Structural and Multidisciplinary Optimization (2020) 61:1187–1213
https://doi.org/10.1007/s00158-019-02413-5

REVIEW PAPER

Surrogate-assisted global sensitivity analysis: an overview

Kai Cheng[1] · Zhenzhou Lu[1] · Chunyan Ling[1] · Suting Zhou[1]

Received: 1 May 2019 / Revised: 26 August 2019 / Accepted: 23 September 2019 / Published online: 17 January 2020

Abstract

Surrogate models are popular tool to approximate the functional relationship of expensive simulation models in multiple scientific and engineering disciplines. Successful use of surrogate models can provide significant savings of computational cost. However, with a variety of surrogate model approaches available in literature, it is a difficult task to select an appropriate one at hand. In this paper, we present an overview of surrogate model approaches with an emphasis of their application for variance-based global sensitivity analysis, including polynomial regression model, high-dimensional model representation, state-dependent parameter, polynomial chaos expansion, Kriging/Gaussian Process, support vector regression, radial basis function, and low rank tensor approximation. The accuracy and efficiency of these approaches are compared with several benchmark examples. The strengths and weaknesses of these surrogate models are discussed, and the recommendations are provided for different types of applications. For ease of implementations, the packages, as well as toolboxes, of surrogate model techniques and their applications for global sensitivity analysis are collected.

Keywords Surrogate model · Uncertainty quantification · Global sensitivity analysis · Sampling strategy · Model selection · Responsible Editor: Shapour Azarm

1 Introduction

For decades, significant advances of computing power allow one to use a variety of sophisticated numerical models to simulate and predict the behavior of complex systems in nearly all fields of engineering and science (Sudret and Mai 2015). Uncertainties are often encountered in these complex systems (Xiao and Lu 2017). These uncertainties arise from various sources (Ghanem et al. 2016), including uncertainty in model input parameters (initial and boundary conditions, geometrical sizes, loads, etc); model discrepancy or inadequacy due to the difference of simulation model; and the true physical system. Limited computational resources restrict the number of model evaluation and supporting analysis computations, numerical computation errors, and so on. Uncertainty quantification (UQ) provides a mathematically rigorous framework for propagating the uncertainties of the model input to a response quantity of interest (QoI) and analyzing the effects of those uncertainties onto the prediction of QoI. UQ is composed of three fundamental steps (Sudret et al. 2017; Konakli and Sudret 2016a): First, the model representing the physical system under consideration is defined, which maps the input variables $\boldsymbol{x}$ to the response QoI y. Second, determine the probabilistic distribution of input variables using available data or expert judgment. The third step is to propagate the uncertainty of input variables to the model response QoI, which usually involves repeated evaluation of the model response for multiple realizations of input variables. Another important section for UQ is sensitivity analysis, which aims at identifying the input variables whose uncertainty has the largest impact on the variability of a quantity of interest; thus, one can rank the input variables with respect to their significances for the response variability. It can be used to screen a relatively small group of important input variables from a great number candidate input variables. Therefore, one can simplify the model at hand by fixing the unimportant input variables to their nominal values without affecting the prediction accuracy or output uncertainty. Meanwhile, sensitivity analysis helps to determine how a reduction of the uncertainty of each input variable will influence the uncertainty in the response QoI (Saltelli 2008; Wei et al. 2015).

Among various sensitivity analysis methods in literature, Sobol variance-based global sensitivity indices have drawn

Responsible Editor: Shapour Azarm

✉ Zhenzhou Lu
zhenzhoulu@nwpu.edu.cn

[1] School of Aeronautics, Northwestern Polytechnical University, Xi'an 710072, People's Republic of China

the most attention in the last few years, which measure the relative importance of input variables by apportioning the total variance of model output to each input variable with the consideration of the interaction effects among input variables (Saltelli 2008; Wei et al. 2015; Garcia-Cabrejo and Valocchi 2014). To estimate these sensitivity indices, traditional Monte Carlo (MC) or Quasi Monte Carlo (QMC) techniques (Saltelli 2002; Saltelli et al. 2010; Sobol and M. I 2001) are unaffordable for computationally expensive models, which motivated researchers to replace the time-consuming computational models with cheap-to-evaluate surrogate models (also known as meta-models). From an abstract point of view, the original computational model can be considered as a "black box" function $y = g(\mathbf{x})$ that maps the input variables $\mathbf{x}$ to the a QoI y. Surrogate model is a popular technique to approximate the "black box" function with an explicit meta-model (Fig. 1) $y = \tilde{g}(\mathbf{x})$ in many disciplines, including statistics, mathematics, computer science, and various engineering disciplines (Wang and Shan 2007).

In general, surrogate model as an important tool is always embedded into UQ. As shown in Fig. 1, the framework of surrogate-assisted UQ includes three parts. Firstly, one needs to generate N samples $\mathbf{X} = \{\mathbf{x}_1, \ldots, \mathbf{x}_N\}^T$ (design of experiment, DoE) according to the joint probability distribution function $f_{\mathbf{X}}(\mathbf{x})$ of input variable, which should capture as much information as possible of the original computational model. Secondly, obtain the corresponding response QoI $\mathbf{Y} = \{y_1, \ldots, y_N\}^T$ by evaluating the expensive computational model repeatedly at the generated sample points. Thirdly, train a surrogate model based on the obtained input-output sample set $\{\mathbf{X}, \mathbf{Y}\}$ as well as check its accuracy by cross-validation or additional test sample set. To enhance the computational efficiency, the whole process is usually accompanied by adaptive sampling strategy or active learning algorithm to generate more points in regions of interest by learning the information from previous surrogate models and training sample set. After a well-trained surrogate model is obtained, one can predict the value of the response QoI y by the established meta-model, and thus obtain the probability distribution $f_Y(y)$ or other statistical information of response QoI to perform UQ accurately and speedily without the need to evaluate the original expensive computational model.

In the past few years, new developments of surrogate model techniques have been continuously coming forth in the literature. Several prior reviews discussed these approaches and their developments. Jin et al. (Jin et al. 2001) discussed the performance of polynomial regression, multivariate adaptive regression splines, radial basis function (RBF), and Kriging model about the efficiency, robustness, and model simplicity. Queipo et al. (Queipo et al. 2005) discussed the selection of loss function and regularization criterion for constructing surrogate model, design of experiment, surrogate selection and identification, sensitivity analysis, and surrogate-based optimization. Barton et al. (Barton and Meckesheimer 2006) studied various surrogate models in support of simulation optimization with a highlight of the local surrogates and global surrogates. Wang et al. (Wang and Shan 2007) studied the meta-model assisted design optimization, including model approximation, design space

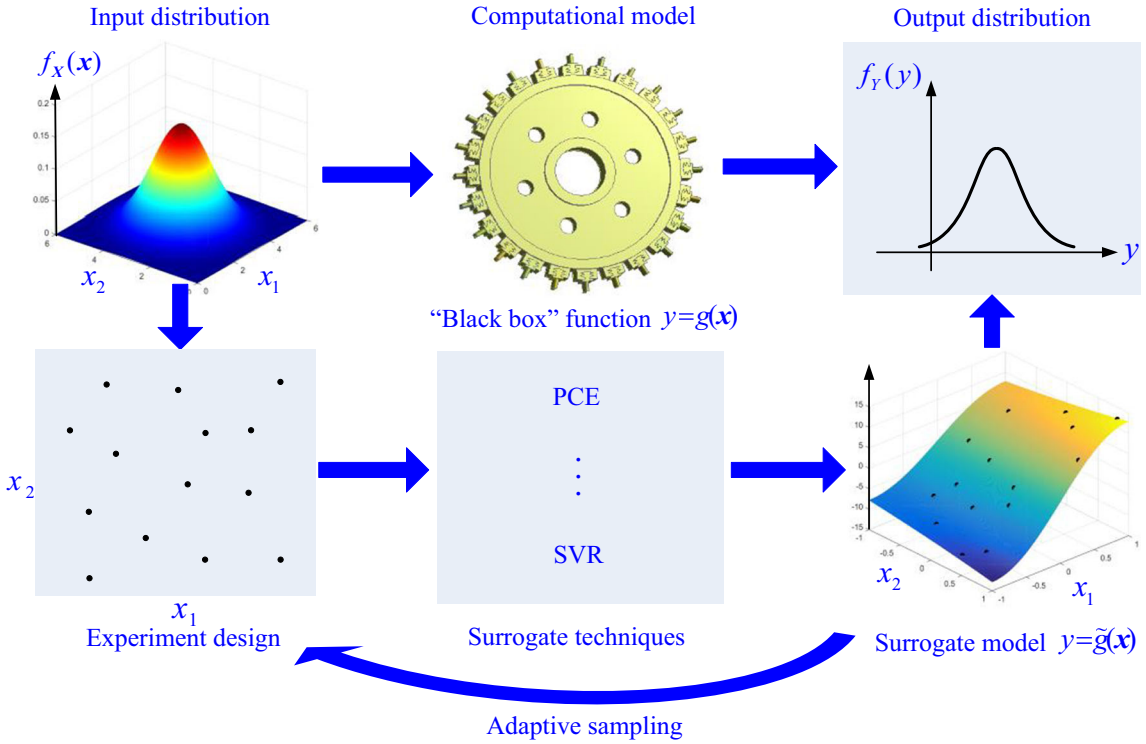

Fig. 1 Surrogate-assisted uncertainty quantification framework

exploration, problem formulation, and solving various types of optimization problems. Motivated by the computational expensive aerospace design problem, Forrester et al. (Forrester and Keane 2009) investigated various meta-model techniques for optimization with focus on sampling strategies, surrogate constructing, and validation method. They also highlighted the methodology for enhancing meta-model accuracy by incorporating gradient information and multi-fidelity analysis. Razavi et al. (Razavi et al. 2012) studied the applications of surrogate model techniques in the field of water resources. Haftka et al. (Haftka et al. 2016) discussed several strategies for surrogate-assisted global optimization from the view of parallelization. Bhosekar (Bhosekar and Ierapetritou 2018) discussed recent advances in the area of surrogate models for problems in modeling, feasibility analysis, and optimization. Recently, Sudret et al. (Sudret et al. 2017) discussed the basics of polynomial chaos expansion (PCE) and low-rank tensor approximation (LRA) with hints on their applications for UQ.

The scope of this overview is confined to the surrogate model techniques and their variants or enhanced versions in the field of UQ and their applications for variance-based global sensitivity analysis (GSA). Figure 2 summarized the available surrogate model methods in literature for UQ, including polynomial regression model, high-dimensional model representation (HDMR), state-dependent parameter (SDP), PCE, LRA, support vector regression (SVR), Kriging/Gaussian process (GP), and radial basis function regression (RBF). We focus on the construction of various surrogate models and the computation of variance-based global sensitivity indices based on these surrogate models. The objective of this article is (a) to incorporate the good practices for surrogate model techniques in the field of UQ with a focus on the computation of variance-based global sensitivity indices, (b) to compare the relative merits of each surrogate model, (c) to explain the link among these surrogate techniques, and (d) to explore the remaining challenges of surrogate models, so as to guide the practitioners to choose the best methods to meet their special requirements and instruct the potential opportunities for further developments of surrogate model techniques for UQ.

The article is organized as follows. Section 2 reviews the Sobol decomposition and variance-based sensitivity indices. Section 3 reviews the construction of various surrogate model techniques and their variants as shown in Fig. 2, and the computation of variance-based global sensitivity indices is provided. Section 4 discusses the sampling strategies in support of surrogate model construction. Section 5 presents the surrogate model accuracy assessment approaches and validation metrics. In Section 6, four benchmark test functions are used to compare the accuracy and efficiency of these surrogate models, and a detailed discussion of their strengths and weaknesses is provided. Finally, in Section 7, we conclude this review and provide the discussion on the current challenges and future perspectives of surrogate models.

2 Sobol decomposition and variance-based global sensitivity indices

Considering a computational model $y = g(\mathbf{x})$ with n-dimensional input variable$\mathbf{x} = (x_1, x_2, \ldots, x_n)$, Sobol sensitivity

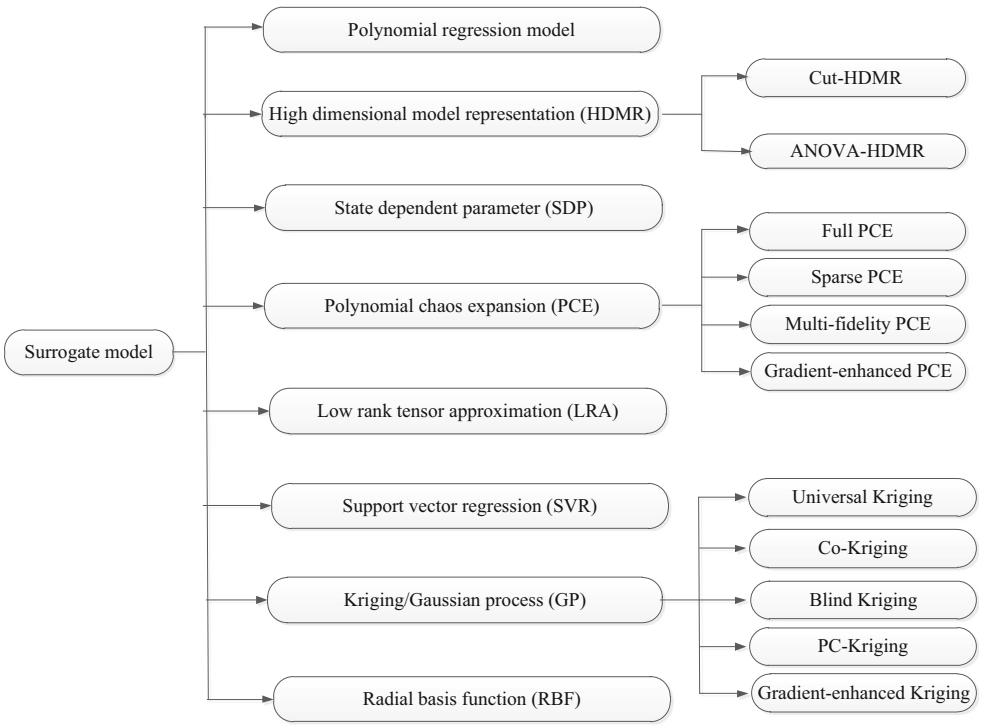

Fig. 2 Various surrogate model techniques reviewed in this article

analysis attributes the total variance of model output y to each input variable with the consideration of the interaction effects among input variables. Assuming that $y = g(\mathbf{x})$ is square-integrable, it can be decomposed uniquely into 2^n orthogonal functional terms of increasing dimensions as (Sobol and M. I 2001; Sobol 1993):

$$g(\mathbf{x}) = g_0 + \sum_{i=1}^{n} g_i(x_i) + \sum_{1\le i\le j\le n} g_{ij}(x_i, x_j) + \ldots + g_{1,2,\ldots,n}(x_1, \ldots, x_n), \tag{1}$$

where

$$\begin{aligned} g_0 &= \int g(\mathbf{x}) \prod_{k=1}^{n} f_{X_k}(x_k) dx_k, \\ g_i(x_i) &= \int g(\mathbf{x}) \prod_{k=1,k\neq i}^{n} f_{X_k}(x_k) dx_k - g_0, \\ &\vdots \\ g_{1,2,\ldots,n}(x_1, \ldots, x_n) &= g(\mathbf{x}) - g_0 - \sum_{i=1}^{n} g_i(x_i) - \sum_{1\le i\le j\le n} g_{ij}(x_i, x_j) - \ldots - \sum_{i=1}^{n} g_{\sim i}(\mathbf{x}_{\sim i}), \end{aligned} \tag{2}$$

in which $f_{X_k}(x_k)$ indicates the probability density function (PDF) of $x_k$and $\mathbf{x}_{\sim i} = (x_1, x_2, \ldots x_{i-1}, x_{i+1}, \ldots, x_n)$. Considering the orthogonality of the component functions in Eq. (1), the total variance of $g(\mathbf{x})$ can be decomposed as:

$$V = \sum_{i=1}^{n} V_i + \sum_{1\le i\le j\le n} V_{ij} + \ldots + V_{1,2,\ldots,n}, \tag{3}$$

where V is the total variance of $g(\mathbf{x})$, V_i is the partial variance contribution of x_i, and $V_{\mathbf{u}}(\mathbf{u} \in [1, \ldots, n])$ is the interactive variance contribution of $\mathbf{x}_{\mathbf{u}}$, which quantifies the interactions among $\mathbf{x}_{\mathbf{u}}$. The Sobol indices are defined as follows

$$S_{\mathbf{u}} = \frac{V_{\mathbf{u}}}{V}. \tag{4}$$

The first-order indices (main indices) S_i measure the individual variance contribution of x_i to the total variance. Furthermore, to value the total effect of a specified input variable, the total sensitivity indices S_{Ti} are defined as the sum of all partial sensitivity indices $S_{\mathbf{u}}$ involving variable i, namely,

$$S_{Ti} = S_i + \sum_{j>i} S_{ij} + \cdots + S_{1,\ldots,n} = 1 - S_{\sim i}, \tag{5}$$

where $S_{\sim i}$ is the sum of all $S_{\mathbf{u}}$ that does not include index i.

3 Surrogate models

In this section, various surrogate model techniques shown in Fig. 2 are reviewed and discussed with a highlight on the computation of variance-based global sensitivity indices.

3.1 Polynomial regression model

Polynomial regression model, also known as response surface method, is one of the oldest and simplest surrogate model in the field of UQ for local and global approximation (Bucher and Bourgund 1990; Guan and Melchers 2001; Youn and Choi 2004). Polynomial regression models have been utilized for GSA in Refs. (Chen and Jin 2004; Hao et al. 2013), where the sensitivity indices are analytically obtained based on the polynomial regression coefficients directly. In general, second-order polynomial regression model is the most widely used response surface method, which is formulated as

$$\widetilde{g}(\mathbf{x}) = \hat{\beta}_0 + \sum_{i=1}^{n} \hat{\beta}_{ii} x_i + \sum_{i=1}^{n} \hat{\beta}_{ii} x_i^2 + \sum_{i=1}^{n} \sum_{j>i}^{n} \hat{\beta}_{ij} x_i x_j, \tag{6}$$

where $\hat{\beta}_0, \hat{\beta}_i, \hat{\beta}_{ii}$ and $\hat{\beta}_{ij}$ are the unknown coefficients. The polynomial regression model in Eq. (6) can be transformed as the tensor product form as (Chen and Jin 2004)

$$\widetilde{g}(\mathbf{x}) = \hat{\beta}_0 + \sum_{\substack{0\le i,j\le n \\ j\neq 0}} \hat{\beta}_{ij} \prod_{l=1}^{n} h_{(i,j)l}, \tag{7}$$

where $\hat{\beta}_{0j} = \hat{\beta}_j$ and

$$h_{(i,j)l} = \begin{cases} 1 & \text{none of } (i,j) = l, \\ x_l & \text{only one of } (i,j) = l, \\ x_l^2 & \text{both of } (i,j) = l. \end{cases} \tag{8}$$

Based on Eq. (76), the variance of a subset of input variables $\mathbf{x}_{\mathbf{U}} = \{x_{i_1}, \ldots, x_{i_s}\} \in \mathbf{x}$ can be computed as

$$\widetilde{V}_{\mathbf{U}} = \sum_{\substack{0\le i_1,j_1\le n \\ j_1\neq 0}} \sum_{\substack{0\le i_2,j_2\le n \\ j_2\neq 0}} \beta_{i_1 j_1} \beta_{i_2 j_2} \prod_{l=1}^{n} C1_{(i_1 j_1)l} C1_{(i_2 j_2)l} \left(\prod_{l\in\mathbf{U}} C2_{(i_1 j_1)(i_2 j_2)l} / C1_{(i_1 j_1)l} C1_{(i_2 j_2)l} - 1 \right), \tag{9}$$

where

$$C1_{(i_1 j_1)l} = \int h_{(i,j)l} f_{X_l}(x_l) dx = \begin{cases} 1 & h_{(i,j)l} = 1 \\ \mu_l & h_{(i,j)l} = x_l \\ \mu_l^2 + \sigma_l^2 & h_{(i,j)l} = x_l^2 \end{cases}, \tag{10}$$

$$C2_{(i_1 j_1)(i_2 j_2)l} = \int h_{(i_1,j_1)l} h_{(i_2,j_2)l} f_{X_l}(x_l) dx = \begin{cases} 1 & h_{(i_1,j_1)l} h_{(i_2,j_2)l} = 1 \\ \mu_l & h_{(i_1,j_1)l} h_{(i_2,j_2)l} = x_l \\ \mu_l^2 + \sigma_l^2 & h_{(i_1,j_1)l} h_{(i_2,j_2)l} = x_l^2 \\ \mu_{l,3} + 3\mu_l\sigma_l^2 + \mu_l^3 & h_{(i_1,j_1)l} h_{(i_2,j_2)l} = x_l^3 \\ \mu_{l,4} + 4\mu_l\mu_{l,3} + 6\mu_l^2\sigma_l^2 + \mu_l^4 & h_{(i_1,j_1)l} h_{(i_2,j_2)l} = x_l^4 \end{cases} \tag{11}$$

in which μ_l and σ_l^2 are the mean and variance of l-th input variable, and $\mu_{l,k}(k = 3, 4)$ is the k-th centered moment of x_l.

With $\mathbf{x}_U = \mathbf{x}$, the variance $\widetilde{V}$ of above second-order polynomial regression model $\widetilde{g}(\mathbf{x})$ can be obtained directly. Therefore, the variance-based global sensitivity indices can be computed directly using analytic formulas based on Eqs. (9)-(11) for any type of distributions.

In theory, the number of the unknown coefficients for second-order polynomial regression model is $(n + 1)(n + 2)/2$. When least square method is used to estimate these coefficients, the training sample size must be greater than the number of unknown coefficients, namely,$N > (n + 1)(n + 2)/2$, which is acceptable for problems with low to medium dimensionality. Moreover, this method is easy to construct and implement. However, the accuracy of this model cannot be guaranteed for highly nonlinear and multimodal problems. In other words, second-order polynomial regression model sacrifices accuracy for efficiency compared to other complex surrogate models.

High-order polynomial regression model has been employed only seldom in the field UQ. The reasons are threefold. Firstly, the proper polynomial order is difficult to determine for a specific problem. Secondly, the number of polynomial coefficients blows up for high-dimensional problems. Thirdly, over-fitting usually occurs when high-order polynomial is used; thus, high-order polynomial regression model is prone to instability.

3.2 High-dimensional model representation

HDMR is developed as a set of quantitative model assessment and analysis tool for capturing high-dimensional input-output system behavior (Sobol 2003; Luo et al. 2014; Li and Wang 2001; Li et al. 2012a; Rabitz and Aliş 1999). It is a particular family of representations where each term in the representation reflects the independent and cooperative contributions of the inputs upon the output response, and its expression is consistent with Eq. (1) in form. Experience shows that the high-order interactions among input variable of HDMR are negligible (Sobol 2003; Luo et al. 2014; Li and Wang 2001; Li et al. 2012a; Rabitz and Aliş 1999); thus, it can be truncated up to two orders, as an accurate approximation of $g(\mathbf{x})$, namely,

$$g(\mathbf{x}) \approx \widetilde{g}(\mathbf{x}) = g_0 + \sum_{i=1}^{n} g_i(x_i) + \sum_{1 \le i \le j \le n} g_{ij}(x_i, x_j), \tag{12}$$

where g_0 indicates a constant term denoting the *0*-th order effect. The first-order component function $g_i(x_i)$ is an univariate function which represents individual contribution to the output $g(\mathbf{x})$. The second-order component function $g_{ij}(x_i, x_j)$ is a bivariate function describing the interactive effect of the input parameters x_i and x_j upon the output response. The total number of summands in Eq. (12) is $1 + n + n(n-1)/2$. Once all the summands are suitably determined, the HDMR can be used as a computationally efficient meta-model for predicting the output. In practice, two types of HDMR meta-models, namely, Cut-HDMR and ANOVA/RS-HDMR, are developed in the literature (Sobol 2003; Luo et al. 2014; Li and Wang 2001; Li et al. 2012a; Rabitz and Aliş 1999; Alış and Rabitz 2001; Li et al. 2006; Li et al. 2008; Ma and Zabaras 2010; Wang et al. 2003) respectively.

3.2.1 Cut-HDMR

The Cut-HDMR is an exact representation of the model $g(\mathbf{x})$ on the lines, planes, and hyper-planes passing through the cut points in the parameter space (Rabitz and Aliş 1999; Rabitz et al. 1999), which can be obtained in an orderly fashion as follows:

$$\begin{aligned} g_0 &= g\left(\overline{\mathbf{x}}\right), \\ g_i(x_i) &= g\left(x_i, \overline{\mathbf{x}}_{-i}\right) - g_0, \\ g_{ij}(x_i, x_j) &= g\left(x_i, x_j, \overline{\mathbf{x}}_{-ij}\right) - g_i(x_i) - g_j(x_j) - g_0, \end{aligned} \tag{13}$$

where $\overline{\mathbf{x}} = \{\overline{x}_1, \ldots, \overline{x}_n\}$ denotes the cut point; $\overline{\mathbf{x}}_{-i}$ and $\overline{\mathbf{x}}_{-ij}$ are $\overline{\mathbf{x}}$ without elements $\overline{x}_i$ and $(\overline{x}_i, \overline{x}_j)$ respectively. $g(\overline{\mathbf{x}})$ is the value of $g(\mathbf{x})$ at $\overline{\mathbf{x}}$; $g(x_i, \overline{\mathbf{x}}_{-i})$ is the model output of all variables evaluated at $\overline{\mathbf{x}}$ except for x_i, etc. Thus, the first-order component functions can be estimated along its axis, and the second-order component functions $g_{ij}(x_i, x_j)$ also can be evaluated in the plane defined by the input variables (x_i, x_j) through the cut point, etc.

The convergence property of Cut-HDMR is rather sensitive to the choice of the cut point, and the authors in Refs (Sobol 2003; Ma and Zabaras 2010) suggested that a good cut point should satisfy:

$$\min \left| g\left(\overline{\mathbf{x}}\right) - E[g(\mathbf{x})] \right|. \tag{14}$$

In practice, the mean of model output $E[g(\mathbf{x})]$ in Eq. (14) is estimated by a moderate number of random samples, then cut point $\overline{x}$ is chosen as the one among the samples whose output is the closest to the estimated mean value. Then, the first and second component functions are evaluated at some discrete points, and these values are used to compute function values at arbitrary points by interpolation (Rabitz et al. 1999; Liu et al. 2016; Chowdhury and Adhikari 2010) technique, namely,

$$\begin{aligned} g_i(x_i) &= \sum_{l=1}^{s} g_i(x_{l,i}) p_l(x_i), \\ g_{ij}(x_i, x_j) &= \sum_{l_1=1}^{s} \sum_{l_2=1}^{s} g_{ij}(x_{l_1,i}, x_{l_2,j}) p_{l_1}(x_i) p_{l_2}(x_j), \end{aligned} \tag{15}$$

where $p_l(x)$ is the interpolation polynomial, s is the number of interpolation nodes $x_{l,\ i}(i = 1, \ldots, n;\ l = 1, \ldots, s)$ in the i-th variable space. The total number of interpolation nodes for constructing a 2-order Cut-HDMR model is

$$\sum_{i=1}^{2} \binom{n}{i} (s-1)^i. \tag{16}$$

When all the component functions are determined, the variance and partial variances can be computed by numerical integration based on the Cut-HDMR meta-model. Therefore, the variance-based global sensitivity indices can be obtained without any additional computational cost (Rabitz et al. 1999; Liu et al. 2016).

3.2.2 ANOVA-HDMR

The ANOVA-HDMR is also known as random-sample HDMR (RS-HDMR), which is equivalent to the Sobol decomposition in Eq. (1) essentially. To estimate the component functions of ANOVA-HDMR, one needs to compute the high-dimensional integrals in Eq. (2), which is usually unaffordable. To circumvent this difficulty, Li et al. (Genyuan Li 2002) proposed efficient approaches to approximate these components functions with analytical orthonormal polynomial basis functions as

$$\begin{aligned} g_i(x_i) &\approx \sum_{l=1}^{m_1^{(i)}} \alpha_l^{(i)} \varphi_l(x_i), \\ g_i(x_i, x_j) &\approx \sum_{l_1=1}^{m_2^{(i)}} \sum_{l_2=1}^{m_2^{(j)}} \beta_{l_1,l_2}^{(ij)} \varphi_{l_1}(x_i) \varphi_{l_2}(x_j), \end{aligned} \tag{17}$$

where $\varphi_l(x)$ is the one-dimensional orthonormal Legendre polynomial basis of order l, $m_1^{(i)}, m_2^{(i)}$ and $m_2^{(j)}$ are the maximum orders of the polynomial bases, and $\alpha_l^{(i)}$ and $\beta_{l_1,l_2}^{(ij)}$ are the unknown coefficients to be determined. Due to the orthonormality of the basis functions, the unknown coefficients in Eq. (17) can be expressed as

$$\begin{aligned} \alpha_l^{(i)} &= \int g(\mathbf{x}) \varphi_l(x_i) d\mathbf{x}, \\ \beta_{l_1,l_2}^{(ij)} &= \int g(\mathbf{x}) \varphi_{l_1}(x_i) \varphi_{l_2}(x_j) d\mathbf{x}. \end{aligned} \tag{18}$$

In practice, g_0 and all these coefficients can be estimated by only a set of random samples as

$$\begin{aligned} g_0 &= \frac{1}{N} \sum_{k=1}^{N} g(\mathbf{x}_k), \\ \alpha_l^{(i)} &= \frac{1}{N} \sum_{k=1}^{N} g(\mathbf{x}_k) \varphi_l(x_{k,i}), \\ \beta_{l_1,l_2}^{(ij)} &= \frac{1}{N} \sum_{k=1}^{N} g(\mathbf{x}_k) \varphi_{l_1}(x_{k,i}) \varphi_{l_2}(x_{k,j}), \end{aligned} \tag{19}$$

where $\mathbf{x}_i (i = 1, \ldots, N)$ are the N realizations of input random variables, and $y_i = g(\mathbf{x}_i) (i = 1, \ldots, N)$ are the corresponding model response. When all the component functions are determined, the variance-based sensitivity indices can be obtained directly from the ANOVA-HDMR meta-model coefficients as

$$S_i = \frac{\sum_{l=1}^{m_1^{(i)}} \left(\alpha_l^{(i)}\right)^2}{\widetilde{V}}, \quad S_{ij} = \frac{\sum_{l_1=1}^{m_2^{(i)}} \sum_{l_2=1}^{m_2^{(j)}} \left(\beta_{l_1,l_2}^{(ij)}\right)^2}{\widetilde{V}}, S_{Ti} = S_i + \sum_{j>i} S_{ij}, \tag{20}$$

where $\widetilde{V}$ is the sum of the square of every coefficient.

To determine the optimal polynomial orders ($m_1^{(i)}, m_2^{(i)}$ and $m_2^{(j)}$) of each component functions, Ziehn et al. (Ziehn and Tomlin 2008) presented an optimization technique based on the least square method. Zuniga et al. (Zuniga et al. 2013) suggested to define the optimal polynomial orders based on the estimated convergence of partial variances. Recently, Lambert and Song et al. (Lambert et al. 2016; Song and Wang 2017) presented an inductive modeling technique, named group method of data handling-neural network (GMDH-NN) method, to construct sparse ANOVA-HDMR. GMDH-NN uses orthonormal polynomials as input neurons, then it generates, validates, and selects many alternative networks of growing complexity until an optimal model has been found. This method is proved to be an efficient method for approximating high-dimensional under-determined functions by selecting important polynomial bases in an adaptive fashion. Readers may refer to (Lambert et al. 2016; Song and Wang 2017) for more details.

3.3 State-dependent parameter

The SDP method was first proposed by Young (Young 1993; Young 2000; Young et al. 2001), and it was applied as surrogate model to estimate the variance-based sensitivity indices by Ratto et al. (Ratto et al. 2007; Ratto et al. 2009). This method is a signal processing and time series analysis tool, in particular for non-stationary and nonlinear signal processing based on the identification and estimation of stochastic models with SDP parameters. The SDP surrogate model takes the truncated HDMR of $g(\mathbf{x})$ as a stochastic non-linear system, and approximates the terms in it using a special recursive fixed interval smoothing (FIS) algorithm (Li et al. 2016a; Li et al. 2011). Here, the SDP method for approximating the first-order HDMR is introduced.

For a dynamical system, the general state-dependent autoregressive with exogenous variables (SDARX) is (Li et al. 2012b)

$$y_t = \mathbf{h}_t^{\mathrm{T}} \mathbf{p}_t + e_t \quad e_t \sim N\left(0, \sigma^2\right), \tag{21}$$

where e_t is a Gaussian white noise, and

$$\begin{aligned} \mathbf{h}_t^{\mathrm{T}} &= \left(-y_{t-1}, -y_{t-2}, \ldots, -y_{t-m_1}, \mathbf{x}_{t-\delta}^{\mathrm{T}}, \mathbf{x}_{t-\delta-1}^{\mathrm{T}}, \ldots, \mathbf{x}_{t-\delta-m_2}^{\mathrm{T}}\right), \\ \mathbf{x}_t^{\mathrm{T}} &= (x_1, \ldots, x_n), \\ \mathbf{p}_t &= (a_1(\mathbf{h}_t), a_2(\mathbf{h}_t), \ldots, a_{m_1}(\mathbf{h}_t), b_0(\mathbf{h}_t), b_1(\mathbf{h}_t), \ldots, b_{m_2}(\mathbf{h}_t))^{\mathrm{T}}, \end{aligned} \tag{22}$$

where t is the time series notation; $\mathbf{x}_t (t = 1, \ldots, N)$ and $y_t (t = 1, \ldots, N)$ are the input and corresponding output at t respectively; δ is a pure time decay, measured in sampling interval; $a_i(\mathbf{h}_t) (i = 1, \ldots, m_1)$ and $b_i(\mathbf{h}_t) (i = 1, \ldots, m_2)$ are the state-dependent parameters, which are the functions of the state vector $\mathbf{h}_t$.

Given the training samples, the first-order HDMR of $y = g(\mathbf{x})$ can be expressed as

$$y_t - g_0 = g_1(x_{1,t}) + g_2(x_{2,t}) + \ldots + g_n(x_{n,t}) + e_t \quad e_t \sim N(0, \sigma^2), \tag{23}$$

where $g_0 = E[y]$, $g_i(x_{i,t}) = E[y|\, x_{i,t}]$, and the higher-order terms of HDMR are assumed to be a Gaussian white noise e_t. It is found that the model in Eq. (23) is deterministic, and no auto-regressive terms of the output response are present, i.e., $m_1 = 0$. Meanwhile, there are no lags or delays in the input variables, i.e., $m_2 = 0$, $\delta = 0$. As a result, we see that $\mathbf{h}_t^{\mathrm{T}} = \mathbf{x}_t^{\mathrm{T}}$ and $\mathbf{p}_t = b_0(\mathbf{h}_t)$. Furthermore, each term in Eq. (18) is a single variable function; thus, each state-dependent parameter $b_{0,i}$ depends only the corresponding input variable x_i, i.e., $p_{i,t} = b_{0,i}(\mathbf{x}_t) = b_{0,i}(x_{i,t})$. Therefore, each term in Eq. (23) can be reformulated as $g_i(x_{i,t}) = b_{0,i}(x_{i,t})x_{i,t} = p_{i,t}x_{i,t}$, and the corresponding SDP model can be rewritten as (Ratto et al. 2007):

$$y_t - g_0 = \mathbf{x}_t^{\mathrm{T}}\mathbf{p}_t + e_t = p_{1,t}x_{1,t} + \ldots + p_{n,t}x_{n,t} \quad e_t \sim N(0, \sigma^2). \tag{24}$$

Therefore, to obtain the first-order HDMR, one needs to estimate the state-dependent parameters $p_{i,t} (i = 1, \ldots, n)$. To this end, it is necessary to characterize the variability of $p_{i,t}$ in some stochastic manner. In SDP model, this is achieved by modeling each state-dependent parameters $p_{i,t}$ by the integrated random walk (IRW) process. Given the IRW characterization of $p_{i,t}$, model (24) can be put into state space form as (Li et al. 2012b)

Observation equation:

$$y_t = \mathbf{x}_t^{\mathrm{T}}\mathbf{p}_t + e_t \tag{25}$$

State equations:

$$\begin{aligned} p_{i,t} &= p_{i,t-1} + d_{i,t-1} \\ d_{i,t} &= d_{i,t-1} + \eta_{i,t} \end{aligned} \tag{26}$$

where $\eta_{i,t} (i = 1, \ldots, n)$ is zero mean white noise with variance $\sigma^2_{\eta_i,t}$ representing the system disturbances. Based on Eqs. (25)-(26), each state-dependent parameter $p_{i,t}$ can be estimated in turn by exploiting backfitting procedure. At each iteration step, a different sorting strategy is used based on the state variable x_i and the current state-dependent parameter $p_{i,t}$ being estimated, and here, $p_{i,t}$ is estimated by the recursive Kalman filter (KF) and fixed interval smoothing (FIS) algorithm. The FIS algorithm yields an estimation $\hat{p}_{i,t|N}$ of $p_{i,t}$ at each sample. Meanwhile, the hyper-parameters (σ^2 and $\sigma^2_{\eta_i,t}$) are optimized by maximum likelihood (ML) estimation using prediction error decomposition.

The similar way can be extended to estimate the interaction terms of HDMR (Ratto et al. 2007). Once the HDMR terms have been obtained, the variance-based sensitivity indices can be estimated by the MCS technique as (Doksum and Samarov 1995)

$$S_i = \frac{\sum_{s=1}^{M}\left(g_i(x_{i,s}) - \overline{g}\right)^2 / M}{\widetilde{V}}, \tag{27}$$

where M is the number of random samples, $\overline{g} = \sum_{s=1}^{M} g_i(x_{i,s}) / M$ and $\widetilde{V}$ is the model variance estimated by the training samples.

The SDP approach has been proven to be efficient for model approximation and sensitivity analysis (Ratto et al. 2007; Ratto et al. 2009; Li et al. 2011; Li et al. 2012b). It usually estimates all the first terms of the HDMR and then the Sobol indices can be obtained with a group of samples, and the computational cost of SDP method for the main indices is almost independent of the input variables dimensionality. When estimating the higher-order terms of HDMR, the computational cost of SDP method is also mild for problems with input variable dimensionality no more than 20 ($n < 20$).

3.4 Polynomial chaos expansion

The classic PCE was first proposed by Wiener (Wiener 1938; Xiu and Karniadakis 2002) in the 30s of last century. The key concept of PCE is to expand the model response onto basis made of multivariate polynomials that are orthogonal with respect to the joint distribution of the input variables. In this setting, characterizing the response probability density function (PDF) is equivalent to evaluate the PC coefficients, i.e., the coordinates of the random response in this basis. The classic PCE of order p for n-dimensional random variable $g(\mathbf{x})$ can be expressed as (Salehi et al. 2018):

$$\widetilde{g}(\mathbf{x}) = \sum_{0 \le |\boldsymbol{\alpha}| \le p} \omega_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\mathbf{x}), |\boldsymbol{\alpha}| = \sum_{i=1}^{n} \alpha_i, \tag{28}$$

where $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_n\} (\alpha_i \ge 0)$ is the multidimensional index notation vector, $\omega_{\boldsymbol{\alpha}}$ is the unknown deterministic coefficients vector, and $\psi_{\boldsymbol{\alpha}}(\mathbf{x})$ is the multivariate polynomial vector. The total number of the expansion terms in the summation of Eq. (28) is:

$$P = \frac{(p+n)!}{p!n!}. \tag{29}$$

Assuming that the input vector $\mathbf{x}$ has independent components x_i with prescribed probability density function (PDF) $f_{X_i}(x_i)$, then the joint PDF of $\mathbf{x}$ can be expressed as:

$$f_{\mathbf{X}}(\mathbf{x}) = \prod_{i=1}^{n} f_{X_i}(x_i). \tag{30}$$

For each x_i, one can construct a family of orthogonal univariate polynomials $\left\{\psi_j^{(i)}, j=0,1,2\ldots\right\}$ with respect to its PDF satisfying:

$$E\left[\psi_j^{(i)}(x_i)\psi_k^{(i)}(x_i)\right] = \int\psi_j^{(i)}(x_i)\psi_k^{(i)}(x_i)f_{X_i}(x_i)dx_i = c_j^{(i)}\delta_{jk}, \tag{31}$$

where δ_{jk} is the Kronecker symbol, and $c_j^{(i)}$ is a constant. The polynomial $\psi_j^{(i)}$ belongs to a specific class according to the PDF of x_i. The types of the orthogonal polynomials used to define the PCE basis functions are shown in Table 1.

Then, the multivariate polynomial can be defined by the tensor product of the univariate polynomials as:

$$\psi_{\boldsymbol{\alpha}}(\mathbf{x}) = \prod_{i=1}^{n}\psi_{\alpha_i}^{(i)}(x_i). \tag{32}$$

PCE was first applied for variance-based global sensitivity analysis in Ref (Sudret 2008), where Sudret proved that ANOVA decomposition of $\widetilde{g}(\mathbf{x})$ can be obtained by reorganization of PCE model as

$$\begin{aligned}\widetilde{g}(\mathbf{x}) = &\sum_{0\le|\boldsymbol{\alpha}|\le p}\omega_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\mathbf{x}) = \omega_0 + \sum_{i=1}^{n}\sum_{\boldsymbol{\alpha}\in\vartheta_i}\omega_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(x_i)\\ &+\sum_{1\le i_1<i_2<n}^{n}\sum_{\boldsymbol{\alpha}\in\vartheta_{i_1,i_2}}\omega_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(x_{i_1},x_{i_2}) + \ldots + \sum_{\boldsymbol{\alpha}\in\vartheta_{1,2,\ldots,n}}\omega_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(x_1,\ldots,x_n),\end{aligned} \tag{33}$$

where $\vartheta_{(\cdot)}$ is defined as follows:

$$\vartheta_{i_1,i_2,\ldots,i_s} = \left\{\boldsymbol{\alpha}: \begin{array}{llll}\alpha_k>0 & \forall k=1,\ldots,n & k\in(i_1,i_2,\ldots,i_s)\\ \alpha_k=0 & \forall k=1,\ldots,n & k\notin(i_1,i_2,\ldots,i_s)\end{array}\right\}. \tag{34}$$

Thus, the variance-based sensitivity indices can be computed by post-processing the PCE coefficients as

$$S_{i_1,\ldots,i_s} = \frac{V_{i_1,\ldots,i_s}}{\widetilde{V}} = \frac{\sum_{\boldsymbol{\alpha}\in\vartheta_{i_1,\ldots,i_s}}\omega_{\boldsymbol{\alpha}}^2E\left[\psi_{\boldsymbol{\alpha}}^2(\mathbf{x})\right]}{\sum_{0\le|\boldsymbol{\alpha}|\le p}\omega_{\boldsymbol{\alpha}}^2E\left[\psi_{\boldsymbol{\alpha}}^2(\mathbf{x})\right]}, \tag{35}$$

where $E\left[\psi_{\boldsymbol{\alpha}}^2(\mathbf{x})\right]$ can be obtained analytically as that in Ref (Sudret 2008).

Table 1 Type of the orthogonal polynomials associated with the random variables with specified PDF

Random variable	Polynomial type	Support
Gaussian	Hermite	$(-\infty,\infty)$
Gamma	Laguerre	$[0,\infty)$
Beta	Jacobi	$[a,b]$
Uniform	Legendre	$[a,b]$
Poisson	Charlier	$\{0,1,2,\cdots\}$
Binomial	Krawtchouk	$\{0,1,2,\cdots,n\}$
Negative binomial	Meixner	$\{0,1,2,\cdots\}$
Hypergeometric	Hahn	$\{0,1,2,\cdots,n\}$

To calculate the PCE coefficients, the traditional projection method and regression method are introduced in Ref (Sudret 2008; Crestaux et al. 2009). Considering the orthogonality of the PCE basis function, the projection approach computes each coefficient of PCE by the multidimensional integral as

$$\omega_{\boldsymbol{\alpha}} = \frac{E[g(\mathbf{x})\psi_{\boldsymbol{\alpha}}(\mathbf{x})]}{E\left[\psi_{\boldsymbol{\alpha}}^2(\mathbf{x})\right]}. \tag{36}$$

This integral problem can be solved by the MCS or Gaussian quadrature scheme (Sudret 2008). The regression method computes the PCE coefficients by the least square method as

$$\boldsymbol{\omega} = \left(\boldsymbol{\Phi}^{\mathrm{T}}\boldsymbol{\Phi}\right)^{-1}\boldsymbol{\Phi}^{\mathrm{T}}\mathbf{Y}, \tag{37}$$

where $\boldsymbol{\Phi}$ is an $N\times P$matrix with entries $\Phi(i,j)=\psi_j(x_i)$. For numerical stability, the size N is usually set to $N=kP$ with $k\in[2,3]$ (Salehi et al. 2018; Blatman and Sudret 2011), which means the computational cost of this method blows up for large n (high-dimensional problem) and p (high-order PCE). Therefore, both of the traditional projection method and regression approaches suffer from the so-called *curse of dimensionality*.

To overcome this issue, Blatman and Sudret (Blatman and Sudret 2011; Blatman and Sudret 2010a; Blatman and Sudret 2010b) proposed several adaptive algorithms to select the most relevant basis functions sequentially from full PCE using only few samples. The selection criterions for retaining the basis functions are based on the determination coefficient R^2 (Blatman and Sudret 2010a; Blatman and Sudret 2010b) and least angle regression technique (Blatman and Sudret 2011). Later, many other attempts also have been made to develop sparse PCE model for UQ (Salehi et al. 2018; Tang et al. 2018; Wan and Karniadakis 2005; Cheng and Lu 2018a; Tang et al. 2016; Shao et al. 2017; Schöbi and Sudret 2017; Lucor and Karniadakis 2005; Ahlfeld et al. 2016; Davis et al. 1997; Pati et al. 1993; Hampton and Doostan 2015; Yang and Karniadakis 2013; Peng et al. 2014; Mathelin and Gallivan 2015; Salehi et al. 2017; Abraham et al. 2017; Cheng and Lu 2018b; Diaz et al. 2018; Mathelin and Gallivan 2012; Jakeman et al. 2015), the common idea in these methods is that the PCE coefficients are sparse (i.e., there are only several dominant coefficients). In general, the dominant PCE coefficients can be recovered by solving the following optimization problem

$$\boldsymbol{\omega} = \underset{\boldsymbol{\omega}}{\operatorname{argmin}}\|\boldsymbol{\omega}\|_1 \text{ subject to} \|\boldsymbol{\Phi}\boldsymbol{\omega}-\mathbf{Y}\|\le\varepsilon, \tag{38}$$

where $\|\boldsymbol{\omega}\|_1$ is the l_1 norm of PCE coefficients, and ε is a tolerance parameter necessitated by the truncation error. To solve the above optimization problem, a large number of efficient algorithms have been proposed, such as some adaptive methods (Tang et al. 2018; Wan and Karniadakis 2005; Cheng

and Lu 2018a; Tang et al. 2016; Shao et al. 2017; Schöbi and Sudret 2017; Lucor and Karniadakis 2005; Ahlfeld et al. 2016; Davis et al. 1997; Pati et al. 1993) and l_1 minimization (Hampton and Doostan 2015; Yang and Karniadakis 2013; Peng et al. 2014; Mathelin and Gallivan 2015; Salehi et al. 2017; Cheng and Lu 2018b; Diaz et al. 2018; Mathelin and Gallivan 2012; Jakeman et al. 2015) methods. The adaptive methods (Tang et al. 2018; Wan and Karniadakis 2005; Cheng and Lu 2018a; Tang et al. 2016; Shao et al. 2017; Schöbi and Sudret 2017; Lucor and Karniadakis 2005; Ahlfeld et al. 2016; Davis et al. 1997; Pati et al. 1993) aim at selecting the significant basis functions from full PCE sequentially using only few samples based on the well-defined selection criterion, such as the correlation criterion in Refs. (Salehi et al. 2018; Davis et al. 1997; Pati et al. 1993) and the variance contribution criterion in Ref. (Cheng and Lu 2018a). The l_1 minimization (Hampton and Doostan 2015; Yang and Karniadakis 2013; Peng et al. 2014; Mathelin and Gallivan 2015; Salehi et al. 2017; Cheng and Lu 2018b; Diaz et al. 2018; Mathelin and Gallivan 2012; Jakeman et al. 2015) techniques aim at minimizing the l_1 norm of PCE coefficients while preserving the fitting accuracy using the compressed sensing algorithm, and it has been shown to be efficient to capture the dominant PCE coefficients.

Recently, other variants of PCE model have also been proposed in literature, such as gradient-enhanced sparse PCE models (Guo et al. 2018; Peng et al. 2016) and multi-fidelity PCE models (Salehi et al. 2018; Bryson and Rumpfkeil 2017; Ng and Eldred 2012; Palar et al. 2018; Palar et al. 2016). The gradient-enhanced methods (Guo et al. 2018; Peng et al. 2016) utilize the derivative information to accelerate the identification of PCE coefficients. Derivative information should be used to enhance the surrogate model only if it is available cheaply. Fortunately, the derivative information for a black-box problem can be efficiently computed using the adjoint approach (Han et al. 2013; Cao et al. 2002; Giles and Pierce 2000), and the computational cost of this method is essentially independent of the input variable dimensionality (Mavriplis 2013). When the derivatives of a function are obtained, the method for incorporating them into surrogate model is essentially the same as that in Eq. (38), and the optimization problem for gradient-enhanced (GE) PCE can be formulated as

$$\boldsymbol{\omega} = \underset{\boldsymbol{\omega}}{\operatorname{argmin}} \|\boldsymbol{\omega}\|_1 \text{ subject to} \left\| \begin{pmatrix} \boldsymbol{\Phi} \\ \boldsymbol{\Phi}_\partial \end{pmatrix} \boldsymbol{\omega} - \begin{pmatrix} \mathbf{Y} \\ \mathbf{Y}_\partial \end{pmatrix} \right\| \leq \varepsilon, \quad (39)$$

with $\mathbf{Y}_\partial = \left(\frac{\partial y_1}{\partial x_1}, \ldots, \frac{\partial y_1}{\partial x_n}, \ldots, \frac{\partial y_N}{\partial x_1}, \ldots, \frac{\partial y_N}{\partial x_n}\right)^{\mathrm{T}}$ and $\boldsymbol{\Phi}_\partial = \left[\frac{\partial \boldsymbol{\Phi}}{\partial x_1}, \ldots, \frac{\partial \boldsymbol{\Phi}}{\partial x_n}\right]$, where $\frac{\partial \boldsymbol{\Phi}}{\partial x_k} (k = 1, \ldots, n)$ is an $N \times P$ matrix with entries $\left[\frac{\partial \boldsymbol{\Phi}}{\partial x_k}\right]_{ij} = \frac{\partial \psi_j(\mathbf{x}_i)}{\partial x_k}$. With this gradient-enhancement, the PCE coefficients can be computed using the similar methods as those of sparse PCE (Guo et al. 2018; Peng et al. 2016).

Another technique to reduce the computational cost of PCE is the multi-fidelity methods (Salehi et al. 2018; Bryson and Rumpfkeil 2017; Ng and Eldred 2012; Palar et al. 2018; Palar et al. 2016). In this method, a large number of low-fidelity samples (cheap data) are coupled with a small number of high-fidelity samples (expensive data) to enhance the accuracy of a surrogate model to approximate an expensive function. The low-fidelity model can be obtained by dimensionality reduction, linearization, and simplification of the physics models, using of coarser discretization or partially converged results (Salehi et al. 2018; Bryson and Rumpfkeil 2017; Palar et al. 2018; Palar et al. 2016; Ng and Willcox 2015; Peherstorfer et al. 2016). In multi-fidelity PCE method, a low-fidelity PCE $\widetilde{g}_l(\mathbf{x})$ is firstly constructed, and this model should capture the global trend of the high-fidelity model response QoI approximately. Then, a small number of expensive samples are utilized to correct the low-fidelity PCE $\widetilde{g}_l(\mathbf{x})$ using a correction PCE (order q, $q<p$) $\widetilde{g}_c(\mathbf{x})$. Finally, a multi-fidelity PCE $\widetilde{g}_h(\mathbf{x})$ can be obtained by combination of the two models as

$$\widetilde{g}_h(\mathbf{x}) = \widetilde{g}_l(\mathbf{x}) + \widetilde{g}_c(\mathbf{x}) = \sum_{0 \leq |\boldsymbol{\alpha}| \leq p} \omega_{\boldsymbol{\alpha},l} \psi_{\boldsymbol{\alpha}}(\mathbf{x}) + \sum_{0 \leq |\boldsymbol{\alpha}| \leq q} \omega_{\boldsymbol{\alpha},c} \psi_{\boldsymbol{\alpha}}(\mathbf{x}). \quad (40)$$

The multi-fidelity PCE provides accurate approximation in cases the low-fidelity model can predict the behavior of the response model sufficiently well. In this regard, Palar et al. (Palar et al. 2018; Palar et al. 2016) suggested that the square of correlation coefficient between low-fidelity model and high-fidelity model should be greater than 0.9 to guarantee the improvement of the prediction accuracy.

The above-mentioned variants and enhanced versions of PCE can be reorganized as that in Eq. (33); thus, the Sobol indices can be obtained in a similar way by post-processing the PCE coefficients in Eq. (35).

3.5 Kriging/Gaussian process

The Kriging model (also known as GP) was first proposed in the field of Geostatistics by Krige (Krige 1953) and Matheron (Matheron 1963). It tends to find the best linear unbiased predictor while minimizes the mean square error of the prediction. The universal Kriging is composed of a polynomial term used for global trend prediction and a Gaussian process term used for local deviation regression, which can be expressed as

$$g_K(\mathbf{x}) = \mathbf{p}^{\mathrm{T}}(\mathbf{x})\boldsymbol{\beta} + Z(\mathbf{x}), \quad (41)$$

where $\mathbf{p}(\mathbf{x}) = [p_1(\mathbf{x}), \ldots, p_M(\mathbf{x})]^{\mathrm{T}}$ is the polynomial basis function, $\boldsymbol{\beta}$ represents the corresponding regression coefficient vector, and $Z(\mathbf{x})$ is a Gaussian process with zero mean and covariance function defined as

$$Cov\left(Z(\mathbf{x}_i), Z\left(\mathbf{x}_j\right)\right) = \sigma^2 R\left(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta}\right), . \quad (42)$$

where σ^2 is the variance of $Z(\mathbf{x})$, and $R(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta})$ is the correlation coefficient between $Z(\mathbf{x}_i)$ and $Z(\mathbf{x}_j)$ with parameters $\boldsymbol{\theta} = [\theta_1, \ldots, \theta_n]^{\mathrm{T}}$. The correlation function controls the smoothness of the Kriging model, and it is usually expressed in a Gaussian form as

$$R(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta}) = \prod_{k=1}^{n} \exp\left[-\theta_k \left(x_i^{(k)} - x_j^{(k)}\right)^2\right]. \tag{43}$$

Given the training sample set $\{\mathbf{X}, \mathbf{Y}\}$, the unknown parameters $\boldsymbol{\gamma} = (\boldsymbol{\beta}, \sigma^2, \boldsymbol{\theta})$ can be estimated by maximizing the likelihood function as

$$\max L(\boldsymbol{\gamma}) = -\frac{N}{2}\ln(2\pi\sigma^2) - \frac{1}{2}\ln(|\mathbf{R}|) - \frac{1}{2\sigma^2}(\mathbf{Y}-\mathbf{F}\boldsymbol{\beta})^{\mathrm{T}}\mathbf{R}^{-1}(\mathbf{Y}-\mathbf{F}\boldsymbol{\beta}), \tag{44}$$

where $\mathbf{R}$ is the $N \times N$ correlation matrix with elements $R_{ij} = R(\mathbf{x}_i, \mathbf{x}_j, \boldsymbol{\theta})$, and $\mathbf{F} = [p(\mathbf{x}_1), \ldots, p(\mathbf{x}_N)]^{\mathrm{T}}$ is the $N \times M$ regression matrix with element $F_{ij} = p_j(\mathbf{x}_i)$. Taking the partial derivative of the log-likelihood equation with respect to $\boldsymbol{\beta}$ and σ^2 to zeros, we obtain

$$\hat{\boldsymbol{\beta}} = \left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{Y}, \tag{45}$$

and

$$\hat{\sigma}^2 = \frac{1}{N}\left(\mathbf{Y}-\mathbf{F}\hat{\boldsymbol{\beta}}\right)^{\mathrm{T}}\mathbf{R}^{-1}\left(\mathbf{Y}-\mathbf{F}\hat{\boldsymbol{\beta}}\right). \tag{46}$$

Then, the parameters $\boldsymbol{\theta}$ can be obtained by solving the following auxiliary optimization problem as

$$\hat{\boldsymbol{\theta}} = \operatorname{argmax}\left(-\frac{N}{2}\ln(\sigma^2) - \frac{1}{2}\ln(|\mathbf{R}|)\right). \tag{47}$$

Kriging is known as an interpolation model; thus, the Kriging predictor $\mu(\mathbf{x})$ can be expressed as a linear combination of the training sample points as

$$\mu(\mathbf{x}) = \mathbf{c}(\mathbf{x})^{\mathrm{T}}\mathbf{Y}, \tag{48}$$

where $\mathbf{c}(\mathbf{x}) = [c_1(\mathbf{x}), \ldots, c_N(\mathbf{x})]^T$ is a vector of weight coefficients associated with the training sample points. After the optimal hyper-parameters $\hat{\boldsymbol{\theta}}$ are obtained, Kriging model can be obtained by minimizing the mean square error (MSE) of prediction under the unbiasedness constraint (interpolation condition), namely,

$$\begin{aligned} &\text{Min} && E\left[(g_K(\mathbf{x}) - \mu(\mathbf{x}))^2\right], \\ &\text{s.t.} && \mathbf{c}(\mathbf{x})^{\mathrm{T}}\mathbf{Y} = \mu(\mathbf{x}). \end{aligned} \tag{49}$$

Solving above optimization problem by Lagrangian theory (Han and Görtz 2012), one can obtain the Kriging predictor as

$$\mu(\mathbf{x}) = \mathbf{p}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}} + \mathbf{r}^{\mathrm{T}}(\mathbf{x})\mathbf{R}^{-1}\left(\mathbf{Y}-\mathbf{F}\hat{\boldsymbol{\beta}}\right). \tag{50}$$

Meanwhile, the Kriging variance can be also obtained as

$$\sigma^2(\mathbf{x}) = \hat{\sigma}^2\left[1 - \mathbf{r}^{\mathrm{T}}(\mathbf{x})\mathbf{R}^{-1}\mathbf{r}(\mathbf{x}) + \left(\mathbf{r}(\mathbf{x})\mathbf{R}^{-1}\mathbf{F} - \mathbf{p}(\mathbf{x})\right)^{\mathrm{T}}\left(\mathbf{F}^{\mathrm{T}}\mathbf{R}^{-1}\mathbf{F}\right)^{-1}\left(\mathbf{r}(\mathbf{x})\mathbf{R}^{-1}\mathbf{F} - \mathbf{p}(\mathbf{x})\right)\right], \tag{51}$$

where $\mathbf{r}(\mathbf{x}) = [R(\mathbf{x}, \mathbf{x}_1), \ldots, R(\mathbf{x}, \mathbf{x}_N)]^{\mathrm{T}}$ represents the correlation vector between $\mathbf{x}$ and the N observed points. The Kriging mean in Eq. (50) is final the Kriging predictor, and the Kriging variance in Eq. (51) can be used to measure the local prediction error.

Kriging model has been applied for global sensitivity analysis in Refs (Chen and Jin 2004; Zhang et al. 2017a; Marrel et al. 2008; Oakley and O'Hagan 2004), where the regression term of Kriging model is assumed to be an unknown constant $\hat{\beta}$ (ordinary Kriging). Chen et al. (Chen and Jin 2004) proved that the ordinary Kriging predictor can be transformed into a sum of multivariate tensor product basis functions. Therefore, the multivariate integrals problem in GSA can be transformed into univariate integrals, and the Sobol indices can be computed analytically based on the ordinary Kriging predictor.

As a more general form, the universal Kriging in Eq. (41) consists of a regression component $\mathbf{p}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}}$ and a correlation component $Z(\mathbf{x})$. The regression term aims at capturing the global trends of the samples data, and it will generally fine-tune the model to give better prediction than ordinary Kriging (Forrester and Keane 2009). In universal Kriging, the regression term always takes form of low-order polynomial regression model (here, we take second-order polynomial regression model as example). In this regard, universal Kriging predictor can be deemed as a combination of a response surface model introduced in Section 3.1 and a GP model. As a consequence, the universal Kriging predictor can be expressed as a combination of two kinds of multivariate tensor product basis functions as

$$\begin{aligned} \widetilde{g}(\mathbf{x}) = \mathbf{p}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}} + \mathbf{r}^{\mathrm{T}}(\mathbf{x})\mathbf{R}^{-1}\left(\mathbf{Y}-\mathbf{F}\hat{\boldsymbol{\beta}}\right) &= \sum_{i=1}^{M}\hat{\beta}_i\prod_{l=1}^{n}p_i(x_l) \\ &+ \sum_{i=1}^{N}\hat{a}_i\prod_{l=1}^{n}\exp\left(-\theta_l\left(x_{i,l}-x_l\right)^2\right), \end{aligned} \tag{52}$$

where $\hat{\mathbf{a}} = [\hat{a}_1, \ldots, \hat{a}_N] = \mathbf{R}^{-1}\left(\mathbf{Y}-\mathbf{F}\hat{\boldsymbol{\beta}}\right)$ and $p_i(\mathbf{x}) = \prod_{l=1}^{n}p_i(x_l)(i = 1, \ldots, M)$ with the univariate function $p_i(x_l) = 1$ or x_l or x_l^2 (Chen and Jin 2004) as that in Eq. (8).

Based on Eq. (52) the variance of a subset of input variables $\mathbf{x}_{\mathbf{U}} = \{x_{i_1}, \ldots, x_{i_s}\} \in \mathbf{x}$ can be computed as

$$\begin{aligned} \widetilde{V}_{\mathbf{U}} = &\sum_{i=1}^{N}\sum_{j=1}^{N}\hat{a}_i\hat{a}_j\prod_{l=1}^{n}C1_{i,l}C1_{j,l}\left(\prod_{l\in\mathbf{U}}C2_{i,j,l}/C1_{i,l}C1_{j,l}-1\right) \\ &+2\sum_{i=1}^{N}\sum_{j=1}^{M}\hat{a}_i\hat{\beta}_j\prod_{l=1}^{n}C1_{i,l}D1_{j,l}\left(\prod_{l\in\mathbf{U}}B_{i,j,l}/C1_{i,l}D1_{j,l}-1\right) \\ &+\sum_{i=1}^{M}\sum_{j=1}^{M}\hat{\beta}_i\hat{\beta}_j\prod_{l=1}^{n}D1_{i,l}D1_{j,l}\left(\prod_{l\in\mathbf{U}}D2_{i,j,l}/D1_{i,l}D1_{j,l}-1\right), \end{aligned} \tag{53}$$

where

$$\begin{aligned} C1_{i,l} &= \int \exp\left(-\theta_l \left(x_{i,l}-x_l\right)^2\right) f_{X_l}(x_l)dx_l, \\ D1_{i,l} &= \int p_i(x_l) f_{X_l}(x_l)dx_l, \\ C2_{i,j,l} &= \int \exp\left(-\theta_l \left[\left(x_{i,l}-x_l\right)^2 + \left(x_{j,l}-x_l\right)^2\right]\right) f_{X_l}(x_l)dx_l, \\ D2_{i,j,l} &= \int p_i(x_l)p_j(x_l) f_{X_l}(x_l)dx_l, \\ B_{i,j,l} &= \int \exp\left(-\theta_l \left(x_{i,l}-x_l\right)^2\right) p_j(x_l) f_{X_l}(x_l)dx_l. \end{aligned} \tag{54}$$

The univariate integrals problem in Eq. (54) can be easily solved analytically without the need of any estimators (Van Steenkiste et al. 2018). For standard normal distributed variable, namely, $x_l \sim N(0, 1)$, we have

$$C1_{i,l} = \int \exp\left(-\theta_l \left(x_{i,l}-x_l\right)^2\right) f_{X_l}(x_l)dx_l = \frac{1}{\sqrt{2\theta_l+1}} \exp\left\{-\frac{\theta_l}{2\theta_l+1} x_{i,l}^2\right\} \tag{55}$$

$$\begin{aligned} C2_{i,j,l} &= \int \exp\left(-\theta_l \left[\left(x_{i,l}-x_l\right)^2 + \left(x_{j,l}-x_l\right)^2\right]\right) f_{X_l}(x_l)dx_l \\ &= \frac{1}{\sqrt{4\theta_l+1}} \exp\left\{\frac{2\theta_l^2}{4\theta_l+1}\left(x_{i,l}+x_{i,l}\right)^2 - \theta_l\left(x_{i,l}^2 + x_{j,l}^2\right)\right\} \end{aligned} \tag{56}$$

$$D1_{i,l} = \int p_i(x_l) f_{X_l}(x_l)dx_l = \begin{cases} 1 & p_i(x_l) = 1 \\ 0 & p_i(x_l) = x_l \\ 1 & p_i(x_l) = x_l^2 \end{cases} \tag{57}$$

$$D2_{i,j,l} = \int p_i(x_l)p_j(x_l) f_{X_l}(x_l)dx_l = \begin{cases} 1 & p_i(x_l)p_j(x_l) = 1 \\ 0 & p_i(x_l)p_j(x_l) = x_l \\ 1 & p_i(x_l)p_j(x_l) = x_l^2 \\ 0 & p_i(x_l)p_j(x_l) = x_l^3 \\ 3 & p_i(x_l)p_j(x_l) = x_l^4 \end{cases} \tag{58}$$

$$B_{i,j,l} = \int \exp\left(-\theta_l \left(x_{i,l}-x_l\right)^2\right) p_j(x_l) f_{X_l}(x_l)dx_l = \begin{cases} \frac{1}{\sqrt{2\theta_l+1}} \exp\left\{-\frac{\theta_l}{2\theta_l+1} x_{i,l}^2\right\} & p_j(x_l) = 1 \\ \frac{2x_l\theta_l}{\sqrt[3]{2\theta_l+1}} \exp\left\{-\frac{\theta_l}{2\theta_l+1} x_{i,l}^2\right\} & p_j(x_l) = x_l \\ \frac{4\theta_l^2 x_{i,l}^2 + 2\theta_l + 1}{\sqrt[5]{2\theta_l+1}} \exp\left\{-\frac{\theta_l}{2\theta_l+1} x_{i,l}^2\right\} & p_j(x_l) = x_l^2 \end{cases} \tag{59}$$

For [0,1] interval uniformly distributed variable, namely, $x_l \sim U(0, 1)$, we have

$$\begin{aligned} C1_{i,l} &= \int \exp\left(-\theta_l \left(x_{i,l}-x_l\right)^2\right) f_{X_l}(x_l)dx_l \\ &= \sqrt{\frac{\pi}{\theta_l}}\left\{\Phi\left(\sqrt{2\theta_l}\left(1-x_{i,l}\right)\right) + \Phi\left(\sqrt{2\theta_l}x_{i,l}\right) - 1\right\} \end{aligned} \tag{60}$$

$$C2_{i,j,l} = \int \exp\left(-\theta_l \left[\left(x_{i,l}-x_l\right)^2 + \left(x_{j,l}-x_l\right)^2\right]\right) f_{X_l}(x_l)dx_l = \sqrt{\frac{\pi}{2\theta_l}} \exp\left(-\frac{\theta_l}{2}\left(x_{i,l}-x_{j,l}\right)^2\right)\left\{\Phi\left(\sqrt{\theta_l}\left(2-x_{i,l}-x_{j,l}\right)\right) + \Phi\left(\sqrt{\theta_l}\left(x_{i,l}+x_{j,l}\right)\right) - 1\right\} \tag{61}$$

$$D1_{i,l} = \int p_i(x_l) f_{X_l}(x_l)dx_l = \begin{cases} 1 & p_i(x_l) = 1 \\ 1/2 & p_i(x_l) = x_l \\ 1/3 & p_i(x_l) = x_l^2 \end{cases} \tag{62}$$

$$D2_{i,j,l} = \int p_i(x_l)p_j(x_l) f_{X_l}(x_l)dx_l = \begin{cases} 1 & p_i(x_l)p_j(x_l) = 1 \\ 1/2 & p_i(x_l)p_j(x_l) = x_l \\ 1/3 & p_i(x_l)p_j(x_l) = x_l^2 \\ 1/4 & p_i(x_l)p_j(x_l) = x_l^3 \\ 1/5 & p_i(x_l)p_j(x_l) = x_l^4 \end{cases} \tag{63}$$

$$B_{i,j,l} = \int \exp\left(-\theta_l \left(x_{i,l}-x_l\right)^2\right) p_j(x_l) f_{X_l}(x_l)dx_l = \begin{cases} C1_{i,l} & p_j(x_l) = 1 \\ \frac{1}{2\theta_l}\left\{\exp\left(-\theta_l x_{i,l}^2\right) - \exp\left(-\theta_l\left(1-x_{i,l}\right)^2\right)\right\} + x_{i,l}C1_{i,l} & p_j(x_l) = x_l \\ \frac{1}{2\theta_l}\left\{x_{i,l}\exp\left(-\theta_l x_{i,l}^2\right) - \left(1+x_{i,l}\right)\exp\left(-\theta_l\left(1-x_{i,l}\right)^2\right)\right\} + \left(\frac{1}{2\theta_l} + x_{i,l}^2\right)C1_{i,l} & p_j(x_l) = x_l^2 \end{cases} \tag{64}$$

With $\mathbf{x_U} = \mathbf{x}$, the variance $\widetilde{V}$ of Kriging predictor $\widetilde{g}(\mathbf{x})$ can be also obtained analytically. Therefore, the variance-based global sensitivity indices can be computed directly using analytic formulas based on universal Kriging predictor.

Similar with PCE, many other enhanced versions of Kriging models have been proposed in literature to improve the performance of the universal Kriging model, including stochastic Kriging (also known as regression Kriging) (Zhang et al. 2017b; Xie et al. 2010; Ankenman et al. 2010; Chen and Kim 2014; Chen et al. 2013; Staum 2009), blind Kriging (Roshan et al. 2008; Couckuyt et al. 2012), PC-Kriging (Kersaudy et al. 2015), Co-Kriging (Kennedy and O'Hagan 2000; Xiao et al. 2018; Gratiet 2015; Gratiet 2012; Parussini et al. 2017; Gratiet 2013; Fernándezgodino et al. 2017), hierarchical Kriging (Han and Görtz 2012; Palar and Shimoyama 2017), and GE Kriging (Han et al. 2013; Liu 2003; AIAA 2009; Bouhlel and Martins 2018; Ulaganathan et al. 2016; Ulaganathan et al. 2015).

Although the interpolation property of Kriging model is somehow an advantage, but it will yield undesired results when dealing with noisy data. In this regard, stochastic Kriging was developed to regress the data instead of interpolation. This is achieved by modeling the noise with a separate Gaussian process with mean 0 and diagonal covariance matrix, which is equivalent to add a regularized constant to the diagonal of the correlation matrix. Therefore, the stochastic Kriging predictor can be expressed as

$$\widetilde{g}(\mathbf{x}) = \mathbf{p}^{\mathrm{T}}(\mathbf{x})\hat{\boldsymbol{\beta}} + \mathbf{r}^{\mathrm{T}}(\mathbf{x})\left(\mathbf{R} + \eta^2\mathbf{I}\right)^{-1}\left(\mathbf{Y} - \mathbf{F}\hat{\boldsymbol{\beta}}\right), \tag{65}$$

where η^2 is the variance of the noise, and it can be estimated by maximum likelihood function.

Considering the polynomial term (first part in Eq. (50) in Kriging sometimes has a great influence on the performance of Kriging predictor, the blind Kriging and PC-Kriging are developed to identify the most relevant basis functions of model response adaptively. Blind Kriging (Roshan et al. 2008; Couckuyt et al. 2012) identifies the basis functions that captures the most variance of the training samples set by Bayesian variable selection technique, which consists of a greedy forward selection procedure to add the most relevant candidate bases iteratively. PC-Kriging (Kersaudy et al. 2015), which combines the advantage of sparse PCE and Kriging, is developed by selecting the most influential orthogonal polynomials to construct the regression term in Kriging model by using least angle regression algorithm. In these two cases, the polynomial term plays a dominant role, and the trend of samples data is almost fully represented by it (the stochastic process only has little influence).

Co-Kriging (also known as multi-fidelity Kriging) (Kennedy and O'Hagan 2000; Xiao et al. 2018; Gratiet 2015; Gratiet 2012; Parussini et al. 2017; Gratiet 2013; Fernándezgodino et al. 2017) aims at exploiting the cross-correlation between low-fidelity and high-fidelity samples, and it will enhance the accuracy of high-fidelity surrogate using supplementary samples of low-fidelity simulation. To make use of the low-fidelity data, one can use a correction Gaussian process to model the difference of the cheap and expensive data, which is formulated as

$$g_h(\mathbf{x}) = \rho g_l(\mathbf{x}) + \delta(\mathbf{x}) \tag{66}$$

where $g_l(\mathbf{x})$ and $g_h(\mathbf{x})$ represent the low-fidelity and high-fidelity Kriging model respectively, $\delta(\mathbf{x})$ is a correction Gaussian process and ρ is a scaling factor. Furthermore, Han and Gortz (Han and Görtz 2012) developed a simplified version of co-Kriging, named hierarchical Kriging. This method neglects the cross-correlation information between low-fidelity model and high-fidelity model, and simply uses the low-fidelity model as a trend to assist the high-fidelity prediction, which is formulated as

$$g_h(\mathbf{x}) = \rho\widetilde{g}_l(\mathbf{x}) + \delta(\mathbf{x}) \tag{67}$$

where $\widetilde{g}_l(\mathbf{x})$ is the low-fidelity Kriging predictor. Admittedly, the hierarchical Kriging method offers several advantages. Firstly, one only needs to build Kriging model of each fidelity level independently, and it does not require to model the cross-correlation between LF and HF samples; thus, it is simpler and computationally cheaper than the classic Co-Kriging model. Secondly, it is more flexible since different correlation kernels can be utilized in Kriging model of different fidelity levels (Abdallah et al. 2018).

In the GE Kriging method, the gradient information is incorporated into the construction of Kriging model to improve the accuracy and efficiency of the prediction as that in gradient-enhanced PCE model. One can refer to (Han et al. 2013; Liu 2003; AIAA 2009; Bouhlel and Martins 2018; Ulaganathan et al. 2016; Ulaganathan et al. 2015) for more details.

When the Gaussian correlation function is used, all of these different versions of Kriging can be transformed into the form of tensor product function as that in Eq. (52); thus, the above derivation of the variance-based sensitivity indices can be easily extended to these method directly to improve the accuracy and efficiency of GSA further.

3.6 Support vector regression

Support-vector regression (SVR) was developed based on statistical learning theory by Vapnik (Vapnik and Vladimir 1997; Vapnik 2008). It projects the original data into a high-dimensional feature space using the kernel function, then searches the best prediction function in the linear feature space. SVR is known for their good generalization performances and their ability to handle nonlinear problems using the kernels technique. The final optimization problem to solve

in SVR is convex, and thus, a unique and global optimal solution can be guaranteed. SVR was firstly applied as meta-model for global sensitivity analysis in Refs (Cheng et al. 2017a; Cheng et al. 2017b).

Given the training sample set $\{\mathbf{X}, \mathbf{Y}\}$, a linear SVR model is formulated as:

$$\widetilde{g}(\mathbf{x}) = \boldsymbol{\omega}\cdot\mathbf{x} + b, \tag{68}$$

where $\boldsymbol{\omega} \in R^n$ is the coefficient vector and $b \in R$ is a constant. The goal of SVR is to find a function $\widetilde{g}(\mathbf{x})$ that can estimate the output response value whose deviation is less than ε from the real targets of the training data (ε-tube) and is as flat as possible.

Flatness in Eq. (68) means that one seeks the biggest tube width γ as shown in Fig. 3, which is equivalent to minimize $\|\boldsymbol{\omega}\|^2$. Therefore, the optimal regression function is determined by solving the following optimization problem:

$$\begin{aligned} \min \quad & \frac{1}{2}\|\boldsymbol{\omega}\|^2, \\ \text{s.t.} \quad & -\varepsilon \le y_i - \boldsymbol{\omega}\cdot\mathbf{x}_i - b \le \varepsilon, \forall i, i \in (1, 2, \ldots N). \end{aligned} \tag{69}$$

Note that the solution of above optimization problem may not actually exist and it is likely that better predictions can be obtained if we allow for some possible outliers. Therefore, the slack variables ξ_i and $\xi_i^*(i = 1, \ldots, N)$ as shown in Fig. 3 are introduced; thus, the above optimization problem turns to

$$\min \quad \frac{1}{2}\|\boldsymbol{\omega}\|^2 + C\sum_{i=1}^{N}(\xi_i + \xi_i^*), \tag{70}$$

$$\begin{aligned} \text{s.t.} \quad & y_i - \boldsymbol{\omega}\cdot\mathbf{x}_i - b \le \varepsilon + \xi_i, \xi_i \ge 0, \\ & \boldsymbol{\omega}\cdot\mathbf{x}_i + b - y_i \le \varepsilon + \xi_i^*, \xi_i^* \ge 0, \quad \forall i, i \in (1, 2, \ldots, N). \end{aligned} \tag{71}$$

The first part in Eq. (70) represents the model complexity, whereas the second part is the empirical risk. The regularization constant C here makes a trade-off between the empirical risk and the model complexity. Low value of C will result in simple (or flat) functions while high value can lead to over-fitting phenomenon. Thus, the proper selection of the parameter is crucial for obtaining high-quality prediction results.

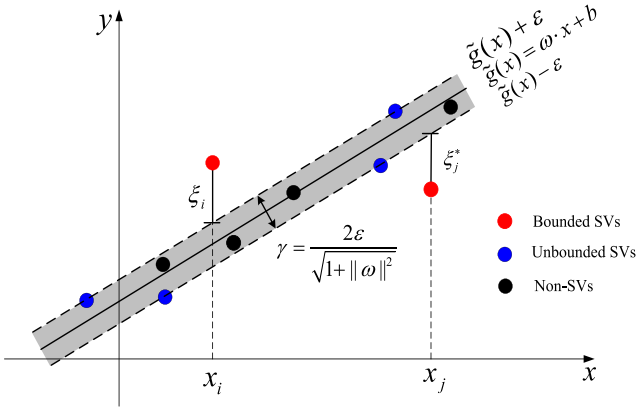

Fig. 3 Geometric interpretation for SVR

According to the Lagrangian theory and the Karush-Kuhn-Tucker condition, the original optimization problem can be transformed into the following dual form:

$$\begin{aligned} \min \quad & \frac{1}{2}\sum_{i=1}^{N}\sum_{j=1}^{N}(\alpha_i - \alpha_i^*)\left(\alpha_j - \alpha_j^*\right)\mathbf{x}_i\cdot\mathbf{x}_j + \varepsilon\sum_{i=1}^{N}(\alpha_i + \alpha_i^*) - \sum_{i=1}^{N} y_i(\alpha_i - \alpha_i^*), \\ \text{s.t.} \quad & \sum_{i=1}^{n}(\alpha_i - \alpha_i^*) = 0, 0 \le \alpha_i, \alpha_i^* \le C, \forall i, i \in (1, 2, \ldots, N). \end{aligned} \tag{72}$$

where $\{\alpha_i, \alpha_i^*\}(i = 1, \ldots, N)$ are Lagrange multipliers. After solving the dual optimization problem above, $\boldsymbol{\omega}$ can be obtained explicitly as $\boldsymbol{\omega} = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{x}_i$, and the final SVR predictor $\widetilde{g}(\mathbf{x})$ is expressed as

$$\widetilde{g}(\mathbf{x}) = \sum_{i=1}^{N}(\alpha_i - \alpha_i^*)\mathbf{x}_i\cdot\mathbf{x} + b. \tag{73}$$

As shown in Fig. 3, the training samples located inside the insensitive tube with $\alpha_i - \alpha_i^* = 0$ are not support vectors (SVs). In other words, only support vectors are relevant to the actual prediction. This leads to a sparse representation of SVR with respect to the training points.

For nonlinear problems, the input variables are mapped into a high-dimensional linear feature space by the nonlinear transform $\mathbf{x} \to \boldsymbol{\varphi}(\mathbf{x})$, then the prediction function in this feature space can be derived as follows:

$$\begin{aligned} \widetilde{g}(\mathbf{x}) &= \sum_{i=1}^{N}(a_i - \alpha_i^*)\boldsymbol{\varphi}(\mathbf{x}_i)\cdot\boldsymbol{\varphi}(\mathbf{x}) + b \\ &= \sum_{i=1}^{N}(a_i - \alpha_i^*)k(\mathbf{x}_i, \mathbf{x}) + b, \end{aligned} \tag{74}$$

where $k(\mathbf{x}_i, \mathbf{x}_j) = \boldsymbol{\varphi}(\mathbf{x}_i)\cdot\boldsymbol{\varphi}(\mathbf{x}_j)$ is the so-called kernel function, which is used to substitute the inner product of the feature variable $\boldsymbol{\varphi}(\mathbf{x})$. Among the various kernel functions in literature (Gaussian Radial Basis Function, Polynomial, Sigmoid, Fourier, etc) (Bourinet et al. 2011), the Gaussian RBF kernel draws the most attention, which is formulated as

$$k(\mathbf{x}_i, \mathbf{x}) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}\|^2/\sigma^2\right), \tag{75}$$

where σ is the hyper-parameter of kernel function. Substituting above Gaussian RBF kernel into Eq. (74), one can find that the SVR predictor can be also transformed into a sum of multivariate tensor product basis functions as:

$$\widetilde{g}(\mathbf{x}) = \sum_{i=1}^{N} \left(a_i - \alpha_i^*\right) k(\mathbf{x}_i, \mathbf{x}) + b$$
$$= \sum_{i=1}^{N} \hat{\beta}_i \prod_{l=1}^{n} \exp\left(-\left(x_{i,l} - x_l\right)^2 / \sigma^2\right) + b, \quad (76)$$

where $\hat{\beta}_i = \left(a_i - \alpha_i^*\right) (i = 1, \ldots, N)$.

Based on Eq. (76), the variance of a subset of input variables $\mathbf{x}_\mathbf{U} = \{x_{i_1}, \ldots, x_{i_s}\} \in \mathbf{x}$ can be computed as

$$\widetilde{V}_\mathbf{U} = \sum_{i=1}^{N} \sum_{j=1}^{N} \hat{\beta}_i \hat{\beta}_j \prod_{l=1}^{n} C1_{i,l} C1_{j,l} \left(\prod_{l \in \mathbf{U}} C2_{i,j,l} / C1_{i,l} C1_{j,l} - 1 \right), \quad (77)$$

where

$$C1_{i,l} = \int \exp\left(-\left(x_{i,l} - x_l\right)^2 / \sigma^2\right) f_{X_l}(x_l) dx_l,$$
$$C2_{i,j,l} = \int \exp\left(-\left[\left(x_{i,l} - x_l\right)^2 + \left(x_{j,l} - x_l\right)^2\right] / \sigma^2\right) f_{X_l}(x_l) dx_l. \quad (78)$$

The similar way can be used to solve above univariate integrals problem as that in Eqs. (55) and (56). Therefore, the variance-based global sensitivity indices can be computed analytically based on SVR predictor.

Cheng and Lu (Cheng and Lu 2018a; Cheng et al. 2017a) proposed an orthogonal polynomial kernel function using the PCE basis function as

$$k\left(\mathbf{x}_i, \mathbf{x}_j\right) = \boldsymbol{\psi}_\boldsymbol{\alpha}(\mathbf{x}_i) \cdot \boldsymbol{\psi}_\boldsymbol{\alpha}\left(\mathbf{x}_j\right) = \sum_{0 \le |\boldsymbol{\alpha}| \le p} \psi_\boldsymbol{\alpha}(\mathbf{x}_i) \psi_\boldsymbol{\alpha}\left(\mathbf{x}_j\right). \quad (79)$$

Based on this kernel function, SVR predictor can be transformed as

$$\widetilde{g}(\mathbf{x}) = \sum_{i=1}^{N} \left(a_i - \alpha_i^*\right) k(\mathbf{x}_i, \mathbf{x}) + b$$
$$= \sum_{i=1}^{N} \left(a_i - \alpha_i^*\right) \boldsymbol{\psi}_\boldsymbol{\alpha}(\mathbf{x}_i) \cdot \boldsymbol{\psi}_\boldsymbol{\alpha}(\mathbf{x}) + b$$
$$= \boldsymbol{\omega}_\boldsymbol{\alpha} \cdot \boldsymbol{\psi}_\boldsymbol{\alpha}(\mathbf{x}) + b, \quad (80)$$

where $\boldsymbol{\omega}_\boldsymbol{\alpha} = \sum_{i=1}^{N} \left(a_i - \alpha_i^*\right) \boldsymbol{\psi}_\boldsymbol{\alpha}(\mathbf{x}_i)$.

In this case, the SVR predictor in Eq. (80) is consistent with the classic PCE model in form. Therefore, the variance-based sensitivity indices can be also computed directly by post-processing the SVR coefficients by reorganizing of Eq. (80) as that in Eq. (33). Similar to the sparse PCE model in Refs (Blatman and Sudret 2011; Blatman and Sudret 2010a; Blatman and Sudret 2010b), an adaptive algorithm is also proposed in Ref (Cheng and Lu 2018a) to select the most important basis functions using the variance contribution criterion to construct a sparse kernel function in Eq. (71). This procedure finally leads to a sparse SVR predictor, and it improves the accuracy and efficiency of SVR model for GSA tremendously.

Enhanced versions of SVR such as gradient-enhanced SVR and multi-fidelity SVR are also available in literature (Zhou and Jiang 2018; Jiang and Zhou 2018; Zhou et al. 2015) to improve the performance of traditional SVR model when gradient information is available cheaply or simulations with multiple levels of fidelity are available. These methods can be applied for GSA in a similar way as introduced above.

The SVR provides better generalization ability compared to other meta-model algorithms, but the performance is sensitive to the hyper-parameters, including the penalty term C, the insensitive tube width ε, and kernel parameter. Generally, the optimal hyper-parameters of SVR model are selected using the general-purpose optimization algorithm to minimize the cross validation (CV) error (Bourinet et al. 2011; Bourinet 2016).

3.7 Radial basis function regression

RBF was originally proposed by Hardy (Hardy 1971) in 1971 to fit the irregular topographic contours of geographical data, and it was used for global sensitivity analysis in Refs (Chen and Jin 2004; Zhang et al. 2017a; Wu et al. 2016a).

Given the training sample set $\{\mathbf{X}, \mathbf{Y}\}$(normalized to the unit interval), the standard RBF regression model is expressed as

$$\widetilde{g}(\mathbf{x}) = \sum_{i=1}^{N} w_i \varphi(\|\mathbf{x} - \mathbf{x}_i\|) \quad (81)$$

where the approximation function $\widetilde{g}(\mathbf{x})$ is represented as a sum of N RBFs, each associated with a different sample point $\mathbf{x}_i$, and weighted by an appropriate coefficient$k_2(\mathbf{v}, x)$. The weight coefficients $\mathbf{w} = [w_1, \ldots, w_N]^\mathrm{T}$ in RBF regression can be computed by solving the following linear system of equations (Majdisova and Skala 2017)

$$\boldsymbol{\Phi}\mathbf{w} = \mathbf{Y}, \quad (82)$$

where $\boldsymbol{\Phi}$ is a $N \times N$ symmetric positive semi-definite matrix (Gram matrix) with elements $\Phi_{ij} = \varphi(\|\mathbf{x}_i - \mathbf{x}_j\|)$. In practice, the following Gaussian basis function is often employed in RBF regression,

$$\varphi(\|\mathbf{x}_i - \mathbf{x}\|) = \exp\left(-\|\mathbf{x}_i - \mathbf{x}\|^2 / c_i^2\right), \quad (83)$$

where c_i is the bandwidth of the i-th Gaussian RBF function.

Therefore, the RBF regression predictor can be also expressed as a sum of multivariate tensor product basis functions as that in ordinary Kriging and SVR (Wu et al. 2016b), namely,

$$\widetilde{g}(\mathbf{x}) = \sum_{i=1}^{N} w_i \varphi(\|\mathbf{x} - \mathbf{x}_i\|) = \sum_{i=1}^{N} w_i \prod_{l=1}^{n} \varphi\left(-\left(x_{i,l} - x_l\right)^2 / c_i^2\right) \quad (84)$$

The only difference of the meta-model expression in Eqs. (76) and (84) is the hyper-parameter (σ in SVR and $c_i (i = 1,$

…, N) in RBF regression). Thus, the variance of a subset of input variables $\mathbf{x}_\mathbf{U} = \{x_{i_1}, \ldots, x_{i_s}\} \in \mathbf{x}$ can be computed as

$$\widetilde{V}_\mathbf{U} = \sum_{i=1}^{N}\sum_{j=1}^{N} w_i w_j \prod_{l=1}^{n} C1_{i,l}C1_{j,l}\left(\prod_{l\in\mathbf{U}} C2_{i,j,l}/C1_{i,l}C1_{j,l} - 1\right) \tag{85}$$

where

$$\begin{aligned} C1_{i,l} &= \int \exp\left(-\left(x_{i,l}-x_l\right)^2/c_i^2\right) f_{X_l}(x_l)dx_l, \\ C2_{i,j,l} &= \int \exp\left(-\left[\left(x_{i,l}-x_l\right)^2/c_i^2 + \left(x_{j,l}-x_l\right)^2/c_j^2\right]\right) f_{X_l}(x_l)dx_l, \end{aligned} \tag{86}$$

Based on Eqs. (85) and (86), the variance-based global sensitivity indices can be obtained analytically using RBF regression predictor. These univariate integrals can be computed as those in Eqs. (55) and (56).

The estimation of the bandwidth $c_i (i = 1, \ldots, N)$ is of great importance for good approximations in RBF regression (Wu et al. 2016a; Amouzgar et al. 2018). To this end, a local density sampling points method was proposed by Wang et al. (Wang et al. 2014a; Wang et al. 2014b) to transform the estimation of numerous bandwidth c_i into the calculation of a singular parameter. The local density of a point is defined as

$$\rho(\mathbf{x}) = \sum_{i=1}^{N} \exp\left(-\|\mathbf{x}_i-\mathbf{x}\|^2/\sigma^2\right), \quad \sigma = \frac{1}{\sqrt[n]{N}} \tag{87}$$

As represented in Fig. 4, the radial basis function with a small bandwidth c_i gives a narrow effect on the surrounding region, and vice versa. Thus, the bandwidth $c_i (i = 1, \ldots, N)$ can be interpreted as the influence extent of the i-th sampling point $\mathbf{x}_i$; thus, the influence volume of the i-th sampling point $\mathbf{x}_i$ can be obtained as $V_i = c_i^n$ (Wang et al. 2014a; Wang et al. 2014b). For good approximation, the influence volume of a sampling point $\mathbf{x}_i$ should be inversely proportional to the local density $\rho(\mathbf{x}_i)$, namely,

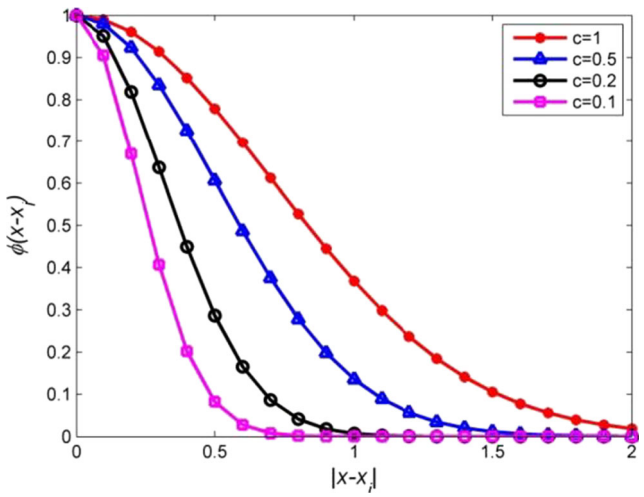

Fig. 4 Influence extent of φ with respect to c

$$\frac{V_i}{V_j} = \frac{\rho(\mathbf{x}_j)}{\rho(\mathbf{x}_i)} \tag{88}$$

The sum of the influence volumes ζ of all the training points can be obtained as $\sum_{i=1}^{N} V_i$, which should cover the whole input variable space. Thus, one only needs to estimate the singular parameter ζ, and the optimal ζ can be estimated by minimizing the cross validation (CV) error (Wang et al. 2014a).

Furthermore, the standard RBF regression in Eq. (81) can theoretically have problems with stability and solvability (Amouzgar and Strömberg 2017; Meczekalski and Podfigurna-Stopa 2000). Thus, a remedy is to combine RBF with a k-order polynomial $p_k(\mathbf{x})$, namely,

$$\widetilde{g}(\mathbf{x}) = \sum_{i=1}^{N} w_i \varphi(\|\mathbf{x}-\mathbf{x}_i\|) + p_k(\mathbf{x}). \tag{89}$$

In practice, the linear polynomial

$$p_1(\mathbf{x}) = \mathbf{a}^\mathrm{T}\mathbf{x} + a_0 \tag{90}$$

is usually used. In this case, the RBF predictor is consistent with the universal Kriging model with linear polynomial regression model.

3.8 Low-rank tensor approximation

Low-rank tensor approximation (LRA) is a meta-model technique developed to tackle the issue of curse-of- for high-dimensional problem (Konakli and Sudret 2016a; Hadigol et al. 2014; Chevreuil et al. 2013; Corveleyn and Vandewalle 2017; Validi 2014; Doostan et al. 2013; Konakli and Sudret 2016b; Konakli and Sudret 2016c), which expresses the model response as a sum of a small number of rank-one tensor functions as

$$\widetilde{g}(\mathbf{x}) = \sum_{i=1}^{R} b_i \prod_{l=1}^{n} v_i^{(l)}(x_l) \tag{91}$$

where R is the rank of the approximation, $v_i^{(l)}$ is a univariate function of x_l in the i-th rank-one component, and $b_i (i = 1, \ldots, R)$ are the normalizing constants. Usually, these univariate functions are expanded onto the polynomial basis function as (Konakli and Sudret 2016a; Doostan et al. 2013),

$$v_i^{(l)}(x_l) = \sum_{k=0}^{p_l} z_{k,i}^{(l)} P_k^{(l)}(x_l) \tag{92}$$

where $P_k^{(l)}$ is the k-th degree univariate polynomial of x_l, P_l is the maximum degree of $P_k^{(l)}$, and $z_{k,i}^{(l)}$ is the coefficient of $P_k^{(l)}$ in the i-th rank-one component.

The number of the unknown coefficients in LRA grows linearly with the input variable dimensionality n; thus, this

technique is promising for dealing with high-dimensional problems. To estimate these coefficients, the alternated least-squares (ALS) minimization method is developed in literature (Konakli and Sudret 2016a; Doostan et al. 2013; Konakli and Sudret 2016b). This method consists in a correction step and an update step. It sequentially solves a least-square minimization problem along each dimension separately, while "freezes" the coefficients in all the remaining dimensions.

To construct a rank-R approximation, a greedy construction algorithm is proposed in Refs (Konakli and Sudret 2016b; Chevreuil et al. 2014). This algorithm iteratively increases the rank of LRA model by adding rank-1 components successively. In each iteration, a correction step is used to determine the coefficients $z_{k,i}^{(l)}(k=0,\ldots,p_l)$ along each dimension by

$$\operatorname{argmin}\left\|M_{r-1}-\left(\prod_{i\neq l}v_r^{(i)}\right)\left(\sum_{k=0}^{p_l}z_{k,i}^{(l)}P_k^{(l)}\right)\right\|,\quad l=1,\ldots,n, \tag{93}$$

where M_{r-1} represents the residual at the training sample set $\{\mathbf{X},\mathbf{Y}\}$ after the completion of the $(r\text{-}1)$-th iteration. Then, an update step is used to compute the normalizing coefficients $b_i(i=1,\ldots,r)$ by

$$\operatorname{argmin}\left\|\mathbf{Y}-\sum_{i=1}^{r}b_i\prod_{l=1}^{n}v_i^{(l)}(x_l)\right\|. \tag{94}$$

Generally, the optimal rank R is determined by using the CV criteria. In Refs (Chevreuil et al. 2013; Konakli and Sudret 2016b), the threefold CV is utilized to select R, and it has been proven to be efficient for the optimal or near optimal rank selection. For more details of LRA, one can refer to Refs (Konakli and Sudret 2016b; Chevreuil et al. 2014).

It is clear that the LRA is formulated as a sum of multivariate tensor product basis functions form; thus, the variance-based global sensitivity indices can be obtained directly (Konakli and Sudret 2016a). Based on Eq. (91), the variance of a subset of input variables $\mathbf{x}_\mathbf{U}=\{x_{i_1},\ldots,x_{i_s}\}\in\mathbf{x}$ can be computed as

$$\widetilde{V}_\mathbf{U}=\sum_{i=1}^{N}\sum_{j=1}^{N}b_ib_j\prod_{l=1}^{n}C1_{i,l}C1_{j,l}\left(\prod_{l\in\mathbf{U}}C2_{i,j,l}/C1_{i,l}C1_{j,l}-1\right) \tag{95}$$

where

$$\begin{aligned}C1_{i,l}&=\int v_i^{(l)}(x_l)f_{X_l}(x_l)dx_l=z_{0,i}^{(l)}E\left[P_0^{(l)}(x_l)\right],\\ C2_{i,j,l}&=\int v_i^{(l)}(x_l)v_j^{(l)}(x_l)f_{X_l}(x_l)dx_l=\sum_{k=0}^{p_l}z_{k,i}^{(l)}z_{k,j}^{(l)}E\left[\left(P_k^{(l)}(x_l)\right)^2\right]\end{aligned} \tag{96}$$

where $E\left[P_0^{(l)}(x_l)\right]$ and $E\left[\left(P_k^{(l)}(x_l)\right)^2\right]$ can be obtained analytically as that in Ref (Sudret 2008).

Since the LRA technique relies on a series of alternated minimizations along separate dimensions, it yields satisfactory results for models with a tensor product structure. For some highly non-linear problems, the performance of this method may be unstable (Konakli and Sudret 2016a).

4 The links among different surrogate models

In this section, we summarize and explain the links of various surrogate model techniques reviewed earlier.

(1) ANOVA-HDMR actually belongs to the methodology of PCE, since both of them utilize the orthogonal polynomial bases to approximate the model response, and one only needs to determine the unknown coefficients of these basis functions. In addition, the coefficients computation method of ANOVA-HDMR in Eq. (18) is the projection approach in PCE essentially. After the basis functions coefficients are determined, the variance-based sensitivity indices can be obtained directly by these coefficients.

(2) In PCE, the response QoI is represented explicitly in a high-dimensional space spanned by the polynomial chaos basis, and one only needs to determine the coordinates (PCE coefficients) of the random response in this basis. In this regard, one can define a kind of kernel function by the inner product of PCE basis as that in Eq. (79) (Cheng and Lu 2018a). This kernel function maps the original data into a high-dimensional linear feature space spanned by the PCE bases. Therefore, one can establish PCE model by SVR technique, namely, PCE and SVR can be combined into a unified framework. Since SVR is developed based on the principal of structural risk minimization, thus it provides better generalization ability

Table 2 Unified surrogate model formulation

Surrogate models	Formulation
Polynomial regression model	$\widetilde{g}(\mathbf{x})=\hat{\beta}_0+\sum_{\substack{0\le i,j\le n\\ j\neq 0}}\hat{\beta}_{ij}\prod_{l=1}^{n}h_{(i,j)l},$
PCE (ANOVA-HDMR)	$\widetilde{g}(\mathbf{x})=\sum_{0\le\lvert\boldsymbol{\alpha}\rvert\le p}\omega_{\boldsymbol{\alpha}}\prod_{l=1}^{n}\psi_{\alpha_l}^{(l)}(x_l)$
Kriging/GP	$\widetilde{g}(\mathbf{x})=\sum_{i=1}^{M}\hat{\beta}_i\prod_{l=1}^{n}p_i(x_l)+\sum_{i=1}^{N}\hat{a}_i\prod_{l=1}^{n}\exp\left(-\theta_l\left(x_{i,l}-x_l\right)^2\right)$
SVR	$\widetilde{g}(\mathbf{x})=\sum_{i=1}^{N}\hat{\beta}_i\prod_{l=1}^{n}\exp\left(-\left(x_{i,l}-x_l\right)^2/\sigma^2\right)+b$
RBF	$\widetilde{g}(\mathbf{x})=\sum_{i=1}^{N}w_i\prod_{l=1}^{n}\exp\left(-\left(x_{i,l}-x_l\right)^2/c_i^2\right)$
LRA	$\widetilde{g}(\mathbf{x})=\sum_{i=1}^{R}b_i\prod_{l=1}^{n}v_i^{(l)}(x_l)$

Table 3 Commonly used surrogate model validation metrics

Validation metric	Formula
Relative root mean square error	$\frac{\sqrt{\sum_{i=1}^{M}\left(y_i-\widetilde{y}_i\right)^2/M}}{\sqrt{\sum_{i=1}^{M}\left(y_i-\overline{y}\right)^2/(M-1)}}$
Relative maximum absolute error	$\max\frac{\lvert y_i-\widetilde{y}_i\rvert}{\sqrt{\sum_{i=1}^{M}\left(y_i-\overline{y}\right)^2/(M-1)}}(i=1,\ldots,M)$
Relative average absolute error	$\frac{\sum_{i=1}^{M}\lvert y_i-\widetilde{y}_i\rvert/M}{\sqrt{\sum_{i=1}^{M}\left(y_i-\overline{y}\right)^2/(M-1)}}$
Coefficient of determination R^2	$1-\frac{\sum_{i=1}^{M}\left(y_i-\widetilde{y}_i\right)^2}{\sum_{i=1}^{M}\left(y_i-\overline{y}\right)^2}$
Mean square error	$\sum_{i=1}^{M}\left(y_i-\widetilde{y}_i\right)^2/M$
Mean absolute error	$\sum_{i=1}^{M}\lvert y_i-\widetilde{y}_i\rvert/M$

and curbs the over-fitting problem when high-order PCE is used.

(3) In Kriging/GP, an important point is the choice of the covariance function (also known as covariance kernel) in Eq. (42). Generally, the covariance function is stationary if it is a function of the sample distance $\mathbf{d} = \mathbf{x}_i - \mathbf{x}_j$, and the correlation between two samples is uniquely determined by their distance. Meanwhile, one can also construct non-stationary covariance function by inner product of PCE basis functions according to Mercer's theorem (Rasmussen and Williams 2006; Cheng et al. 2019)

$$Cov\left(Z(\mathbf{x}_i),Z(\mathbf{x}_j)\right)=\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\mathbf{x}_i)\cdot\boldsymbol{\psi}_{\boldsymbol{\alpha}}(\mathbf{x}_j)=\sum_{0\le|\boldsymbol{\alpha}|\le p}\psi_{\boldsymbol{\alpha}}(\mathbf{x}_i)\psi_{\boldsymbol{\alpha}}(\mathbf{x}_j). \tag{97}$$

Based on this covariance function, one can also establish PCE model by GP regression technique (Cheng et al. 2019; Yan et al. 2018). In this setting, the obtained PCE model provides the prediction variance (Kriging variance) to measure the local predict risk.

(4) Although developed from different branch of research, multiple surrogate models reviewed above have the unified formulation. Here, we summarize the expressions of above reviewed various surrogate models, where the Kriging/GP covariance function, SVR kernel function, and RBF basis function are all Gaussian form.

From Table 2, one can see that these surrogate models can be expressed as a sum of multivariate tensor product basis functions form. In the field of UQ, an important section is the estimation of statistical moments of a QoI, which usually utilizes the numerical integral techniques to perform multivariate integrals. However, based on above unified formulation, one can transform the multivariate integrals problem into univariate integrals; thus, the statistical moments and variance-based sensitivity indices can be computed analytically based on these surrogate models. Moreover, although these surrogate models listed in Table 2 developed based on different theory from different perspectives, one can conclude that all the surrogate models can be expressed as a linear combination of the of coefficients with the corresponding "basis functions."

5 Sampling strategy

The accuracy and efficiency of surrogate models heavily depend on the sampling strategy. To maintain the quality of surrogate model without incurring excessive samples, the sampling strategy of constructing surrogate model is of immense significance (Bhosekar and Ierapetritou 2018; Picheny et al. 2012; Garud et al. 2017; Yondo et al. 2018).

Generally speaking, the sampling strategies are broadly classified into two categories (Bhosekar and Ierapetritou 2018; Garud et al. 2017; Liu et al. 2018): static/one-shot sampling and adaptive/sequential sampling. The static/one-shot sampling consists of methods that determine the sample size

Table 4 Four benchmark test functions

Function	Dimensionality	Expression	Variable space
Ishigami function	3	$g(\mathbf{x})=\sin x_1+7\sin^2 x_2+0.1x_3^4\sin x_1$	$x_i\sim U(-\pi,\pi)$
Borehole function	8	$g(\mathbf{x})=\frac{2\pi x_3(x_2-x_6)}{\ln(x_1/x_4)\left(1+\frac{2x_7x_3}{\ln(x_1/x_4)x_4^2x_8}+\frac{x_3}{x_5}\right)}$	$x_1\sim U(100,50000),x_2\sim U(990,1110)$ $x_3\sim U(63070,115600),x_6\sim U(700,820)$ $x_5\sim U(63.1,116),x_8\sim U(9855,12045)$ $x_7\sim U(1120,1680),x_4\sim U(0.05,0.15)$
Sobol function	8	$g(\mathbf{x})=\prod_{i=1}^{8}\frac{\lvert 4x_i-2\rvert+a_i}{1+a_i},\mathbf{a}=[1,2,5,10,20,50,100,500]$	$x_i\sim U(0,1)$
Morris function	20	$g(\mathbf{x})=\sum_{i=1}^{20}w_i+\sum_{i<j}^{20}w_iw_j+\sum_{i<j<l}^{10}w_iw_jw_l+\sum_{i<j<l<s}^{5}w_iw_jw_lw_s,$ $w_i=\begin{cases}1.1x_i/(x_i+0.1)-0.5 & \text{if } i=3,5,7\\ x_i-0.5 & \text{else}\end{cases}$	$x_i\sim U(0,1)$

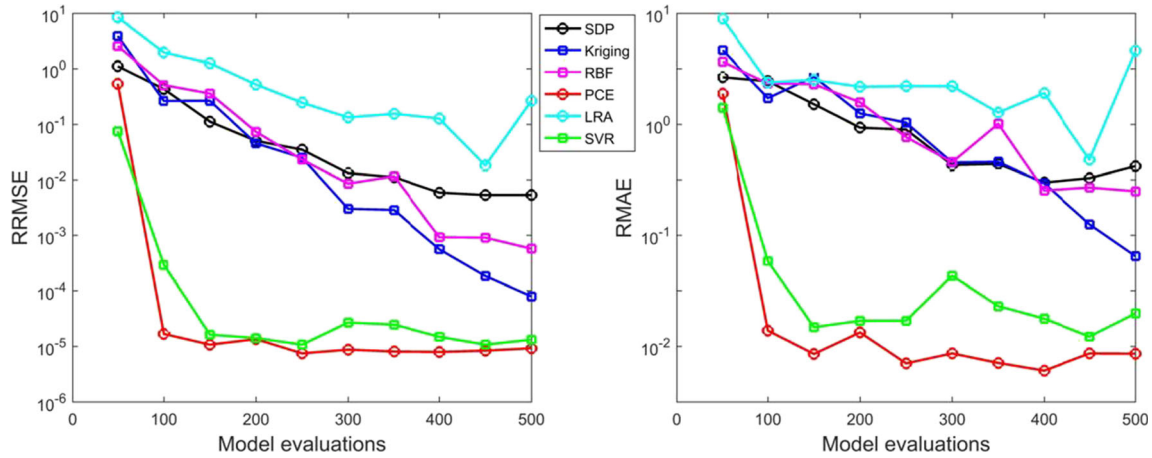

Fig. 5 Comparisons of the accuracy of different surrogate model techniques of the Ishigami function

at once, but these approaches can result in under/oversampling problem. Some of the widely used static/one-shot sampling strategies are Monte Carlo and stratified Monte Carlo sampling (Shlomo and Shaul 2011), Quasi-Monte Carlo sampling (Sobol 1967; Halton 1964), space-filling sampling (Johnson et al. 1990), Latin hypercube sampling (Mckay et al. 2000), and orthogonal array sampling (Owen 1992). Compared to the static/one-shot sampling technique, the adaptive/sequential sampling strategy allows one to choose informative points actively via the meta-model that it learns, and consequently, performs better with fewer points. In adaptive/sequential sampling method, the informative points are selected by some criteria to tackle the trade-off between local exploitation and global exploration. Local exploitation aims at finding the interesting regions with large prediction errors, while the global exploration aims at escaping from the local regions and discovering more interesting regions that have not been detected.

In this survey, we mainly compare the accuracy and efficiency of different surrogate model techniques. Therefore, for a test function, all the surrogate models reviewed in Section 3 are constructed with the same training samples for fair comparisons. Here, the training samples are generated by Sobol Quasi-Monte Carlo approach for its good space-filling property.

6 Surrogate model accuracy validation

Assessing the accuracy of surrogate model is of significant importance for meta-model construction and hyper-parameters selection, and an inaccurate surrogate will waste the computational resources and lead to a wrong GSA results. Therefore, surrogate model validation is a vital process to assess the reliability of meta-model.

To validate the surrogate model accuracy without using extra samples, the usually used methods are resampling strategies, such as CV and bootstrapping. In k-fold CV, the samples set $\{\mathbf{X}, \mathbf{Y}\}$ is randomly split into k mutually exclusive subsets with equal size (Bourinet 2016), then samples from (k-1) subsets are used as training set to fit a meta-model and the

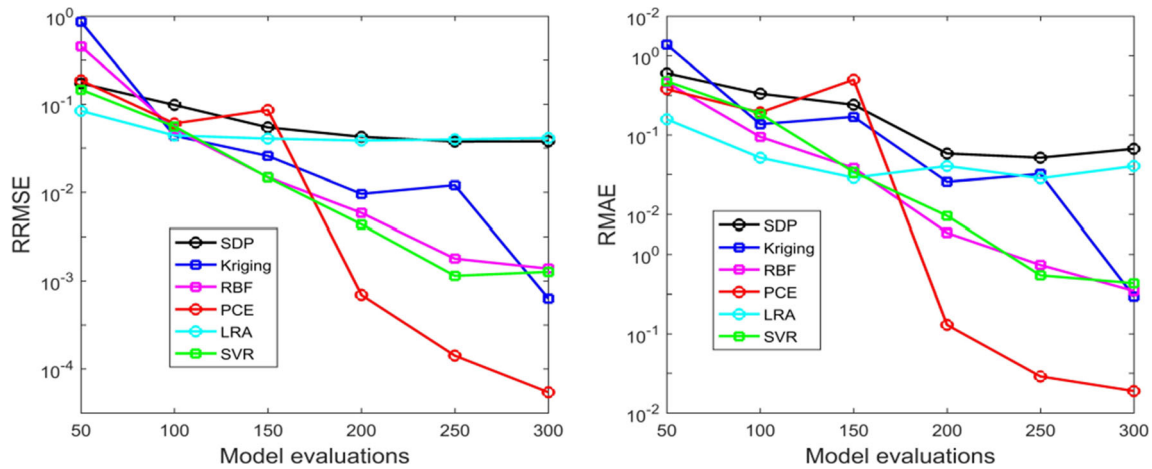

Fig. 6 Comparisons of the accuracy of different surrogate model techniques for borehole function

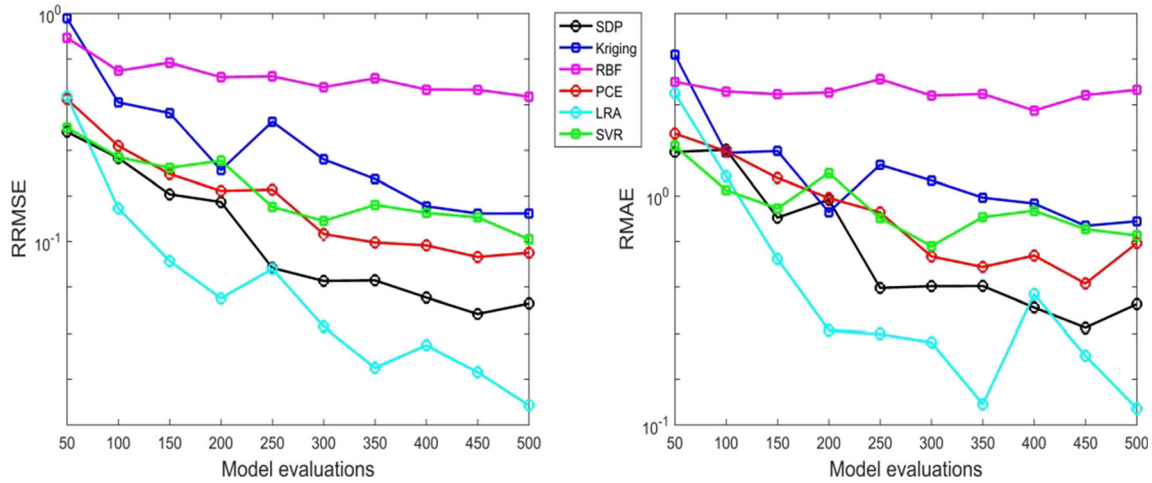

Fig. 7 Comparisons of the accuracy of different surrogate model techniques for the Sobol function

remaining subsets are used as the test set to validate the meta-model accuracy. This procedure is repeated for each of the k subsets and the average of the k obtained validation errors provides an estimate of the meta-model accuracy. The limit case corresponding k=N is known as the leave-one-out CV. A similar resampling technique is bootstrap method (Marelli and Sudret 2018; Johnson 2001; Efron 1982; Kohavi 1995), which allows repeated samples in the training set to build meta-model. This method consists in selecting N samples randomly from the original samples set $\{\mathbf{X}, \mathbf{Y}\}$ by resampling with replacement. Since the dataset is sampled with replacement, the probability of a sample not being chosen after N sampling is $(1 - 1/N)^N \approx e^{-1} = 0.368$. Therefore, there will be about $0.368N$ samples that can be used as the test samples set to assess the surrogate model accuracy.

To assess the surrogate model accuracy, various validation metrics can be used to measure the error of meta-model from different perspectives. Table 3 lists some commonly used surrogate model accuracy validation metrics, including relative root mean square error (RRMSE), relative maximum absolute error (RMAE), relative average absolute error (RAAE), coefficient of determination (R^2), mean square error (MSE), mean absolute error (MAE), where $y, \widetilde{y}, \overline{y}$ and M denote true model response value, surrogate model predicted value, mean of the true model response value, and number of test samples respectively. The RRMSE, RAAE, R^2, and MSE can be usually used to gauge the overall accuracy of the meta-model, while the RMAE and MAE can be adopted to measure the local accuracy of surrogate model.

7 Comparisons of the performance of various surrogate techniques

This section is dedicated to the comparisons of different surrogate model techniques. Four test functions listed in Table 4, which are widely used as benchmarks (Konakli and Sudret 2016a; Sudret 2008; Blatman and Sudret 2010a; Cheng and

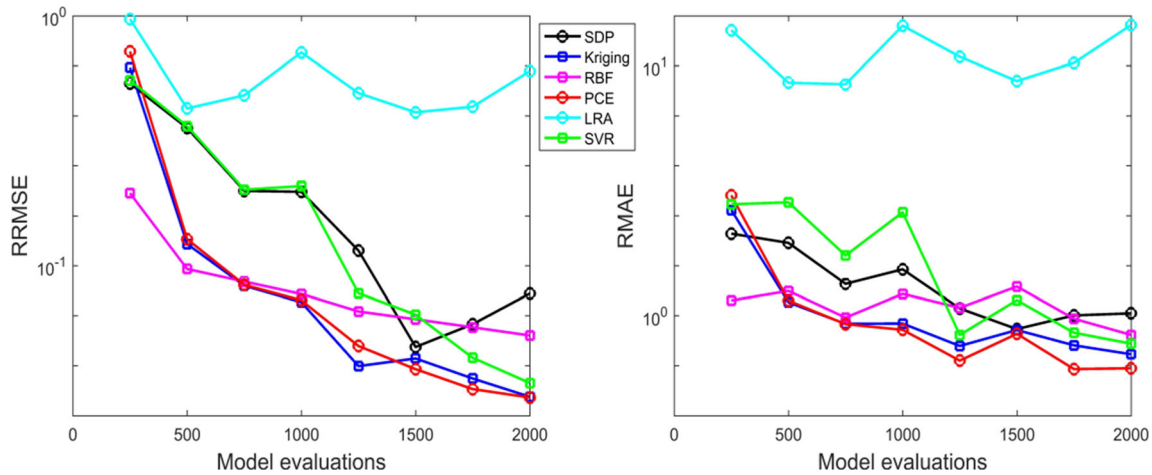

Fig. 8 Comparisons of the accuracy of different surrogate model techniques for the Morris function

Table 5 The results of sensitivity analysis for the Ishigami function with 100 samples

Sensitivity indices	References	SDP	Relative error	Kriging	Relative error	RBF	Relative error
S_1	0.314	0.349	11.23%	0.312	0.70%	0.323	2.92%
S_2	0.442	0.450	1.77%	0.461	4.19%	0.360	18.65%
S_3	0	0	–	0.001	–	0.006	–
Sensitivity indices	References	PCE	Relative error	LRA	Relative error	SVR	Relative error
S_1	0.314	0.314	0.04%	0.617	96.7%	0.314	0.06%
S_2	0.442	0.443	0.03%	0.135	69.6%	0.444	0.3%
S_3	0	0	–	0.090	–	0	–

Lu 2018a; Cheng et al. 2017b) to validate sensitivity analysis method, are used to compare the accuracy and efficiency the various surrogate model techniques, including SDP (Ratto et al. 2007), sparse PCE (Blatman and Sudret 2011), universal Kriging (2-order polynomial) (Sn Lophaven and Søndergaard 2002), SVR, RBF (Wu et al. 2016c), and LRA (Konakli and Sudret 2016a; Konakli and Sudret 2016b). To illustrate the performance of different surrogate model techniques comprehensively, we compare the accuracy of each surrogate model as well as the corresponding main sensitivity indices. Here, the relative root mean square error (RRMSE) and relative maximum absolute error (RMAE) are used as accuracy metrics to measure the global accuracy and local accuracy of these surrogate models with 5000 test samples, and the relative error is used to measure the accuracy of the sensitivity indices.

Along with the increase of the training sample size, the convergence curves of RRMSE and RMAE obtained by various surrogate models for the four test functions are shown in Figs. 5, 6, 7, and 8. Meanwhile, the main sensitivity indices obtained by these surrogate models are listed in Tables 5, 6, 7, and 8. With the increase of the training sample, all the surrogate models are re-constructed and corresponding hyperparameters in each surrogate model are updated. For the Ishigami function and Sobol function, the reference values of sensitivity indices are obtained analytically (Sudret 2008; Blatman and Sudret 2010a). For the Borehole function and Morris function, the reference values of sensitivity indices are computed by MC simulation method (Wei et al. 2015).

Based on the comparison results of the four benchmarks from Figs. 4, 5, 6, 7, and 8 and corresponding sensitivity indices in Tables 4, 5, 6, 7, and 8, we see that an accurate surrogate model generally provides accurate sensitivity analysis results. Indeed, since the variance-based sensitivity indices reflect the second moment information of a stochastic function, an accurate surrogate model leads to accurate

Table 6 The results of sensitivity analysis for borehole function with 150 samples

Sensitivity indices	References	SDP	Relative error	Kriging	Relative error	RBF	Relative error
S_1	0.8309	0.8866	6.70%	0.8319	0.12%	0.8302	0.08%
S_2	0.0000	0.0000	–	0.0000	–	0.0000	–
S_3	0.0000	0.0000	–	0.0000	–	0.0000	–
S_4	0.0395	0.0479	21.27%	0.0416	5.32%	0.04108	4.00%
S_5	0.0000	0.0001	–	0.0000	–	0.0000	–
S_6	0.0417	0.0375	10.07%	0.0406	2.64%	0.0414	0.72%
S_7	0.0386	0.0418	8.29%	0.0382	1.04%	0.0392	1.55%
S_8	0.0097	0.0110	13.4%	0.0096	1.03%	0.0096	1.03%
Sensitivity indices	References	PCE	Relative error	LRA	Relative error	SVR	Relative error
S_1	0.8309	0.8305	0.048%	0.8684	4.51%	0.8301	0.096%
S_2	0.0000	0.0000	–	0.0000	–	0.0000	–
S_3	0.0000	0.0000	–	0.0000	–	0.0000	–
S_4	0.0395	0.0415	5.06%	0.0523	32.4%	0.0414	4.81%
S_5	0.0000	0.0000	–	0.0000	–	0.0000	–
S_6	0.0417	0.0410	1.68%	0.0552	32.37%	0.0414	0.72%
S_7	0.0386	0.0391	1.30%	0.0512	32.64%	0.0386	0.00%
S_8	0.0097	0.0095	2.06%	0.0130	34.02%	0.0096	1.03%

sensitivity indices unquestionably. From above comparisons and the references, here, we summarize the pros and cons of these surrogate model methods.

SDP is very efficient when interaction terms of the response function are very weak, e.g., test functions 2–4. This technique can be used to deal with non-smooth or even discontinuous response function, e.g., test function 2, which is a unique property compared to other methods. However, when the response exhibits strong interaction effect, the performance of SDP is relatively poor.

Sparse PCE is a promising meta-model technique in the field of UQ. In all the test functions, sparse PCE exhibits excellent performance. Benefitting from the polynomial basis function, PCE can capture the global behavior of a function efficiently. For high-dimensional problem, sparse PCE developed with some variable selection techniques (such as least angle regression, orthogonal matching pursuit, etc) fits a model very fast even the training samples size is relatively large. However, high-order PCE is prone to over-fitting, which degrades the generalization ability of PCE. Furthermore, the total number ($(p + n)\ !\ /p\ !\ n!$) of the basis function rapidly becomes too large for high-dimensional problem (large n) and high-order expansion (large p). Although some truncation strategies are introduced (such as the hyperbolic index in Ref (Blatman and Sudret 2011)) to reduce the bases, some of the important terms may be discarded.

Kriging exhibits excellent performance for low-medium dimensional problem. For high-dimensional problems, e.g., test functions 4, the polynomial term actually plays a dominant role, and the stochastic process only has little influence to the predictor. Kriging model provides the prediction variance to measure the local prediction error. Thanks to the Kriging variance, some active learning strategies or sequentially sampling techniques can be performed to update Kriging model adaptively. However, for high-dimensional problems, the number of samples needed to train Kriging is very large. In this case, the correlation matrix of Kriging is prone to be ill conditioned, and the estimation of the Kriging hyper-parameters by maximum likelihood estimation procedure is computational demanding. For large training samples set (about 5000), it may take several hours to train a Kriging model.

LRA provides more accurate prediction results compared to other methods for high-dimensional problem and function with a tensor product structure, e.g., test functions 2 and 3. Since LRA relies on a series of small-size minimization problems solved using ordinary least-squares, it is highly efficient to fit the model even for large sampling size. However, since the alternated least-squares minimization is employed, convergence problem occurs for fitting highly nonlinear problem, e.g., test functions 1 and 4.

SVR provides excellent performance for nearly all the test functions. It is robust and practical for high-dimensional problems and works well with few training samples. Due to the principle of structural risk minimization, over-fitting is infrequent even the high-order polynomial kernel in Eq. (71) is used. The computational cost of training SVR model is efficient since the operator associated with the training samples in SVR formulation is dot product (Razavi et al. 2012). Unfortunately, the hyper-parameters tuning process of SVR is a computationally demanding process (Yondo et al. 2018). However, the sparseness of the SVR predictor leads to high prediction speed of SVR model, which make it a good candidate for prediction of very

Table 7 The results of sensitivity analysis for the Sobol function with 200 samples

Sensitivity indices	References	SDP	Relative error	Kriging	Relative error	RBF	Relative error
S_1	0.604	0.636	5.43%	0.626	3.72%	0.528	12.57%
S_2	0.268	0.266	1.02%	0.241	10.31%	0.192	28.26%
S_3	0.067	0.068	0.84%	0.067	0.26%	0.060	10.22%
S_4	0.020	0.024	21.64%	0.023	17.07%	0.017	16.17%
S_5	0.006	0.005	8.92%	0.006	6.38%	0.009	62.01%
S_6	0.001	0	100%	0.002	100%	0.001	0.00%
S_7	0	0	–	0	–	0.001	–
S_8	0	0	–	0	–	0.001	–
Sensitivity indices	References	PCE	Relative error	LRA	Relative error	SVR	Relative error
S_1	0.604	0.631	4.48%	0.601	0.43%	0.612	1.41%
S_2	0.268	0.266	1.01%	0.269	0.38%	0.271	1.10%
S_3	0.067	0.059	11.69%	0.070	3.78%	0.067	0.90%
S_4	0.020	0.014	28.17%	0.019	5.94%	0.019	3.47%
S_5	0.006	0.004	25.05%	0.005	6.54%	0.006	0.80%
S_6	0.001	0	100%	0.001	0.00%	0	100%
S_7	0	0	–	0	–	0	–
S_8	0	0	–	0	–	0	–

Table 8 The results of sensitivity analysis for the Morris function with 1000 samples

Sensitivity indices	References	SDP	Relative error	Kriging	Relative error	RBF	Relative error
S_1	0.051	0.059	15.69%	0.051	0.00%	0.056	9.80%
S_2	0.051	0.052	1.96%	0.050	1.96%	0.052	1.96%
S_3	0.014	0.021	50.00%	0.013	7.14%	0.013	7.14%
S_4	0.051	0.056	9.80%	0.051	0.00%	0.049	3.92%
S_5	0.014	0.019	35.71%	0.014	0.00%	0.013	7.14%
S_6	0.051	0.055	7.84%	0.052	1.96%	0.052	1.96%
S_7	0.014	0.016	14.29%	0.015	7.14%	0.013	7.14%
S_8	0.051	0.065	27.45%	0.056	9.80%	0.056	9.80%
S_9	0.051	0.067	31.37%	0.054	5.88%	0.054	5.88%
S_{10}	0.050	0.064	28.00%	0.050	0.00%	0.053	6.00%
S_{11}	0.037	0.054	45.95%	0.042	13.51%	0.039	5.41%
S_{12}	0.037	0.045	21.62%	0.037	0.00%	0.038	2.70%
S_{13}	0.037	0.053	43.24%	0.038	2.70%	0.038	2.70%
S_{14}	0.037	0.047	27.03%	0.041	10.81%	0.041	10.81%
S_{15}	0.037	0.051	37.84%	0.043	16.22%	0.040	8.11%
S_{16}	0.037	0.043	16.22%	0.037	0.00%	0.039	5.41%
S_{17}	0.037	0.048	29.73%	0.040	8.11%	0.038	2.70%
S_{18}	0.037	0.055	48.65%	0.038	2.70%	0.040	8.11%
S_{19}	0.037	0.048	29.73%	0.042	13.51%	0.041	10.81%
S_{20}	0.037	0.053	43.24%	0.040	8.11%	0.041	10.81%
Sensitivity indices	References	PCE	Relative error	LRA	Relative error	SVR	Relative error
S_1	0.051	0.052	1.96%	0.042	17.65%	0.055	7.84%
S_2	0.051	0.048	5.88%	0.037	27.45%	0.051	0.00%
S_3	0.014	0.014	0.00%	0.009	35.71%	0.015	7.14%
S_4	0.051	0.053	3.92%	0.040	21.57%	0.051	0.00%
S_5	0.014	0.013	7.14%	0.009	35.71%	0.017	21.43%
S_6	0.051	0.051	0.00%	0.039	23.53%	0.052	1.96%
S_7	0.014	0.015	7.14%	0.007	50.00%	0.013	7.14%
S_8	0.051	0.054	5.88%	0.055	7.84%	0.057	11.76%
S_9	0.051	0.054	5.88%	0.043	15.69%	0.053	3.92%
S_{10}	0.050	0.050	0.00%	0.049	2.00%	0.053	6.00%
S_{11}	0.037	0.042	13.51%	0.032	13.51%	0.043	16.22%
S_{12}	0.037	0.038	2.70%	0.028	24.32%	0.041	10.81%
S_{13}	0.037	0.039	5.41%	0.029	21.62%	0.042	13.51%
S_{14}	0.037	0.040	8.11%	0.035	5.41%	0.045	21.62%
S_{15}	0.037	0.042	13.51%	0.032	13.51%	0.041	10.81%
S_{16}	0.037	0.036	2.70%	0.026	29.73%	0.040	8.11%
S_{17}	0.037	0.040	8.11%	0.031	16.22%	0.040	8.11%
S_{18}	0.037	0.040	8.11%	0.026	29.73%	0.047	27.03%
S_{19}	0.037	0.040	8.11%	0.043	16.22%	0.040	8.11%
S_{20}	0.037	0.041	10.81%	0.023	37.84%	0.046	24.32%

large data sets. In addition, as one of the most popular machine learning technique, SVR is supported by many machine learning libraries and can be accelerated by specialized hardware.

RBF provides desired approximation for nearly all the smooth benchmarks. It is relatively straightforward, and no parameters need to be specified by a user. It can be used for low-medium dimensional and highly nonlinear problem. Since there is only a single parameter that needs to be tuned in the training process, RBF can handle large number of training samples efficiently. A disadvantage of RBF regression is the large and usually ill-conditioned Gram matrix.

Table 9 Packages and software of surrogate models and GSA

Name	Source	Description
UQLab	Ref. (Marelli and Sudret 2014)	A free Matlab-based toolbox for uncertainty quantification, include PCE, SVR, Kriging, LRA and PC-Kriging meta-models, and the Sobols indices computed by PCE and LRA meta-models are also provided.
GUI-HDMR	Ref. (Ziehn and Tomlin 2009)	A software tool for variance-based sensitivity indices, where the Sobols indices are computed with ANOVA-HDMR meta-model.
SDP	Ref. (Ratto et al. 2007)	A Matlab package for constructing SDP meta-model and computing Sobols indices.
UQTk	Ref. (Debusschere et al. 2015)	A toolbox for uncertainty quantification, where the PCE meta-model is provided to compute Sobols indices.
SIMLAB	Ref. (Tarantola and Becker 2016)	A free framework for global sensitivity analysis, include various sampling strategies for estimating Sobols indices.
OpenTURNS	Ref. (Dutfoy et al. 2009)	An open-source software for uncertainty quantification, where the PCE meta-model is provided to compute Sobols indices.
OpenCossan	Ref. (Patelli et al. 2014)	A multidisciplinary software for uncertainty quantification and risk analysis, include PCE and Kriging meta-model.
UQ-PyL	Ref. (Wang et al. 2016)	A GUI software for uncertainty quantification, include polynomial, Kriging, support vector machine meta-models.
dace	Ref. (Sn Lophaven and Søndergaard 2002)	A Matlab package of Kriging meta-model.
ooDACE	Ref. (Couckuyt et al. 2014)	A Matlab toolbox that includes Kriging and other variants (Blind Kriging, Co-Kriging, etc).
Libsvm	Ref. (Chang and Lin 2011)	A package of support vector machine and support vector regression.

8 Conclusions, recommendations, and prospects

This paper provides an overview of various surrogate model techniques in the field of UQ, including polynomial regression model, HDMR, SDP, PCE and their variants, Kriging and their enhanced versions, SVR, RBF regression, and LRA. Their applications for variance-based global sensitivity analysis are emphasized. The sampling strategy for constructing surrogate model and meta-model accuracy validation scheme are briefly discussed. The accuracy and efficiency of these meta-model methods are compared using four test functions, and the strengths and weaknesses of each method are discussed in depth.

The following remarks summarize some suggestions and future research directions of surrogate model techniques for UQ.

The choice of the surrogate model should be based on the problem at hands. Overall, PCE, Kriging and SVR are the most widely used meta-model techniques in the field UQ due to their excellent performance. However, no single type of surrogate model outperforms all other types for different problems. Recent researches have suggested that mixtures of different surrogate models in a weighted average way may provide better prediction than single one. Actually, blind Kriging and PC-Kriging can be viewed as the mixtures of polynomial with Kriging; thus, they usually yield more desirable prediction than single meta-model. Use of multiple surrogate models provides the possibility to emphasize more on good surrogates and put less emphasis on bad surrogates, and this idea could provide more robust predictor as well as identify regions with large prediction risk.

Considering the problem of "curse of dimensionality," future developments and deepening should focus on dimensionality reduction techniques, such as active subspace method (Constantine et al. 2013; Tripathy et al. 2016) , partial least square scheme (Bouhlel et al. 2016), or sliced inverse regression technique (Li et al. 2016b; Wu 2008; Yeh et al. 2009; Li 1991; Pan and Dias 2017; Li and Nachtsheim 2006; Lin et al. 2016). These approaches project the original high-dimensional inputs onto a low-dimensional subspace that captures most of the variation information contained in the original input; thus, it is possible to develop a surrogate model in the low-dimensional subspace to predict a high-dimensional function efficiently. The recent advances of deep learning algorithm (Zhu and Zabaras 2018; Yeo and Melnyk 2019; Tripathy and Bilionis 2018; Sirignano and Spiliopoulos 2018; Pathirage et al. 2018), such as deep neural network, offer a promising research direction for tackling the problem of "curse of dimensionality." By representing the high-dimensional functions through a hierarchy of features with increasing complexity, one can learn a surrogate model through a series of nonlinear projections of the high-dimensional inputs into the latent space to automatically extract their features. The powerful non-linear approximation capabilities of deep learning algorithm allow one to approximate an underlying high-dimensional function of arbitrary complexity. However, these algorithms usually involve time-consuming iterative training to tune hundreds of hyper-parameters. More importantly, deep learning algorithm performance

is sensitive to the hyper-parameters, and thus requires an appropriate and typically time-consuming calibration procedure. Therefore, these promising research orientations still should be brought to full maturity and should prove to be applicable to complex UQ problems.

Many surrogate model packages and software have been well developed in multiple disciplines. For ease of application, we summarize these packages and software in Table 9 with details.

Funding information This work was supported by the National Natural Science Foundation of China (Grant No. NSFC 51775439), National Science and Technology Major Project (Grant No. 2017-IV-0009-0046), and "Innovation Foundation for Doctor Dissertation of Northwestern Polytechnical University" with project code of CX201933.

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

Abdallah I, Lataniotis C, Sudret B, (2018) Parametric hierarchical kriging for multi-fidelity aero-servo-elastic simulators—application to extreme loads on wind turbines. Probabilistic Engineering Mechanics

Abraham S et al (2017) A robust and efficient stepwise regression method for building sparse polynomial chaos expansions. J Comput Phys 332:461–474

Ahlfeld R, Belkouchi B, Montomoli F (2016) SAMBA: sparse approximation of moment-based arbitrary polynomial chaos. J Comput Phys 320:1–16

AIAA (2009) Efficient uncertainty quantification using gradient-enhanced Kriging

Alış ÖF, Rabitz H (2001) Efficient implementation of high dimensional model representations. J Math Chem 29(2):127–142

Amouzgar K, Strömberg N (2017) Radial basis functions as surrogate models with a priori bias in comparison with a posteriori bias. Struct Multidiscip Optim 55(4):1453–1469

Amouzgar K, Bandaru S, Ng AHC (2018) Radial basis functions with a priori bias as surrogate models: a comparative study. Eng Appl Artif Intell 71:28–44

Ankenman B, Nelson BL, Staum J (2010) Stochastic Kriging for simulation metamodeling. Oper Res 58(2):371–382

Barton RR, Meckesheimer M (2006) Chapter 18 Metamodel-based simulation optimization, in Handbooks in Operations Research and Management Science, Henderson SG and Nelson BL, Editors, Elsevier. p. 535–574

Bhosekar A, Ierapetritou M (2018) Advances in surrogate based modeling, feasibility analysis, and optimization: a review. Comput Chem Eng 108:250–267

Blatman G, Sudret B (2010a) Efficient computation of global sensitivity indices using sparse polynomial chaos expansions. Reliab Eng Syst Saf 95(11):1216–1229

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

Blatman G, Sudret B (2011) Adaptive sparse polynomial chaos expansion based on least angle regression. J Comput Phys 230(6):2345–2367

Bouhlel MA, Martins JRRA (2018) Gradient-enhanced kriging for high-dimensional problems. Eng Comput, p. 1–17

Bouhlel MA et al (2016) Improving kriging surrogates of highdimensional design models by partial least squares dimension reduction. Struct Multidiscip Optim 53(5):935–952

Bourinet JM (2016) Rare-event probability estimation with adaptive support vector regression surrogates. Reliab Eng Syst Saf 150:210–221

Bourinet JM, Deheeger F, Lemaire M (2011) Assessing small failure probabilities by combined subset simulation and support vector machines. Struct Saf 33(6):343–353

Bryson DE, Rumpfkeil MP (2017) All-at-once approach to multifidelity polynomial chaos expansion surrogate modeling. Aerosp Sci Technol 70:121–136

Bucher CG, Bourgund U (1990) A fast and efficient response surface approach for structural reliability problems. Struct Saf 7(1):57–66

Cao Y et al (2002) Adjoint sensitivity analysis for differential-algebraic equations: the Adjoint DAE system and its numerical solution. SIAM J Sci Comput 24(3):1076–1089

Chang CC, Lin CJ (2011) LIBSVM: a library for support vector machines. 2(3): p. 1–27

Chen W, Jin R (2004) Analytical variance-based global sensitivity analysis in simulation-based design under uncertainty. J Mech Des 127(5):953–962

Chen X, Kim KK (2014) Stochastic kriging with biased sample estimates. Acm Transactions on Modeling & Computer Simulation 24(2):1–23

Chen X, Wang K, Yang F (2013) Stochastic kriging with qualitative factors. Winter Simulation Conference

Cheng K, Lu Z (2018a) Adaptive sparse polynomial chaos expansions for global sensitivity analysis based on support vector regression. Comput Struct 194:86–96

Cheng K, Lu Z (2018b) Sparse polynomial chaos expansion based on DMORPH regression. Appl Math Comput 323:17–13

Cheng K et al (2017a) Global sensitivity analysis using support vector regression. Appl Math Model

Cheng K et al (2017b) Mixed kernel function support vector regression for global sensitivity analysis. Mech Syst Signal Process 96:201–214

Cheng K, Lu Z, Zhen Y (2019) Multi-level multi-fidelity sparse polynomial chaos expansion based on Gaussian process regression. Comput Methods Appl Mech Eng

Chevreuil M et al (2013) A least-squares method for sparse low rank approximation of multivariate functions. 3(1)

Chevreuil M, Rai P, Nouy A (2014) Sampling based tensor approximation method for uncertainty propagation. Icossar Org

Chowdhury R, Adhikari S (2010) High dimensional model representation for stochastic finite element analysis. ApplMathModel 34(12): 3917–3932

Constantine PG, Dow E, Wang Q (2013) Active subspace methods in theory and practice: applications to kriging surfaces. SIAM J Sci Stat Comput, 36(4)

Corveleyn S, Vandewalle S (2017) Computation of the output of a function with fuzzy inputs based on a low-rank tensor approximation. Fuzzy Sets Syst 310:74–89

Couckuyt I et al (2012) Blind Kriging: implementation and performance analysis. Adv Eng Softw 49(1):1–13

Couckuyt I, Dhaene T, Demeester P (2014) ooDACE toolbox: a flexible object-oriented Kriging implementation. J Mach Learn Res 15(1): 3183–3186

Crestaux T, Le Maıˆtre O, Martinez J-M (2009) Polynomial chaos expansion for sensitivity analysis. Reliab Eng Syst Saf 94(7):1161–1172

Davis G, Mallat S, Avellaneda M (1997) Adaptive greedy approximations. Constr Approx 13(1):57–98

Debusschere B et al (2015) Uncertainty quantification toolkit (UQTk)

Diaz P, Doostan A, Hampton J (2018) Sparse polynomial chaos expansions via compressed sensing and D-optimal design. Comput Methods Appl Mech Eng 336:640–666

Doksum K, Samarov A (1995) Nonparametric estimation of global functionals and a measure of the explanatory power of covariates in regression. Ann Stat 23(5):1443–1473

Doostan A, Validi A, Iaccarino G (2013) Non-intrusive low-rank separated approximation of high-dimensional stochastic models. Comput Methods Appl Mech Eng 263:42–55

Dutfoy A et al (2009) OpenTURNS, an Open Source initiative to Treat Uncertainties, Risks'N Statistics in a structured industrial approach. 41èmes Journées De Statistique Sfds Bordeaux

Efron B (1982) P The Jackknife, the bootstrap and other resampling plans. Siam Monograph, 38(384)

Fernándezgodino MG et al (2017) Review of multi-fidelity models

Forrester AIJ, Keane AJ (2009) Recent advances in surrogate-based optimization. Prog Aerosp Sci 45(1):50–79

Garcia-Cabrejo O, Valocchi A (2014) Global sensitivity analysis for multivariate output using polynomial chaos expansion. Reliab Eng Syst Saf 126:25–36

Garud SS, Karimi IA, Kraft M (2017) Design of computer experiments: a review. Comput Chem Eng 106:71–95

Genyuan Li A (2002) Shengwei Wang, and Herschel Rabitz, practical approaches to construct RS-HDMR component functions. J Phys Chem A 106(37):8721–8733

Ghanem R, Higdon D, Owhadi H (2016) Handbook of Uncertainty Quantification

Giles MB, Pierce NA (2000) An introduction to the adjoint approach to design. Flow Turbulence & Combustion 65(3–4):393–415

Gratiet LL (2012) Recursive co-kriging model for design of computer experiments with multiple levels of fidelity with an application to hydrodynamic. 4(5)

Gratiet LL (2013) Multi-fidelity Gaussian process regression for computer experiments

Gratiet LL (2015) Recursive co-kriging model for design of computer experiments with multiple levels of fidelity 4(5)

Guan XL, Melchers RE (2001) Effect of response surface parameter variation on structural reliability estimates. Struct Saf 23(4):429–444

Guo L, Narayan A, Zhou T (2018) A gradient enhanced ℓ1-minimization for sparse approximation of polynomial chaos expansions. J Comput Phys

Hadigol M et al (2014) Partitioned treatment of uncertainty in coupled domain problems: a separated representation approach. Comput Methods Appl Mech Eng 274(6):103–124

Haftka RT, Villanueva D, Chaudhuri A (2016) Parallel surrogateassisted global optimization with expensive functions—a survey. Struct Multidiscip Optim 54(1):3–13

Halton J (1964) Radical-inverse quasi-random point sequence [G5]. Commun ACM 7

Hampton J, Doostan A (2015) Compressive sampling of polynomial chaos expansions: convergence analysis and sampling strategies. J Comput Phys 280:363–386

Han ZH, Görtz S (2012) Hierarchical Kriging model for variable-fidelity surrogate modeling. AIAA J 50(9):1885–1896

Han Z-H, Görtz S, Zimmermann R (2013) Improving variable-fidelity surrogate modeling via gradient-enhanced kriging and a generalized hybrid bridge function. Aerosp Sci Technol 25(1):177–189

HaoW LZ, Li L (2013) A new interpretation and validation of variance based importance measures for models with correlated inputs. Comput Phys Commun 184(5):1401–1413

Hardy RL (1971) Multiquadric equations of topography and other irregular surfaces. J Geophys Res 76(8):1905–1915

Jakeman JD, EldredMS SK (2015) Enhancing ℓ1-minimization estimates of polynomial chaos expansions using basis selection. J Comput Phys 289:18–34

Jiang T, Zhou X (2018) Gradient/hessian-enhanced least square support vector regression. Inf Process Lett 134:1–8

Jin R, Chen W, Simpson TW (2001) Comparative studies of metamodelling techniques under multiple modelling criteria. Struct Multidiscip Optim 23(1):1–13

Johnson RW (2001) An introduction to the bootstrap. Teach Stat 23(2): 49–54

Johnson ME, Moore LM, Ylvisaker D (1990) Minimax and maximin distance designs ☆. J Stat Plan Inference 26(2):131–148

Kennedy MC, O'Hagan A (2000) Predicting the output from a complex computer code when fast approximations are available. Biometrika 87(1):1–13

Kersaudy P et al (2015) A new surrogate modeling technique combining Kriging and polynomial chaos expansions—application to uncertainty analysis in computational dosimetry. J Comput Phys 286: 103–117

Kohavi R (1995) A study of cross-validation and bootstrap for accuracy estimation andmodel selection. International Joint Conference on Artificial Intelligence

Konakli K, Sudret B (2016a) Global sensitivity analysis using low-rank tensor approximations. Reliab Eng Syst Saf 156:64–83

Konakli K, Sudret B (2016b) Polynomial meta-models with canonical low-rank approximations: numerical insights and comparison to sparse polynomial chaos expansions. J Comput Phys 321:1144–1169

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

Krige DG (1953) A statistical approach to some basic mine valuation problems on the Witwatersrand. OR 4(1):18–18

Lambert RSC et al (2016) Global sensitivity analysis using sparse high dimensional model representations generated by the group method of data handling. Math Comput Simul 128(C):42–54

Li KC (1991) Sliced inverse regression for dimension reduction. PublAm Stat Assoc 86(414):316–327

Li L, Nachtsheim CJ (2006) Sparse sliced inverse regression. Technometrics 48(4):503–510

Li G, Wang SW (2001) High dimensional model representations generated from low dimensional data samples. I mp-Cut-HDMR J Math Chem 30(1):1–30

Li G et al (2006) Random sampling-high dimensional model representation (RS-HDMR) and orthogonality of its different order component functions. J Phys Chem A 110(7):2474–2485

Li G et al (2008) Regularized random-sampling high dimensional model representation (RS-HDMR). J Math Chem 43(3):1207–1232

Li L, Lu Z, Zhou C (2011) Importance analysis for models with correlated input variables by the state dependent parameters method. Comput Math Appl 62(12):4547–4556

Li E, Wang H, Li G (2012a) High dimensional model representation (HDMR) coupled intelligent sampling strategy for nonlinear problems. Comput Phys Commun 183(9):1947–1955

Li L et al (2012b) Moment-independent importance measure of basic variable and its state dependent parameter solution. Struct Saf 38: 40–47

Li L, Lu Z, Chen C (2016a) Moment-independent importance measure of correlated input variable and its state dependent parameter solution. Aerosp Sci Technol 48:281–290

Li W, Lin G, Li B (2016b) Inverse regression-based uncertainty quantification algorithms for high-dimensional models: theory and practice. J Comput Phys 321:259–278

Lin Q, Zhao Z, Liu JS (2016) Sparse sliced inverse regression for high dimensional data

Liu W (2003) Development of gradient-enhanced kriging approximations for multidisciplinary design optimization. Univ of Notre Dame, Notre Dame Indiana, pp 177

Liu Y, Yousuff Hussaini M, Ökten G (2016) Accurate construction of high dimensional model representation with applications to uncertainty quantification. Reliab Eng Syst Saf 152:281–295

Liu H, Ong Y-S, Cai J (2018) A survey of adaptive sampling for global metamodeling in support of simulation-based complex engineering design. Struct Multidiscip Optim 57(1):393–416
Lucor D, Karniadakis GE (2005) Adaptive generalized polynomial chaos for nonlinear random oscillators. Soc Ind Appl Math :720–735
Luo X, LuZ XX (2014) Reproducing kernel technique for high dimensional model representations (HDMR). Comput Phys Commun 185(12):3099–3108
Ma X, Zabaras N (2010) An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. J Comput Phys 229(10):3884–3915
Majdisova Z, Skala V (2017) Radial basis function approximations: comparison and applications. Appl Math Model 51:728–743
Marelli S, Sudret B (2014) UQLab: a framework for uncertainty quantification in Matlab. Int Conf on Vulnerability, Risk Analysis and Management
Marelli S, Sudret B (2018) An active-learning algorithm that combines sparse polynomial chaos expansions and bootstrap for structural reliability analysis. Struct Saf 75:67–74
Marrel A et al (2008) Calculations of Sobol indices for the Gaussian process metamodel. Reliab Eng Syst Saf 94(3):742–751
Mathelin L, Gallivan KA (2012) A compressed sensing approach for partial differential equations with random input data. Communications in Computational Physics 12(4):919–954
Mathelin L, Gallivan KA (2015) A compressed sensing approach for partial differential equations with random input data. Communications in Computational Physics 12(4):919–954
Matheron G (1963) Principles of geostatistics. Econ Geol 58(8):1246–1266
Mavriplis D (2013) A discrete adjoint-based approach for optimization problems on three-dimensional unstructured meshes
Mckay MD, Beckman RJ, Conover WJ (2000) A comparison of three methods for selecting values of input variables in the analysis of output from a computer code. Technometrics 21(2):239–245
Meczekalski B, Podfigurna-Stopa A (2000) Global optimization of costly nonconvex functions using radial basis functions. Optim Eng 1(4):373–397
Ng WT, Eldred M (2012) Multifidelity uncertainty quantification using non-intrusive polynomial chaos and stochastic collocation
Ng LWT, Willcox KE (2015) Multifidelity approaches for optimization under uncertainty. Int J Numer Methods Eng 100(10):746–772
Oakley JE, O'Hagan A (2004) Probabilistic sensitivity analysis of complex models: a Bayesian approach. J R Stat Soc 66(3):751–769
Owen A (1992) Orthogonal arrays for computer experiments, integration and visualization. Stat Sin 2(2):439–452
Palar PS, Shimoyama K (2017) Multi-fidelity uncertainty analysis in CFD using hierarchical Kriging. Aiaa Applied Aerodynamics Conference
Palar PS, Tsuchiya T, Parks GT (2016) Multi-fidelity non-intrusive polynomial chaos based on regression. Comput Methods Appl Mech Eng 305:579–606
Palar PS et al (2018) Global sensitivity analysis via multi-fidelity polynomial chaos expansion. Reliab Eng Syst Saf 170:175–190
Pan Q, Dias D (2017) Sliced inverse regression-based sparse polynomial chaos expansions for reliability analysis in high dimensions. Reliab Eng Syst Saf 167:484–493
Parussini L et al (2017) Multi-fidelity Gaussian process regression for prediction of random fields. J Comput Phys 336(C):36–50
Patelli E et al (2014) OpenCossan: an efficient open tool for dealing with epistemic and aleatory uncertainties. International Conference on Vulnerability and Risk Analysis and Management
Pathirage CSN et al (2018) Structural damage identification based on autoencoder neural networks and deep learning. Eng Struct 172:13–28
Pati YC, Rezaiifar R, Krishnaprasad PS (1993) Orthogonal matching pursuit: recursive function approximation with applications to wavelet decomposition. Proceedings of 27th Asilomar Conference on Signals, Systems and Computers
Peherstorfer B et al (2016) Multifidelity importance sampling. Comput Methods Appl Mech Eng 300:490–509
Peng J, Hampton J, Doostan A (2014) A weighted ℓ1-minimization approach for sparse polynomial chaos expansions. J Comput Phys 267:92–111
Peng J, Hampton J, Doostan A (2016) On polynomial chaos expansion via gradient-enhanced ℓ 1 -minimization. J Comput Phys 310(C):440–458
Picheny V et al (2012) Adaptive designs of experiments for accurate approximation of a target region. J Mech Des 132(7):461–471
Queipo NV et al (2005) Surrogate-based analysis and optimization. Prog Aerosp Sci 41(1):1–28
Rabitz H, Aliş ÖF (1999) General foundations of high-dimensional model representations. J Math Chem 25(2–3):197–233
Rabitz H et al (1999) Efficient input—output model representations. Comput Phys Commun 117(1–2):11–20
Rasmussen C, Williams C (2006) Gaussian processes for machine learning. MIT Press :69–106
Ratto M, Pagano A, Young P (2007) State dependent parameter metamodelling and sensitivity analysis. Comput Phys Commun 177(11):863–876
Ratto M, Pagano A, Young PC (2009) Non-parametric estimation of conditional moments for sensitivity analysis. Reliab Eng Syst Saf 94(2):237–243
Razavi S, Tolson BA, Burn DH (2012) Review of surrogate modeling in water resources. Water Resour Res 48(7):7401
Roshan V, Ying H, Sudjianto A (2008) Blind Kriging: a new method for developing metamodels. JMech Des 130(3):350–353
Salehi S et al (2017) Efficient uncertainty quantification of stochastic CFD problems using sparse polynomial chaos and compressed sensing. Comput Fluids 154:296–321
Salehi S et al (2018) An efficient multifidelity ℓ1-minimization method for sparse polynomial chaos. Comput Methods Appl Mech Eng 334:183–207
Saltelli A (2002) Making best use of model evaluations to compute sensitivity indices. Comput Phys Commun 145(2):280–297
Saltelli A (2008) Global sensitivity analysis : the primer, John Wiley
Saltelli A et al (2010) Variance based sensitivity analysis of model output. Design and estimator for the total sensitivity index. Comput Phys Commun 181(2):259–270
Schöbi R, Sudret B (2017) Uncertainty propagation of p-boxes using sparse polynomial chaos expansions. J Comput Phys 339:307–327
Shao Q et al (2017) Bayesian sparse polynomial chaos expansion for global sensitivity analysis. Comput Methods Appl Mech Eng 318:474–496
Shlomo M, Shaul M (2011) Applications of Monte CarloMethods in Science and Engineering
Sirignano J, Spiliopoulos K (2018) DGM: a deep learning algorithm for solving partial differential equations. J Comput Phys 375:1339–1364
Sn Lophaven HN, Søndergaard J (2002) DACE –AMATLABKriging Toolbox – Version 2.0
Sobol IM (1967) On the distribution of points in a cube and the approximate evaluation of integrals. USSR Comput Math Math Phys 7(4):86–112
Sobol IM (1993) Sensitivity estimates for nonlinear mathematical models. Math model comput exp 1(1):112–118
Sobol IM (2003) Theorems and examples on high dimensional model representation. Reliab Eng Syst Saf 79(2):187–193
Sobol IM (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Elsevier Science Publishers B. V. 271–280
Song S, Wang L (2017) Modified GMDH-NN algorithm and its application for global sensitivity analysis. J Comput Phys 348

Staum J (2009) Better Simulation metamodeling: the why, what, and how of stochastic kriging. Simulation Conference
Sudret B (2008) Global sensitivity analysis using polynomial chaos expansions. Reliab Eng Syst Saf 93(7):964–979
Sudret B, Mai CV (2015) Computing derivative-based global sensitivity measures using polynomial chaos expansions. Reliab Eng Syst Saf 134:241–250
Sudret B, Marelli S, Wiart J (2017) Surrogate models for uncertainty quantification: an overview. European Conference on Antennas and Propagation
Tang K, Congedo PM, Abgrall R (2016) Adaptive surrogate modeling by ANOVA and sparse polynomial dimensional decomposition for global sensitivity analysis in fluid simulation. J Comput Phys 314(3):557–589
Tang K et al (2018) An adaptive least-squares global sensitivity method and application to a plasma-coupled combustion prediction with parametric correlation. J Comput Phys 361:167–198
Tarantola S, Becker W (2016) SIMLAB Software for uncertainty and sensitivity analysis
Tripathy RK, Bilionis I (2018) Deep UQ: learning deep neural network surrogate models for high dimensional uncertainty quantification. J Comput Phys 375:565–588
Tripathy R, Bilionis I, Gonzalez M (2016) Gaussian processes with builtin dimensionality reduction: applications to high-dimensional uncertainty propagation. J Comput Phys 321:191–223
Ulaganathan S et al (2015) Performance study of multi-fidelity gradient enhanced kriging. Struct Multidiscip Optim 51(5):1–17
Ulaganathan S et al (2016) High dimensional Kriging metamodelling utilising gradient information. Appl Math Model 40(9):5256–5270
Validi AA (2014) Low-rank separated representation surrogates of highdimensional stochastic functions: application in Bayesian inference. J Comput Phys 260(2):37–53
Van Steenkiste T et al (2018) Sequential sensitivity analysis of expensive black-box simulators with metamodelling. Appl Math Model 61: 668–681
Vapnik VN (2008) Statistical learning theory. Encyclopedia of the Sciences of Learning 41(4):3185–3185
Vapnik VN (1997) The nature of statistical learning theory. IEEE Trans Neural Netw 38(4):409–409
Wan X, Karniadakis GE (2005) An adaptive multi-element generalized polynomial chaos method for stochastic differential equations. J Comput Phys 209(2):617–642
Wang GG, Shan S (2007) Review of metamodeling techniques in support of engineering design optimization. ASME 2006 International Design Engineering Technical Conferences and Computers and Information in Engineering Conference
Wang SW et al (2003) Random sampling-high dimensional model representation (RS-HDMR) with nonuniformly distributed variables: application to an integrated multimedia/multipathway exposure and dose model for trichloroethylene. J Phys Chem A 107(23):4707–4716
Wang D et al (2014a) A CAD/CAE integrated framework for structural design optimization using sequential approximation optimization. Adv Eng Softw 76(3):56–68
Wang D et al (2014b) Structural design employing a sequential approximation optimization approach. Comput Struct 134(4):75–87
Wang C et al (2016) A GUI platform for uncertainty quantification of complex dynamical models. EnvironModel Softw 76:1–12
Wei P, Lu Z, Song J (2015) Variable importance analysis: a comprehensive review. Reliab Eng Syst Saf 142:399–432
Wiener N (1938) The homogeneous chaos. Am J Math 60(4):897–936
Wu HM (2008) Kernel sliced inverse regression with applications to classification. J Comput Graph Stat 17(3):590–610
Wu Z et al (2016a) Global sensitivity analysis using a Gaussian radial basis function metamodel. Reliab Eng Syst Saf 154:171–179
Wu Z et al (2016b) Unified estimate of Gaussian kernel width for surrogate models. Neurocomputing 203:41–51
Wu Z et al (2016c) Global sensitivity analysis using a Gaussian radial basis function metamodel. Reliab Eng Syst Saf 154:171–179
Xiao S, Lu Z (2017) Structural reliability sensitivity analysis based on classification of model output. Aerosp Sci Technol 71
Xiao M et al (2018) Extended co-Kriging interpolation method based on multi-fidelity data. Appl Math Comput 323:120–131
Xie W, Nelson BL, Staum J (2010) The influence of correlation functions on stochastic kriging metamodels. Simulation Conference
Xiu D, Karniadakis GE (2002) TheWiener–Askey polynomial chaos for stochastic differential equations. Siam J Sci Comput
Yan L et al (2018) Gaussian processes and polynomial chaos expansion for regression problem: linkage via the RKHS and comparison via the KL divergence. Entropy 20(3):191
Yang X, Karniadakis GE (2013) Reweighted ℓ1 minimization method for stochastic elliptic differential equations. J Comput Phys 248:87–108
Yeh YR, Huang SY, Lee YJ (2009) Nonlinear dimension reduction with kernel sliced inverse regression. IEEE Transactions on Knowledge & Data Engineering 21(11):1590–1603
Yeo K, Melnyk I (2019) Deep learning algorithm for data-driven simulation of noisy dynamical system. J Comput Phys 376:1212–1231
Yondo R, Andrés E, Valero E (2018) A review on design of experiments and surrogate models in aircraft real-time and many-query aerodynamic analyses. Prog Aerosp Sci 96:23–61
Youn BD, Choi KK (2004) A new response surface methodology for reliability-based design optimization. Comput Struct 82(2):241–256
Young P (1993) Time variable and state dependent modelling of nonstationary and nonlinear time series. Developments in Time
Young P (2000) Stochastic, dynamic modelling and signal processing: time variable and state dependent parameter estimation. Cambridge University Press, Cambridge, pp 74–114
Young P, McKenna P, Bruun J (2001) Identification of non-linear stochastic systems by state dependent parameter estimation. Int J Control 74(18):1837–1857
Zhang K et al (2017a) Analytical variance based global sensitivity analysis for models with correlated variables. Appl Math Model 45:748–767
Zhang J et al (2017b) Estimation of the Pareto front in stochastic simulation through stochastic Kriging. Simul Model Pract Theory 79:69–86
Zhou X, Jiang T (2018) An effective way to integrate ε-support vector regression with gradients. Expert Syst Appl 99:126–140
Zhou Q et al (2015) An adaptive global variable fidelity metamodeling strategy using a support vector regression based scaling function. Simul Model Pract Theory 59:18–35
Zhu Y, Zabaras N (2018) Bayesian deep convolutional encoder–decoder networks for surrogate modeling and uncertainty quantification. J Comput Phys 366:415–447
Ziehn T, Tomlin AS (2008) Global sensitivity analysis of a 3D street canyon model—part I: the development of high dimensional model representations. Atmos Environ 42(8):1857–1873
Ziehn T, Tomlin AS (2009) GUI–HDMR—a software tool for global sensitivity analysis of complex models. Environ Model Softw 24(7):775–785
Zuniga MM, Kucherenko S, Shah N (2013) Metamodelling with independent and dependent inputs. Comput Phys Commun 184(6): 1570–1580