HAL
archives-ouvertes.fr

Adaptive surrogate modeling by ANOVA and sparse polynomial dimensional decomposition for global sensitivity analysis in fluids simulation

Kunkun Tang, Pietro Marco Congedo, Remi Abgrall

▶ To cite this version:

Kunkun Tang, Pietro Marco Congedo, Remi Abgrall. Adaptive surrogate modeling by ANOVA and sparse polynomial dimensional decomposition for global sensitivity analysis in fluids simulation. [Research Report] RR-8758, Inria Bordeaux Sud-Ouest. 2015. hal-01178398

HAL Id: hal-01178398
https://hal.inria.fr/hal-01178398

Submitted on 20 Jul 2015

Adaptive surrogate modeling by ANOVA and sparse polynomial dimensional decomposition for global sensitivity analysis in fluids simulation

Kunkun Tang, Pietro M. Congedo , Rémi Abgrall

ISSN 0249-6399　　ISRN INRIA/RR--8758--FR+ENG

Adaptive surrogate modeling by ANOVA and sparse polynomial dimensional decomposition for global sensitivity analysis in fluids simulation

Kunkun Tang [*], Pietro M. Congedo [*], Rémi Abgrall[†]

Project-Team Cardamom

[*] INRIA Bordeaux Sud-Ouest, Team Cardamom, 200 avenue de la Vieille Tour, 33405 Talence, France
[†] Institut für Mathematik, Universität Zürich, Winterthurerstrasse 190, CH-8057 Zürich, Switzerland

Abstract: The polynomial dimensional decomposition (PDD) is employed in this work for the global sensitivity analysis and uncertainty quantification (UQ) of stochastic systems subject to a moderate to large number of input random variables. Due to the intimate structure between the PDD and the Analysis of Variance (ANOVA) approach, PDD is able to provide a simpler and more direct evaluation of the Sobol' sensitivity indices, when compared to the Polynomial Chaos expansion (PC). Unfortunately, the number of PDD terms grows exponentially with respect to the size of the input random vector, which makes the computational cost of standard methods unaffordable for real engineering applications. In order to address the problem of the curse of dimensionality, this work proposes essentially variance-based adaptive strategies aiming to build a cheap meta-model (i.e. surrogate model) by employing the sparse PDD approach with its coefficients computed by regression. Three levels of adaptivity are carried out in this paper: 1) the *truncated dimensionality* for ANOVA component functions, 2) the *active dimension* technique especially for second- and higher-order parameter interactions, and 3) the *stepwise regression* approach designed to retain only the most influential polynomials in the PDD expansion. During this adaptive procedure featuring stepwise regressions, the surrogate model representation keeps containing few terms, so that the cost to resolve repeatedly the linear systems of the least-square regression problem is negligible. The size of the finally obtained sparse PDD representation is much smaller than the one of the full expansion, since only significant terms are eventually retained. Consequently, a much less number of calls to the deterministic model is required to compute the final PDD coefficients.

Key-words: Uncertainty quantification; Global sensitivity analysis; Analysis of Variance (ANOVA); Polynomial dimensional decomposition (PDD); Regression approach; Adaptive sparse polynomial surrogate model; Atmospheric reentry

Adaptive surrogate modeling by ANOVA and sparse polynomial dimensional decomposition for global sensitivity analysis in fluids simulation

Résumé : Pas de résumé

Mots-clés : Pas de motclef

Contents

1 Introduction

Compared to local sensitivity analysis methods, global sensitivity analysis (GSA) has the advantage of taking into account the overall influence of input parameters and their interactions onto the output quantity of interest, by considering the entire input space and not depending on a specific nominal point (see for example [8]).

However, the main difficulty encountered when employing the global methods is the required high cost of numerical computations, since the Monte Carlo simulation (MC) or a quasi Monte Carlo method (QMC) is usually applied to estimate the sensitivity indices. For complex problems in the real life, relying upon MC or QMC can be very expensive from a computational point of view.

Traditionally, GSA is performed using methods based on the decomposition of the output variance [36], *i.e.* Analysis of Variance (ANOVA), which is nowadays one of the most commonly used GSA techniques in the literature. Indeed, ANOVA relies on a functional decomposition that incorporates component functions involving a single uncertain parameter, or a group of parameters, and the computation of the sensitivity measures of each component function usually requires MC or QMC simulations.

In 2001, Sobol' used this formulation to define global sensitivity indices [36], displaying the relative variance contributions of different ANOVA terms. In [27, 28], they introduced two High-Dimensional Model Representation (HDMR) techniques to capture input-output relationships of physical systems with many input variables. These techniques are based on ANOVA-type decompositions. Since it usually requires a large number of function evaluations to perform this analysis, several techniques have been developed to compute the different so-called sensitivity indices at low cost [33]. In [37, 14, 4], the generalized Polynomial Chaos expansions (gPC) are used to build surrogate models for computing the Sobol' indices analytically as a post-processing of the PC coefficients. In [15], they combine the multi-element polynomial chaos with the ANOVA functional decomposition to enhance the convergence rate of polynomial chaos in high dimensions and in problems with low stochastic regularity. The use of adaptive ANOVA decomposition is investigated in [49] as an effective dimension-reduction technique in modeling incompressible and compressible flows with a high dimensionality of random space. In Sudret [4], the sparse Polynomial Chaos (PC) expansions are introduced in order to compute the global sensitivity indices. An adaptive algorithm allows to build a PC-based meta-model that only contains the significant terms whereas the PC coefficients are computed by the least-square regression. Other approaches are developed if the assumption of independence of the input parameters is not valid. New indices have been proposed to address the dependence [46, 47], but these attempts are limited to a linear correlation. In [7], they introduce a global sensitivity indicator which looks at the influence of input uncertainty on the entire output distribution without reference to a specific moment of the output (moment independence) and which can also be defined in the presence of correlations among the parameters. In [11], a gPC methodology to address the GSA for this kind of problems is introduced. A moment-independent sensitivity index that suits problems with dependent parameters is reviewed. Recently, in [9], a numerical procedure is set-up for moment-independent sensitivity methods.

Recently, the anchored ANOVA method (i.e. cut-HDMR) has been widely used in the literature (see for instance [22, 42, 23, 16, 50, 51, 49, 38]). In particular, the method proposed in [38] is based on the covariance decomposition of the output variance to obtain accurate results which are less sensitive to the choice of the reference point, and meanwhile preserves the main advantage of the anchored method, i.e. a much less number of deterministic simulations is needed compared to the standard ANOVA method combined with MC or QMC. A disadvantage of this anchored based global method is that a surrogate model cannot be built in a straightforward

way.

The objective of this work consists in building an efficient UQ and GSA method featuring a surrogate model representation that is affordable for complex numerical simulation problems. An accurate surrogate representation is useful for accelerating the model evaluations when using, for instance, the MC type techniques. We emphasize again that in [4], the generalized Polynomial Chaos (gPC) is combined with the ANOVA approach to perform the global analysis and to build a surrogate sparse representation. We recall a traditional gPC decomposition of a N-dimensional function $f(\xi_1, \xi_2, \cdots, \xi_N)$ of degree not exceeding p can be written as follows

$$
\begin{aligned}
f(\xi_1, \xi_2, \cdots, \xi_N) = \quad & a_0\zeta_0 + \sum_{i_1=1}^{N} a_{i_1}\zeta_1(\xi_{i_1}) + \sum_{i_1=1}^{N}\sum_{i_2=1}^{i_1} a_{i_1 i_2}\zeta_2(\xi_{i_1}, \xi_{i_2}) \\
& + \cdots + \sum_{i_1=1}^{N}\sum_{i_2=1}^{i_1}\cdots\sum_{i_p=1}^{i_{p-1}} a_{i_1 i_2\cdots i_p}\zeta_p(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_p}),
\end{aligned}
\tag{1}
$$

where $\zeta_n(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_n})$ denote the orthogonal polynomials of order n from the Askey scheme (see [44, 45]) in terms of the multivariate random variables $(\xi_{i_1}, \xi_{i_2}, \cdots, \xi_{i_n})$. On the one hand, note the gPC formulation (1) is organized with respect to an increasing degree of multivariate polynomials, and not to an increasing order of parameter interactions. For instance, the group of polynomials of order n can contain polynomials with random variables subject to a dimensionality equal to or less than n. For this reason in order to compute the Sobol' indices, one needs to additionally reorder the PC terms according to the random variables they depend on. This task can be annoying in our practice for complex multivariate problems. On the other hand, the gPC expansions are known to succumb to the curse of dimensionality for high-dimensional problems. To address these issues in the present work, we employ the polynomial dimensional decomposition (PDD) (introduced and developed by Rahman and coworkers in e.g. [29, 30, 32, 48, 31]), instead of PC, to combine with the ANOVA decomposition due to their direct link between each other. Indeed, the PDD expansion of the model output relies on the ANOVA functional decomposition in a very direct way: a T-dimensional ANOVA component function is represented by a summation of T-dimensional multivariate polynomials from lowest to highest degree using a specific polynomial basis subject to the input parameter distributions. Thus, the PDD gives priority to exploit the low order parameter interactions, which matches perfectly the principle of the ANOVA decomposition where we suppose that the low order component functions are dominant for most engineering cases. Consequently, even with a shortly truncated PDD expansion, one can take advantage of employing relatively high-order single- or multi-variate polynomials. Similar to the gPC approach, an important inevitable task is to determine the polynomial coefficients of PDD. The least-square regression approach reveals itself as an efficient tool for this purpose, by minimizing the error of the surrogate model representation in the mean square sense (see e.g. [12, 37, 4]). Compared to the projection approach (see e.g. [19, 24, 3]) where each polynomial coefficient is obtained by computing a multi-dimensional integral, we find the regression approach more flexible for problems involving a moderate number of uncertain parameters. It is known that the number of ANOVA component functions increases exponentially with respect to the uncertain parameter dimensionality, and meanwhile the imposed polynomial order for the PDD expansion involves a polynomial increase of the number of PDD terms for each component function. This phenomenon causes one main limitation of the regression approach even for a truncated low-order ANOVA expansion, namely the required high number of deterministic model evaluations for problems characterized by a moderate to large number of uncertainties; indeed, for the regression problem to be well posed, the number of deterministic model evaluations is necessary to be larger than the total polynomial expansion size [4, 5]. In this respect, this paper proposes

to combine the *active dimension* strategy in the framework of the adaptive ANOVA method [49, 38] and the *stepwise regression* technique [4, 5] to obtain an efficient sparse surrogate model representation. This method has similarities with the adaptive sparse PC approach presented in [4, 5], but differs by the use of the PDD expansion, the truncation order / active dimension technique used in the ANOVA type methods, and the selection criteria allowing to retain the most important polynomial terms. Once the sparse surrogate model is built, the Sobol' sensitivity indices can be easily obtained by manipulating the polynomial coefficients without any reordering task as required for the PC expansion.

The paper is organized as follows: the classical ANOVA functional decomposition and the related variance-based global sensitivity indices are summarized in Section 2. The polynomial dimensional decomposition (PDD) expansion, the determination of its coefficients, and its link with the sensitivity analysis are presented in Section 3. The proposed adaptive sparse meta-modeling approach is then presented in Section 4 by using a combination of the active dimension technique of the ANOVA expansion and the stepwise regression approach. As far as the strategy of selecting the most significant polynomial terms is concerned, the variance-based selection criterion is detailed (Criterion 1). Two academic benchmark problems as well as an atmospheric reentry spacecraft problem are carried out in Section 5. Conclusions follow. A is devoted to the presentation of the estimator of accuracy used in this work and of the Criterion 2 allowing to retain the most important terms by comparing the surrogate representation error.

2 ANOVA dimensional decomposition of the model response

Let us introduce some notations. The upper-case letters, $\mathbf{X} = (X_1, \cdots, X_N)$ and Y, denote a list of independent input random variables (random vector) and a scalar random output, respectively; the lower-case letters $\mathbf{x}$ and y represent the realizations. $\{\mathbf{X}\}$ is used to represent a set whose members are the random variables contained in $\mathbf{X}$, i.e. $\{\mathbf{X}\} = \{X_1, \cdots, X_N\}$. We assume the N input random variables $\mathbf{X}$ admit a joint probabilty density function (pdf)

$$p_{\mathbf{X}}(\mathbf{x}). \tag{2}$$

The assumption of independence of random variables in the set with respect to each other implies:

$$p_{\mathbf{X}}(\mathbf{x}) = \prod_{i}^{N} p_{X_i}(x_i)$$

where $p_{X_i}(x_i)$ is the marginal pdf of X_i.

The expectation and variance of an integrable function of random vector $\mathbf{X}$, $g(\mathbf{X})$, denoted by $E(g(\mathbf{X}))$ and $\mathrm{Var}(g(\mathbf{X}))$ respectively, are given by

$$
\begin{aligned}
E(g(\mathbf{X})) &= \int_{\mathbb{R}^N} g(\mathbf{x}) p_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x}, \\
\mathrm{Var}(g(\mathbf{X})) &= \int_{\mathbb{R}^N} (g(\mathbf{x}) - E(g(\mathbf{X})))^2 p_{\mathbf{X}}(\mathbf{x}) \, \mathrm{d}\mathbf{x},
\end{aligned}
$$

where $\mathrm{d}\mathbf{x} = \prod_{i=1}^{N} \mathrm{d}x_i$.

Let us suppose that the response of a given system of interest can be represented by a N-dimensional function $y = f(\mathbf{x})$

$$y = f(\mathbf{x}) = f(x_1, x_2, \cdots, x_N). \tag{3}$$

We consider (3) in its functional expansion form as follows

$$y = f_0 + \sum_{1 \leqslant i \leqslant N}^{N} f_i(x_i) + \sum_{1 \leqslant i < j \leqslant N}^{N} f_{ij}(x_i, x_j) + \cdots + f_{1,2,\cdots,N}(x_1, x_2, \cdots, x_N), \qquad (4)$$

or in compact form using a multi index system:

$$y = f_{s_0} + \sum_{j=1}^{2^N-1} f_{s_j}(\mathbf{x}_{s_j}). \qquad (5)$$

The multi indices s_j are defined such as

$$
\begin{array}{rcl}
s_0 & = & (0,0,0,\cdots,0) \\
s_1 & = & (1,0,0,\cdots,0) \\
s_2 & = & (0,1,0,\cdots,0) \\
& \vdots & \\
s_N & = & (0,0,0,\cdots,1) \\
s_{N+1} & = & (1,1,0,\cdots,0) \\
s_{N+2} & = & (1,0,1,\cdots,0) \\
& \vdots & \\
s_{\mathscr{N}} & = & (1,1,1,\cdots,1).
\end{array}
\qquad (6)
$$

where

$$\mathscr{N} = 2^N - 1. \qquad (7)$$

The representation (5) is called ANOVA (Analysis Of Variance) decomposition [1] of $f(\mathbf{x})$, if for any $j \in \{1, \cdots, \mathscr{N}\}$,

$$\int_{\mathbb{R}} f_{s_j}(\mathbf{x}_{s_j}) p_{X_i}(x_i) \, \mathrm{d}x_i = 0 \quad \text{for} \quad x_i \in \{\mathbf{x}_{s_j}\}. \qquad (8)$$

It follows from (8) the orthogonality of ANOVA component terms, namely

$$E(f_{s_j} f_{s_k}) = 0 \quad \text{for} \quad j \neq k, \qquad (9)$$

Meanwhile, we obviously have

$$E(f_{s_j}) = 0 \quad \text{for} \quad j = 1, \cdots, \mathscr{N}.$$

Note that the terms in the ANOVA decomposition can be expressed as integrals of $f(\mathbf{x})$. Indeed, we have

$$
\begin{array}{rcl}
E(Y) & = & f_0, \\
E(Y|x_i) & = & f_0 + f_i(x_i), \\
E(Y|x_i, x_j) & = & f_0 + f_i(x_i) + f_j(x_j) + f_{ij}(x_i, x_j),
\end{array}
\qquad (10)
$$

and so on, where $E(Y|\cdot)$ denotes the conditional expectation with respect to the conditional pdf $p_{\mathbf{X}|\mathbf{X}_{s_j}}(\mathbf{x}|x_{s_j})$ defined in the standard way. For instance,

$$
\begin{array}{rl}
E(Y|x_i) & = \int_{\mathbb{R}^{N-1}} f(\mathbf{x}) p_{\mathbf{X}|X_i}(\mathbf{x}|x_i) \, \mathrm{d}\mathbf{x}_{-i} = \int_{\mathbb{R}^{N-1}} f(\mathbf{x}) \frac{p_{\mathbf{x}}(\mathbf{x})}{p_{X_i}(x_i)} \, \mathrm{d}\mathbf{x}_{-i} \\
& = \int_{\mathbb{R}^{N-1}} f(\mathbf{x}) \left(\prod_{j=1, j \neq i}^{N} p_{X_j}(x_j) \right) \, \mathrm{d}\mathbf{x}_{-i},
\end{array}
\qquad (11)
$$

where $\mathrm{d}\mathbf{x}_{-i} = \prod_{j=1, j\neq i}^{N} \mathrm{d}x_j$. We thus observe from (10) the ANOVA terms can be computed as follows:

$$f_{s_j}(\mathbf{x}_{s_j}) = E(Y|\mathbf{x}_{s_j}) - \sum_{s_k \subset s_j} f_{s_k}, \tag{12}$$

with s_k a subset multi-index of s_j (Here we define $s_k \subset s_j \Leftrightarrow \{\mathbf{x}_{s_k}\} \subset \{\mathbf{x}_{s_j}\}$).

By taking the variance operator to the formulation (12), and keeping in mind the orthogonality between component functions, we can obtain the component variance of the term $f_{s_j}(\mathbf{x}_{s_j})$:

$$\mathrm{Var}(f_{s_j}) = \mathrm{Var}_{\mathbf{X}_{s_j}}(E(Y|\mathbf{X}_{s_j})) - \sum_{s_k \subset s_j} \mathrm{Var}(f_{s_k}). \tag{13}$$

The notation

$$\mathrm{Var}_{\mathbf{X}_{s_j}}(E(Y|\mathbf{X}_{s_j}))$$

in (13) is used to emphasize its meaning of the variance of the conditional expectation of Y given $\mathbf{X}_{s_j}$.

On the other hand, by integrating f^2 and exploiting the orthogonality property of component functions, the output variance of f can be written as follows:

$$V(Y) = \sum_{j=1}^{\mathscr{N}} E(f_{s_j}^2(\mathbf{X}_{s_j})) = \sum_{j=1}^{\mathscr{N}} \mathrm{Var}(f_{s_j}), \tag{14}$$

which is in fact the sum of the variances of all the decomposition terms. Note (14) is a special case of (13) when $s_j = s_{\mathscr{N}}$.

2.1 Variance-based global sensitivity analysis

The ANOVA decomposition is closely related to the global sensitivity indices [35, 36] which are defined by the ratios

$$\mathcal{S}_{s_j} = \frac{V_{s_j}}{V(Y)} = \frac{\mathrm{Var}_{\mathbf{X}_{s_j}}(E(Y|\mathbf{X}_{s_j})) - \sum_{s_k \subset s_j} V_{s_k}}{V(Y)} \quad \text{for} \quad j = 1, \cdots, \mathscr{N}. \tag{15}$$

For simplicity, we have denoted here $V_{s_j} = \mathrm{Var}(f_{s_j})$. In particular, we note the first-order sensitivity indices are defined by

$$\mathcal{S}_{X_i} = \frac{V_i}{V(Y)} = \frac{\mathrm{Var}_{X_i}(E(Y|X_i))}{V(Y)} \quad \text{for} \quad i = 1, \cdots, N$$

where $V_i = \mathrm{Var}(f_i)$. From (14), all the $\mathcal{S}_{s_j}$ are non-negative and their sum equals unity:

$$\sum_{j=1}^{\mathscr{N}} \mathcal{S}_{s_j} = 1. \tag{16}$$

Furthermore, the total effects of the variable X_i is estimated by

$$\begin{aligned} \mathcal{S}_i^T &= \frac{V(Y) - \left(V_{\mathbf{X}_{-i}} + \sum_{\{\mathbf{X}_{s_k}\} \subset \{\mathbf{X}_{-i}\}} V_{s_k}\right)}{V(Y)} \\ &= 1 - \frac{\mathrm{Var}_{\mathbf{X}_{-i}}(E(Y|\mathbf{X}_{-i}))}{V(Y)} \quad \text{for} \quad i = 1, \cdots, N \end{aligned} \tag{17}$$

which is indeed the sum of all sensitivity indices containing X_i. Here $\mathbf{X}_{-i} = (X_1, \cdots, X_{i-1}, X_{i+1}, \cdots, X_N)$.

3 Polynomial dimensional decomposition (PDD) of the model response

The previous section deals with the functional decomposition of a model response aiming to compute the moments and sensitivity indices. However, this approach does not provide any strategy to build a meta-model. For time-consuming numerical simulations, the meta-modeling is of importance to approximate and thus to accelerate the model evaluations. To seek an efficient way for this purpose, the polynomials can be used to represent the component functions in the ANOVA expansion. Roughly speaking, two techniques are widely used in the literature: polynomial chaos (PC) [44], and polynomial dimensional decomposition (PDD) [29]. We prefer to employ the PDD representation in this work to take advantage of the close dimensional structure between the PDD and ANOVA. Note however the PC can also be used in a similar way, and no additional difficulty appears.

3.1 PDD representation

Let us consider an orthogonal system of polynomials in the Hilbert space $\mathscr{L}_2$, denoted by $\{\psi_j(x_i); j = 0, 1, \cdots\}$, which is characterized by the following relation

$$\int_{\mathbb{R}} \psi_j(x_i)\psi_k(x_i)p_{X_i}(x_i)\,\mathrm{d}x_i = \gamma_{j,X_i}\delta_{jk}, \tag{18}$$

where $\delta_{jk} = 1$ if $j = k$, otherwise $\delta_{jk} = 0$. The normalization constant γ_{j,X_i} can obviously be determined as follows

$$\gamma_{j,X_i} = \int_{\mathbb{R}} \psi_j^2(x_i)p_{X_i}(x_i)\,\mathrm{d}x_i. \tag{19}$$

As well known in the literature, common distributions can be associated to specific families of polynomials [44]. For instance, a uniform distribution can be associated to Legendre polynomials, and a Gaussian to Hermite polynomials.

Let us consider a general T-dimensional component function ($1 \leq T \leq N$) of the ANOVA decomposition

$$f_{i_1,i_2,\cdots,i_T}(x_{i_1}, x_{i_2}, \cdots, x_{i_T}). \tag{20}$$

Due to the assumption of independence between member variables of the random vector $\mathbf{X}$, it can be proved that

$$\Psi_{\mathbf{J}_T}(x_{i_1}, x_{i_2}, \cdots, x_{i_T}) = \prod_{k=1}^{T} \psi_{j_k}(x_{i_k}) \tag{21}$$

is a multivariate basis in the T-dimensional space. j_k is the order of polynomial for the variable x_{i_k}, and $\mathbf{J}_T = \{j_1, j_2, \cdots, j_T\}$.

Keeping in mind its zero mean property, the component function (20) can be expanded as done in [29]:

$$f_{i_1,i_2,\cdots,i_T}(x_{i_1}, x_{i_2}, \cdots, x_{i_T}) = \sum_{j_T=1}^{\infty} \cdots \sum_{j_1=1}^{\infty} C_{i_1,i_2,\cdots,i_T}^{j_1,j_2,\cdots,j_T} \prod_{k=1}^{T} \psi_{j_k}(x_{i_k}). \tag{22}$$

Here $C_{i_1,i_2,\cdots,i_T}^{j_1,j_2,\cdots,j_T}$ is the coefficient. As will be discussed later, this coefficient will be determined by the regression approach in this work.

Due to the fact of

$$E(\Psi_{\mathbf{J}_S}(X_{i_1}, X_{i_2}, \cdots, X_{i_S})\Psi_{\mathbf{J}_T}(X_{i_1}, X_{i_2}, \cdots, X_{i_T})) = 0, \text{ with } S \neq T,$$

the orthogonality property of ANOVA component functions (9) is guaranteed.

In practice, the expansion with an infinite number of terms in (22) must be truncated. For the sake of simplicity, as done in [29], we truncate (22) by m terms for each dimension, i.e.

$$f_{i_1, i_2, \cdots, i_T}(x_{i_1}, x_{i_2}, \cdots, x_{i_T}) = \sum_{j_T=1}^{m} \cdots \sum_{j_1=1}^{m} C_{i_1, i_2, \cdots, i_T}^{j_1, j_2, \cdots, j_T} \prod_{k=1}^{T} \psi_{j_k}(x_{i_k}). \tag{23}$$

In particular, the first-order, second-order, and third-order component functions can be expressed respectively by

$$
\begin{aligned}
f_i(x_i) &= \sum_{j=1}^{m} C_i^j \psi_j(x_i), \\
f_{i_1, i_2}(x_{i_1}, x_{i_2}) &= \sum_{j_2=1}^{m} \sum_{j_1=1}^{m} C_{i_1, i_2}^{j_1, j_2} \psi_{j_1}(x_{i_1}) \psi_{j_2}(x_{i_2}), \\
f_{i_1, i_2, i_3}(x_{i_1}, x_{i_2}, x_{i_3}) &= \sum_{j_3=1}^{m} \sum_{j_2=1}^{m} \sum_{j_1=1}^{m} C_{i_1, i_2, i_3}^{j_1, j_2, j_3} \psi_{j_1}(x_{i_1}) \psi_{j_2}(x_{i_2}) \psi_{j_3}(x_{i_3}).
\end{aligned}
\tag{24}
$$

In conclusion, the polynomial dimensional decomposition of order m of the model output $f(\mathbf{x})$ can be written as

$$
\begin{aligned}
f(\mathbf{x}) \approx f_m(\mathbf{x}) \quad = \quad & f_0 + \sum_{i=1}^{N} \sum_{j=1}^{m} C_i^j \psi_j(x_i) + \sum_{i_1 < i_2} \sum_{j_2=1}^{m} \sum_{j_1=1}^{m} C_{i_1, i_2}^{j_1, j_2} \psi_{j_1}(x_{i_1}) \psi_{j_2}(x_{i_2}) \\
+ \quad & \sum_{i_1 < i_2 < i_3} \sum_{j_3=1}^{m} \sum_{j_2=1}^{m} \sum_{j_1=1}^{m} C_{i_1, i_2, i_3}^{j_1, j_2, j_3} \psi_{j_1}(x_{i_1}) \psi_{j_2}(x_{i_2}) \psi_{j_3}(x_{i_3}) \\
+ \quad & \cdots + \sum_{i_1 < \cdots < i_N} \sum_{j_N=1}^{m} \cdots \sum_{j_1=1}^{m} C_{i_1, i_2, \cdots, i_N}^{j_1, j_2, \cdots, j_N} \prod_{k=1}^{N} \psi_{j_k}(x_{i_k}).
\end{aligned}
\tag{25}
$$

Hence, the total size P of the m-th order full PDD expansion of an N-dimensional function is found to be the following

$$P = 1 + Nm + \binom{N}{2} m^2 + \cdots + \binom{N}{N} m^N = (1 + m)^N. \tag{26}$$

Note in the particular case of $m = 1$, the PDD representation has the same expansion size as ANOVA.

The following section is devoted to the discussion of the expansion coefficient computation.

3.2 PDD coefficient computation by regression

The computation of the coefficients involved in (25) can be obtained by projection [19, 18]. Indeed, we have

$$C_{i_1, i_2, \cdots, i_T}^{j_1, j_2, \cdots, j_T} = \frac{E\left(f(\mathbf{X})\Psi_{\mathbf{J}_T}(X_{i_1}, X_{i_2}, \cdots, X_{i_T})\right)}{E\left(\Psi_{\mathbf{J}_T}^2(X_{i_1}, X_{i_2}, \cdots, X_{i_T})\right)}. \tag{27}$$

As it is well-known, the formulation (27) can be expensive to evaluate in case of a large number of input variables, since high-dimensional integrations are required to be computed. Typically, one can employ random sampling type methods (e.g. Monte Carlo (MC), Latin Hypercube) which are costly in general, or Gauss quadrature methods which are prohibitive in the high-dimensional case even when using a sparse grid.

As done in [5, 4], in order to take advantage of the flexibility property, we use instead the regression approach in this work to determine the expansion coefficients.

The regression approach in this work can be regarded as a response surface aiming to provide an optimized PDD expansion to the considered model problem.

Let us write the finite PDD expansion (25) in a vector form as follows

$$f_m(\mathbf{x}) = \mathbf{C}_\alpha^T \psi_\alpha(\mathbf{x}) \tag{28}$$

where $\mathbf{C}_\alpha = (C_{\alpha_0}, \cdots, C_{\alpha_{P-1}})^T$ is a vector containing all the coefficients, and $\psi_\alpha(\mathbf{x}) = (\psi_{\alpha_0}, \cdots, \psi_{\alpha_{P-1}})^T$ gathering all the multivariate basis polynomials including the unity basis $\psi_{\alpha_0} = 1$. Here P is the total size of PDD expansion, given by (26).

When using a regression method to determine the expansion coefficients, it is necessary to choose a set of realizations of input random vector (i.e. an experimental design [5, 4], for instance by a Sobol' quasi-random sequence [34]), denoted by

$$\mathcal{X} = \{\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^Q\},$$

whose size is necessary to be larger than the PDD expansion size. As indicated in [6], in order for the regression problem to be stable enough, the experimental design size Q, in practice, is usually taken as

$$Q = kP \text{ with } 2 \leq k \leq 3. \tag{29}$$

We denote the corresponding model outputs by

$$\mathcal{Y} = \{y^1, y^2, \cdots, y^Q\}.$$

The idea is to determine the coefficients $\mathbf{C}_\alpha$ by minimizing the projection error in $\mathscr{L}_2$ norm. That is,

$$\widetilde{\mathbf{C}}_\alpha = \arg \min_{\mathbf{C}_\alpha \in \mathbb{R}^P} \sum_{i=1}^Q \left(y^i - \mathbf{C}_\alpha^T \psi_\alpha(\mathbf{x}^i) \right)^2. \tag{30}$$

The solution of the least-square regression problem (30) can be easily found by using a variational approach

$$\frac{\partial}{\partial C_{\alpha_j}} \left[\sum_{i=1}^Q \left(y^i - \mathbf{C}_\alpha^T \psi_\alpha(\mathbf{x}^i) \right)^2 \right] = 0, \text{ for } j = 0, \cdots, P-1,$$

and can be obtained by solving the following linear system

$$(\mathbf{\Psi}^T \mathbf{\Psi}) \widetilde{\mathbf{C}}_\alpha = \mathbf{\Psi}^T \mathcal{Y}. \tag{31}$$

$\mathbf{\Psi}$ represents the following matrix involving basis polynomials evaluated at the realizations in the experimental design:

$$\mathbf{\Psi}_{ij} = \psi_{\alpha_j}(\mathbf{x}^i), \quad i = 1, \cdots, Q, \quad j = 0, \cdots, P-1 \tag{32}$$

It is well known the real matrix $\mathbf{\Psi}^T \mathbf{\Psi}$ is symmetric and positive-definite. If this matrix is well-conditioned, the linear system (31) can be in general resolved efficiently by Cholesky factorization.

We emphasize the design size Q is directly linked to the global computational cost of uncertainty quantification for numerical simulations. As a consequence from (29) for the choice of experimental design size, the main objective of this work is to minimize the PDD expansion size P, by including only the most influential polynomials terms.

3.3 PDD based global sensitivity indices

Once the coefficients $\mathbf{C}_\alpha$ are determined by the regression approach for the PDD expansion (28), the second-order moment and the global sensitivity indices can be obtained in a straightforward way.

Indeed, keeping in mind

$$E(f_m(\mathbf{X})) = C_{\alpha_0},$$

the approximated variance of the model output is then

$$\mathrm{Var}(f_m(\mathbf{X})) = \sum_{j=1}^{P-1} C_{\alpha_j}^2 \gamma_{\alpha_j},$$

with the multivariate normalization constant determined by

$$\gamma_{\alpha_j} = E\left[\psi_{\alpha_j}^2(\mathbf{X})\right].$$

If one employs normalized basis polynomials, i.e. $\gamma_{\alpha_j} = 1$, the output variance formulation can be further simplified to

$$\mathrm{Var}(f_m(\mathbf{X})) = \sum_{j=1}^{P-1} C_{\alpha_j}^2.$$

3.3.1 ANOVA and PDD-based sensitivity estimates

It is now straightforward to write the variance-based global sensitivity indices by using the PDD expansion. Indeed,

$$\mathcal{S}_{s_j} = \frac{V_{s_j}}{\mathrm{Var}(f_m(\mathbf{X}))} = \frac{1}{\mathrm{Var}(f_m(\mathbf{X}))} \sum_{\alpha_j \in s_j} C_{\alpha_j}^2 \gamma_{\alpha_j} \tag{33}$$

The total effects of an input variable can be written as done in Section 2.

4 Variance-based dimension reduction for the model representation

For practical problems, in particular for the ones with a large number of stochastic parameters, the size of the PDD representation given in Section 3 must be reduced to make the uncertainty analysis feasible.

The purpose of this section is to present our adaptive technique belonging to the family of *stepwise regression*. The method used in this work is indeed a variant of the one used in [5, 4].

Apart from applying a truncated dimensionality widely used for the ANOVA component functions, we consider additionally two levels of dimension reduction in this section, both of which relies on the relative importance of expansion functions in terms of their variance contribution compared to the total variance. This criterion is different from [5, 4] where the adaptive strategy is based on the numerical error of the model representation. We think a variance-based criterion is more reliable if one cares more about the statistical properties of the model output.

4.1   Adaptive ANOVA – retaining active dimensions

As already done in our previous work [38], we use the adaptive strategy presented in [49] in order to retain only the most important dimensions for interaction terms in ANOVA.

4.1.1   Truncation dimension

The low order interactions of input variables often have the main impact upon the output [36]. Thus, the full ANOVA expansion (4) can be approximated by

$$f(\mathbf{x}) = f_0 + \sum_{T=1}^{\nu} \sum_{i_1 < \cdots < i_T}^{N} f_{i_1, i_2, \cdots, i_T}(x_{i_1}, x_{i_2}, \cdots, x_{i_T}), \text{ with } \nu \ll N. \tag{34}$$

Here, ν is called the truncation (or effective) dimension representing the highest dimension of the ANOVA component functions.

4.1.2   Adaptive ANOVA

For problems featuring a high dimensionality N, ANOVA decomposition method is still very expensive even when we only choose a truncation dimension $\nu = 2$. An efficient way to solve this problem is to use the adaptive ANOVA decomposition. To this purpose, we replace the approximation (34) by

$$f(\mathbf{x}) = f_0 + \sum_{T=1}^{\nu} \sum_{i_1 < \cdots < i_T}^{D_T} f_{i_1, i_2, \cdots, i_T}(x_{i_1}, x_{i_2}, \cdots, x_{i_T}), \text{ with } \nu \ll N, \ D_T \leqslant N. \tag{35}$$

Here D_T is the active dimension for ANOVA component functions of order T. For problems considered in this work, we take $D_1 = N$. The active dimension for higher order component functions will be determined by the criterion presented below (see [49]).

In this work. we use the variance-based criterion [49] for choosing the active dimension D_2 and further selecting the most important second- and higher-order terms. It is assumed $\mathrm{Var}(f_i)$ $(i \in [1, N])$ of first-order terms are monotonically decreasing with respect to i.

D_2 is evaluated using the sum of variances of first-order terms:

$$\sum_{i=1}^{D_2} \mathrm{Var}(f_i) \geqslant p \sum_{i=1}^{N} \mathrm{Var}(f_i), \tag{36}$$

where p is a proportionality constant in $(0, 1)$, and is very close to 1.

For simplicity in our applications, we set

$$D_T = D_2, \text{ for } T \geq 3$$

in this work. Note however D_T can certainly be further reduced depending on problems and objectives.

4.2   Adaptive PDD algorithm – eliminating non-important polynomials

In section 4.1, we have talked about how to reduce the size of ANOVA expansion. However, even with a sparse ANOVA expansion, if applying the classical PDD expansion to the component function (see the formulation (23) and (24)), the required computational cost still remains very

high. Indeed, it reveals that, for a large number of engineering problems, the contribution of many polynomial terms is negligible when regarding the accuracy of the constructed meta-model [5, 4]. In this work on the other hand, as will be shown in our numerical results, if we eliminate those polynomials whose variance is negligible, we can also build a very sparse PDD representation without compromising the accuracy of the meta-model.

Considering the adaptive ANOVA technique presented in Section 4.1, let us describe our adaptive algorithm for stepwise regression from a practical point of view as follows.

1. First of all, we construct a full set of PDD representation (given m) for all first-order ANOVA component functions, namely

$$f(\mathbf{x}) \approx f_m(\mathbf{x}) = f_0 + \sum_{i=1}^{N} f_i(x_i) = f_0 + \sum_{i=1}^{N} \left(\sum_{j=1}^{m} C_i^j \psi_j(x_i) \right). \quad (37)$$

We then compute the total first-order variance by

$$\mathrm{Var}(f_m(\mathbf{X})) = \sum_{i=1}^{N} \sum_{j=1}^{m} (C_i^j)^2 \gamma_i^j.$$

The first-order global sensitivity indices can be obtained by a simple rearrangement:

$$\mathcal{S}_i = \sum_{j=1}^{m} (C_i^j)^2 \gamma_i^j \Big/ \mathrm{Var}(f_m(\mathbf{X})).$$

Let us assume that the sensitivity indices $\{\mathcal{S}_i\}$ are monotonically decreasing with respect to i (thus, a re-ordering task is generally required), so we choose the active dimension D_2 for second-order ANOVA functions in such a way that

$$\sum_{i=1}^{D_2} \mathcal{S}_i \geqslant p,$$

with p a constant close to the unity (e.g. $p = 0.999$). Note for the sake of simplicity in this work, we emphasize again the same active dimension is employed for second- and also higher-order component functions if applicable. However, one can further reduce this dimension if necessary.

2. The objective of this step is to reduce the size of the first-order PDD expansion as expressed in (37). The principle remains similar as in the previous step: we eliminate those non-important polynomial terms by measuring their variance contribution. More in detail, if

$$(C_i^j)^2 \gamma_i^j \Big/ \mathrm{Var}(f_m(\mathbf{X})) < \theta, \quad (38)$$

the corresponding polynomial is then to be removed from the expansion. Here θ is a pre-defined threshold (e.g. $\theta = 10^{-5}$). The resultant first-order model representation contains only significant components, and thus is more concise. Let us denote this first-order PDD base by $\{\psi_{\alpha^1}\}$.

3. Starting from the concise first-order PDD representation, the task of this step is to enrich the model representation by adding significant second and higher order PDD polynomials.

After choosing a truncation dimension ν and an active dimension for each order $(D_2, \cdots, D_\nu)$, the full set of second and higher order PDD polynomial bases $\{\psi_{\alpha^{2+}}\}$ can be constructed easily from the tensor product rule (see the formulation (24) or (25)).

The selection process of important polynomials from $\{\psi_{\alpha^{2+}}\}$ can be explained by the simple stepwise algorithm 1.

Algorithm 1 Adaptive PDD by stepwise regression (**Criterion 1**)

1: initialization of multivariate PDD Base: $\{\psi^w\} = \{\psi_{\alpha^1}\}$
2: **for** $\psi_{\alpha_i} \in \{\psi_{\alpha^{2+}}\}$ **do**
3:     add ψ_{α_i} into $\{\psi^w\}$, i.e. $\{\psi^w\} = \{\psi^w, \psi_{\alpha_i}\}$
4:     depending on the size P^w of $\{\psi^w\}$, adjust, if necessary, the size Q^w of the experimental design (see formulation (29))
5:     solve the regression system (31) to determine the PDD expansion coefficients $\widetilde{\mathbf{C}^w}$.
6:     compute the total variance: $\mathrm{Var}(f^w(\mathbf{X})) = \sum_k (C_{\alpha_k})^2 \gamma_{\alpha_k}$
7:     **for** $\psi_{\alpha_j} \in \{\psi^w\}$ **do**
8:         **if** $(C_{\alpha_j})^2 \gamma_{\alpha_j} / \mathrm{Var}(f^w(\mathbf{X})) < \theta$ **then**
9:             eliminate this polynomial: $\{\psi^w\} = \{\psi^w\} \setminus \psi_{\alpha_j}$
10:         **end if**
11:     **end for**
12: **end for**
13: solve the final regression system based on the constructed base $\{\psi^F\}$
14: compute the final total variance using the obtained PDD coefficients
15: compute the global (total) sensitivity indices

Let us mention in this algorithm the cost of the recursive resolution of the regression linear system is negligible compared to the one of deterministic model evaluations.

If denoting the size of the finally obtained sparse PDD representation by P_{sparse}, we define the *sparsity* of our adaptive PDD approach as follows

$$\text{sparsity} = \frac{P_{sparse}}{(1+m)^N}. \tag{39}$$

Note that the formulation (39) will be employed in our numerical applications for assessing the efficiency of the proposed approach.

4. Several polynomial chaos error estimators are presented in [5, 4]. These estimators can be directly used for the sparse PDD expansions in this work. We present the estimator used in this work in A.

4.2.1 Criterion 1: by comparing the component variance of the concerned polynomial term

We name the operation 8–9 in the Algorithm 1 (see also (38)) as the criterion 1 whose objective is to eliminate non-important polynomial terms. This is achieved by discarding any component polynomial whose variance is negligible compared to the total variance. The threshold θ is predefined by the user.

	$\mathcal{S}_{s_i}$	$\mathcal{S}_i^T$
X_1	0.3138	0.5574
X_2	0.4424	0.4424
X_3	0	0.2436
X_1, X_2	0	
X_1, X_3	0.2436	
X_2, X_3	0	
X_1, X_2, X_3	0	
$V(Y)$	13.845	

Table 1: Analytical variance and variance-based sensitivity indices for Ishigami function.

4.2.2 Criterion 2: by comparing the model accuracy of the concerned polynomial term

The formulation (38) and the operation 8–9 in the Algorithm 1 can be replaced by a second criterion (already used in [4]) who relies on the estimator used to evaluate the accuracy of the model representation. Briefly speaking, one eliminates those polynomials whose contribution to the model accuracy can be negligible. The technical details of the estimator of accuracy and how to implement this criterion can be found in A.

5 Numerical results

This section is devoted to the presentation of our numerical results. Two academic functions will be studied, and we will also investigate one CFD application example.

5.1 Ishigami function

Let us consider the Ishigami function [17] which has been thoroughly studied in our previous work [38] by making use of the Covariance-based Sensitivity Analysis:

$$Y = \sin X_1 + a \sin^2 X_2 + b X_3^4 \sin X_1, \tag{40}$$

where the input random variables $\mathbf{X} = (X_1, X_2, X_3)$ are uniformly distributed over $[-\pi, \pi]$. The constants are set to $a = 7, b = 0.1$, as done in [37, 21].

As presented in [17, 37, 38], the total output variance and the component variances based on standard ANOVA expansion can be obtained analytically:

$$
\begin{aligned}
V(Y) &= \frac{a^2}{8} + \frac{b\pi^4}{5} + \frac{b^2\pi^8}{18} + \frac{1}{2}, \\
V_1 &= \frac{b\pi^4}{5} + \frac{b^2\pi^8}{50} + \frac{1}{2}, \ V_2 = \frac{a^2}{8}, \ V_3 = 0, \\
V_{12} &= V_{23} = 0, \ V_{13} = \frac{8b^2\pi^8}{225}, \\
V_{123} &= 0.
\end{aligned}
\tag{41}
$$

Thus, the variance-based sensitivity indices can be gathered in Table 1.

We set the truncation dimension $\nu = 3$; i.e. all interactions are considered. We use the quasi-random sampling design based on Sobol' sequences for solving the regression system. For the sake of comparison with [4], let us choose the experimental design size $Q = 200$, and set the

	$\mathcal{S}_{s_i}$	$\mathcal{S}_i^T$
X_1	0.3139	0.5573
X_2	0.4427	0.4427
X_3	0	0.2434
X_1, X_2	0	
X_1, X_3	0.2434	
X_2, X_3	0	
X_1, X_2, X_3	0	
$V(Y)$	13.8304	
model evaluations	200	
sparse polynomial base	15	
PDD order m	10	
sparsity	$15/1331 \approx 0.01127$	
model accuracy Q^2		
(A)	0.999965	

Table 2: Numerical results for the Ishigami test case.

PDD order $m = 10$. The active dimension selection technique in Section 4.1 is not adopted for this relatively low dimensional test case. The numerical results with the proposed approach are reported in Table 2.

Comparing our results with the ones reported in [4, Table 2, last column, pp. 1223], sensitivity indices are obtained with a similar accuracy. With our approach, only 15 polynomial terms are necessary to obtain the model accuracy of 0.999965, while, in [4], 77 terms are needed to have an accuracy of 0.9999. We write the constructed surrogate polynomial model as follows to approximate the Ishigami function:

$$
\begin{aligned}
Y &= 3.50078 + 2.81152\tilde{L}e_1(X_1) - 3.41737\tilde{L}e_3(X_1) + 0.649461\tilde{L}e_5(X_1) \\
&\quad -1.33092\tilde{L}e_2(X_2) - 5.85130\tilde{L}e_4(X_2) + 4.90081\tilde{L}e_6(X_2) \\
&\quad -1.39774\tilde{L}e_8(X_2) + 0.212954\tilde{L}e_{10}(X_2) \\
&\quad +5.29778\tilde{L}e_1(X_1)\tilde{L}e_2(X_3) + 2.09215\tilde{L}e_1(X_1)\tilde{L}e_4(X_3) \\
&\quad -6.46581\tilde{L}e_3(X_1)\tilde{L}e_2(X_3) - 2.60564\tilde{L}e_3(X_1)\tilde{L}e_4(X_3) \\
&\quad +1.23633\tilde{L}e_5(X_1)\tilde{L}e_2(X_3) + 0.495326\tilde{L}e_5(X_1)\tilde{L}e_4(X_3).
\end{aligned}
\tag{42}
$$

where $\tilde{L}e_j(x_i)$ represents the $j-$th order shifted Legendre polynomial for the variable x_i with respect to the weight function $w(x_i) = 1/(2\pi)$ for $x_i \in [-\pi, \pi]$.

Because the underlyng function in (40) is even with respect to the variable X_2 and X_3, the odd polynomials related to these variables are found to be zero in (42) as expected. For the same reason, the even polynomials linked to the variable X_1 have zero coefficients.

As it is known, by the analytical analysis using the standard ANOVA (see e.g. [38]), the maximum interaction order for Ishigami function is 2, as is also found in the surrogate model (42). Note thus if setting the truncation dimension $\nu = 2$ before constructing the surrogate model, one finds the same results as in (42).

5.1.1 Sensitivity to the PDD order m with the variance criterion 1

In this section, we study the sensitivity of the proposed approach with respect to the single-variable polynomial order m, by reporting the sparsity of the surrogate representation and showing the convergence of the method in terms of model accuracy.

Note we fix the size of our experimental design to $Q = 200$. The dimension reduction technique presented in Section 4.1 is not adopted; only the adaptive PDD algorithm in Section 4.2 is employed here. Note the variance selection threshold is set to $\theta = 10^{-5}$ for Criterion 1. The polynomial order m varies from 5 to 9. We report the results in Table 3.

SI	Exact	$m = 5$	$m = 6$	$m = 7$	$m = 8$	$m = 9$
$\mathcal{S}_1$	0.3138	0.2104	0.3141	0.3113	0.3136	0.3164
$\mathcal{S}_2$	0.4424	0.4720	0.4099	0.4040	0.4427	0.4379
$\mathcal{S}_3$	0	0	0	0	0	0
$\mathcal{S}_{12}$	0	0	0	0	0	0
$\mathcal{S}_{13}$	0.2436	0.1482	0.2532	0.2583	0.2437	0.2407
$\mathcal{S}_{23}$	0	0	0	0	0	0
$\mathcal{S}_{123}$	0	0.1527	0.02	0.024	0	0
$\mathcal{S}_1^T$	0.5574	0.5181	0.5898	0.5953	0.5573	0.5616
$\mathcal{S}_2^T$	0.4424	0.6414	0.4326	0.4304	0.4428	0.4430
$\mathcal{S}_3^T$	0.2436	0.3108	0.2736	0.2834	0.2437	0.2452
$V(Y)$	13.845	22.386	13.392	14.241	13.887	14.075
sparse base		88	83	104	15	67
sparsity		$\frac{88}{216} \approx 0.407$	$\frac{83}{343} \approx 0.242$	$\frac{104}{512} \approx 0.203$	$\frac{15}{729} \approx 0.021$	$\frac{67}{1000} \approx 0.067$
model accuracy Q^2		0.618	0.988	0.996	0.99987	0.99928

Table 3: Ishigami test. Numerical results using criterion 1 by varying the polynomial order m. The experimental design size $Q = 200$. The model accuracy is estimated by the cross validation method.

By looking at the Table 3 together with the Table 2, we notice, for this specific case, the adaptive PDD approach with the even number of m provides better results in general than with the odd m. However, the convergence tendency when increasing m can be clearly observed.

5.1.2 Sensitivity to the target accuracy Q_{tgt}^2 with the error criterion 2

In order to evaluate the model accuracy by using the error criterion 2, we vary the target accuracy Q_{tgt}^2 to compute the sensitivity indices. The results are reported in Table 4.

SI	Exact	$Q_{tgt}^2 = 0.9$	$Q_{tgt}^2 = 0.99$	$Q_{tgt}^2 = 0.999$
$\mathcal{S}_1$	0.3138	0.2796	0.3171	0.3139
$\mathcal{S}_2$	0.4424	0.4607	0.4381	0.4424
$\mathcal{S}_3$	0	0	0	0
$\mathcal{S}_{12}$	0	0	0	0
$\mathcal{S}_{13}$	0.2436	0.2597	0.2448	0.2437
$\mathcal{S}_{23}$	0	0	0	0
$\mathcal{S}_{123}$	0	0	0	0
$\mathcal{S}_1^T$	0.5574	0.5393	0.5619	0.5576
$\mathcal{S}_2^T$	0.4424	0.4607	0.4381	0.4424
$\mathcal{S}_3^T$	0.2436	0.2597	0.2448	0.2437
$V(Y)$	13.845	13.026	13.671	13.851
model evaluations		75	390	430
PDD degree m		6	7	8
sparse base		10	14	15
sparsity		$\frac{10}{343} \approx 0.029$	$\frac{14}{512} \approx 0.027$	$\frac{15}{729} \approx 0.021$
model accuracy Q^2		0.9428	0.9908	0.9996

Table 4: Ishigami test. Numerical results using criterion 2 by varying the target accuracy Q_{tgt}^2. The model accuracy is estimated by the cross validation method. $\epsilon_{Q^2} = 10^{-7}$.

The convergence of the sensitivity indices to the reference results is clearly verified when we increase the target model accuracy.

In the authors' experience, we observe it is relatively difficult to obtain an accuracy superior to 0.9999 for this case when using the criterion 2. However, the numerical results are found to be sufficiently accurate with the conditions in the last column in Table 4.

By comparing these results with the ones in Table 3, we notice that, with Criterion 2, a larger number of model evaluations are usually required to obtain the sensitivity indices with a comparable accuracy. On the other hand, it shows, with the criterion 2, we have sparser polynomial bases than with the criterion 1.

5.2   8-dimensional Sobol' function

The second test case is devoted to an eight-dimensional Sobol' function (see [4, Section 6.2]):

$$Y = f(\mathbf{X}) = \prod_{k=1}^{N} f^{(k)}(X_k), \ X_k \sim \mathscr{U}(0,1), \tag{43}$$

where

$$
\begin{aligned}
N &= 8, \\
f^{(k)}(X_k) &= \frac{|4X_k - 2| + a_k}{1 + a_k}, \\
\mathbf{a} &= \{1, 2, 5, 10, 20, 50, 100, 500\}^T.
\end{aligned}
\tag{44}
$$

Note the members of the input random vector $\mathbf{X}$ are uniformly distributed over $[0,1]$. $\{a_k, k = 1, \cdots, N\}$ are positive coefficients whose values are gathered in the vector $\mathbf{a}$.

In the computation by the proposed approach, let us set the ANOVA interaction order equal to 2 for simplicity. We first set the active dimension $D_1 = D_2 = 8$ here (by imposing a large enough p). The experimental design size is set to 150 by a quasi-random Sobol' sequence, for the sake of comparison with the results obtained with the adaptive sparse polynomial chaos method and the Monte Carlo method reported in [4]. Our numerical results with the variance criterion 1 (the predefined threshold $\theta = 2 \times 10^{-4}$) are shown in Table 5 compared to the ones in the reference [4].

It is noticed, by using a half number of polynomial terms with the highest degree equal to 10, the corresponding model representation by the present approach is able to provide more accurate sensitivity indices than the method presented in [4] using the sparse PC of degree 6. On the other hand, we clearly observe the advantage of our method compared to the classical MC method when looking at the corresponding number of model evaluations.

5.2.1   Sensitivity to the p constant for the selection of active dimension

In this section, we test the method sensitivity to the p constant (see Section 4.1.2) which allows the efficient selection of the active dimension, and thus reducing the order of modeling difficulty.

The results by varying p from 0.9 to 0.999 are reported in Table 6. We can conclude, from this analysis, that the using of only the 3 most important variables (by setting $p = 0.9$) responsible for the second-order interactions are sufficient to provide very accurate results regarding both the sensitivity indices and the meta-model representation whose error is measured by Q^2.

5.3   Application example

The method proposed in this paper is applied to the uncertainty quantification of an atmospheric entries spacecraft case. Numerical simulation solves a set of governing equations including mod-

SI	Exact	Sparse PDD (present work)	Sparse PC [4]	Crude MC [4]
$\mathcal{S}_1$	0.60	0.60	0.56	0.57
$\mathcal{S}_2$	0.27	0.25	0.22	0.29
$\mathcal{S}_3$	0.07	0.06	0.05	0.06
$\mathcal{S}_4$	0.02	0.02	0.02	0.03
$\mathcal{S}_5$	0.01	0.01	0.01	0.01
$\mathcal{S}_6$	0.00	0.00	0.00	0.01
$\mathcal{S}_7$	0.00	0.00	0.00	0.00
$\mathcal{S}_8$	0.00	0.00	0.00	0.00
$\mathcal{S}_1^T$	0.63	0.62	0.59	0.66
$\mathcal{S}_2^T$	0.29	0.27	0.26	0.27
$\mathcal{S}_3^T$	0.08	0.08	0.10	0.08
$\mathcal{S}_4^T$	0.02	0.03	0.05	0.01
$\mathcal{S}_5^T$	0.01	0.01	0.03	0.01
$\mathcal{S}_6^T$	0.00	0.00	0.04	0.00
$\mathcal{S}_7^T$	0.00	0.00	0.03	0.00
$\mathcal{S}_8^T$	0.00	0.00	0.03	0.00
model evaluations		150	150	100,000
sparse polynomial base		38	76	
ANOVA order		2		
max polynomial degree		10 $(m = 5)$	6	
sparsity		$\frac{38}{1679616} \approx$2.3e-5	$\frac{76}{3003} \approx$0.02	
model accuracy Q^2		0.993	0.99	

Table 5: 8-dimensional Sobol' test case. $D_2 = 8$. $\theta = 2 \times 10^{-4}$ as the predefined threshold for the variance Criterion 1. Numerical results with the proposed sparse PDD approach are compared to the ones obtained in [4] (sparse PC and Crude MC).

elization of rarefied gas effects, aerothermochemistry, radiation, and the response of thermal protection materials to extreme conditions. A global overview over this problem has been studied in [40].

Here, the focus is in predicting stagnation-point pressure and heat flux from freestream conditions, and is described by a physico-chemical model and solved by suitable numerical methods proposed by Barbante [2, 43].

We use a set of physico-chemical models to simulate high temperature reacting flows, including 2D axisymmetric Navier Stokes equations and gas/surface interaction equations (see Ref. [2]). Indeed, the wall of the spacecraft acts as a catalyzer and promotes recombination of atoms. This phenomenon is modeled by a catalytic wall at radiative equilibrium, where the so-called effective catalytic recombination coefficient γ represents the proportion of gas impinging the body that will recombine. A mixture of 5 species of air is used, namely N, O, N_2, O_2, and NO, with chemical mechanism due to Park [26]. Input data for the forward model are the freestream pressure p_∞ and Mach number M_∞, the effective catalytic recombination coefficient γ, and the gas reaction rate coefficients k_r of the chemical reactions r.

The code COSMIC developed by Barbante [2] is used, which was designed to approximate hypersonic flow models where chemical non-equilibrium effects need to be accounted for. It includes a Hybrid Upwind Splitting (HUS) scheme [13], which is an interesting attempt of combining, in a mathematically rigorous way, Flux Vector Splitting (FVS) and Flux Difference Splitting (FDS)

SI	Exact	$p = 0.9$	$p = 0.99$	$p = 0.999$
$\mathcal{S}_1$	0.60	0.61	0.61	0.60
$\mathcal{S}_2$	0.27	0.27	0.25	0.24
$\mathcal{S}_3$	0.07	0.06	0.06	0.07
$\mathcal{S}_4$	0.02	0.03	0.02	0.02
$\mathcal{S}_5$	0.01	0.01	0.01	0.01
$\mathcal{S}_6$	0.00	0.00	0.00	0.00
$\mathcal{S}_7$	0.00	0.00	0.00	0.00
$\mathcal{S}_8$	0.00	0.00	0.00	0.00
$\mathcal{S}_1^T$	0.63	0.63	0.64	0.62
$\mathcal{S}_2^T$	0.29	0.29	0.28	0.28
$\mathcal{S}_3^T$	0.08	0.07	0.07	0.08
$\mathcal{S}_4^T$	0.02	0.03	0.03	0.03
$\mathcal{S}_5^T$	0.01	0.01	0.01	0.01
$\mathcal{S}_6^T$	0.00	0.00	0.01	0.01
$\mathcal{S}_7^T$	0.00	0.00	0.00	0.00
$\mathcal{S}_8^T$	0.00	0.00	0.00	0.01
Active dimension D_2		3	6	7
sparse polynomial base		25	40	43
ANOVA order		2	2	2
max polynomial degree		9 $(m = 5)$	9 $(m = 5)$	10 $(m = 5)$
sparsity		$\frac{25}{1679616} \approx$1.5e-5	$\frac{40}{1679616} \approx$2.4e-5	$\frac{43}{1679616} \approx$2.6e-5
model accuracy Q^2		0.982	0.987	0.992

Table 6: 8-dimensional Sobol' test case. Experimental design size $Q = 150$. $\theta = 2 \times 10^{-4}$ as the predefined threshold for the variance criterion 1. Numerical results with the proposed sparse PDD approach by varying p.

schemes. The design principle combines the robustness of FVS schemes in the capture of nonlinear waves and the accuracy of some FDS schemes in the resolution of linear waves. In particular, COSMIC uses the hybridization of the Van Leer scheme [20] and the Osher scheme [25] and includes a carbuncle fix.

The boundary conditions are illustrated in the left panel of Fig. 1: an axisymmetric condition is imposed on the y axis (horizontal axis on Fig. 1), while the wall of the body is modeled by a partially catalytic wall at radiative equilibrium. The mesh used for the computations is given in the right panel of Fig. 1. Pressure and temperature iso-contours of the flow around the European EXPerimental Reentry Test-bed (EXPERT) vehicle obtained with COSMIC for input data mean values are shown in Fig. 2. Note that a specific point of the trajectory of EXPERT is considered [39]. The trajectory point corresponds roughly to the chemical non-equilibrium flow conditions of Table 7.

Flow conditions	Altitude [km]	T_∞ [K]	p_∞ [Pa]	M_∞ [-]
Chemical non-equilibrium	60	245.5	20.3	15.5

Table 7: Freestream conditions for one trajectory point of the EXPERT vehicle.

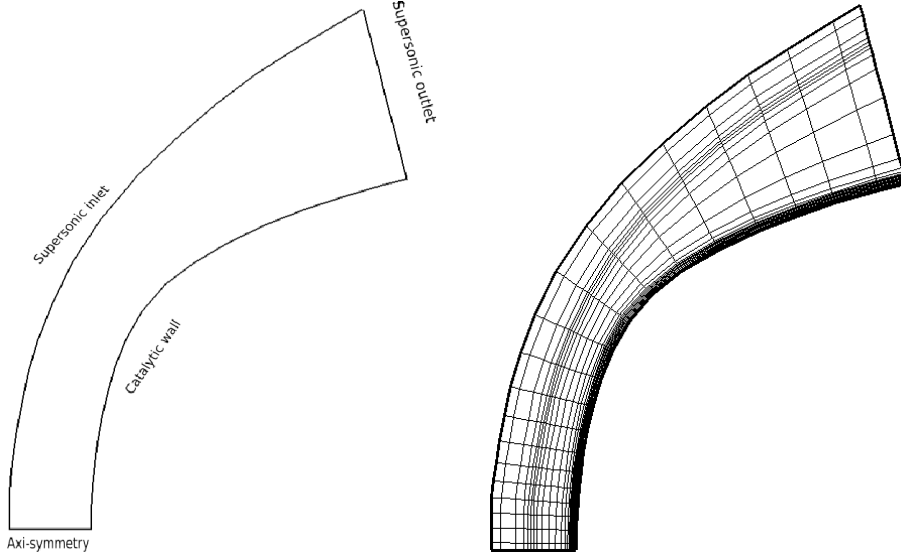Figure 1: Boundary conditions (left) and mesh (right)

Uncertainties are considered on p_∞, M_∞, and γ, with uniform distributions detailed in Table 8. Concerning p_∞ and M_∞, only a priori ranges of plausible values are known. Concerning γ, the mean value corresponds roughly to the EXPERT material, while the 33% error have been previously determined [41].

Variable	Distribution	Min	Max
$X_1(p_\infty)$ [Pa]	Uniform	16.3	24.3
$X_2(M_\infty)$ [-]	Uniform	13.7	17.3
$X_3(\gamma)$ [-]	Uniform	0.001	0.002

Table 8: Distributions of M_∞, p_∞, and γ

Uncertainty is also considered on the gas reaction rate coefficients k_r of four chemical reactions of the dissociation reaction. For the trajectory point investigated, the dissociation reaction of molecular oxygen and nitric oxide was found important. Following the suggestion of Bose *et al.* [10], the uncertainty concerns only the pre-exponential factor A_r of the Arrhenius rate equation: $k_r = A_r T^{b_r} \exp(-E_r/RT)$. Since the uncertainties on k_r can be quite large, it is appropriate to consider them on a logarithmic scale; in particular, $\log_{10}(k_r/k_{r,0})$, with $k_{r,0}$ the recommended rate constant, is commonly assumed to vary following a normal distribution,

$$P(k_r) \propto \exp\left[-\frac{1}{2}\left(\frac{\log_{10}(k_r/k_{r,0})}{\sigma_{rmr}}\right)^2\right] \tag{45}$$

where $\pm 2\sigma_r$ (reported in Table 9) defines the 95% confidence limits symmetrically bounding $k_{r,0}$.

Note that the quantities of interest are the pressure p_{st} and heat flux q_{st} at the stagnation point.
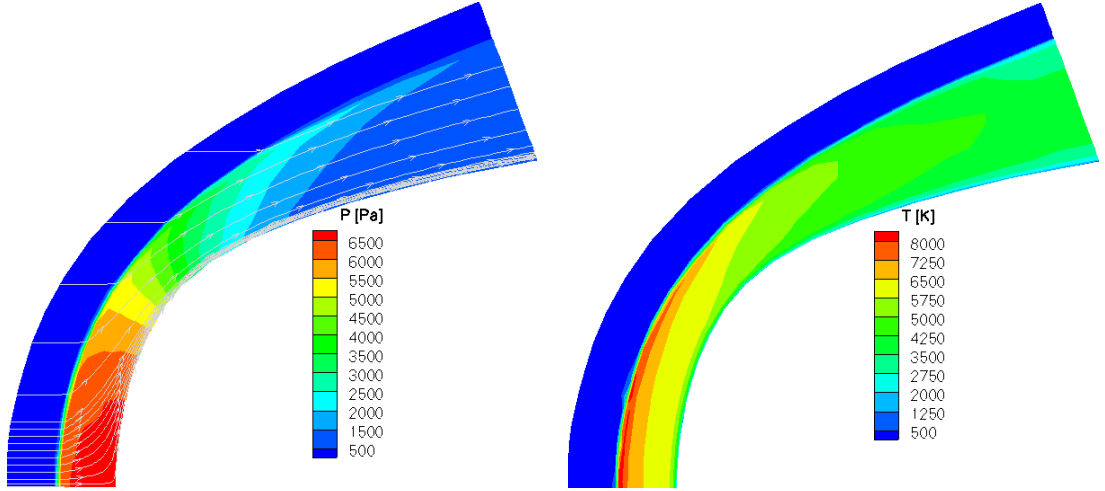
Figure 2: Pressure and temperature iso-contours for input data mean values

Gas reaction	Distribution of $\log_{10} k_r$	σ_r
$X_4(\mathrm{NO} + \mathrm{O} \rightarrow \mathrm{N} + \mathrm{O} + \mathrm{O})$	Normal	0.12
$X_5(\mathrm{NO} + \mathrm{N} \rightarrow \mathrm{N} + \mathrm{O} + \mathrm{N})$	Normal	0.12
$X_6(\mathrm{O}_2 + \mathrm{N}_2 \rightarrow 2\mathrm{O} + \mathrm{N}_2)$	Normal	0.10
$X_7(\mathrm{O}_2 + \mathrm{O} \rightarrow 2\mathrm{O} + \mathrm{O})$	Normal	0.10

Table 9: Distributions of $\log_{10} k_r$

5.4 Uncertainty quantification results

As far as this application test is concerned, 1,000 resolutions of the deterministic code are used to recursively solve linear regression systems. Let us first consider the pressure p_{st} as the quantity of interest and set the highest ANOVA interaction order equal to $\nu = 2$ (the truncation dimension). We further set $p = 0.99999$ for the selection of the active dimensions for the second-order ANOVA interaction terms. For the sake of comparison with the UQ results presented in the reference [40], let us vary the PDD polynomial order from $m = 2$ to $m = 4$. Both criteria for the selection of the most important polynomial terms are tested. The computed mean, variance and sensitivity indices are reported in Table 10. Note that all first-order and total indices are presented here; concerning the second-order sensitivity indices, for conciseness, we only present the most important one measuring the interaction between p_∞ and M_∞. The moment results from Table 10 show the method is convergent when increasing the PDD order m for both of the two selection criteria. Meanwhile, for all cases considered here, the obtained model accuracy is very high (all superior to 0.9995). p_∞ and M_∞ are found to be the two most important parameters whose first-order sensitivity measures are largely superior to those of other parameters. The second-order interaction between these two parameters is found to be non-negligible compared to other first-order sensitivity estimates. γ reveals its negligible influence on p_{st}. We further observe the two criteria provide similar sensitivity estimates and moments. Following the numerical conditions used for this case, the Criterion 2 generally produces a sparser polynomial basis for a similar model accuracy. Note nevertheless the Criterion 2, based essentially upon the model error estimates for the selection of polynomial terms, requires more resolutions of regression problems than the Criterion 1.

	p_{st} (Criterion 1)			p_{st} (Criterion 2)		
	$m=2$	$m=3$	$m=4$	$m=2$	$m=3$	$m=4$
$E(p_{st})$	6499.31	6499.24	6498.83	6499.31	6499.27	6499.30
$V(p_{st})$	0.133743E+07	0.133791E+07	0.133790E+07	0.133761E+07	0.133791E+07	0.133787E+07
$\mathcal{S}_1$	0.411332E+00	0.411383E+00	0.411377E+00	0.411411E+00	0.411422E+00	0.411390E+00
$\mathcal{S}_2$	0.581267E+00	0.581192E+00	0.580948E+00	0.581190E+00	0.581178E+00	0.581124E+00
$\mathcal{S}_3$	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00
$\mathcal{S}_4$	0.937678E-06	0.141060E-05	0.508373E-06	0.936357E-06	0.763570E-06	0.000000E+00
$\mathcal{S}_5$	0.239850E-05	0.238962E-05	0.399189E-05	0.239324E-05	0.230749E-05	0.153589E-05
$\mathcal{S}_6$	0.746241E-06	0.826894E-06	0.795075E-06	0.741934E-06	0.855743E-06	0.780410E-06
$\mathcal{S}_7$	0.110647E-06	0.164247E-05	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00
$\mathcal{S}_{12}$	0.739341E-02	0.738204E-02	0.741574E-02	0.739252E-02	0.738666E-02	0.741212E-02
$\mathcal{S}_1^T$	0.418726E+00	0.418771E+00	0.418842E+00	0.418804E+00	0.418809E+00	0.418803E+00
$\mathcal{S}_2^T$	0.588663E+00	0.588578E+00	0.588393E+00	0.588585E+00	0.588569E+00	0.588547E+00
$\mathcal{S}_3^T$	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00
$\mathcal{S}_4^T$	0.937678E-06	0.383797E-04	0.121003E-03	0.936357E-06	0.957319E-05	0.575515E-04
$\mathcal{S}_5^T$	0.518330E-05	0.238962E-05	0.126563E-03	0.474173E-05	0.230749E-05	0.148786E-04
$\mathcal{S}_6^T$	0.746241E-06	0.826894E-06	0.795075E-06	0.741934E-06	0.855743E-06	0.780410E-06
$\mathcal{S}_7^T$	0.110647E-06	0.290228E-04	0.182279E-03	0.000000E+00	0.506326E-05	0.607349E-04
D_2	3	4	5	3	4	5
sparse base	14	34	79	11	14	21
sparsity	$\frac{14}{2187}$ ≈6.4e-3	$\frac{34}{16384}$ ≈2.1e-3	$\frac{79}{78125}$ ≈1.0e-3	$\frac{11}{2187}$ ≈5.0e-3	$\frac{14}{16384}$ ≈8.5e-4	$\frac{21}{78125}$ ≈2.7e-4
accuracy Q^2	0.999678	0.999718	0.999728	0.999680	0.999722	0.999743

Table 10: Application test. Uncertainty quantification results for p_{st} using criterion 1 and criterion 2 by varying the PDD polynomial order m. The truncation dimension $\nu = 2$. The experimental design size $Q = 1000$. The model accuracy is estimated by the cross validation method (see A). For Criterion 1, the threshold $\theta = 10^{-7}$; for Criterion 2, the threshold $\epsilon_{Q^2} = 10^{-7}$.

When the heat flux quantity q_{st} is considered, the same type of UQ analysis is realized and the results are reported in Table 11. We set the maximum interaction order $\nu = 3$. When

	q_{st} (Criterion 1)			q_{st} (Criterion 2)		
	$m=2$	$m=3$	$m=4$	$m=2$	$m=3$	$m=4$
$E(q_{st})$	286927.0	286945.0	285907.0	286898.0	286947.0	287233.0
$V(q_{st})$	0.296806E+10	0.303006E+10	0.387263E+10	0.296506E+10	0.301310E+10	0.311091E+10
$\mathcal{S}_1$	0.104034E+00	0.102941E+00	0.838386E-01	0.104603E+00	0.103322E+00	0.101710E+00
$\mathcal{S}_2$	0.874650E+00	0.872884E+00	0.712564E+00	0.876280E+00	0.877072E+00	0.872107E+00
$\mathcal{S}_3$	0.946742E-02	0.961702E-02	0.656856E-02	0.924117E-02	0.908686E-02	0.876759E-02
$\mathcal{S}_4$	0.366610E-03	0.367870E-03	0.611833E-02	0.000000E+00	0.000000E+00	0.000000E+00
$\mathcal{S}_5$	0.219524E-02	0.192399E-02	0.189310E-02	0.225662E-02	0.195543E-02	0.923366E-03
$\mathcal{S}_6$	0.200434E-03	0.139166E-02	0.151648E-02	0.000000E+00	0.000000E+00	0.000000E+00
$\mathcal{S}_7$	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00
$\mathcal{S}_{12}$	0.588863E-02	0.616165E-02	0.624935E-02	0.588868E-02	0.597284E-02	0.642160E-02
$\mathcal{S}_{23}$	0.140176E-02	0.151489E-02	0.805645E-03	0.133857E-02	0.135844E-02	0.104616E-02
$\mathcal{S}_{24}$	/	/	0.121030E-02	/	/	0.117491E-02
$\mathcal{S}_{13}$	0.176827E-03	0.945794E-03	0.337197E-03	0.000000E+00	0.532325E-03	0.000000E+00
$\mathcal{S}_{14}$	/	/	0.486300E-02	/	/	0.000000E+00
$\mathcal{S}_{34}$	/	/	0.218688E-01	/	/	0.000000E+00
$\mathcal{S}_{123}$	0.161953E-02	0.225210E-02	0.734882E-02	0.392357E-03	0.699826E-03	0.178231E-02
$\mathcal{S}_{124}$	/	/	0.293194E-01	/	/	0.144664E-02
$\mathcal{S}_{234}$	/	/	0.597520E-01	/	/	0.000000E+00
$\mathcal{S}_{134}$	/	/	0.557466E-01	/	/	0.462055E-02
$\mathcal{S}_1^T$	0.111719E+00	0.112301E+00	0.187703E+00	0.110884E+00	0.110527E+00	0.115981E+00
$\mathcal{S}_2^T$	0.883560E+00	0.882812E+00	0.817249E+00	0.883899E+00	0.885103E+00	0.883978E+00
$\mathcal{S}_3^T$	0.126655E-01	0.143298E-01	0.152428E+00	0.109721E-01	0.116774E-01	0.162166E-01
$\mathcal{S}_4^T$	0.366610E-03	0.367870E-03	0.178878E+00	0.000000E+00	0.000000E+00	0.724210E-02
$\mathcal{S}_5^T$	0.219524E-02	0.192399E-02	0.189310E-02	0.225662E-02	0.195543E-02	0.923366E-03
$\mathcal{S}_6^T$	0.200434E-03	0.139166E-02	0.151648E-02	0.000000E+00	0.000000E+00	0.000000E+00
$\mathcal{S}_7^T$	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00	0.000000E+00
D_2, D_3	3	3	4	3	3	4
sparse base	19	31	122	11	14	18
sparsity	$\frac{19}{2187}$ ≈8.7e-3	$\frac{31}{16384}$ ≈1.9e-3	$\frac{122}{78125}$ ≈1.6e-3	$\frac{11}{2187}$ ≈5.0e-3	$\frac{14}{16384}$ ≈8.5e-4	$\frac{18}{78125}$ ≈2.3e-4
accuracy Q^2	0.783770	0.793848	0.805251	0.785050	0.796990	0.813363

Table 11: Application test. Uncertainty quantification results for q_{st} using criterion 1 and criterion 2 by varying the PDD polynomial order m. The truncation dimension $\nu = 3$. The experimental design size $Q = 1000$. The model accuracy is estimated by the cross validation method (see A). For Criterion 1, the threshold $\theta = 10^{-4}$; for Criterion 2, the threshold $\epsilon_{Q^2} = 10^{-7}$.

the p constant is modified to $p = 0.995$ for both criteria, 3 active dimensions are retained for second- and third-order interactions when the PDD order is set as $m = 2$ and $m = 3$, whilst 4 active dimensions are retained for $m = 4$. Concerning the computed expectation and variance, the convergence is less good than for the quantity of p_{st} regarding both criteria. For Criterion 1, p_∞ and M_∞ are still found to be the most significant parameters, while the gas reaction $O_2 + O \rightarrow 2O + O$ (X_7) is negligible. Indeed, this reaction uncertainty is not taken into consideration in the final meta-model polynomial representation, since its total sensitivity effect is also zero. Moreover, Table 11 shows that all second- and third-order interactions are non-negligible: the orders of magnitude of these sensitivity indices are comparable to those of first-order contributions (X_3, X_4, X_5, X_6). In particular, in the case of $m = 4$, the third-order interactions are found more important than the second-order ones. It is also shown that the sensitivity measure of a parameter or a group of parameters can vary significantly when a different number of active dimensions is employed. For instance, the total sensitivity indice of X_4 is about 500 times bigger with $D_2, D_3 = 4$ than with $D_2, D_3 = 3$. Finally, the accuracy of the PDD model representation for q_{st} is found to be less good than for p_{st}, as also mentioned in the reference [40]. As far as the Criterion 2 is concerned, in order to obtain a similar model accuracy, a less number of polynomial terms are required. However, the model representation includes fewer uncertain parameters. For instance, X_6 and X_7 are excluded when using $m = 4$, and X_4 is additionally neglected if we set $m = 2$ or $m = 3$.

6   Conclusions

This paper aims to deal with engineering and physical problems featuring a moderate to large number of uncertain input parameters. The purpose is to identify the relative importances of these uncertainties onto a given quantity of interest. This is achieved in this work by performing global sensitivity analysis, and in particular by combining the Analysis of Variance technique (ANOVA) and the polynomial dimensional decomposition approach (PDD). Complexities present in pracitical problems usually make the global methods prohibitive due to the large uncertainty in the model output. In this paper, we have employed three levels of adaptivity to reduce the meta-modeling difficulty which read as follows:

1. We set a truncation dimension (the maximum interaction order) in the ANOVA expansion.

2. Resulting from the solution of the regression system including only the PDD terms of the first-order ANOVA component functions, a rank of importances can be established quantitatively for all the input parameters. Hence, we can retain so-called active dimensions (the most influential parameters) for the PDD terms of the second- and higher-order ANOVA components.

3. Starting from the PDD polynomials of the first-order components, we enrich our surrogate model representation by adding only significant polynomials of second- and higher-order functions. Two selection criteria have been utilized in this work for this purpose. We emphasize that recursive resolutions of regression problems are required for this task.

The resulting surrogate polynomial approximation is a very sparse representation of the deterministic model. Since the surrogate model size is updated recursively with respect to the resolutions of regression problems subject to the enrichment of the polynomial basis, the number of the required deterministic model evaluations is well controlled, and its final value is significantly smaller than when employing a standard Monte Carlo or quasi Monte Carlo method. The

computation of the global sensitivity indices simply requires a post-processing of the polynomial coefficients.

The proposed approach has been tested on two well-known benchmark problems. The results show the method is convergent, and can be more accurate and efficient using the PDD than the sparse polynomial chaos expansion [4]. On the other hand concerning the selection criterion for retaining the significant polynomial terms, the Criterion 1 based on the variance contribution is found to be very efficient, but can result in a slightly larger basis than the Criterion 2, based on the model error comparison, for a similar meta-model accuracy. The satisfactory uncertainty quantification results for the application example show that our approach is capable of treating complex engineering problems. Note in particular for this application case, the finally obtained surrogate model does not include all the input parameters, which means throughout the meta-modeling approach, non-important parameters are well identified, which reduces the modeling difficulty.

A   Estimator of accuracy Q^2

The *Leave-one-out cross validation* estimator presented in [4] is directly used in this work of which we summarize the main ingredients as follows. We keep the same notations as in [4].

Let $f_{\mathcal{X}}(\mathbf{x})$ denote the surrogate model representation obtained by adaptive sparse polynomial dimensional decomposition (PDD) with the coefficients determined by the regression approach with the experimental design $\mathcal{X} = \{\mathbf{x}^1, \mathbf{x}^2, \cdots, \mathbf{x}^Q\}$ (see Sections 3 and 4). We remind that $\mathcal{Y} = \{y^1, y^2, \cdots, y^Q\}$ represents the corresponding experimental design outputs.

Let $f_{\mathcal{X} \setminus i}(\mathbf{x})$ denote the sparse PDD representation constructed from the experimental design $\mathcal{X} \setminus \{\mathbf{x}^i\}$.

The following *empirical mean-square predicted residual* [4] is computed to estimate the approximation error:

$$I_{\mathcal{X}}^*[f_{\mathcal{X}}] = \frac{1}{Q} \sum_{i=1}^{Q} \left(y^i - f_{\mathcal{X} \setminus i}(\mathbf{x}^i) \right)^2. \tag{46}$$

In the case of linearly parameterized regression, we have the following result [4] useful for the computation of (46):

$$y^i - f_{\mathcal{X} \setminus i}(\mathbf{x}^i) = \frac{y^i - f_{\mathcal{X}}(\mathbf{x}^i)}{1 - h_i} \tag{47}$$

where h_i is the i-th diagonal term of the *projection matrix*

$$\mathbf{\Psi}(\mathbf{\Psi}^T \mathbf{\Psi})^{-1} \mathbf{\Psi}^T$$

where $\mathbf{\Psi}$ is expressed as in (32). Hence, (46) can be expressed as

$$I_{\mathcal{X}}^*[f_{\mathcal{X}}] = \frac{1}{Q} \sum_{i=1}^{Q} \left(\frac{y^i - f_{\mathcal{X}}(\mathbf{x}^i)}{1 - h_i} \right)^2. \tag{48}$$

Note the use of the formulation (48) avoids any additional resolution of regression problems concerning $f_{\mathcal{X} \setminus i}(\mathbf{x})$.

The *determination coefficient* then writes as follows:

$$Q^2[f_{\mathcal{X}}] = 1 - \frac{I_{\mathcal{X}}^*[f_{\mathcal{X}}]}{\hat{\mathbb{V}}[\mathcal{Y}]}, \tag{49}$$

where

$$\hat{\mathbb{V}}[\mathcal{Y}] = \frac{1}{Q-1}\sum_{i=1}^{Q}\left(y^i - \bar{y}\right)^2, \quad \bar{y} = \frac{1}{Q}\sum_{i=1}^{Q}y^i. \tag{50}$$

The quantity Q^2 is exclusively used in this work to estimate the accuracy of the surrogate model representation. $Q^2 = 1$ indicates a perfect fit, while $Q^2 \approx 0$ or $Q^2 < 0$ reflects a poor model accuracy.

A.1 Criterion 2: select the most important polynomials by comparing the model accuracy

The algorithm 1 can be modified to algorithm 2, by considering the Criterion 2 based on the model accuracy Q^2 (in particular, see the operations 13–14 in algorithm 2).

Algorithm 2 Adaptive PDD by stepwise regression (**Criterion 2**)

1: initialization of multivariate PDD Base: $\{\psi^w\} = \{\psi_{\alpha^1}\}$
2: **for** $\psi_{\alpha_i} \in \{\psi_{\alpha^{2+}}\}$ **do**
3:      add ψ_{α_i} into $\{\psi^w\}$, i.e. $\{\psi^w\} = \{\psi^w, \psi_{\alpha_i}\}$
4:      depending on the size P^w of $\{\psi^w\}$, adjust, if necessary, the size Q^w of the experimental design (see formulation (29))
5:      solve the regression system (31) to determine the PDD expansion coefficients $\widetilde{\mathbf{C}^w}$.
6:      evaluate the model accuracy Q_i^2
7:      **if** $Q_i^2 \geq Q_{tgt}^2$ **then**
8:          exit
9:      **end if**
10:      **for** $\psi_{\alpha_j} \in \{\psi^w\}$ **do**
11:          solve the regression system with the polynomial base $\{\psi^w\} \setminus \psi_{\alpha_j}$, i.e. by excluding ψ_{α_j}
12:          evaluate the model accuracy $Q_{i \setminus \alpha_j}^2$
13:          **if** $Q_i^2 - Q_{i \setminus \alpha_j}^2 < \epsilon_{Q^2}$ **then**
14:             eliminate this polynomial: $\{\psi^w\} = \{\psi^w\} \setminus \psi_{\alpha_j}$
15:          **end if**
16:      **end for**
17: **end for**
18: solve the final regression system based on the constructed base $\{\psi^F\}$
19: compute the final total variance using the obtained PDD coefficients
20: compute the global (total) sensitivity indices

Q_{tgt}^2 in the operation 7 of the algorithm 2 is a predefined target accuracy (e.g. 0.999). Once the algorithm reaches this target accuracy, we stop and use the constructed base. This target accuracy can also be set for the algorithm 1. ϵ_{Q^2} in the operation 13 is a predefined threshold (e.g. 10^{-5}). If the decrease of the model accuracy is smaller than ϵ_{Q^2} when excluding the polynomial term in consideration, we then eliminate this term from our working base.

Note that the initialization of multivariate PDD base (operation 1) can be realized by Criterion 2 in algorithm 2.

The operation 11 in algorithm 2 involves the additional resolution of regression systems compared to the algorithm 1. Thus, the algorithm 2 is in general more expensive than the algorithm 1.

References

[1] G. E. B. Archer, A. Saltelli, and I. M. Sobol'. Sensitivity measures, anova-like techniques and the use of bootstrap. *Journal of Statistical Computation and Simulation*, 58:99–120, 1997.

[2] P.F. Barbante. *Accurate and efficient modelling of high temperature non-equilibrium air flows*. PhD thesis, von Karman Institute, Rhode-Saint-Genèse, Belgium, 2001.

[3] Philip S. Beran, Chris L. Pettit, and Daniel R. Millman. Uncertainty quantification of limit-cycle oscillations. *Journal of Computational Physics*, 217(1):217–247, 2006.

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

[5] Géraud Blatman and Bruno Sudret. An adaptive algorithm to build up sparse polynomial chaos expansions for stochastic finite element analysis. *Probabilistic Engineering Mechanics*, 25(2):183–197, April 2010.

[6] Géraud Blatman and Bruno Sudret. Adaptive sparse polynomial chaos expansion based on least angle regression. *Journal of Computational Physics*, 230(6):2345–2367, March 2011.

[7] E. Borgonovo. A new uncertainty importance measure. *Reliability Engineering & System Safety*, 92(6):771–784, June 2007.

[8] E Borgonovo, G E Apostolakis, S Tarantola, and A Saltelli. Comparison of global sensitivity analysis techniques and importance measures in PSA. *Reliability Engineering & System Safety*, 79(2):175–185, 2003.

[9] Emanuele Borgonovo and Stefano Tarantola. Advances in sensitivity analysis. *Reliability Engineering & System Safety*, 107:1–2, November 2012.

[10] D. Bose, M. Wright, and T. Gokçen. Uncertainty and Sensitivity Analysis of Thermochemical Modeling for Titan Atmospheric Entry. *37th AIAA Thermophysics Conference, Portland, Oregon*, 2004.

[11] J Y Caniou and B Sudret. Distribution-based global sensitivity analysis in case of correlated input parameters using polynomial chaos expansions. In *ICASP2011*, 2011.

[12] Seung-Kyum Choi, Ramana V. Grandhi, Robert A. Canfield, and Chris L. Pettit. Polynomial Chaos Expansion with Latin Hypercube Sampling for Estimating Response Variability. *AIAA Journal*, 42(6):1191–1198, 2004.

[13] F. Coquel and M.S. Liou. Hybrid Upwind Splitting Scheme by a Field-by-Field Decomposition. *NASA Tech. Memorandum 106823*, 1995.

[14] Thierry Crestaux, Olivier Le Maître, and Jean-Marc Martinez. Polynomial chaos expansion for sensitivity analysis. *Reliability Engineering & System Safety*, 94(7):1161–1172, July 2009.

[15] Jasmine Foo and George Em Karniadakis. Multi-element probabilistic collocation method in high dimensions. *Journal of Computational Physics*, 229(5):1536–1557, March 2010.

[16] Zhen Gao and Jan S. Hesthaven. On ANOVA expansions and strategies for choosing the anchor point. *Applied Mathematics and Computation*, 217(7):3274–3285, December 2010.

[17] T. Ishigami and T. Homma. An importance quantification technique in uncertainty analysis for computer models. *ISUMA'90, First International Symposium on Uncertainty Modeling and Analysis*, pages 398–403, 1990.

[18] O. Le Maître and O. M. Knio. *Spectral Methods for Uncertainty Quantification*. Springer, 2010.

[19] O. Le Maître, M. T. Reagan, H. N. Najm, R. G. Ghanem, and O. M. Knio. A Stochastic Projection Method for Fluid Flow II. Random Process. *Journal of Computational Physics*, 181(1):9–44, September 2002.

[20] B. Van Leer. Towards the ultimate conservative difference scheme. V. A second-order sequel to godunov's method. *Journal of Computational Physics*, 32(1):101–136, 1979.

[21] G. Li, H. Rabitz, P. E. Yelvington, O. O. Oluwole, F. Bacon, C. E. Kolb, and J. Schoendorf. Global sensitivity analysis for systems with independent and/or correlated inputs. *The journal of physical chemistry. A*, 114(19):6022–32, May 2010.

[22] Genyuan Li, Carey Rosenthal, and Herschel Rabitz. High Dimensional Model Representations. *The journal of physical chemistry. A*, 105(33), 2001.

[23] X. Ma and N. Zabaras. An adaptive high-dimensional stochastic model representation technique for the solution of stochastic partial differential equations. *Journal of Computational Physics*, 229(10):3884–3915, May 2010.

[24] Hermann G. Matthies and Andreas Keese. Galerkin methods for linear and nonlinear elliptic stochastic partial differential equations. *Computer Methods in Applied Mechanics and Engineering*, 194(12-16):1295–1331, 2005.

[25] S. Osher and F. Solomon. Upwind Difference Schemes for Hyperbolic Systems of Conservation Laws. *Mathematics of Computation*, 38(158):339–374, 1982.

[26] C. Park, R.L. Jaffe, and H. Partridge. Chemical-Kinetic Parameters of Hyperbolic Earth Entry. *Journal of Thermophysics and Heat Transfer*, 15(1):76–90, 2001.

[27] Herschel Rabitz and Omer F. Alis. General foundations of high-dimensional model representations. *Journal of Mathematical Chemistry*, 25(2-3):197–233, 1999.

[28] Herschel Rabitz, Ömer F. Alis, Jeffrey Shorter, and Kyurhee Shim. Efficient input-output model representations. *Computer Physics Communications*, 117(1-2):11–20, 1999.

[29] Sharif Rahman. A polynomial dimensional decomposition for stochastic computing. *International Journal for Numerical Methods in Engineering*, 76:2091–2116, 2008.

[30] Sharif Rahman. Global sensitivity analysis by polynomial dimensional decomposition. *Reliability Engineering & System Safety*, 96(7):825–837, July 2011.

[31] Sharif Rahman and Xuchun Ren. Novel computational methods for high-dimensional stochastic sensitivity analysis. *International Journal for Numerical Methods in Engineering*, 2014.

[32] Sharif Rahman and Vaibhav Yadav. Orthogonal Polynomial Expansions for Solving Random Eigenvalue Problems. *International Journal for Uncertainty Quantification*, 1(2):163–187, 2011.

[33] A Saltelli. Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. *Comput. Phys. Commun.*, 181:259–270, 2010.

[34] I. M. Sobol. Uniformly distributed sequences with an additional uniform property. *USSR Computational Mathematics and Mathematical Physics*, 16(5):236–242, 1976.

[35] I. M. Sobol'. On sensitivity estimation for nonlinear mathematical models. *Matem. Mod.*, 2(1):112–118, 1990.

[36] I. M. Sobol'. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. *Mathematics and Computers in Simulation*, 55(1-3):271–280, February 2001.

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

[38] Kunkun Tang, Pietro M Congedo, and Rémi Abgrall. Sensitivity analysis using anchored ANOVA expansion and high order moments computation. *International Journal for Numerical Methods in Engineering*, pages 1–31, 2015.

[39] J. Thoemel, J. Muylaert, F. Ratti, and J. Gavira. In-flight testing of critical technologies and experimentation of aerothermodynamic phenomena. In *Proceedings of the 16th AIAA/DLR/DGLR International Space Planes and Hypersonic Systems and Technologies Conference*, Bremen, Germany, 2009. American Institute of Aeronautics and Astronautics.

[40] J. Tryoen, P. M. Congedo, R. Abgrall, N. Villedieu, and T. E. Magin. Bayesian-based method with metamodels for rebuilding freestream conditions in atmospheric entry flows. *AIAA Journal*, 52(10):2190–2197, 2014.

[41] N. Villedieu, F. Panerai, O. Chazot, and T.E. Magin. Uncertainty quantification for gas-surface interaction in plasmatron. *Specialists Meeting on Catalytic Gas Surface Interactions (AVT-199), Rhode-Saint-Genèse, Belgium*, 2012.

[42] X. Wang. On the approximation error in high dimensional model representation. In *Proceedings of the 2008 Winter Simulation Conference*, number x, pages 453–462, 2008.

[43] M. J. Wright, D. Bose, G. E. Palmer, and E. Levin. Recommended collision integrals for transport property computations part 1: Air species. *AIAA Journal*, 43:2558–2564, 2005.

[44] Dongbin Xiu and George Em Karniadakis. The Wiener–Askey Polynomial Chaos for Stochastic Differential Equations. *SIAM Journal on Scientific Computing*, 24(2):619–644, January 2002.

[45] Dongbin Xiu and George Em Karniadakis. Modeling uncertainty in flow simulations via generalized polynomial chaos. *Journal of Computational Physics*, 187(1):137–167, May 2003.

[46] C Xu and G Gertner. Extending a global sensitivity analysis technique to models with correlated parameters. *Computational Statistics & Data Analysis*, 51:5579–5590, 2007.

[47] C Xu and G Gertner. Uncertainty and sensitivity analysis for models with correlated parameters. *Reliability Engineering & System Safety*, 93:1563–1573, 2008.

[48] Vaibhav Yadav and Sharif Rahman. Adaptive-sparse polynomial dimensional decomposition methods for high-dimensional stochastic computing. *Computer Methods in Applied Mechanics and Engineering*, 274:56–83, June 2014.

[49] X. Yang, M. Choi, G. Lin, and G. E. Karniadakis. Adaptive ANOVA decomposition of stochastic incompressible and compressible flows. *Journal of Computational Physics*, 231(4):1587–1614, February 2012.

[50] Z. Zhang, M. Choi, and G. E. Karniadakis. *Spectral and High Order Methods for Partial Differential Equations*, volume 76 of *Lecture Notes in Computational Science and Engineering*. Springer Berlin Heidelberg, Berlin, Heidelberg, 2011.

[51] Z. Zhang, M. Choi, and G. E. Karniadakis. Error estimates for the ANOVA method with Polynomial Chaos interpolation: Tensor product functions. *SIAM J. Sci. Comput.*, 34(2):A1165–A1186, 2012.