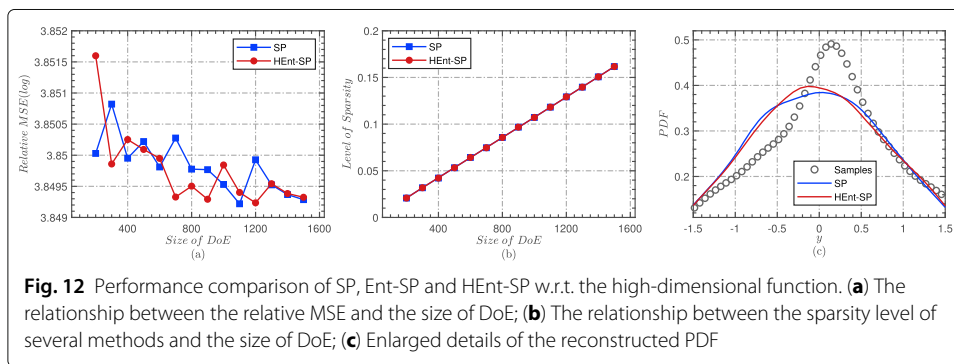
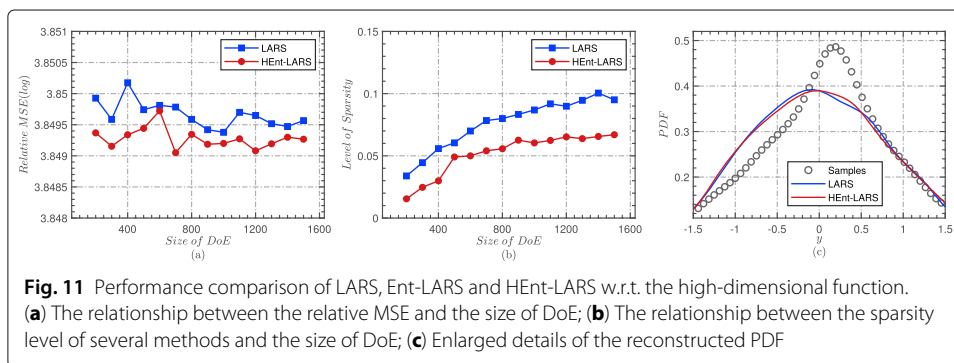
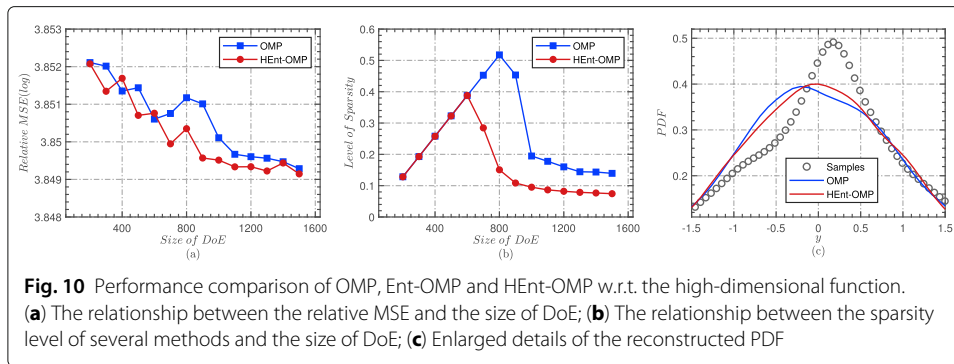


Table 2 Reference value of uncertainty factors in experiments

Uncertainty Factors	Reference Value
Mach number	0.729
Angle of attack	2.31
Free-stream temperature	288.15
Reynolds number	6.5×10^6
Reynolds length	1.0

4.5 Case 4: RAE2822 airfoil

The final application involves the RAE2822 airfoil, a supercritical airfoil, which is a challenging case used by various researchers [24, 51] to test for quantification of uncertainty. The reference values of uncertainty factors in the test are shown in Table 2. We specify that all random variables are subject to a uniform distribution, with the upper and lower boundaries at a reference value of $1 \pm 5\%$.

For deterministic CFD solutions, the Spalart-Allmaras one-equation turbulence model is used for turbulence modeling, and SU2 simulation software is used as a black-box solver. The model of 5th order, using different sparse PCE methods is constructed with 20 samples. The results (pressure coefficient) were obtained with four classical sparse PCE methods and two improved methods proposed above. The mean, standard deviation, the CDF figure of one random point in the RAE2822 airfoil and relative error are shown in Fig. 14 (OMP), Fig. 15 (LARS), Fig. 16 (SP), Fig. 17 (BCS).

The airfoil has upper and lower surfaces, which are marked in (a). For easier presentation, the results are expanded in (b)-(d) with the line $x = 0$ as the dividing line, with the lower surface on the left and the upper surface on the right. For Figs. 14–17, it can be observed that the optimized results are in acceptable agreement with those of the classical sparse models and the mean of the sampling points (shown in (a)). The main difference between classical sparse models and corresponding optimized models is the standard

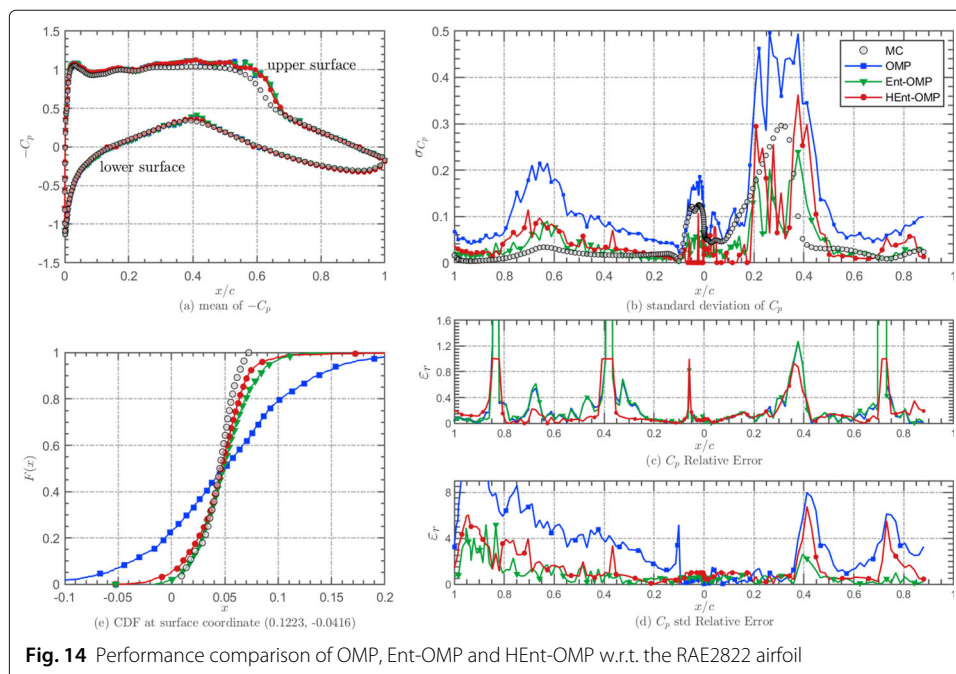
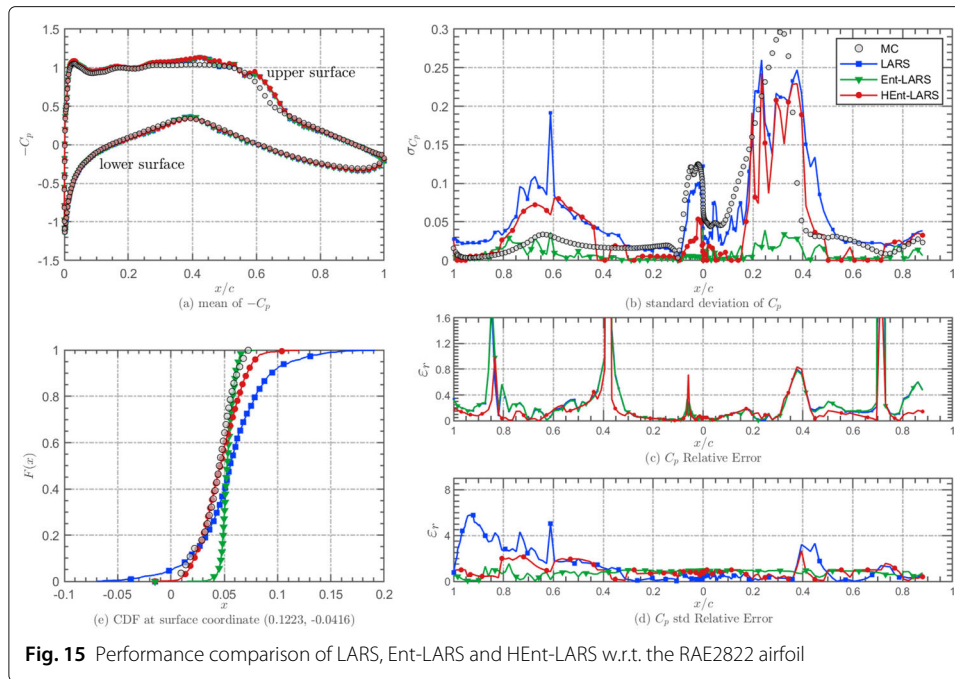
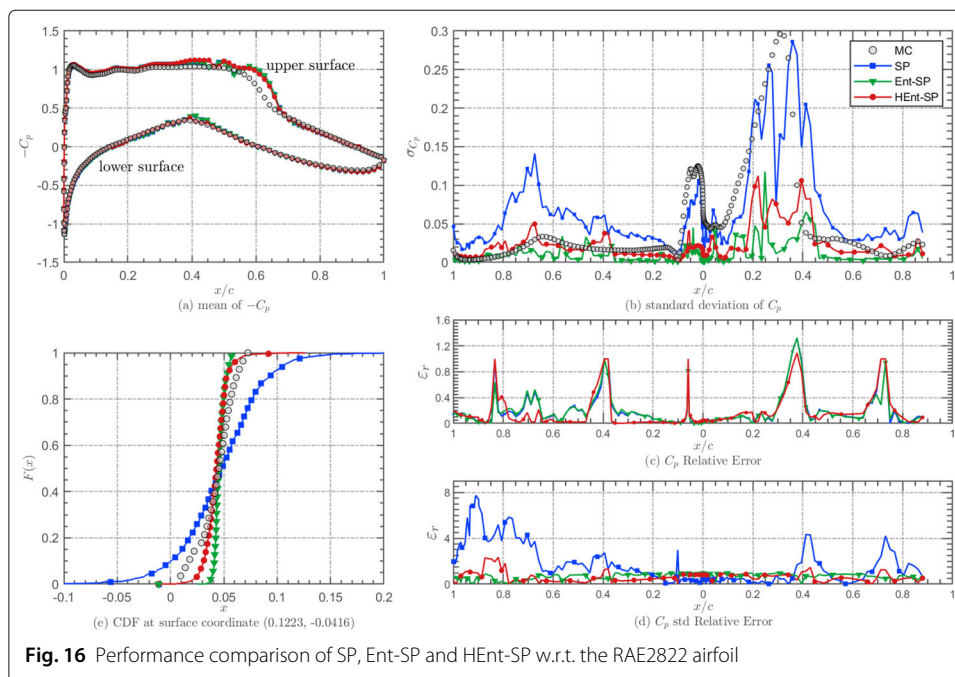


Fig. 14 Performance comparison of OMP, Ent-OMP and HEnt-OMP w.r.t. the RAE2822 airfoil



deviation (shown in (b)). Each position will form a standard deviation, because of the sensitivity of the RAE2822 airfoil model. The standard deviations of the classical sparse models are usually greater than the actual results, while optimization can reduce the standard deviation and achieve better results by adjusting the parameters. However, the least squares used in Ent-PCE will make the standard deviation too small. Moreover, although the dimensionality is not high, one RAE2822 case needs to build hundreds of PCE models (each position of the airfoil needs a model, the number is related to the density of grid



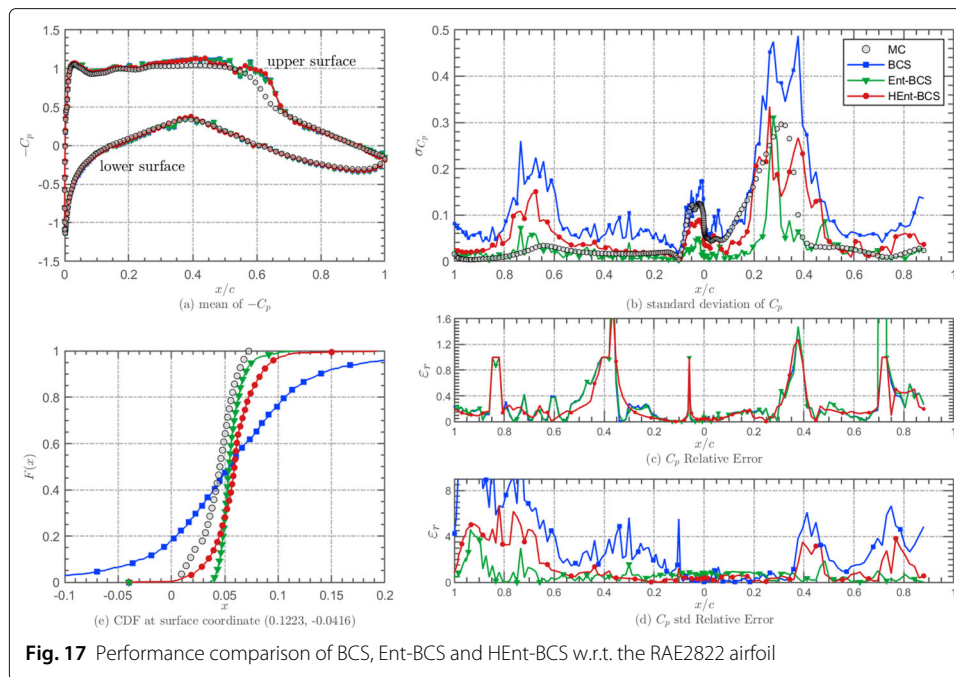


Fig. 17 Performance comparison of BCS, Ent-BCS and HEnt-BCS w.r.t. the RAE2822 airfoil

nodes), which takes longer calculation time and cost for Ent-PCE. Thus, HEnt-PCE is a better choice in such a situation. Although the relative error of the mean value remains hardly changed, the relative error of the standard deviation of the optimized results is reduced (shown in (c) and (d)), so the results will be closer to the original distribution. CDF figures were evaluated for the distribution results at a random point on the airfoil (shown in (e)). It can be seen that optimized models greatly improve the distribution of the model based on the classical models, which is in line with the inference made above.

5 Conclusion

In this paper, an effective framework of sparse PCE is realized and verified. The main contribution of this paper is the development of two adaptive regression methods for optimizing sparse polynomial chaos expansion, Ent-PCE and HEnt-PCE. The former is essentially a followed-up optimization of the results that have been trained by classical sparse methods, and can be easily implemented and extended to various sparse methods; the latter embeds a penalty term into any known sparse algorithms, and allows important higher-order features that may be discarded by classical sparse methods to be retained.

The advantages of the proposed methods are twofold. First, the regularization method in the classical sparse PCE model is followed, which improves accuracy and reduces sparsity while avoiding over-fitting problems. Secondly, considering the uncertainty propagation in the model, the distribution rules of model input are passed, so the output retains uncertainty to the greatest extent.

Furthermore, selection of regularization parameters is changed, because of inconsistencies in the standards for feature selection of different classical sparse PCE models. To achieve the best optimization effect, a set of parameter selection rules was formed. For verification, several applications were considered. Two benchmark analysis functions were initially studied. The results show that the two proposed algorithms can achieve a

certain optimization effect. Among them, Ent-PCE is simpler and more universal in low-dimensional situations. In the case of high dimensions, as HEnt-PCE can retain important higher-order features and does not need secondary calculation, the results are more accurate and the calculation speed is faster. Additionally, the algorithm effect was analyzed and compared through the wind tunnel application. The experimental results show that Ent-PCE has greater limitations, but it can be used independently and can easily optimize small-scale models. HEnt-PCE is more efficient and can achieve higher accuracy, so it is an effective method that is worth being promoted at the industrial level to establish a high-quality sparse PCE model.

Abbreviations

PCE: Polynomial Chaos Expansion; CFD: Computational Fluid Dynamics; OLS: Orthogonal Least Square; NP: Non-deterministic Polynomial; MSE: Mean Squared Error; MEP: Maximum Entropy Principle; OMP: Orthogonal Matching Pursuit; LARS: Least Angle Regression; SP: Subspace Pursuit; BCS: Bayesian Compressive Sensing; DoE: Design of Experiment; LOO: Leave-One-Out; PDF: Probability Density Function; CDF: Cumulative Distribution Function; Ent-PCE: Entropy-based re-optimization sparse PCE; HEnt-PCE: Hybrid entropy-based comprehensive sparse PCE; LHS: Latin Hypercube Sampling

Acknowledgements

The authors are thankful to the reviewers for their valuable comments to improve the quality of the manuscript. In addition, we are grateful to Mingze Qi for his constructive suggestion about paper writing and organization.

Authors' contributions

All authors read and approved the final manuscript.

Funding

This work was funded by the National Numerical Wind tunnel project (NNW2019ZT7-B23) and National Natural Science Foundation of China (1771450). The authors gratefully acknowledge this funding.

Availability of data and materials

The datasets used or analyzed during the current study are available from the corresponding author on reasonable request.

Declarations

Competing interests

The authors declare that they have no competing interests.

Author details

¹College of Liberal Arts and Sciences, National University of Defense Technology, Changsha, China. ²China Aerodynamics Research and Development Center, Mianyang, China.

Received: 17 July 2021 Accepted: 19 October 2021

Published online: 17 January 2022

References

1. Tang T, Zhou T (2015) Recent developments in high order numerical methods for uncertainty quantification. *Sci Sin Math* 45(7):891–928
2. Hu X, Parks GT, Chen X, Seshadri P (2016) Discovering a one-dimensional active subspace to quantify multidisciplinary uncertainty in satellite system design. *Adv Space Res* 57(5):1268–1279
3. Vu KK, d'Ambrosio C, Hamadi Y, Liberti L (2017) Surrogate-based methods for black-box optimization. *Int Trans Oper Res* 24(3):393–424
4. Dutta S, Gandomi AH (2020) Design of experiments for uncertainty quantification based on polynomial chaos expansion metamodels. In: *Handbook of probabilistic models*. pp 369–381
5. Lüthen N, Marelli S, Sudret B (2021) Sparse polynomial chaos expansions: Literature survey and benchmark. *SIAM/ASA J Uncertain Quantif* 9:593–649
6. Gaspar B, Teixeira AP, Soares CG (2014) Assessment of the efficiency of Kriging surrogate models for structural reliability analysis. *Probabilistic Eng Mech* 37:24–34
7. Zhang L, Lu Z, Pan W (2015) Efficient structural reliability analysis method based on advanced Kriging model. *Appl Math Model* 39(2):781–793
8. Sen O, Gaul NJ, Choi KK, Jacobs G, Udaykumar HS (2017) Evaluation of Kriging based surrogate models constructed from mesoscale computations of shock interaction with particles. *J Comput Phys* 336:235–260
9. Song K, Zhang Y, Zhuang X, Yu X, Song B (2020) An adaptive failure boundary approximation method for reliability analysis and its applications. *Eng Comput*:1–16
10. Dai H, Hao Z, Rasmussen K, Wei W (2015) Wavelet density-based adaptive importance sampling method. *Struct Saf* 52:161–169

11. Li S, Yang B, Qi F (2016) Accelerate global sensitivity analysis using artificial neural network algorithm: Case studies for combustion kinetic model. *Combust Flame* 168:53–64
12. Tripathy RK, Ilias B (2018) Deep UQ: Learning deep neural network surrogate models for high dimensional uncertainty quantification. *J Comput Phys* 375:565–588
13. Bhowsekar A, Ierapetritou M (2018) Advances in surrogate based modeling, feasibility analysis, and optimization: A review. *Comput Chem Eng* 108:250–267
14. Xu J, Kong F (2018) A cubature collocation based sparse polynomial chaos expansion for efficient structural reliability analysis. *Struct Saf* 74:24–31
15. Mohammadi A, Raisee M (2019) Efficient uncertainty quantification of stochastic heat transfer problems by combination of proper orthogonal decomposition and sparse polynomial chaos expansion. *Int J Heat Mass Transfer* 128:581–600
16. Tarakanov A, Elsheikh AH (2019) Regression-based sparse polynomial chaos for uncertainty quantification of subsurface flow models. *J Comput Phys* 399:108909
17. Ghanem RG, Spanos PD (1991) *Stochastic finite elements: a spectral approach*. Springer-Verlag, New York
18. Xiu D, Karniadakis GE (2002) The wiener–askey polynomial chaos for stochastic differential equations. *SIAM J Sci Comput* 24(2):619–644
19. Xiu D, Karniadakis GE (2003) Modeling uncertainty in flow simulations via generalized polynomial chaos. *J Comput Phys* 187(1):137–167
20. Schaefer J, Hosder S, West T, Rumsey C, Carlson J-R, Kleb W (2017) Uncertainty quantification of turbulence model closure coefficients for transonic wall-bounded flows. *AIAA J* 55(1):195–213
21. Zhang W, Wang X, Yu J, Yan C (2018) Uncertainty quantification analysis in hypersonic aerothermodynamics due to freestream. *J Beijing Univ Aeronaut Astronaut* 44(5):1102–1109
22. Avdonin A, Jaensch S, Silva CF, Češnovar M, Polifke W (2018) Uncertainty quantification and sensitivity analysis of thermoacoustic stability with non-intrusive polynomial chaos expansion. *Combust Flame* 189:300–310
23. Eldred MS, Burkardt J (2009) Comparison of non-intrusive polynomial chaos and stochastic collocation methods for uncertainty quantification. In: 47th AIAA Aerospace Sciences Meeting Including The New Horizons Forum and Aerospace Exposition. AIAA 2009-976
24. Kumar D, Raisee M, Lacor C (2016) An efficient non-intrusive reduced basis model for high dimensional stochastic problems in CFD. *Comput Fluids* 138:67–82
25. Huan Z, Gao Z, Xu F, Zhang Y, Huang J (2019) An efficient adaptive forward-backward selection method for sparse polynomial chaos expansion. *Comput Methods Appl Mech Eng* 355:456–491
26. Blatman G, Sudret B (2008) Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach. *C R Mécanique* 336(6):518–523
27. Blatman G, Sudret B (2011) Adaptive sparse polynomial chaos expansion based on least angle regression. *J Comput Phys* 230(6):2345–2367
28. Hampton J, Doostan A (2015) Compressive sampling of polynomial chaos expansions: Convergence analysis and sampling strategies. *J Comput Phys* 280:363–386
29. Salehi S, Raisee M, Cervantes MJ, Nourbakhsh A (2017) Efficient uncertainty quantification of stochastic CFD problems using sparse polynomial chaos and compressed sensing. *Comput Fluids* 154:296–321
30. Liu Z, Lesselier D, Sudret B, Wiart J (2020) Surrogate modeling based on resampled polynomial chaos expansions. *Reliab Eng Syst Saf* 202:107008
31. Vermet F (2018) Statistical learning methods. In: *Big data for insurance companies*, Wiley. <https://doi.org/10.1002/9781119489368.ch3>
32. Louizos C, Welling M, Kingma DP (2017) Learning sparse neural networks through ℓ_0 regularization. arXiv preprint arXiv:1712.01312
33. Liu Z, Sun F, McGovern DP (2017) Sparse generalized linear model with L_0 approximation for feature selection and prediction with big omics data. *Biodata Mining* 10:39
34. Jakeman JD, Eldred MS, Sargsyan K (2015) Enhancing ℓ_1 -minimization estimates of polynomial chaos expansions using basis selection. *J Comput Phys* 289:18–34
35. Marelli S, Sudret B (2014) UQLab: a framework for uncertainty quantification in MATLAB. In: *SIAM Conference on Uncertainty Quantification*. pp 2554–2563
36. Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27(3):379–423
37. Jaynes ET (1957) Information theory and statistical mechanics. *Phys Rev* 108(2):171–190
38. Grechuk B, Zabarankin M (2019) Regression analysis: likelihood, error and entropy. *Math Program* 174:145–166
39. Wang X, Tao D, Zhu L (2010) Entropy controlled laplacian regularization for least square regression. *Sign Process* 90(6):2043–2049
40. Bai L, Liang J (2020) Sparse subspace clustering with entropy-norm. In: *International Conference on Machine Learning (ICML 2020)*
41. Obuchi T, Nakanishi-Ohno Y, Okada M, Kabashima Y (2018) Statistical mechanical analysis of sparse linear regression as a variable selection problem. *J Stat Mech Theory Exp* 2018(10):103401
42. Murari A, Peluso E, Cianfrani F, Gaudio P, Lungaroni M (2019) On the use of entropy to improve model selection criteria. *Entropy* 21(4):394
43. Bruckstein AM, Donoho DL, Elad M (2009) From sparse solutions of systems of equations to sparse modeling of signals and images. *SIAM Rev* 51(1):34–81
44. Tribus M, McIrvine EC (1971) Energy and information. *Sci Am* 224:178–184
45. Jones DR (2008) Direct global optimization algorithm. In: Floudas CA, Pardalos PM (eds). *Encyclopedia of optimization*. Springer, Boston
46. Jones DR, Martins J (2021) The DIRECT algorithm—25 years later. *J Glob Optim* 79:521–566
47. Rezaifar YCP, Krishnaprasad PS (2002) Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. In: *Conference on Signals, Systems & Computers*

48. Wei D, Milenkovic O (2009) Subspace pursuit for compressive sensing signal reconstruction. *IEEE Trans Inf Theory* 55(5):2230–2249
49. Sargsyan K, Safta C, Najm HN, Debusschere BJ, Ricciuto D, Thornton P (2014) Dimensionality reduction for complex models via bayesian compressive sensing. *Int J Uncertain Quantif* 4(1):63–93
50. Oakley JE, O'Hagan A (2010) Probabilistic sensitivity analysis of complex models: a bayesian approach. *J R Stat Soc* 66(3):751–769
51. Witteveen J, Doostan A, Pecnik R, Iaccarino D (2009) Uncertainty quantification of the transonic flow around the RAE 2822 airfoil. *Annual Research Briefs, Center for Turbulence Research, Stanford University*. pp 93–104

Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more biomedcentral.com/submissions

