Bayesian sparse polynomial chaos expansion for global sensitivity analysis*

Qian Shao[a,b], Anis Younes[c,d,e], Marwan Fahs[c], Thierry A. Mara[a]

[a] *PIMENT, EA 4518, Université de La Réunion, FST, 15 Avenue René Cassin, 97715 Saint-Denis, Réunion*
[b] *School of Civil Engineering, Wuhan University, 8 South Road of East Lake, Wuchang, 430072 Wuhan, PR China*
[c] *LHyGeS, UMR-CNRS 7517, Université de Strasbourg/EOST, 1 rue Blessig, 67084 Strasbourg, France*
[d] *IRD UMR LISAH, F-92761 Montpellier, France*
[e] *LMHE, Ecole Nationale d'Ingénieurs de Tunis, Tunisie*

Abstract

Polynomial chaos expansions are frequently used by engineers and modellers for uncertainty and sensitivity analyses of computer models. They allow representing the input/output relations of computer models. Usually only a few terms are really relevant in such a representation. It is a challenge to infer the best sparse polynomial chaos expansion of a given model input/output data set. In the present article, sparse polynomial chaos expansions are investigated for global sensitivity analysis of computer model responses. A new Bayesian approach is proposed to perform this task, based on the Kashyap information criterion for model selection. The efficiency of the proposed algorithm is assessed on several benchmarks before applying the algorithm to identify the most relevant inputs of a double-diffusive convection model.

keywords: Global sensitivity analysis, Sobol' indices, Sparse polynomial chaos expansion, Bayesian model averaging, Kashyap information criterion, Double-diffusive convection

*email: shaoq88@gmail.com, younes@unistra.fr, fahs@unistra.fr, mara@univ-reunion.fr

*Q. Shao, A. younes, M. Fahs, T.A. Mara (2017), *Bayesian sparse polynomial chaos expansion for global sensitivity analysis*, Computer Methods in Applied Mechanics and Engineering, 318, 474–496, doi:10.1016/j.cma.2017.01.033

Contents

1 Introduction **3**

2 Sobol' decomposition **4**

3 Polynomial chaos representation of the model response **6**
3.1 Full PC expansion 6
3.2 Computing the PC coefficients 7
3.3 Sparse polynomial chaos expansion 8
3.4 PC-based global sensitivity indices 8

4 The Bayesian Model Averaging Framework **9**

5 Bayesian sparse PCE **11**
5.1 Our working assumptions 11
5.2 Post-processing 13
5.3 The proposed algorithm 14

6 Synthetic mathematical examples **16**
6.1 Ishigami function 16
6.2 Sobol' function 18
6.3 Morris function 20

7 Application to double diffusive convection in porous media **21**
7.1 Problem statement 21
7.2 Sensitivity analysis 23

8 Conclusion **24**

1. Introduction

Mathematical models are widely used in many scientific disciplines to explain and understand the observed real world. Their translation into computer models allows studying different scenarios by exploring the input space. However, the use of computer models for specific applications is usually hampered by the inherent uncertainties about the input values and the model itself. Hence, good modelling practice requires that uncertainties be acknowledged and taken into account by modellers. Uncertainty and sensitivity analyses should be routinely implemented both in the modelling process and in the operational use of the model [1].

For this purpose, polynomial chaos expansion (PCE) has received much attention over the last two decades (*e.g.,* [2, 3, 4, 5, 6, 7, 8, 9]). In computer models in engineering, PCE has been proven useful for the analyses of uncertainty, sensitivity and risk of failure. A 'good' PCE representation contains all the salient features of the model response within the input space in which it has been built. The challenge is to obtain such a 'good' representation. There are typically two approaches to building a PCE: *intrusive* and *non-intrusive*. The *intrusive* approach requires modifying the computer model, whereas the *non-intrusive* approach only needs input/output samples. In this paper, we discuss *non-intrusive* approaches.

The *non-intrusive* computation of the PCE coefficients is often conducted with one of two methods, projection or regression. The latter is efficient when dealing with a moderate number of input variables. However, with a large number of input variables or a high-degree polynomial, the number of coefficients increases dramatically. In that case, a large number of model evaluations is required to compute the overall set of PCE coefficients (see [7]). Moreover, a large number of coefficients poses the problem of overfitting when regression-based methods are employed. To circumvent this problem, one needs to decrease the number of coefficients in the PCE. To this end, several approaches have been developed to construct a sparse PCE, where only basis functions and coefficients that make significant contributions to the model response of interest are retained. As far as our knowledge extends, the original idea of a sparse PCE came from Blatman and Sudret [10, 11, 12], where they developed an iterative forward–backward algorithm to construct a sparse PCE based on a stepwise regression technique. Later, a least angle regression algorithm was proposed by [13]. Hu and Youn [14] presented a sparse iterative scheme using the projection technique. More recently, Fajraoui et al. [15] developed a simple strategy to construct a sparse PCE from a heuristic rule.

In the present paper, a new algorithm based on Bayesian model averag-

ing (BMA) is proposed to construct sparse PCEs. BMA, relying on Bayes' theorem, is a well known statistical approach to perform quantitative comparisons of competing models [16, 17]. The difficulty of BMA lies in the evaluation of a quantity referred to as the 'Bayesian model evidence' (BME), which involves an integral over the whole input space, so it generally has no analytical expression. The Kashyap information criterion (KIC) was derived from BMA under some assumptions regarding the posterior probability distribution [18]. In the present paper, the KIC is employed to select the best sparse PCE for a given input/output sample.

The sparse PCE is employed in the present paper for the global sensitivity analysis of computer model responses. For this purpose, variance-based sensitivity indices are of interest. These sensitivity indices (also called Sobol' indices) are defined in Section 2. In Section 3, the polynomial chaos representation of a multi-dimensional function is recalled. In particular, it is given in detail how to determine the Sobol' indices from the PCE coefficients. Then, in Section 4, the BMA and KIC are defined before proposing, in § 5, our algorithm for inferring the optimal sparse PCE from a given data set. The performance of the proposed algorithm is assessed on different well-known benchmarks for global sensitivity analysis in Section 6. Finally, an application to a porous medium is proposed in Section 7, before concluding (§ 8).

2. Sobol' decomposition

Let us consider a mathematical model $Y = \mathcal{M}(\boldsymbol{X})$ having an independent input vector $\boldsymbol{X} = (X_1, \ldots, X_n)^T$ and a scalar output Y. We denote by $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)^T$ the isoprobabilistic transformed vector of $\boldsymbol{X}$, namely,

$$\begin{cases} x_1 = F_1(X_1) \\ \quad \vdots \\ x_n = F_n(X_n) \end{cases} \tag{1}$$

where F_i is the cumulative distribution function of X_i, that is, $X_i \sim p_i(X_i) = \mathrm{d}F_i(X_i)/\mathrm{d}X_i$. Such a transformation is convenient because the input vector $\boldsymbol{x}$ contains independent random parameters uniformly distributed in the n-dimensional unit hypercube $\mathbb{K}^n$. In the sequel, we assume that Y is square integrable, that is, $Y \in \mathcal{L}^2$.

The Sobol' decomposition represents any square-integrable function $\mathcal{M}(\boldsymbol{X})$ into a sum of terms of increasing dimensions:

$$\mathcal{M}(\boldsymbol{X}) \equiv \mathcal{M}_0 + \sum_{i_1=1}^{n} \mathcal{M}_{i_1}(x_{i_1}) + \sum_{i_2>i_1}^{n} \mathcal{M}_{i_1 i_2}(x_{i_1}, x_{i_2}) + \cdots + \mathcal{M}_{12\ldots n}(\boldsymbol{x}) \tag{2}$$

such that

$$\int_0^1 \mathcal{M}_{i_1\dots i_s}(x_{i_1},\dots,x_{i_s})\mathrm{d}x_{i_k} = 0 \quad \text{if} \quad k \in \{1,\dots,s\} \tag{3}$$

On the one hand, Eq. (3) ensures the uniqueness of the decomposition Eq. (2), and on the other hand, ensures the pairwise orthogonality of the summands in the following sense:

$$\int_{\mathbb{K}^n} \mathcal{M}_{i_1\dots i_s}(x_{i_1},\dots,x_{i_s})\mathcal{M}_{j_1\dots j_t}(x_{j_1},\dots,x_{j_t})\mathrm{d}\boldsymbol{x} = 0 \quad \text{for} \quad \{i_1\dots i_s\} \neq \{j_1,\dots,j_t\} \tag{4}$$

where $\mathrm{d}\boldsymbol{x} = \mathrm{d}x_1\dots\mathrm{d}x_n$ for the sake of simplicity.

Moreover, with the above properties, each term in Eq. (2) can be derived analytically. For example, the constant term, the univariate term, and the bivariate terms can be written, respectively, as follows:

$$\mathcal{M}_0 = \mathbb{E}[\mathcal{M}(\boldsymbol{X})] \equiv \int_{\mathbb{K}^n} \mathcal{M}(\boldsymbol{X})\mathrm{d}\boldsymbol{x} \tag{5}$$

$$\mathcal{M}_{i_1}(x_{i_1}) = \int_{\mathbb{K}^{n-1}} \mathcal{M}(\boldsymbol{X})\mathrm{d}\boldsymbol{x}_{\sim i_1} - \mathcal{M}_0 \tag{6}$$

$$\mathcal{M}_{i_1 i_2}(x_{i_1},x_{i_2}) = \int_{\mathbb{K}^{n-2}} \mathcal{M}(\boldsymbol{X})\mathrm{d}\boldsymbol{x}_{\sim\{i_1,i_2\}} - \mathcal{M}_{i_1}(x_{i_1}) - \mathcal{M}_{i_2}(x_{i_2}) - \mathcal{M}_0 \tag{7}$$

In these expressions, $\int_{\mathbb{K}^{n-1}} \mathrm{d}\boldsymbol{x}_{\sim i_1}$ denotes the integration over all variables except x_{i_1}. Similarly, $\int_{\mathbb{K}^{n-2}} \mathrm{d}\boldsymbol{x}_{\sim\{i_1,i_2\}}$ denotes the integration over all parameters except x_{i_1} and x_{i_2}.

As $\boldsymbol{X}$ is a vector of random variables, the model response $Y = \mathcal{M}(\boldsymbol{X})$ is also a random variable, with variance D:

$$D = \mathbb{V}[\mathcal{M}(\boldsymbol{X})] \equiv \int_{\mathbb{K}^n} \mathcal{M}^2(\boldsymbol{X})\mathrm{d}\boldsymbol{x} - \mathcal{M}_0^2 \tag{8}$$

Due to the orthogonality property in Eq. (4), the total variance Eq. (8) can be decomposed as follows:

$$D = \sum_{i_1=1}^{n} D_{i_1} + \sum_{i_2>i_1}^{n} D_{i_1 i_2} + \dots + D_{12\dots n} \tag{9}$$

where $D_{i_1\dots i_s}$ is the partial variance:

$$D_{i_1\dots i_s} = \int_{\mathbb{K}^s} \mathcal{M}^2_{i_1\dots i_s}(x_{i_1},\dots,x_{i_s})dx_{i_1}\dots,dx_{i_s} \tag{10}$$

Thereby, the partial sensitivity indices (Sobol' indices) due to the cooperative effect of the input random variables $\{x_{i_1}, \ldots, x_{i_s}\}$ can be defined in the following form:

$$S_{i_1 \ldots i_s} = \frac{D_{i_1 \ldots i_s}}{D} \in [0, 1] \tag{11}$$

Hence, the first-order sensitivity index S_i represents the amount of variance of the model response due to x_i alone. The higher S_i, the more Y is sensitive to the variable x_i. S_{ij} measures the amount of variance of Y due to the cooperative effect (also called the interaction) of x_i and x_j. To further evaluate the whole contribution of x_i to the variance of Y, the total sensitivity index S_i^T is introduced [19]:

$$S_i^T = \sum_{\boldsymbol{u}: i \in \boldsymbol{u}} S_{\boldsymbol{u}} \tag{12}$$

There have been a plethora of methods proposed in the literature to assess the Sobol' indices. They can be classified as spectral methods [7, 20, 21, 22], non-parametric methods [23, 24, 25, 26], emulator-based methods [27, 28, 29], among others. In the present work, we focus on the spectral method called the polynomial chaos (PC) expansion. In this approach, the model response is cast onto an orthonormal polynomial basis of $\mathcal{L}^2$. Thanks to the nature of the polynomial chaos basis, the sensitivity indices can be computed simply, as analytical functions of the PC coefficients [7].

3. Polynomial chaos representation of the model response

3.1. Full PC expansion

The model response can be expanded as follows in terms of a polynomial basis:

$$Y = \mathcal{M}(\boldsymbol{X}) \equiv \sum_{\boldsymbol{\alpha} \in \mathbb{N}^n} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \tag{13}$$

where $\boldsymbol{\alpha} = \alpha_1 \ldots \alpha_n$ (with $\alpha_i \geqslant 0$) is an n-dimensional index, and the $a_{\boldsymbol{\alpha}}$'s are the PC coefficients.

The multidimensional polynomial $\psi_{\alpha_1 \ldots \alpha_n}$ is the tensor product of univariate *standardized* shifted-Legendre polynomials:

$$\psi_{\alpha_1 \ldots \alpha_n}(\boldsymbol{x}) = \prod_{i=1}^{n} \psi_{\alpha_i}(x_i) \tag{14}$$

We recall that the first shifted *standardized* Legendre polynomials are: $\psi_0 = 1$, $\psi_1(x) = \sqrt{3}(2x - 1)$, $\psi_2(x) = \frac{3\sqrt{5}}{2}(2x - 1)^2 - \frac{\sqrt{5}}{2}$ and so on.

Eq. (13) is usually referred to as the polynomial chaos expansion (PCE) of Y. For computational purposes, the PCE is usually truncated to retain only

a finite number of terms. One commonly retains those polynomials whose total degree $|\boldsymbol{\alpha}| \equiv \sum_{i=1}^{n} \alpha_i$ does not exceed a given degree p:

$$Y \simeq \mathcal{M}_p(\boldsymbol{x}) \equiv \sum_{\boldsymbol{\alpha} \in \mathcal{A}^{p,n}} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}), \quad \mathcal{A}^{p,n} \equiv \{\boldsymbol{\alpha} \in \mathbb{N}^n : |\boldsymbol{\alpha}| \leq p\} \tag{15}$$

With such a truncation, the problem of characterizing the random response Y is reduced to evaluating a finite set of unknown coefficients. The total number of unknown coefficients P can be calculated from the maximal degree p and the number n of input variables as follows:

$$P = \binom{n+p}{p} = \frac{(n+p)!}{n!p!} \tag{16}$$

where P increases polynomially with both p and n. The expression Eq. (15) is called the full PC representation of degree p of the model response Y.

3.2. Computing the PC coefficients

There are typically two ways to compute the PC coefficients: 1) by projection (see [30] among others) and 2) by regression (*e.g.,* [7]). Projection-based methods exploit the orthonormality of the PC basis elements by assessing the integral

$$a_{\boldsymbol{\alpha}} = \int_{\mathbb{R}^n} \mathcal{M}(\boldsymbol{X}) \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \mathrm{d}\boldsymbol{x} \tag{17}$$

using a numerical integration scheme.

The regression-based methods minimize some distance between the model responses and the truncated PCE,

$$\boldsymbol{a}_p = \underset{\boldsymbol{a}_p}{\operatorname{argmin}} \left(L(\mathcal{M}(\boldsymbol{X}) | \mathcal{M}_p) \right) \tag{18}$$

with $\boldsymbol{a}_p = \{a_{\boldsymbol{\alpha}}, 0 \leqslant |\boldsymbol{\alpha}| \leqslant p\}$ the vector of PC coefficients. $L(\mathcal{M}(\boldsymbol{X}) | \mathcal{M}_p)$ defines the distance to be minimized between $\mathcal{M}$ and $\mathcal{M}_p$. In a probabilistic framework, $L(\cdot)$ represents a probability function that measures how likely the identified PCE fits the model response.

With the regression-based methods considered in this paper, a problem arises from the dramatic increase of P when increasing the maximal degree p or the number of input variables n. Indeed, a large number of model evaluations N is required in this context and the evaluation of Eq. (18) can be hampered by overfitting issues. To prevent overfitting, a Bayesian based algorithm is proposed in Section 5 to build a sparse PCE, which retains only a small number of basis functions and PC coefficients to capture the main stochastic features of the model response. Thus, a small number of model evaluations may be sufficient to compute the sparse PC coefficients.

3.3. Sparse polynomial chaos expansion

Let $\mathcal{A}$ be a non-empty finite subset of $\mathbb{N}^n$, with which the truncated PCE can be defined by

$$\mathcal{M}_{\mathcal{A}}(\boldsymbol{x}) \equiv \sum_{\boldsymbol{\alpha}\in\mathcal{A}} a_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \tag{19}$$

The common truncation scheme in Eq. (15) corresponds to the choice $\mathcal{A} = \mathcal{A}^{p,n}$, which is referred to as the full PCE. Since the large cardinality of this set may lead to the computational issues previously discussed, the determination of truncation sets $\mathcal{A}$ of small cardinality is of interest. Thus, we define that if the following condition is verified, the truncated PCE is sparse:

$$IS = \frac{\mathrm{card}(\mathcal{A})}{\mathrm{card}(\mathcal{A}^{p,n})} \ll 1, \quad p \equiv \max_{\boldsymbol{\alpha}\in\mathcal{A}}(|\boldsymbol{\alpha}|) \tag{20}$$

In the present paper, a new algorithm, based on Bayesian model selection, is proposed to build sparse PCEs.

3.4. PC-based global sensitivity indices

Once the sparse PCE of a model response is built, a global sensitivity analysis can be carried out at a negligible additional computational cost by analytically computing the Sobol' indices. Let us consider the PCE in Eq. (19). A subset of multidimensional indices $\mathcal{I}_{i_1\ldots i_s}$ in $\mathcal{A}$ is defined such that

$$\mathcal{I}_{i_1\ldots i_s} = \left\{ \boldsymbol{\alpha}\in\mathcal{A}: \begin{array}{llll} \alpha_k > 0 & k\in(i_1,\ldots,i_s), & \forall k = 1,\ldots,n \\ \alpha_k = 0 & k\notin(i_1,\ldots,i_s), & \forall k = 1,\ldots,n \end{array} \right\} \tag{21}$$

Using this notation, the sparse PCE can be rewritten in the form of the Sobol' decomposition:

$$\begin{aligned} \mathcal{M}_{\mathcal{A}} =& a_{\mathbf{0}} + \sum_{i_1=1}^{n}\sum_{\boldsymbol{\alpha}\in\mathcal{I}_{i_1}} a_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(x_{i_1}) + \sum_{i_2>i_1}^{n}\sum_{\boldsymbol{\alpha}\in\mathcal{I}_{i_1 i_2}} a_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(x_{i_1},x_{i_2}) \\ &+\cdots+\sum_{i_s>\cdots>i_1}^{n}\sum_{\boldsymbol{\alpha}\in\mathcal{I}_{i_1,\ldots,i_s}} a_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(x_{i_1}\ldots,x_{i_s}) \\ &+\cdots+\sum_{\boldsymbol{\alpha}\in\mathcal{I}_{1,\ldots,n}} a_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(\boldsymbol{x}) \end{aligned} \tag{22}$$

where each summand in Eq. (2) can be identified in the above equation as follows: $\mathcal{M}_{i_1\ldots i_s}(x_{i_1,\ldots,i_s}) = \sum_{\boldsymbol{\alpha}\in\mathcal{I}_{i_1\ldots i_s}} a_{\boldsymbol{\alpha}}\psi_{\boldsymbol{\alpha}}(x_{i_1},\ldots,x_{i_s})$.

Due to the orthonormal property of the polynomial basis, the total and partial variances can be derived analytically from the sparse PCE representation as follows:

$$D^{\mathcal{A}} = \sum_{\boldsymbol{\alpha} \in \mathcal{A} \setminus \{\mathbf{0}\}} a_{\boldsymbol{\alpha}}^2, \quad D_{i_1 \dots i_s}^{\mathcal{A}} = \sum_{\boldsymbol{\alpha} \in \mathcal{I}_{i_1 \dots i_s}} a_{\boldsymbol{\alpha}}^2 \tag{23}$$

Now it is easy to compute the partial sensitivity indices for the subset of input variables $\{x_{i_1}, \dots, x_{i_s}\}$ from the above equations:

$$S_{i_1 \dots i_s}^{\mathcal{A}} = \frac{D_{i_1 \dots i_s}^{\mathcal{A}}}{D^{\mathcal{A}}} \tag{24}$$

The total sensitivity index of an input variable x_i is thus given by the sum of all the partial sensitivity indices involving i:

$$S_i^{T,\mathcal{A}} = \sum_{\boldsymbol{\alpha}: \alpha_i > 0} S_{\boldsymbol{\alpha}}^{\mathcal{A}} \tag{25}$$

In the numerical exercises in Section 6, the estimated sensitivity indices are compared to analytical values for some benchmarks functions. The efficiency of the algorithm proposed in Section 5 is assessed by evaluating the following errors with respect to the sample size (*i.e.,* computational cost):

$$e_1 = \sum_{i=1}^{n} |S_i^{\text{ex}} - S_i^{\mathcal{A}}| \tag{26}$$

$$e_T = \sum_{i=1}^{n} |S_i^{T,\text{ex}} - S_i^{T,\mathcal{A}}| \tag{27}$$

where the superscript "ex" stands for the analytical value.

4. The Bayesian Model Averaging Framework

Let us consider N_m plausible competing sparse PCE models $\mathcal{M}_{\mathcal{A}_k}$:

$$\mathcal{M}_{\mathcal{A}_k} \equiv \sum_{\boldsymbol{\alpha} \in \mathcal{A}_k} a_{\boldsymbol{\alpha}} \psi_{\boldsymbol{\alpha}}(\boldsymbol{x}), \quad k = 1, \dots, N_m \tag{28}$$

The above equation can be written in the vector form $\mathcal{M}_{\mathcal{A}_k} = \boldsymbol{a}_k \boldsymbol{\psi}_k$ with the parameter vector $\boldsymbol{a}_k$ and the vector of polynomial terms $\boldsymbol{\psi}_k$. Let $\mathcal{X} = \{\boldsymbol{X}^{(1)}, \dots, \boldsymbol{X}^{(N)}\}$ be a set of input data and $\mathcal{Y} = \{Y^{(1)}, \dots, Y^{(N)}\}^T$ be the set of output data such that $Y^{(r)} = \mathcal{M}(\boldsymbol{X}^{(r)})$. The challenge is to scrutinize each model's ability to reproduce the data set $(\mathcal{X}, \mathcal{Y})$ and further to pinpoint

the best sparse PC representation among the set $\{\mathcal{M}_{\mathcal{A}_k}, k = 1, \ldots, N_m\}$. To this end, Bayesian model averaging (BMA), a formal statistical approach based on Bayes' theorem, is introduced to realize an objective ranking and a quantitative comparison of the proposed alternative models. The BMA approach combines prior information of each model with the observed data to estimate the posterior probability for each individual model to be the best one among the competing models. The posterior probabilities $\mathcal{P}(\mathcal{M}_{\mathcal{A}_k}|\mathcal{Y})$ are given by Bayes' theorem:

$$\mathcal{P}(\mathcal{M}_{\mathcal{A}_k}|\mathcal{Y}) = \frac{\mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k})\mathcal{P}(\mathcal{M}_{\mathcal{A}_k})}{\sum_{i=1}^{N_m} \mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_i})\mathcal{P}(\mathcal{M}_{\mathcal{A}_i})} \tag{29}$$

where $\mathcal{P}(\mathcal{M}_{\mathcal{A}_k})$ is the prior probability that model $\mathcal{M}_{\mathcal{A}_k}$ is the best one from the set of considered models before any data is collected. The equally likely prior $\mathcal{P}(\mathcal{M}_{\mathcal{A}_k}) = 1/N_m$ is usually used if the prior information is vague. $\mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k})$ is the likelihood of the observed data, expressing the preference shown by the data for different models. The denominator in Eq. (29) is the normalization constant, which ensures that the posterior distribution on the left-hand side is a valid probability density and integrates to one. Neglecting the normalizing constant, Bayes' theorem can be written in the following way:

$$\mathcal{P}(\mathcal{M}_{\mathcal{A}_k}|\mathcal{Y}) \propto \mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k})\mathcal{P}(\mathcal{M}_{\mathcal{A}_k}) \tag{30}$$

Therefore the key to obtaining the posterior probability is to define the term $\mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k})$, which is known as the Bayesian model evidence (BME). This term quantifies the likelihood of the observed data integrated over each model's parameter space with the following expression:

$$\mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k}) = \int_{\mathbb{R}^{P_k}} \mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k}, \boldsymbol{a}_k)\mathcal{P}(\boldsymbol{a}_k|\mathcal{M}_{\mathcal{A}_k})\mathrm{d}\boldsymbol{a}_k \tag{31}$$

where P_k is the number of parameters for the model $\mathcal{M}_{\mathcal{A}_k}$. $\mathcal{P}(\boldsymbol{a}_k|\mathcal{M}_{\mathcal{A}_k})$ denotes the prior distribution of the parameter set $\boldsymbol{a}_k$. $\mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k}, \boldsymbol{a}_k)$ is the likelihood function, which expresses how probable is the observed data for different settings of the parameter vector $\boldsymbol{a}_k$ of the model $\mathcal{M}_{\mathcal{A}_k}$.

The integral in Eq. (31) over the full parameter space of the model is not easy to calculate analytically, especially for high-dimensional parameter spaces. A mathematical approximation, *e.g.,* a Taylor series expansion followed by a Laplace approximation is thus used to render the integration computationally tractable. The Laplace approximation assumes that the posterior distribution of the parameters is Gaussian and highly peaked around its local maximum with the expression $\mathcal{P}(\boldsymbol{a}_k|\mathcal{Y}, \mathcal{M}_{\mathcal{A}_k}) \sim \mathcal{N}(\tilde{\boldsymbol{a}}_k, \boldsymbol{C}_{\tilde{a}\tilde{a}})$. The mean of the posterior $\tilde{\boldsymbol{a}}_k$ is referred to as the maximum a posteriori estimate

(MAP), which represents the most likely parameter set for model $\mathcal{M}_{\mathcal{A}_k}$, considering both the prior belief about the parameters and the fitting of the observed data. The covariance matrix $\boldsymbol{C}_{\tilde{a}\tilde{a}}$ is estimated at the MAP solution (*i.e.*, $\boldsymbol{a}_k = \tilde{\boldsymbol{a}}_k$).

Conducting a Taylor expansion of $\ln \mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k})$ centred on the posterior mode $\tilde{\boldsymbol{a}}_k$, with third- and higher-order terms neglected, then taking the exponential of the resulting expansion and computing the integration with the Laplace approximation yields

$$\mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k}) \simeq \mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k}, \tilde{\boldsymbol{a}}_k)\mathcal{P}(\tilde{\boldsymbol{a}}_k|\mathcal{M}_{\mathcal{A}_k})(2\pi)^{P_k/2}|\tilde{\boldsymbol{\Sigma}}|^{-1/2} \tag{32}$$

where $\tilde{\boldsymbol{\Sigma}}$ is the $P_k \times P_k$ Hessian matrix of second derivatives of the negative log posterior defined by

$$\left[\tilde{\boldsymbol{\Sigma}}\right]_{ij} = - \left. \frac{\partial^2 \ln \mathcal{P}(\tilde{\boldsymbol{a}}_k|\mathcal{Y}, \mathcal{M}_{\mathcal{A}_k})}{\partial a_i \partial a_j} \right|_{\boldsymbol{a}_k = \tilde{\boldsymbol{a}}_k} \tag{33}$$

Eq. (32) yields

$$-2\ln \mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k}) \simeq -2\ln \mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k}, \tilde{\boldsymbol{a}}_k) - 2\ln \mathcal{P}(\tilde{\boldsymbol{a}}_k|\mathcal{M}_{\mathcal{A}_k}) - P_k \ln(2\pi) + \ln|\tilde{\boldsymbol{\Sigma}}| \tag{34}$$

By assuming that the posterior distribution is virtually Gaussian around the MAP (Laplace approximation), one can set $\tilde{\boldsymbol{\Sigma}} = \boldsymbol{C}_{\tilde{a}\tilde{a}}^{-1}$, which leads to the Kashyap information criterion (KIC, [18]):

$$KIC_k = -2\ln \mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k}, \tilde{\boldsymbol{a}}_k) - 2\ln \mathcal{P}(\tilde{\boldsymbol{a}}_k|\mathcal{M}_{\mathcal{A}_k}) - P_k \ln(2\pi) - \ln|\boldsymbol{C}_{\tilde{a}\tilde{a}}| \tag{35}$$

Evaluating KIC_k is a computationally feasible alternative to directly computing BME. KIC_k reduces the computational effort by considering the most likely parameter set instead of integrating over the entire parameter space. Note that $\tilde{\boldsymbol{a}}_k$ and $\boldsymbol{C}_{\tilde{a}\tilde{a}}$ in Eq. (35) are usually estimated by optimization algorithms.

5. Bayesian sparse PCE

5.1. Our working assumptions

Evaluating the KIC_k of each competing model $\mathcal{M}_{\mathcal{A}_k}$ allows finding the best model, the one corresponding to the smallest KIC. Let us define the degree and the interaction order of any index $\boldsymbol{\alpha}$ respectively as follows:

$$p_{\boldsymbol{\alpha}} \equiv |\boldsymbol{\alpha}| = \sum_{i=1}^{n} \alpha_i, \quad q_{\boldsymbol{\alpha}} \equiv \sum_{i=1}^{n} \mathbf{1}_{\alpha_i > 0} \tag{36}$$

where $\mathbf{1}_{\alpha_i>0} = 1$ if $\alpha_i > 0$ and 0 otherwise.

Moreover, let $y = (Y - \mathbb{E}[Y])/\mathbb{V}[Y]$ be the standardized model response variable and $\boldsymbol{y}$ the vector of standardized model responses. Our strategy to build a sparse PCE for y relies on the first assumption that the model error,

$$\epsilon_k = y - \mathcal{M}_{\mathcal{A}_k}(\boldsymbol{x}) \tag{37}$$

is a homoscedastic Gaussian variable, that is, $\epsilon_k \sim \mathcal{N}(0, \sigma_k^2)$. This leads to a Gaussian likelihood function $\mathcal{P}(\mathcal{Y}|\mathcal{M}_{\mathcal{A}_k}, \boldsymbol{a}_k, \sigma_k^2) \sim \mathcal{N}(\mathcal{M}_{\mathcal{A}_k}, \sigma_k^2)$. The matrix formulation of the previous equation also becomes

$$\boldsymbol{\epsilon}_k = \boldsymbol{y} - \boldsymbol{\psi}_k \boldsymbol{a}_k \tag{38}$$

Because $\mathcal{M}_{\mathcal{A}_k}$ is an approximation of y (centred and reduced), it is expected that the PCE coefficients are close to zero. Consequently, we further assume that

$$\mathcal{P}(\boldsymbol{a}_k|\mathcal{M}_{\mathcal{A}_k}) \sim \mathcal{N}(\mathbf{0}, \boldsymbol{C}_{aa}) \tag{39}$$

with

$$\boldsymbol{C}_{aa} = \begin{bmatrix} \sigma^2_{\boldsymbol{\alpha}_1} & 0 & \dots & \dots \\ 0 & \ddots & 0 & \dots \\ \vdots & 0 & \ddots & 0 \\ 0 & \dots & 0 & \sigma^2_{\boldsymbol{\alpha}_{P_k}} \end{bmatrix}$$

where $\sigma^2_{\boldsymbol{\alpha}_i} = (p_{\boldsymbol{\alpha}_i} + q_{\boldsymbol{\alpha}_i} - 1) q^2_{\boldsymbol{\alpha}_i}$ stands for the variance assigned to the ith term in $\mathcal{A}_k$. This prior assigns a high variance to the PC element with a high level of interactions and a high polynomial degree. This has the effect of favouring low degree and interaction terms by assigning smaller weights to those elements. Note that because $\boldsymbol{y}$ is a standardized vector, $a_{\mathbf{0}}$ is always equal to zero. Thus, in the following, it is supposed that $a_{\mathbf{0}} \notin \mathcal{A}_k$. In the present work, Gaussian priors are assigned to the PCE coefficients but it is worth mentioning that Laplace priors can also be a good choice [31, 32]. Finally, it is worth mentioning that, except for the model selection criterion, our proposed approach is equivalent to the ridge regression approach [33] and that assuming Laplace priors is tantamount to the Least Absolute Shrinkage and Selection Operator approach (LASSO, see [34]).

By noting that the PCE is linear with respect to its coefficients, the analytical expression of the posterior distribution is given by

$$\mathcal{P}(\boldsymbol{a}_k|\boldsymbol{y}, \mathcal{M}_{\mathcal{A}_k}, \tilde{\sigma}_k^2) \sim \mathcal{N}(\tilde{\boldsymbol{a}}_k, \boldsymbol{C}_{\tilde{a}\tilde{a}}) \tag{40}$$

and

$$\mathcal{P}(\sigma_k^{-2}|\boldsymbol{y}, \mathcal{M}_{\mathcal{A}_k}, \tilde{\boldsymbol{a}}_k) \sim \Gamma\left(\frac{N+2}{2}, \frac{2}{N}\tilde{\sigma}_k^{-2}\right) \tag{41}$$

with

$$\tilde{\boldsymbol{a}}_k = \frac{\boldsymbol{C}_{\tilde{a}\tilde{a}}\boldsymbol{\psi}_k^T\boldsymbol{y}}{\tilde{\sigma}_k^2} \tag{42}$$

$$\boldsymbol{C}_{\tilde{a}\tilde{a}} = \left(\frac{\boldsymbol{\psi}_k^T\boldsymbol{\psi}_k}{\tilde{\sigma}_k^2} + \boldsymbol{C}_{aa}^{-1}\right)^{-1} \tag{43}$$

$$\tilde{\sigma}_k^2 = \frac{(\boldsymbol{y} - \boldsymbol{\psi}_k\tilde{\boldsymbol{a}}_k)^T(\boldsymbol{y} - \boldsymbol{\psi}_k\tilde{\boldsymbol{a}}_k)}{N} \tag{44}$$

Eq. (44) gives the MAP estimate of the current model error σ_k^2. Solving Eqs. (42–44) requires an iterative process that starts by setting $\tilde{\sigma}_k^2$ (say, $\tilde{\sigma}_k^2 = 1$). Then, after inferring $\tilde{\boldsymbol{a}}_k$ and $\boldsymbol{C}_{\tilde{a}\tilde{a}}$ from Eq. (42) and Eq. (43) respectively, $\tilde{\sigma}_k^2$ is updated from Eq. (44). The calculations are repeated until $\tilde{\sigma}_k^2$ converges within a given relative precision (say 10^{-2}). In this context, we can take advantage of these analytical expressions to evaluate KIC_k and infer the best sparse PCE for $\boldsymbol{y}$. In the sequel, the Bayesian sparse PCE is denoted by $\mathcal{M}_{\mathcal{A}}$, and the associated vector of coefficients and the optimal error variance are denoted by, respectively, $\boldsymbol{a}_{\mathcal{A}}$ and $\sigma_{\mathcal{A}}^2$. Similarly, the PCE degree and level of interaction are denoted $p_{\mathcal{A}}$ and $q_{\mathcal{A}}$, respectively, while $P_{\mathcal{A}}$ corresponds to the number of coefficients in the sparse PCE.

5.2. Post-processing

Eqs. (40–41) indicate that $\boldsymbol{a}_{\mathcal{A}}$ and $\sigma_{\mathcal{A}}^2$ are random variables. Consequently, the Sobol' indices estimated with the best sparse PCE should also be treated like random variables as well as any statistics computed with $\mathcal{M}_{\mathcal{A}}$. Hence, it is possible to assign a credible interval to the Sobol' indices estimate. This is achieved by randomly sampling draws of $\boldsymbol{a}_{\mathcal{A}}$ and $\sigma_{\mathcal{A}}^2$ according to their posterior densities and evaluating the total variance and partial variances (from Eq. (23)) for each draw before estimating the Sobol' indices (Eq. (24)). In this way, one obtains a sample of sensitivity indices from which, for instance, the 95% credible interval of each statistic can be extracted. In the sequel, Latin hypercube samples of size 100,000 are generated to evaluate the credible intervals. It is worth mentioning that these calculations are computationally cheap with sparse PCEs.

Since KIC-based PCE selection is a compromise between model simplicity and goodness of fit, overfitting is avoided. Thereby one can rely on the relative training error to gauge whether the identified sparse PCE is an accurate or poor representation of the original model $\mathcal{M}$. Because the vector

of model responses is standardized, the relative training error is nothing but $\sigma^2_{\mathcal{A}}$. However, the validity of this statement relies on the assumption that the model error is Gaussian, which is not always true. One can confirm this assumption by checking whether the residual (*i.e.,* the vector of model error at the MAP) follows the expected distribution, namely, $\mathcal{N}(0, \sigma^2_{\mathcal{A}})$. If it is not the case, it is likely that the identified sparse PCE is not reliable. In this situation, it is recommended to increase the sample size where the Laplace approximation is more likely to hold [17].

5.3. The proposed algorithm

Suppose that we have generated an experimental design $\mathcal{X} = \{\boldsymbol{X}^{(1)}, \ldots, \boldsymbol{X}^{(N)}\}$ with N realizations, *e.g.*, a random design based on Monte Carlo sampling, Latin Hypercube sampling or quasi-random low discrepancy sequences. The quasi-Monte Carlo (QMC) method of [35] is adopted for generating samples in this work due to its space filling property. After running the model at the design points, the model responses of interest are gathered into the vector $\mathcal{Y} = \{Y^{(1)}, \ldots, Y^{(N)}\}^T$. Using the concept of BMA, it is now possible to devise an algorithm that selects the optimal sparse PCE model from the data set $(\mathcal{X}, \mathcal{Y})$. The algorithm is outlined in the following:

Step 1 (Initialization): The data $(\mathcal{X}, \mathcal{Y})$ are transformed into standardized vectors $(\boldsymbol{x}, \boldsymbol{y})$. Then, the initial degree and interaction order of the PCE are defined; it is recommended to choose either $(p = 2, q = 1)$ or $(p = 4, q = 2)$, depending on the features of the model response of interest. Further, the following subset is created: $\mathcal{A}^{p,q} = \{\boldsymbol{\alpha} \in \mathbb{N}^n : p_{\boldsymbol{\alpha}} \leqslant p, q_{\boldsymbol{\alpha}} \leqslant q\}/\{\boldsymbol{0}\}$.

Step 2 (Ranking via correlation coefficient): Set $P = \mathrm{Card}\,(\mathcal{A}^{p,q})$ and define the polynomial basis functions $\boldsymbol{\psi} = (\psi_1, \psi_2, \ldots, \psi_P)$ associated to $\mathcal{A}^{p,q}$. Then, calculate the Pearson correlation coefficient r_j between each polynomial term $\psi_j(\boldsymbol{x})$ $\forall j = 1, \ldots, P$ and the model response vector $\boldsymbol{y}$ as follows:

$$r_j = \frac{\mathbb{COV}\left[\boldsymbol{y}, \psi_j(\boldsymbol{x})\right]}{\sqrt{\mathbb{V}[\boldsymbol{y}]\mathbb{V}[\psi_j(\boldsymbol{x})]}} \tag{45}$$

where $\mathbb{COV}$ is the covariance operator. Then, sort the array $(r_1^2, r_2^2, \ldots, r_P^2)$ in descending order and rearrange the polynomial basis functions accordingly in a new vector $\hat{\boldsymbol{\psi}} = (\hat{\psi}_1, \ldots, \hat{\psi}_j, \hat{\psi}_{j+1}, \ldots, \hat{\psi}_P)$ such that $r_j^2 \geqslant r_{j+1}^2$.

Step 3 (Ranking via partial correlation coefficient): Compute the partial correlation coefficient $\hat{r}_{j|1,\ldots,j-1}$ between each basis function $\hat{\psi}_j(\boldsymbol{x})$ and $\boldsymbol{y}$ for $j = 1, \ldots, P$ with the following equation:

$$\hat{r}_{j|1,\ldots,j-1} = \frac{\mathbb{COV}\left[\boldsymbol{y}, \hat{\psi}_j(\boldsymbol{x}) | \hat{\psi}_1(\boldsymbol{x}), \ldots, \hat{\psi}_{j-1}(\boldsymbol{x})\right]}{\sqrt{\mathbb{V}\left[\boldsymbol{y} | \hat{\psi}_1(\boldsymbol{x}), \ldots, \hat{\psi}_{j-1}(\boldsymbol{x})\right] \mathbb{V}\left[\hat{\psi}_j(\boldsymbol{x}) | \hat{\psi}_1(\boldsymbol{x}), \ldots, \hat{\psi}_{j-1}(\boldsymbol{x})\right]}} \tag{46}$$

where $\mathbb{COV}[\cdot|\cdot]$ is the conditional covariance operator and $\mathbb{V}[\cdot|\cdot]$ the conditional variance operator. As in step 2, sort the array $(\hat{r}^2_1, \hat{r}^2_{2|1}, \hat{r}^2_{3|1,2}, \ldots, \hat{r}^2_{P|1,\ldots,P-1})$ in descending order, based on which we update the vector of PC basis elements $\tilde{\boldsymbol{\psi}} = (\tilde{\psi}_1, \ldots, \tilde{\psi}_j, \tilde{\psi}_{j+1}, \ldots, \tilde{\psi}_P)$ such that $\hat{r}^2_{j|1,\ldots,j-1} \geqslant \hat{r}^2_{j+1|1,\ldots,j}$. Initialize $KIC_1 = +\infty$, $\boldsymbol{\psi}_{\mathcal{A}} = \tilde{\psi}_1$ and $k = 2$.

Step 4 (Identification of the current sparse PCE): Define a sparse PCE model $\mathcal{M}_{\mathcal{A}_k}$ with the polynomial basis $\boldsymbol{\psi}_k = (\boldsymbol{\psi}_{\mathcal{A}}, \tilde{\psi}_k)$. The BMA approach is used to estimate the current sparse PCE model $\mathcal{M}_{\mathcal{A}_k}$. Evaluate the MAP estimates $\tilde{\boldsymbol{a}}_k$ and $\boldsymbol{C}_{\tilde{a}\tilde{a}}$ from Eqs. (42–43) as well as the KIC_k assigned to the current model $\mathcal{M}_{\mathcal{A}_k}$ (Eq. (35)). If $KIC_k \leqslant KIC_{k-1}$, set $\boldsymbol{\psi}_{\mathcal{A}} = \boldsymbol{\psi}_k$ and $\boldsymbol{a}_{\mathcal{A}} = \tilde{\boldsymbol{a}}_k$, otherwise set $KIC_k = KIC_{k-1}$. Then, set $k = k + 1$ and repeat this step until $k = P$.

Step 5 (Enriching $\mathcal{A}^{p,q}$): Write $\mathcal{M}_{\mathcal{A}} = \boldsymbol{a}_{\mathcal{A}}\boldsymbol{\psi}_{\mathcal{A}}$ for the identified sparse PCE whose elements belongs to the subset $\mathcal{A}$. If the subset $\mathcal{A}$ contains (i) elements of degree $p-1$ or p, then set $p = p + 2$ or (ii) elements of level of interaction q, then set $q = q + 1$ and resume from Step 2 after setting $\mathcal{A}^{p,q} = \mathcal{A}$ and enriching the subset from elements of degree $p-1$ and p as well as elements of level of interaction q. Otherwise, stop the calculation.

The algorithm starts by considering all the PC basis elements of low degree (typically $p = 2$) and low level of interactions ($q = 1$). This ensures that the initial number of elements to be analyzed with the KIC is small. Before evaluating the latter, the basis elements are reordered by order of importance in the next two following steps. First, in Step 2, the Pearson correlation coefficient r_j between the model response $\boldsymbol{y}$ and each basis element ψ_j is computed. The Pearson correlation coefficient measures the strength of the linear relation between ψ_j and $\boldsymbol{y}$. Consequently, r_j^2 measures how ψ_j is a relevant basis element for the investigated sparse PCE. Because there may have been spurious correlations in the experimental design $\mathcal{X}$, the Pearson correlation coefficients can be misleading in highlighting the relevant terms. Consequently, a second reordering is performed in Step 3 on the basis of the estimated partial correlation coefficients. Step 4 proceeds to the identification of the optimal sparse PCE for the basis elements belonging to the current subset $\mathcal{A}^{p,q}$. Then, the latter is enriched in Step 5 if the optimal sparse PCE contains elements of degree exceeding $p-1$ or level of interactions q. The enrichment is made using basis elements of higher degrees ($p+1$, $p+2$) and higher levels of interactions ($q+1$). In this case, the procedure is resumed for the new subset (from Step 2). If the current identified sparse PCE does not contain basis elements of degree exceeding $p-1$ or level interaction q, then the current sparse PCE is deemed the Bayesian sparse PCE $\mathcal{M}_{\mathcal{A}}$.

Note that for high-dimensional problems ($n > 10$), considering all the

possible interaction levels $q+1$ in Step 5 can provide a very high number of terms in $\mathcal{A}^{p,q}$. Consequently, during this phase, only the interactions that involve the relevant inputs of the current iteration are considered. If we consider the multi-dimensional index $\boldsymbol{\alpha} = \alpha_1 \dots \alpha_n$, the relevant inputs are those with at least one nonzero index in the subset $\mathcal{A}^{p,q}$. Thus, $\forall \boldsymbol{\alpha} \in \mathcal{A}^{p,q}$, if α_{i_1} always equals zero, then x_{i_1} is deemed irrelevant and is not further considered for higher interaction levels.

Although the proposed enrichment strategy reduces the computational cost, it may result in the identification of a sparse PCE with poor performance. The quality of the identified PCE model $\mathcal{M}_{\mathcal{A}}$ is measured a posteriori by assessing the relative training error $\sigma^2_{\mathcal{A}}$. Our experience suggests that a good sparse PCE satisfies $\sigma^2_{\mathcal{A}} < 0.05$. Otherwise, it is recommended to restart the PCE identification procedure with higher initial values of p and q. If the identified PCE is still not satisfactory, then the experimental design $\mathcal{X}$ should be enriched and the model response vector $\mathcal{Y}$ updated before restarting the identification of the optimal sparse PCE.

Regarding the parameter estimation, we assume that the prior parameter and error distributions are both Gaussian. Thus the parameters can be estimated by the analytical expression of the MAP as described in Section 4. Generally, calculating the MAP analytically allows avoiding the problem of a poorly conditioned matrix usually encountered with the regression method for small sample sizes. However, a larger data set basically yields a more accurate estimation of the sparse PCE, as the Laplace approximation is expected to hold in situations where a relatively large sample size is available.

6. Synthetic mathematical examples

6.1. Ishigami function

Let us consider a popular benchmark in global sensitivity analysis, the Ishigami function [36, 37]:

$$Y = \sin X_1 + a \sin^2 X_2 + b X_3^4 \sin X_1 \tag{47}$$

where the inputs are independent random variables uniformly distributed over $[-\pi, \pi]$. This function is smooth, nonlinear and non-monotonic. We note that $x_i = \frac{X_i}{2\pi} + \frac{1}{2}$. The total and partial variances of Y can be calculated analytically:

$$D = \frac{a^2}{8} + \frac{b\pi^4}{5} + \frac{b^2\pi^8}{18} + \frac{1}{2},$$

$$D_1 = \frac{b\pi^4}{5} + \frac{b^2\pi^8}{50} + \frac{1}{2}, \quad D_2 = \frac{a^2}{8}, \quad D_3 = 0,$$

$$D_{12} = D_{23} = 0, \quad D_{13} = \frac{8b^2\pi^8}{225}, \quad D_{123} = 0.$$

The coefficients in the function take the numerical values $a = 7, b = 0.1$. Note that the function is sparse in nature since there are three independent variables in the model but the maximum interaction order is 2. Moreover, the function is even with respect to the variables X_2 and X_3, hence the odd polynomials of these variables should be zero in the PCE. This means that to identify a good sparse PCE for this function, one must initialize $p = 4$ and $q = 2$. Indeed, choosing $p = 2$ and $q = 1$ will discard the input variable X_3 in the investigated sparse PCE (see the discussion in the previous section).

The sparse PCE coefficients are evaluated using quasi-random sequences of different sizes $N = 2^j$ ($j = 5, 6, \ldots, 13$). These sequences are used to study the convergence of the proposed algorithm for building a sparse PCE. The degree and interaction level of the PCE are iteratively increased from $p = 4$ and $q = 2$ as described in Section 5.3. The obtained results from the sparse PCEs are reported in Table 1.

Table 1 reveals that $N = 2^6$ model evaluations are sufficient to obtain relatively accurate estimates with a sparse PCE. We note that the relative training error is very small, $\sigma^2_{\mathcal{A}} \simeq 1.2 \times 10^{-3}$. The sensitivity indices in this case show a discrepancy around 2% with respect to the reference solution. With $N = 2^6$ model evaluations, the sparse PCE produced by the iterative procedure has a maximum degree $p_{\mathcal{A}} = 9$ but contains only $P_{\mathcal{A}} = 13$ terms, whereas the corresponding full PCE would contain $P = \binom{3+9}{9} = 220$ terms (the index of sparsity $IS = 13/220 \approx 0.059$) and would thus require at least 220 model evaluations to compute the whole set of coefficients.

Now let us investigate the sensitivity of the sparse PCE estimation to the number of model evaluations. In Table 1 it is observed that the accuracy of the estimates increases with the number of model evaluations. For instance, using $N = 2^5$ model evaluations leads to a quite large relative error ($\sigma^2_{\mathcal{A}} \simeq 0.31$) while the latter reduces from $N = 2^6$ model evaluations. In particular, a two-digit accuracy is obtained for all the sensitivity indices for $N = 2^j$, $j \geqslant 7$. It is also observed that all the sparse PCEs contain relatively low numbers of terms compared to their full counterparts (with indices of sparsity that range from 2.6% to 6.0%), which reflect the sparse structure of the model response.

The various 95% credible intervals of both the first-order and total sensitivity indices estimates are plotted in Figure 1 together with the analytical reference values. From this figure, it clearly appears that the credible intervals obtained from $N = 2^6$ model runs are quite narrow and well centred around the exact values, which indicates a low bias of the sparse PCE-based

estimates.

Furthermore, the rate of convergence of the proposed procedure for building a sparse PCE is studied. In Figure 2, the absolute errors of the first-order and total sensitivity indices are reported as functions of the number of model evaluations. The results of this figure show a high convergence rate of the sensitivity indices with respect to the number of model evaluations for the sparse PCE. This high convergence rate is due to the smoothness of the Ishigami function.

All the statistics computed with the identified Bayesian sparse PCEs for different sample sizes are reliable under the assumption of Gaussian model errors. To check this assumption a posteriori, the empirical cumulative distribution function (CDF) of the residual evaluated at the MAP is compared with the expected normal CDF. This comparison is carried out for different sample sizes (see Figure 3). It can be inferred that, except for $N = 2^5$, the residuals are virtually normally distributed. Hence, assuming a Gaussian error between the Ishigami function and the Bayesian sparse PCE seems a reasonable assumption.

To test the stability of the algorithm, for different sample sizes we repeat one hundred times the Bayesian sparse PCE identification from different draws. To do so, the quasi-random sequences are generated by sampling from different points in the input space. The results are depicted in Figure 4 in the form of box-and-whisker plots. The latter show relatively large variations of the Sobol' indices estimates, especially at low sample sizes ($N = 2^5, 2^6$). We note that, as compared to the uncertainty ranges provided by the Bayesian sparse PCE (see Figure 1), the effect of the experimental design is non-negligible at low sample sizes. Thereby, the uncertainties computed with the optimal sparse PCE are underestimated. To obtain more reliable uncertainties one could consider all probable sparse PCEs instead of considering only the *best* one. We did not consider this alternative in the present work.

6.2. Sobol' function

Let us consider the Sobol' function [23]:

$$Y = \prod_{i=1}^{n} \frac{|4X_i - 2| + b_i}{1 + b_i} \tag{48}$$

where the input variables X_i, $i = 1, \ldots, n$ are uniformly distributed over $[0, 1]$. This function is non-smooth and non-monotonic. The analytical expressions of the total and partial variances are

$$D = \prod_{i=1}^{n}(D_i + 1) - 1, \quad D_i = \frac{1}{3(1 + b_i)^2}, \quad D_{i_1 \ldots i_s} = \prod_{k=1}^{s} D_{i_k}.$$

For numerical application, we set the number of input variables $n = 9$ and $b_i = (i-1)/4$. This function is a challenging one due to the presence of the absolute value which slows down the convergence of the PCE. Moreover, the level of interactions is very high, as revealed by the differences between the first-order and total Sobol' indices (see Table 2, column #2).

For the same reason as before, the initial degree and interaction order of the sparse PCE are $p = 4$ and $q = 2$. Several sample sizes are used to assess the efficiency of the algorithm $N = 2^j$, $(j = 5, \ldots, 13)$. Inaccurate results were obtained for sample sizes less than $N = 2^8$. The results are listed in Table 2 for $N = 2^8, \ldots, 2^{13}$ and depicted in Figure 5.

It can be observed that with the increase of the number of model evaluations, the accuracy of the Bayesian sparse PCE slightly improves. The relative training error $\sigma^2_{\mathcal{A}}$ never decreases below 10^{-2} (see the last row of Table 2). With $N = 2^{12}$ model evaluations, the sparse PCE yields estimates with an accuracy similar to the reference solution (discrepancy less than 8%). In this case, the sparse PCE has a maximum degree $p_{\mathcal{A}} = 18$ and total terms $P_{\mathcal{A}} = 174$, revealing a noticeably small sparsity index of 3.7×10^{-5} with respect to a full PCE of the same degree. Indeed, the corresponding full PCE of degree $p = 18$ would contain $P = 4,686,825$ terms. This would be computationally unaffordable because of the large number of model evaluations required.

The credible intervals assigned to the estimated sensitivity indices are depicted in Figure 5. The uncertainty bounds are particularly large at low sample sizes and become very narrow from $N \geqslant 2^{11}$. Yet, despite of the relatively large uncertainties at $N = 2^8$ and $N = 2^9$, the first three inputs are identified as significantly more important than the remainder. Surprisingly, for $N = 2^{10}$ the subset of inputs (x_3, x_4, x_5) is found of equal importance. Although at $N = 2^{11}$ the uncertainties assigned to the estimated total sensitivity indices are very narrow, they do not encompass the analytical values (meaning that the estimated values are slightly biased). This indicates that the proposed approach may require a lot of model runs to accurately capture the structure of non-smooth functions.

Figure 6 shows the rate of convergence of the proposed procedure for building a sparse PCE in terms of the absolute errors of the first-order and total sensitivity indices as functions of the number of model evaluations. The results show that the convergence rate of the proposed method for the Sobol' function is relatively slow compared to the results for the Ishigami function. This low convergence rate is due to the loss of the spectral convergence of the PCE for non-smooth functions.

The Gaussian error assumption is checked a posteriori by comparing the CDF of the residual with the expected target CDF (see Figure 7). It can

be noted that, albeit some slight discrepancies, the CDFs match surprisingly well despite the non-smoothness of the Sobol' function. Hence, it can be concluded that the Gaussian error assumption is acceptable in the present exercise.

The stability of the algorithm is again tested by replicating one hundred times the Bayesian sparse PCE identification at different sample sizes. The results are depicted in Figure 8. We note that the total sensitivity indices estimates are more sensitive to the experimental design (bottom plot) than the first-order sensitivity indices (top plot). The influence of the experimental design at low sample sizes is very important on the uncertainty range and on the bias. They decrease with the increase of the sample size.

6.3. Morris function

To assess the proposed method for large dimensional problems, we consider now the so-called Morris function [38]:

$$Y = \beta_0 + \sum_{i=1}^{20} \beta_i X_i + \sum_{i<j}^{20} \beta_{ij} X_i X_j + \sum_{i<j<k}^{20} \beta_{ijk} X_i X_j X_k + \sum_{i<j<k<l}^{20} \beta_{ijkl} X_i X_j X_k X_l \tag{49}$$

where

$$X_i = \begin{cases} 2(1.1x_i/(x_i + 0.1) - 0.5) & \text{if } i = 3, 5, 7 \\ 2(x_i - 0.5) & \text{otherwise} \end{cases} \tag{50}$$

and the x_is are uniformly distributed over [0,1]. The coefficients β_i are assigned as follows:

$$\begin{cases} \beta_i = 20 & \text{for } i = 1, \ldots, 10 \\ \beta_{ij} = -15 & \text{for } i, j = 1, \ldots, 6 \\ \beta_{ijk} = -10 & \text{for } i, j, k = 1, \ldots, 5 \\ \beta_{ijkl} = 5 & \text{for } i, j, k, l = 1, \ldots, 4 \end{cases} \tag{51}$$

The remaining coefficients are zero.

The sensitivity indices are computed by post-processing the sparse PCE obtained by setting the initial degree and interaction order to $p = 2$ and $q = 1$. This choice is reasonable because, unlike the previous cases, the Morris function is neither even nor odd, and its dimensionality is high. Hence, initializing $p = 2$ and $q = 1$ allows reducing the computational time of the postprocessing. Different sizes of experimental designs $N = 2^j$, $(j = 5, 6, \ldots, 13)$ are used to build the sparse PCE. Because there are no reference values for this function, we assess the quality of the estimation by comparing the sensitivity indices and their credible intervals.

Figure 9 shows the ten greatest total sensitivity indices with their uncertainty bounds for the different sample sizes. It is found that the credible intervals with a small number of model evaluations ($N = 2^5$ and $N = 2^7$) are relatively wide, which indicates the necessity of enlarging the number of model evaluations to get more accurate results. Much narrower uncertainty bounds are found using large numbers of model evaluations ($N = 2^9, 2^{11}$ and 2^{13}) in Figure 9. The values of the total sensitivity indices are greater than those of the first-order sensitivity indices, which demonstrates that the model is non-additive and contains interactions between the parameters. Moreover, we can observe in Figure 9 that S_1^T, S_2^T and S_4^T are the three greatest values and have almost the same importance. S_3^T, S_5^T, S_6^T, S_8^T, S_9^T and S_{10}^T have intermediate values. From the sample size $N = 2^9$, the following classification of the input variables can be made:

- a group of important variables: x_1, x_2, x_4;
- variables with intermediate significance: x_3, x_5, x_6, x_8-x_{10};
- one variable with small importance: x_7;
- a group of non-significant variables: x_{11}-x_{20}.

7. Application to double diffusive convection in porous media

7.1. Problem statement

In this section, the methodology developed for sensitivity analysis is applied to the problem of double diffusive convection (DDC) in saturated porous media. DDC occurs when the saturating fluid contains several constituents and when the density gradients inducing natural convection are caused simultaneously by the temperature and compositional effects. This configuration has received considerable attention in recent years due to the wide range of its environmental and energetic applications. DDC in porous media involves multiple physical processes related to the flow, heat transfer, mass transfer, and buoyancy forces. For this reason, it is an appropriate concrete problem for testing new methods of sensitivity analysis.

The problem under consideration is that of a square porous cavity with horizontal mass and thermal gradients (Figure 10). The left and right vertical walls of the cavity are subjected to normalized temperatures and concentrations $T_L = C_L = 1$ and $T_R = C_R = 0$, respectively. The horizontal surfaces are assumed to be adiabatic and impermeable. The heat and mass gradients generate buoyancy forces and yield to a rotating unit cell within the cavity. This problem has played a key role in the investigation of the DDC in porous

media. However, as far as our knowledge extends, no global sensitivity analysis has ever been conducted to identify the most relevant model inputs for DDC.

To model DDC in porous media, the fluid flow in the cavity is assumed to comply with Darcy's law and the Boussinesq approximation. The porous medium is in local thermal equilibrium with the fluid. Under these conditions, the steady state governing equations of the fluid flow as well as the heat and mass transfer inside the porous cavity can be written in the following non-dimensional form:

$$\frac{\partial u}{\partial x} + \frac{\partial v}{\partial y} = 0 \tag{52}$$

$$\frac{H^2}{K} u = -\frac{\partial p_t}{\partial x} \tag{53}$$

$$\frac{H^2}{K} v = -\frac{\partial p_t}{\partial y} + Gr_T \left(T + Nr.C\right) \tag{54}$$

$$u\frac{\partial T}{\partial x} + v\frac{\partial T}{\partial y} = \frac{R_k}{Pr}\left(\frac{\partial^2 T}{\partial x^2} + \frac{\partial^2 T}{\partial y^2}\right) \tag{55}$$

$$u\frac{\partial C}{\partial x} + v\frac{\partial C}{\partial y} = \frac{1}{Le.Pr}\left(\frac{\partial^2 C}{\partial x^2} + \frac{\partial^2 C}{\partial y^2}\right) \tag{56}$$

where H is the size of the square cavity, u and v are the velocity components in the $x \in [0,1]$ and $y \in [0,1]$ directions, p_t is the total pressure of the system including the fluid pressure and gravitational head, T and C are the temperature and concentration, respectively, R_k is the ratio of the effective thermal diffusivity of the porous medium to that of the fluid, Pr, Le and Gr_T are the Prandtl, Lewis and thermal Grashof numbers, respectively, Nr is the buoyancy ratio, and K is the permeability of the stratified porous medium, defined by

$$K(x,y) = K_0 e^{\zeta_1 Hx + \zeta_2 Hy} \tag{57}$$

Here, K_0 is the permeability at the origin, while ζ_1 and ζ_2 are the rates of change of $\ln(K)$ in the x and y directions. In the simulation, the average permeability and Rayleigh number are defined for the heterogeneous porous medium as follows:

$$\overline{K} = \int_0^1 \int_0^1 K(x,y)\mathrm{d}x\mathrm{d}y \tag{58}$$

$$\overline{Ra} = \frac{\overline{K}.Pr.Gr_T}{H^2} \tag{59}$$

The system is solved using the highly accurate Fourier–Galerkin (FG) method described in [39, 40]. The FG method is one of the most popular

spectral methods, used extensively in large-scale computations to solve partial differential equations because of its ability to achieve high precision with a relatively small number of degrees of freedom. The method consists in expanding the unknowns (stream function, temperature and concentration) into the appropriate Fourier series truncated at given orders. As shown in [40], the number of Fourier modes required to obtain stable solutions with the FG method depends on the average Rayleigh number ($\overline{Ra}$) and the level of heterogeneity (ζ_1 and ζ_2). In this paper, the sensitivity analysis requires several evaluations of the model. Hence, moderate ranges of $\overline{Ra}$ and (ζ_1, ζ_2) are used in order to obtain accurate solutions in an affordable CPU time (see Table 3). The level of truncation orders of the Fourier series is fixed to be 90 and 30 in the x and y directions, respectively. Thus a total number of $3 \times 90 \times 30 = 8100$ Fourier coefficients are determined in the computation.

The model parameters are listed in Table 3. The size of the cavity H and the ratio of thermal diffusivity R_k are considered as deterministic. The Prandtl number, the average Rayleigh number, the average permeability, the rates of change of the permeability in the x and y directions, the Lewis number, and the buoyancy ratio are assumed to be random parameters. Their properties are specified in Table 3. These parameters are gathered into a random vector $\boldsymbol{X} = (Pr, \overline{Ra}, \overline{K}, \zeta_1, \zeta_2, Le, Nr)^T$ of dimension $n = 7$. All seven random variables are supposed to be uniformly and independently distributed.

The model responses of interest are the average Nusselt and Sherwood numbers ($\overline{Nu}, \overline{Sh}$), as well as the maximum velocity components (u_{max}, v_{max}) in the x and y directions. These variables are commonly used to investigate DDC problems because they provide a quantitative idea of the fluid circulation velocity, as well as the heat and mass fluxes. The average Nusselt number (resp. Sherwood number) represents the ratio of the total rate of heat transfer (resp. mass transfer) to the rate of conductive heat transfer (resp. diffusive mass transport) across the boundary. They are defined by

$$\overline{Nu} = \int_0^1 \left.\frac{\partial T}{\partial x}\right|_{x=0} \mathrm{d}y \tag{60}$$

$$\overline{Sh} = \int_0^1 \left.\frac{\partial C}{\partial x}\right|_{x=0} \mathrm{d}y \tag{61}$$

7.2. Sensitivity analysis

The sensitivity indices of the input parameters ($Pr, \overline{Ra}, \overline{K}, \zeta_1, \zeta_2, Le, Nr$) are estimated using the Bayesian sparse PCE approach. The initial size of the quasi-random experimental design is $N = 2^5$. The experimental design is progressively enriched until a satisfactory solution is reached ($\sigma^2_{\mathcal{A}} < 5 \times 10^{-2}$).

In this example, a satisfactory sparse PCE is obtained with a sample size of $N = 2^7$, for which we found $\sigma^2_{\mathcal{A}}(\overline{Nu}) = 3.1 \times 10^{-3}$, $\sigma^2_{\mathcal{A}}(\overline{Sh}) = 3.9 \times 10^{-3}$, $\sigma^2_{\mathcal{A}}(u_{max}) = 2.1 \times 10^{-2}$, and $\sigma^2_{\mathcal{A}}(v_{max}) = 1.5 \times 10^{-2}$. The 95% credible interval of each sensitivity index is derived using the scheme outlined in Section 5.2.

The estimates of the first-order and total sensitivity indices are sketched in Figure 11. In this figure, the credible intervals are quite narrow for all the sensitivity indices, which demonstrates the validity of the sparse PCE estimation. It also appears that the variances of the model responses are due to distinct random variables. For instance, for the first model response $\overline{Nu}$, the variance is mainly explained by the average Rayleigh number $\overline{Ra}$ and buoyancy ratio Nr. As a matter of fact, the Nusselt number represents the dimensionless thermal diffusive flux through the hot wall, which is proportional to the temperature gradient. The latter is in turn proportional to the thickness of the thermal boundary layer generated by the fluid circulation within the porous cavity. The origin of this flow lies in the buoyancy forces, which are mainly controlled by the Rayleigh number and the buoyancy ratio for the thermal effects. The same behaviour is noted for the average Sherwood number $\overline{Sh}$, which represents the solute diffusive flux at the salted wall. However, for $\overline{Sh}$, we can notice the additional influence of the Lewis number Le, which is the main parameter controlling the mass diffusivity. For both $\overline{Nu}$ and $\overline{Sh}$, Figure 11 shows that $\overline{Ra}$ has the most significant contribution. Besides, for both $\overline{Nu}$ and $\overline{Sh}$, the total sensitivity indices are close to the first-order sensitivity indices, which means that the relations between the model responses $(\overline{Nu}, \overline{Sh})$ and the parameters are practically additive (negligible interactions). Regarding the model responses u_{max} and v_{max}, it is shown in Figure 11 that the influence of Pr and $\overline{Ra}$ are both significant. This may be explained by the fact that Pr and $\overline{Ra}$ are related to the viscous and buoyancy effects, which play the key roles in the fluid flow process. The total sensitivity indices of Pr and $\overline{Ra}$ are greater than their first-order sensitivity indices, which indicates the existence of interactions between them.

The identified sparse PCEs only contain 13, 15, 16 and 15 terms in the expansions for $\overline{Nu}$, $\overline{Sh}$, u_{max}, and v_{max}, respectively, which indicates a high sparsity. Taking the model response u_{max}, for example, the maximum degree of the sparse PCE is $p_{\mathcal{A}} = 6$. Thus a full PCE of the same degree would contain $P = 10,296$ terms, which yields a sparsity index of 1.5×10^{-3}.

8. Conclusion

In this paper, a new algorithm is proposed to build sparse PCEs of computer model responses. They are then used to compute Sobol' indices for global sensitivity analysis. The new algorithm relies on the statistical ap-

proach called BMA, which performs quantitative comparisons of competing models. In particular, the model selection criterion KIC is adopted to identify the best sparse PCE from a given data set. We propose a new algorithm to construct a sparse PCE that only contains the significant polynomial terms for the data set at hand. Thus, a reduced number of coefficients is estimated. Using the analytical expression of the MAP, the retained coefficients are computed with a low number of model evaluations (as compared to the full PCE), avoiding the problem of a poorly conditioned matrix encountered with the regression method.

The proposed algorithm is tested on four examples of applications, including three analytical functions, namely the Ishigami function (smooth), the Sobol' function (non-smooth), and the Morris function (high-dimensional). The fourth example is a Fourier–Galerkin model for double-diffusive convection in a heterogeneous porous cavity. High sparsity is found in these examples, with a sparsity index varying from 0% to 6%. Using Bayesian sparse PCE, good estimates of the sensitivity indices can be obtained with a rather low number of model evaluations compared to full PCE.

Moreover, thanks to the Bayesian framework, the posterior distribution of the coefficients in the sparse PCE can be easily computed. Accordingly, it is possible to assign credible interval to each sensitivity index with little computational effort. Furthermore, the accuracy of the sparse PCE is evaluated by both the residual and the size of the credible intervals. Although our study reveals that these intervals can be underestimated, they allow for gauging the degree of importance of the input variables. The numerical exercises also show that the proposed Bayesian sparse PCE can yield estimates of the sensitivity indices with low bias and variance.

Acknowledgement

This work has been funded by the French National Research Agency through the research project RESAIN (n° ANR-12-BS06-0010-02).

References

[1] A. Saltelli, M. Ratto, T. Andres, F. Campolongo, J. Cariboni, D. Gatelli, M. Saisana, S. Tarantola, Global Sensitivity Analysis: The Primer, Probability and Statistics, John Wiley and Sons, Chichester, 2008.

[2] R. G. Ghanem, S. P. Spanos, Stochastic finite elements: A spectral approach, Springer-Verlag, Berlin, 1991.

[3] M. A. Tatang, W. W. Pan, R. G. Prin, G. J. McRae, An efficient method for parametric uncertainty analysis of numerical geophisical model, J. Geophysics Research 102 (1998) 21925–21932.

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

[5] O. Le Maître, M. T. Reagan, H. N. Najm, R. G. Ghanem, O. M. Knio, A stochastic projection method for fluid flow: II. Random process, Journal of Computational Physics 181 (9–44).

[6] H. G. Matthies, A. Keese, Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations, Computer Methods in Applied Mechanics & Engineering (2005) 1295–1331.

[7] B. Sudret, Global sensitivity analysis using polynomial chaos expansions, Reliability Engineering and System Safety 93 (2008) 964–979.

[8] N. Fajraoui, F. Ramasomanana, A. Younes, T. A. Mara, P. Ackerer, A. Guadagnini, Use of global sensitivity analysis and polynomial chaos expansion for interpretation of nonreactive transport experiments in laboratory-scale porous media, Water Resources Research 47, W02521 (2011) doi:10.1029/2010WR009639.

[9] M. Riva, A. Guadagnini, A. Dell'Oca, Probabilistic assessment of seawater intrusion under multiple sources of uncertainty, Advances in Water Resources 75 (2015) 93–104.

[10] G. Blatman, B. Sudret, Sparse polynomial chaos expansions and adaptive stochastic finite elements using a regression approach, Comptes Rendus de Mécanique 336 (2008) 518–523.

[11] G. Blatman, B. Sudret, Efficient computation of global sensitivity indices using sparse polynomial chaos expansions, Reliability Engineering and System Safety 95 (11) (2010) 1216–1229.

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

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

[14] C. Hu, B. D. Youn, Adaptive-sparse polynomial chaos expansion for reliability and design of complex engineering systems, Structural and Multidisciplinary Optimization 43 (2010) 419–442.

[15] N. Fajraoui, T. A. Mara, A. Younes, R. Bouhlila, Reactive transport parameter estimation and global sensitivity analysis using sparse polynomial chaos expansion, Water, Air and Soil Pollution 223 (2012) 4183–4197.

[16] S. P. Neuman, Maximum likelihood Bayesian averaging of uncertain model predictions, Stochastic Environmental Research and Risk Assessment 17 (5) (2003) 291–305.

[17] A. Schöniger, T. Wöhling, L. Samaniego, W. Novak, Model selection on solid ground: Rigourous comparison of nine ways to evaluate Bayesian evidence, Water Resources Research 50 (2014) 9484–9513.

[18] R. L. Kayshap, Optimal choice of AR and MA parts in autoregressive moving average models, IEEE Trans. Pattern Anal. Machine Intell. 4 (2) (1982) 99–104.

[19] T. Homma, A. Saltelli, Importance measures in global sensitivity analysis of nonlinear models, Reliability Engineering and System Safety 52 (1996) 1–17.

[20] R. I. Cukier, C. M. Fortuin, K. E. Shuler, A. G. Petschek, J. H. Schaibly, Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I. theory, J. Chemical Physics 59 (1973) 3873–3878.

[21] A. Saltelli, S. Tarantola, K. Chan, A quantitative model independent method for global sensitivity analysis of model output, Technometrics 41 (1999) 39–56.

[22] T. A. Mara, Extension of the rbd-fast method to the computation of global sensitivity indices, Reliability Engineering and System Safety 94 (2009) 1274–1281.

[23] I. M. Sobol', Sensitivity estimates for nonlinear mathematical models, Math. Mod. and Comput. Exp. 1 (1993) 407–414.

[24] M. J. J. Jansen, Analysis of variance designs for model output, Computer Physics Communication 117 (1999) 35–43.

[25] A. Saltelli, Making best use of model evaluations to compute sensitivity indices, Computational Physics Communications 145 (2002) 280–297.

[26] H. Rabitz, O. Alis, J. Shorter, K. Shim, Efficient input-output model representations, Computer Physics Communications 117 (1999) 11–20.

[27] J. E. Oakley, A. O'Hagan, Probabilistic sensitivity analysis of complex models: a Bayesian approach, J. Royal Statist. Soc. B 66 (2004) 751–769.

[28] M. Ratto, A. Pagano, P. Young, State dependent parameter metamodelling and sensitivity analysis, Computer Physics Communications 117 (11) (2007) 863–876.

[29] G. T. Buzzard, D. Xiu, Variance-based global sensitivity analysis via sparse-grid interpolation and cubature, Communications in Computational Physics 9 (2011) 542–567.

[30] T. Crestaux, O. L. Maître, J.-M. Martinez, Polynomial chaos expansion for sensitivity analysis, Reliability Engineering and System Safety 94 (2009) 1161–1172.

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

[32] S. D. Babacan, R. Molina, A. K. Katsaggelos, Bayesian compressive sensing using Laplace priors, IEEE Trans. on Image Process. (2010) 53–63.

[33] A. E. Hoerl, R. Kennerd, Ridge regression: Biased estimation for nonorthogonal problems, Technometrics 12 (1970) 55–67.

[34] R. Tibshirani, Regression shrinkage and selection via the LASSO, Journal of the Royal Statistical Society, Serie B 58 (1) (1996) 267–288.

[35] I. M. Sobol', V. I. Turchaninov, Y. L. Levitan, B. V. Shukman, Quasi-random sequence generator (routine LPTAU51), Keldysh Institute of Applied Mathematics, Russian Academy of Sciences (1992).

[36] A. Saltelli, K. Chan, E. M. Scott, Sensitivity analysis, John Wiley and Sons, Chichester, 2000.

[37] T. Ishigami, T. Homma, An importance quantification technique in uncertainty analysis for computer models, in: First International Symposium on Uncertainty Modeling and Analysis, IEEE, 1990, pp. 398–403.

Table 1: Ishigami function—Sensitivity indices estimated with Bayesian sparse PCE versus the number of model runs

	Ref. Val.	$N = 2^5$		$N = 2^6$		$N = 2^7$		$N = 2^8$	
		Est.	Err. (%)	Est.	Err. (%)	Est.	Err. (%)	Est.	Err. (%)
S_1	0.31	0.35	12.8	0.31	-1.7	0.31	-0.7	0.31	0.0
S_2	0.44	0.65	46.0	0.45	2.2	0.44	0.3	0.44	0.0
S_3	0.00	0.00	-	0.00	-	0.00	-	0.00	-
S_1^T	0.56	0.35	-36.5	0.55	-1.8	0.56	-0.2	0.56	0.0
S_2^T	0.44	0.65	46.0	0.45	2.2	0.44	0.2	0.44	0.0
S_3^T	0.24	0.00	-100.0	0.24	-1.9	0.24	0.5	0.24	0.0
$p_{\mathcal{A}}$		6		9		9		12	
$q_{\mathcal{A}}$		1		2		2		2	
$P_{\mathcal{A}}$		5		13		13		17	
IS		6.0×10^{-2}		6.0×10^{-2}		6.0×10^{-2}		3.7×10^{-2}	
$\sigma^2_{\mathcal{A}}$		3.1×10^{-1}		1.2×10^{-3}		2.1×10^{-3}		1.2×10^{-5}	

$N = 2^9$		$N = 2^{10}$		$N = 2^{11}$		$N = 2^{12}$		$N = 2^{13}$	
Est.	Err. (%)	Est.	Err. (%)	Est.	Err. (%)	Est.	Err. (%)	Est.	Err. (%)
0.31	0.0	0.31	0.0	0.31	0.0	0.31	0.0	0.31	0.0
0.44	0.0	0.44	0.0	0.44	0.0	0.44	0.0	0.44	0.0
0.00	-	0.00	-	0.00	-	0.00	-	0.00	-
0.56	0.0	0.56	0.0	0.56	0.0	0.56	0.0	0.56	0.0
0.44	0.0	0.44	0.0	0.44	0.0	0.44	0.0	0.44	0.0
0.24	0.0	0.24	0.0	0.24	0.0	0.24	0.0	0.24	0.0
12		14		14		14		16	
2		2		2		2		2	
20		22		22		22		25	
4.4×10^{-2}		3.2×10^{-2}		3.2×10^{-2}		3.2×10^{-2}		2.6×10^{-2}	
1.9×10^{-8}		1.5×10^{-8}		1.5×10^{-8}		1.5×10^{-8}		9.0×10^{-12}	

[38] M. D. Morris, Factorial sampling plans for preliminary computational experiments, Technometrics 33 (1991) 161–174.

[39] Q. Shao, M. Fahs, A. Younes, A. Makradi, A high-accurate solution for darcy–brinkman double-diffusive convection in saturated porous media, Numerical Heat Transfer, Part B: Fundamentals 69 (2016) 26–47.

[40] M. Fahs, A. Younes, A. Makradi, A reference benchmark solution for free convection in a square cavity filled with a heterogeneous porous medium, Numerical Heat Transfer, Part B: Fundamentals 67 (2015) 437–462.

Table 2: Sobol' function—Sensitivity indices estimated with Bayesian sparse PCE versus the number of model runs

	Ref. Val.	$N = 2^8$		$N = 2^9$		$N = 2^{10}$		$N = 2^{11}$		$N = 2^{12}$		$N = 2^{13}$	
		Est.	Err. (%)	Est.	Err. (%)	Est.	Err. (%)	Est.	Err. (%)	Est.	Err. (%)	Est.	Err. (%)
S_1	0.19	0.26	33.8	0.21	6.1	0.23	19.0	0.21	8.5	0.20	3.0	0.20	2.1
S_2	0.12	0.18	43.4	0.16	26.2	0.16	25.5	0.14	8.4	0.13	3.7	0.13	2.7
S_3	0.09	0.13	55.9	0.11	26.7	0.08	-7.3	0.10	10.3	0.09	1.9	0.09	1.7
S_4	0.06	0.05	-14.6	0.08	22.5	0.08	33.2	0.07	4.8	0.07	6.8	0.06	1.7
S_5	0.05	0.05	5.7	0.05	5.2	0.06	29.6	0.05	9.1	0.05	3.1	0.05	-0.8
S_6	0.04	0.02	-53.1	0.03	-22.3	0.04	6.3	0.04	5.2	0.04	7.6	0.04	2.0
S_7	0.03	0.03	-14.8	0.03	-3.4	0.03	6.7	0.03	8.7	0.03	4.1	0.03	2.0
S_8	0.03	0.02	-3.4	0.03	31.7	0.03	2.1	0.03	10.7	0.03	4.6	0.03	3.5
S_9	0.02	0.03	34.5	0.02	-9.2	0.02	6.1	0.02	15.1	0.02	4.5	0.02	2.6
S_1^T	0.40	0.41	4.2	0.39	-1.5	0.37	-5.9	0.39	-1.8	0.39	-0.3	0.39	-0.3
S_2^T	0.28	0.29	3.6	0.19	4.0	0.29	4.4	0.26	-6.0	0.27	-2.8	0.28	-0.2
S_3^T	0.20	0.22	6.1	0.21	4.5	0.15	-28.7	0.20	-2.4	0.19	-5.6	0.20	-2.4
S_4^T	0.16	0.07	-56.2	0.17	7.1	0.16	4.3	0.14	-11.7	0.16	0.8	0.15	-2.3
S_5^T	0.12	0.10	-20.0	0.08	-30.6	0.14	14.2	0.11	-8.0	0.11	-6.4	0.11	-8.0
S_6^T	0.10	0.03	-70.0	0.08	-22.6	0.10	-0.1	0.08	-16.3	0.10	-2.2	0.09	-5.1
S_7^T	0.08	0.04	-50.3	0.06	-30.3	0.07	-16.9	0.07	-16.9	0.08	-4.5	0.08	-4.4
S_8^T	0.07	0.04	-40.4	0.05	-31.7	0.05	-30.8	0.06	-7.3	0.06	-4.2	0.07	0.0
S_9^T	0.06	0.03	-48.6	0.05	-17.5	0.07	27.6	0.05	-5.2	0.05	-6.0	0.06	-2.1
$p_{\mathcal{A}}$		4		8		18		16		18		16	
$q_{\mathcal{A}}$		2		4		7		8		8		8	
$P_{\mathcal{A}}$		19		30		45		99		174		321	
IS		2.7×10^{-2}		1.2×10^{-3}		9.6×10^{-6}		4.8×10^{-5}		3.7×10^{-5}		1.6×10^{-4}	
$\sigma^2_{\mathcal{A}}$		2.1×10^{-1}		1.8×10^{-1}		1.6×10^{-1}		5.8×10^{-2}		3.3×10^{-2}		1.9×10^{-2}	

Table 3: Parameters of the double-diffusive convection model

Parameter	Notation	Type of PDF	Range of values
Size of cavity	H	Deterministic	1.0
Ratio of thermal diffusivity	R_k	Deterministic	1.0
Prandtl number	Pr	Uniform	[0.5, 2.0]
Average Rayleigh number	$\overline{Ra}$	Uniform	[1.0, 100.0]
Average permeability	$\overline{K}$	Uniform	$[10^{-9}, 10^{-7}]$
Rate of change in x direction	ζ_1	Uniform	[0.0, 2.0]
Rate of change y direction	ζ_2	Uniform	[0.0, 2.0]
Lewis number	Le	Uniform	[1.0, 5.0]
Buoyancy ratio	Nr	Uniform	[0.0, 2.0]

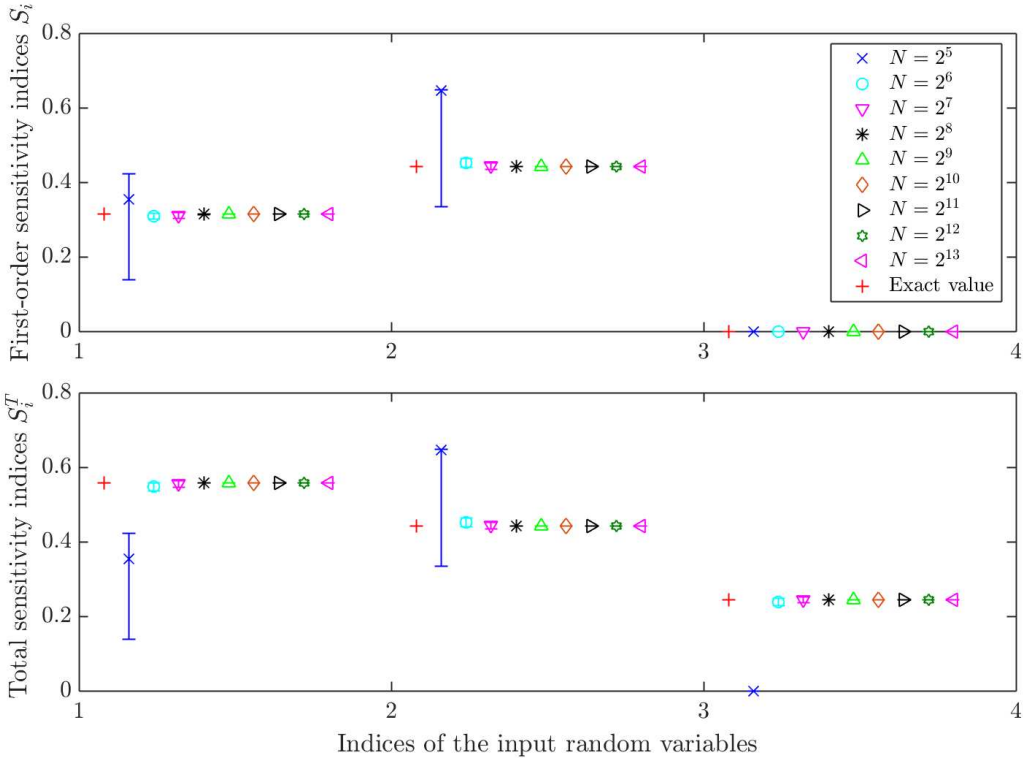


Figure 1: Ishigami function—95% confidence intervals of the first-order and total sensitivity indices computed from Bayesian sparse PCE with different numbers of model evaluations

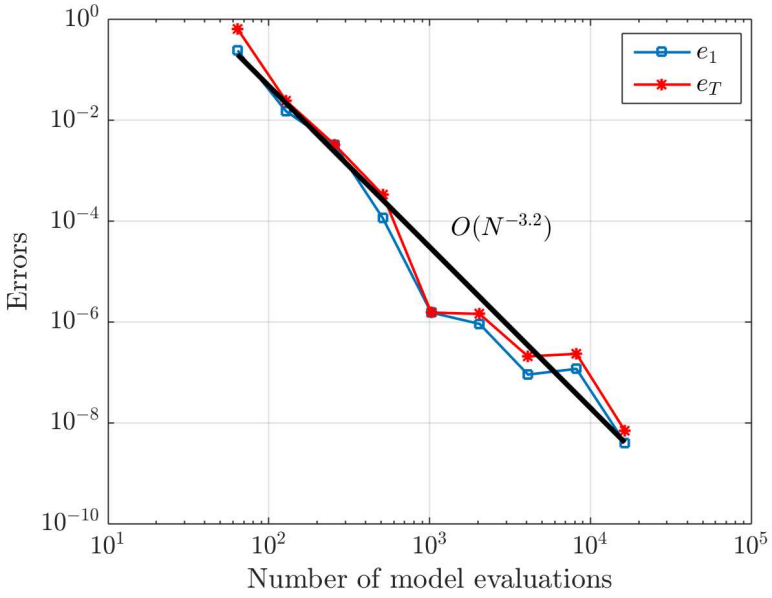


Figure 2: Ishigami function—Absolute errors of first-order and total sensitivity indices computed as functions of the number of model evaluations

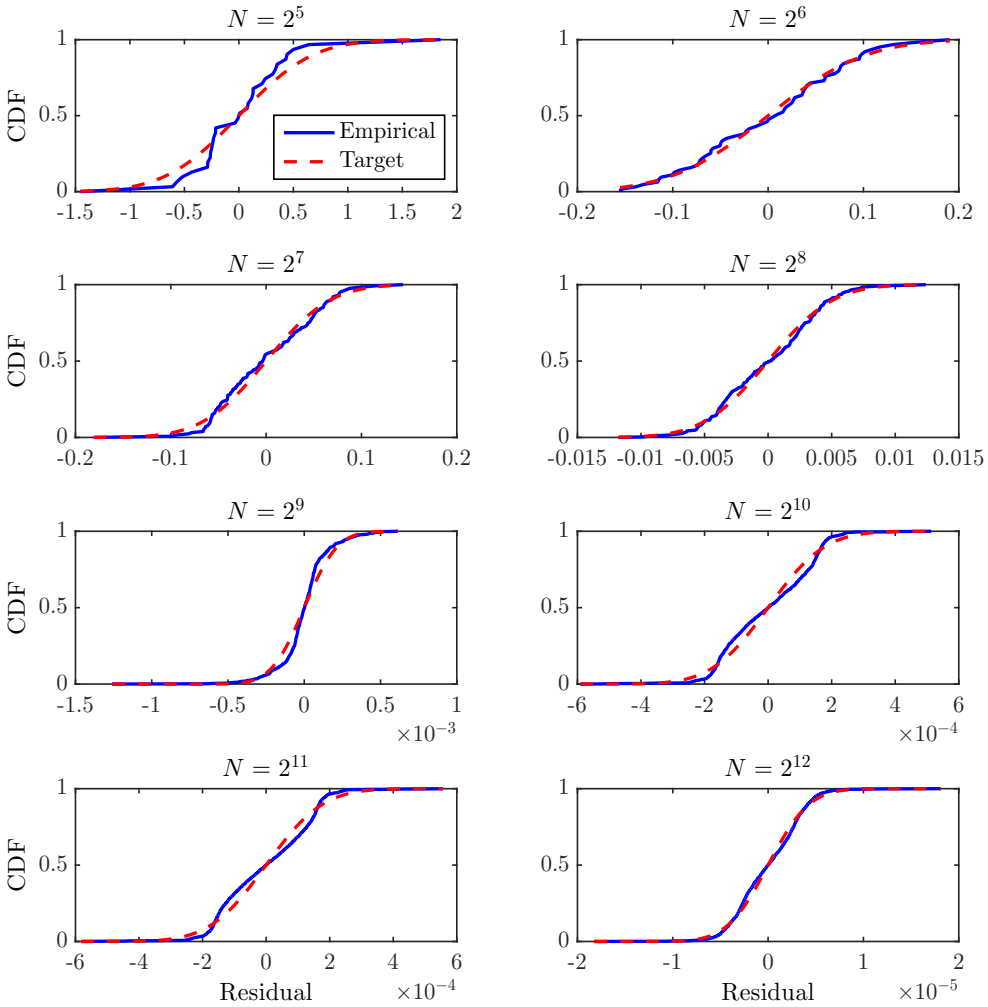


Figure 3: Ishigami function—Comparison of the empirical CDF of the residual with the target CDF at different sample sizes

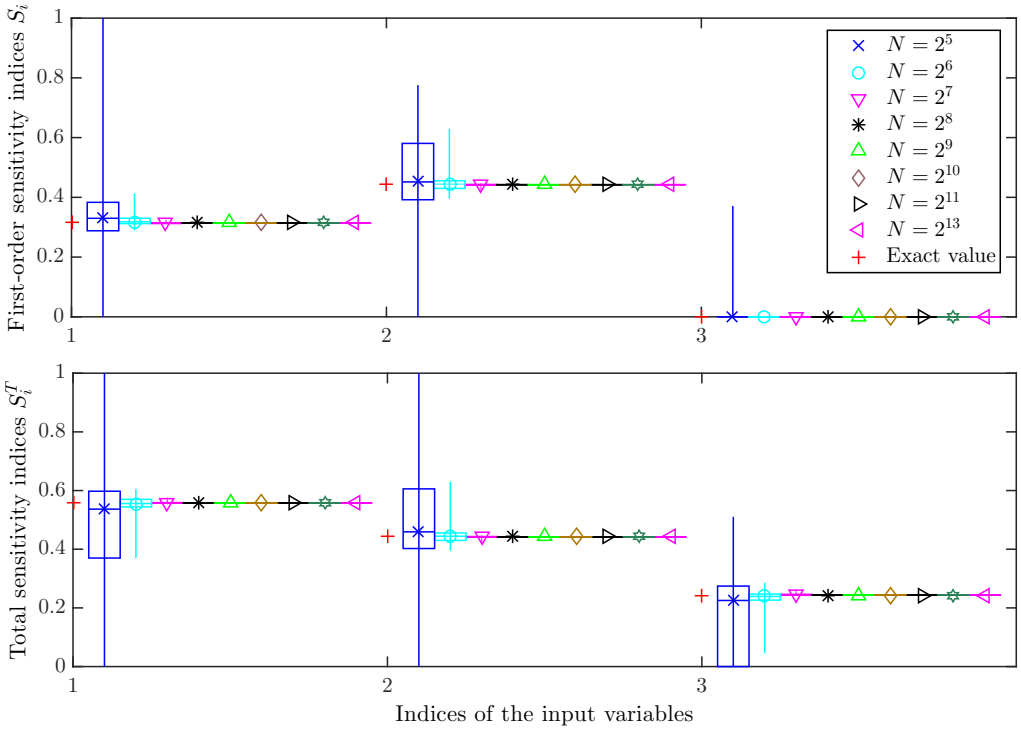


Figure 4: Ishigami function—Effect of the experimental design onto the sensitivity indices estimate. The symbols represent the median value over 100 replicate estimates. The whisker represents the range of variation of the estimates. The bottom and top of the boxes are the first and third quartiles

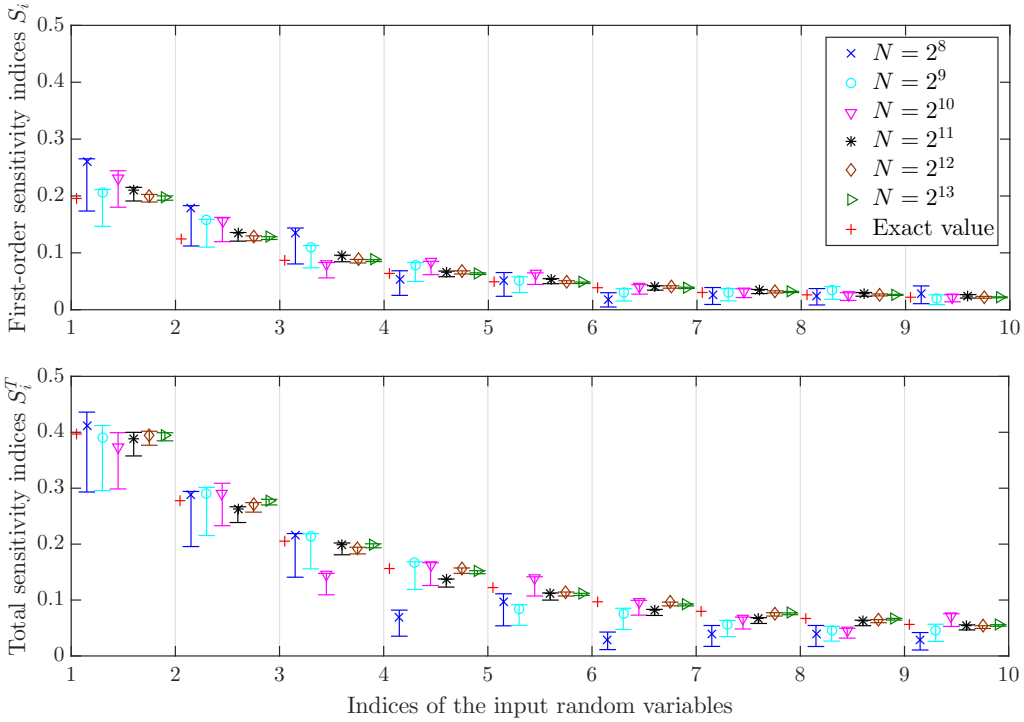


Figure 5: Sobol' function—95% credible intervals of the first-order and total sensitivity indices computed from Bayesian sparse PCE with different numbers of model evaluations

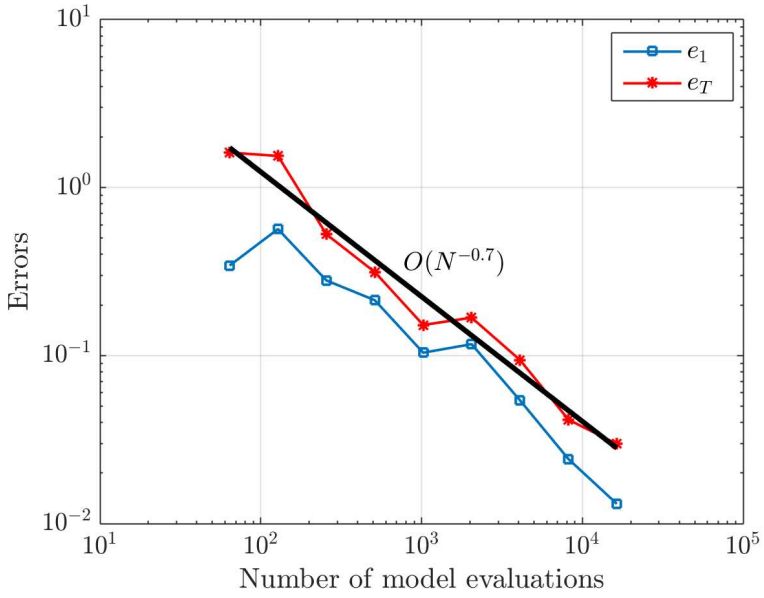


Figure 6: Sobol' function—Absolute errors of first-order and total sensitivity indices computed as functions of the number of model evaluations

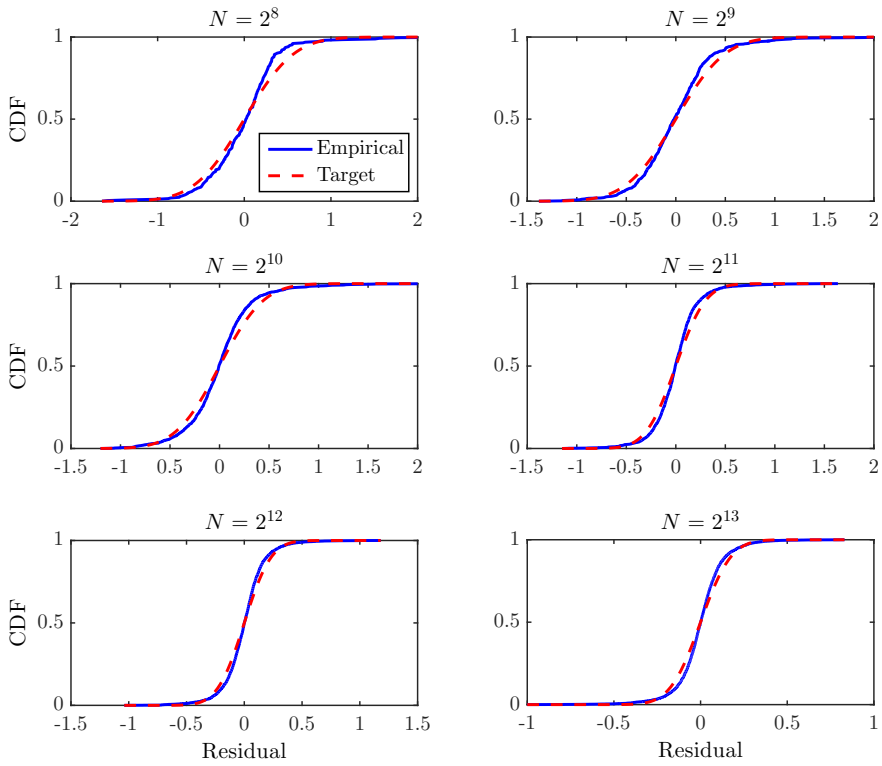


Figure 7: Sobol' function—Comparison of the empirical CDF of the residual with the target CDF at different sample sizes

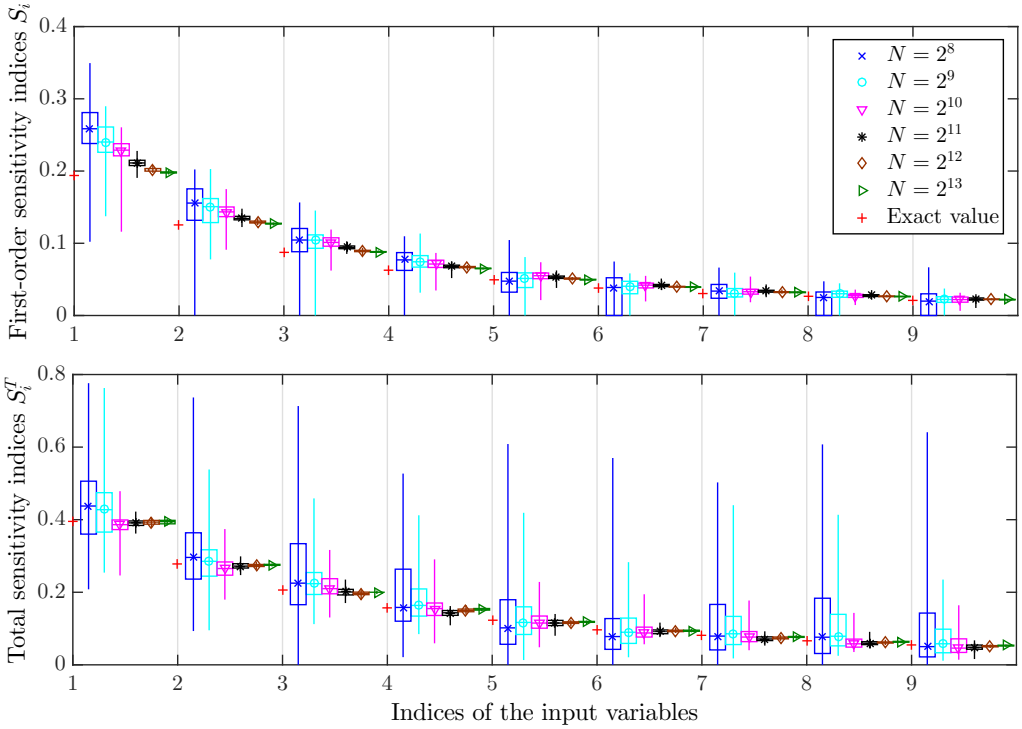


Figure 8: Sobol' function—Effect of the experimental design onto the sensitivity indices estimate (see Figure 4 for details)

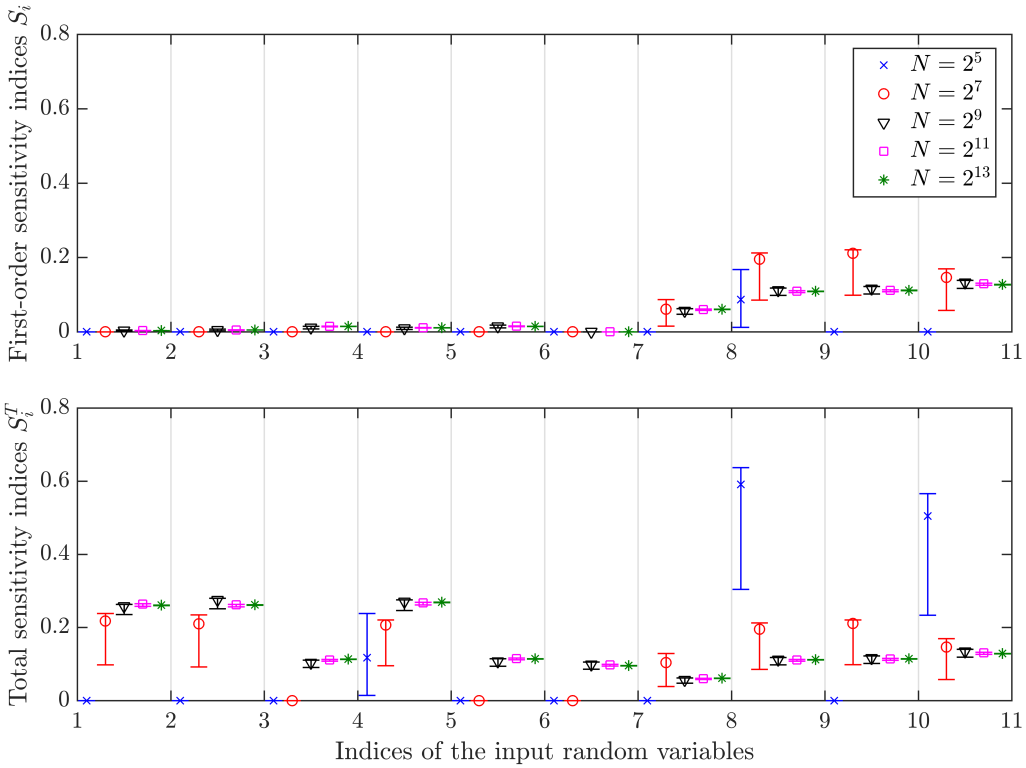


Figure 9: Morris function—95% credible intervals of the first-order and total sensitivity indices computed from Bayesian sparse PCE with different numbers of model evaluations

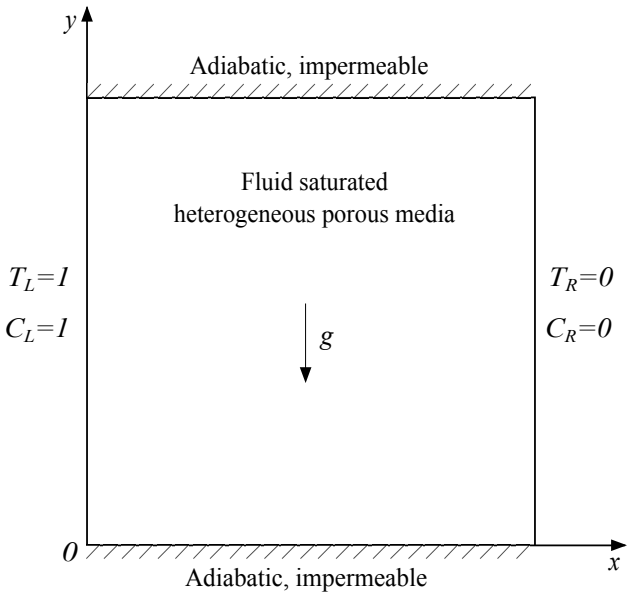


Figure 10: The double-diffusive natural convection problem in the heterogeneous porous cavity

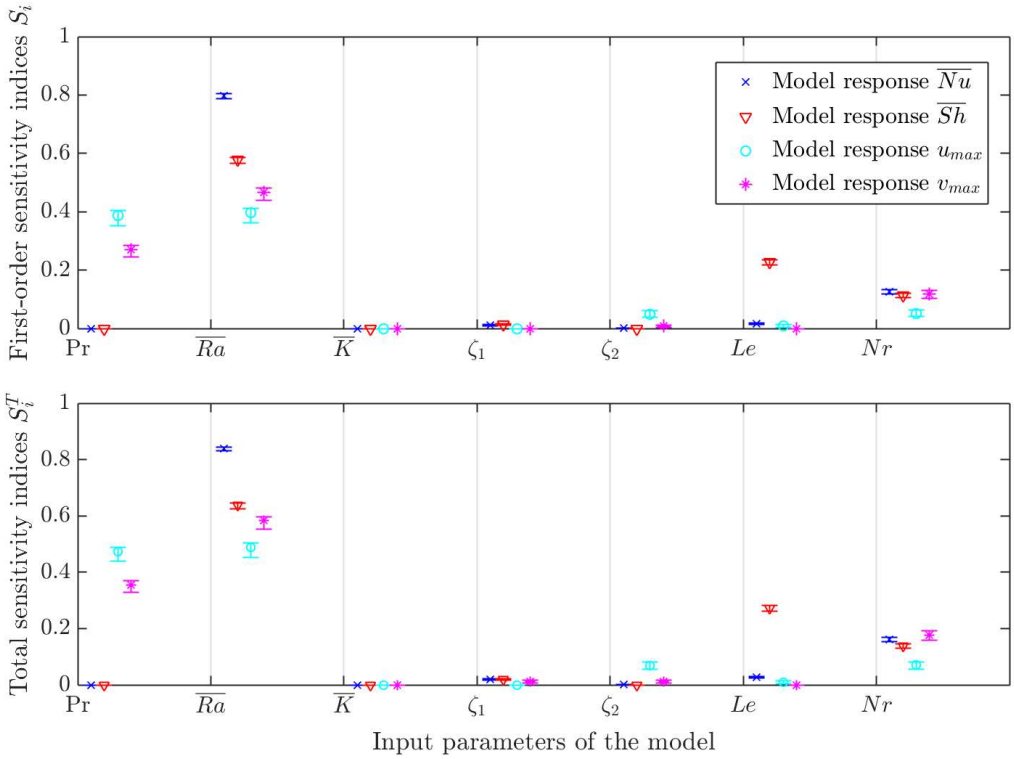


Figure 11: Double-diffusive convection—95% confidence intervals of the first-order and total sensitivity indices computed from Bayesian sparse PCE for different model responses