Global Uncertainty Assessments by High Dimensional Model Representations (HDMR)

ARTICLE *in* CHEMICAL ENGINEERING SCIENCE · NOVEMBER 2002

Impact Factor: 2.34 · DOI: 10.1016/S0009-2509(02)00417-7

5 AUTHORS, INCLUDING:

Genyuan Li
Princeton University
54 PUBLICATIONS **1,548** CITATIONS
SEE PROFILE

Herschel Rabitz
Princeton University
947 PUBLICATIONS **23,653** CITATIONS
SEE PROFILE

Sookyun Wang
Pukyong National University
27 PUBLICATIONS **590** CITATIONS
SEE PROFILE

Peter R. Jaffe
Princeton University
134 PUBLICATIONS **3,674** CITATIONS
SEE PROFILE

PERGAMON

Chemical Engineering Science 57 (2002) 4445–4460

Chemical Engineering Science

www.elsevier.com/locate/ces

Global uncertainty assessments by high dimensional model representations (HDMR)

Genyuan Li[a], Sheng-Wei Wang[a,1], Herschel Rabitz[a,*], Sookyun Wang[b], Peter Jaffé[b]

[a] *Department of Chemistry, Princeton University, Princeton, NJ 08544, USA*
[b] *Department of Civil and Environmental Engineering, Princeton University, Princeton, NJ 08544, USA*

Received 1 January 2002; received in revised form 4 June 2002; accepted 15 July 2002

Abstract

A general set of quantitative model assessment and analysis tools, termed high-dimensional model representations (HDMR), have been introduced recently for high dimensional input–output systems. HDMR are a particular family of representations where each term in the representation reflects the independent and cooperative contributions of the inputs upon the output. When data are randomly sampled, a RS(random sampling)-HDMR can be constructed, which is an efficient tool to provide a fully global statistical analysis of a model. The individual RS-HDMR component functions have a direct statistical correlation interpretation. This relation permits the model output variance σ^2 to be decomposed into its input contributions $\sigma^2 = \sum_i \sigma_i^2 + \sum_{i<j} \sigma_{ij}^2 + \cdots$ due to the independent variable action σ_i^2, the pair correlation action σ_{ij}^2, etc. The information gained from this decomposition can be valuable for attaining a physical understanding of the origins of output uncertainty as well as suggesting additional laboratory/field studies or model refinements to best improve the quality of the model. To reduce sampling effort, the RS-HDMR component functions are approximately represented by orthonormal polynomials. Only one randomly sampled set of input–output data is needed to determine all σ_i, σ_{ij}, etc. and a few hundred samples may give reliable results. This paper presents its methodology and applications on an atmospheric photochemistry model and a trace metal bioremediation model.

Keywords: Global uncertainty analysis; High dimensional model representation; Monte Carlo method

1. Introduction

In all areas of modeling the issue of predictive quality, in view of model input uncertainty, has been a topic of prime concern. Systematic uncertainty assessments can identify the inputs with the most influence and their action patterns on the predictions. The resultant information can provide guidance for new experiments or parameter refinements to reduce the output uncertainty. Essentially two approaches have been taken to treat this problem: sensitivity analysis (Iman & Conover, 1980; Gardner, 1983; Yetter, Eslava, Dryer, & Rabitz, 1984; Rabitz, 1989) and Monte Carlo sampling (Stolarski, Butler, & Rundle, 1978; Doll & Freeman, 1986; Sobol, 1993). In sensitivity analysis output uncertainties are often explicitly represented as functions of input uncertainties and the method is used for small input ranges. Traditional Monte Carlo methods can be used to treat implicit models, but they can be prohibitively expensive when the system has a large number of inputs. Even though different improvements have been suggested (Cukier, Levine, & Schuler, 1978; Iman & Conover, 1980; McRae, Tilden, & Seinfeld, 1981; McRae, Goodin, & Seinfeld, 1982; Tatang, Pan, Prinn, & McRae, 1997; Saltelli & Bolado, 1998; Saltelli, Tarantola, & Chan, 1999; Isukapalli, Roy, & Georgopoulos, 2000), global uncertainty assessments for high dimensional systems (with a large number of inputs) are still a challenge.

A general set of quantitative model assessment and analysis tools, termed high dimensional model representations (HDMR), have been introduced recently for improving the efficiency of deducing high dimensional input–output (IO) system behavior (Rabitz, Alis, Shorter, & Shim, 1999; Alis & Rabitz, 1999, 2000; Shorter, Precila, & Rabitz, 1999; Shorter & Rabitz, 2000; Li, Rosenthal, & Rabitz, 2001a). HDMR techniques are based on optimization and projection operator theory, which can dramatically reduce the sampling

* Corresponding author. Tel.: +1-609-258-3917; fax: +1-609-258-0967.
E-mail address: hrabitz@chemvax.princeton.edu (H. Rabitz).
[1] Environmental and Occupational Health Sciences Institute, 170 Frelinghuysen Road, Piscataway, NJ 08854, USA.

0009-2509/02/$ - see front matter © 2002 Published by Elsevier Science Ltd.
PII: S0009-2509(02)00417-7

effort for learning the IO behavior of high dimensional systems (i.e., a reduction of effort from exponential scaling to only polynomic complexity). HDMR can be applied for different purposes: construction of a computational model directly from laboratory/field data, creating an efficient fully equivalent operational model (FEOM) to replace an existing time-consuming mathematical model, identification of key model variables, global uncertainty assessments, efficient quantitative risk assessments, etc. For global uncertainty assessments, with only a modest sampling effort, HDMR can provide reliable information, especially by decomposing the model output variance σ^2 into its different contributions: the independent input variable action σ_i^2, the pair correlation action σ_{ij}^2, etc. This information is most valuable for attaining a physical understanding of the origins of output uncertainty as well as providing suggestions for additional laboratory/field studies or parameter refinements to best improve the quality of the model predictions. In contrast, except of σ_i^2 the current statistical approaches cannot easily determine high order correlated actions (Saltelli & Bolado, 1998; Saltelli, et al., 1999).

In this paper, we present the methodology and the results of global uncertainty assessments given by RS-HDMR for different models. For a better understanding of the methodology of HDMR, in Section 2 the principles of HDMR are first briefly presented, and then RS-HDMR, its approximation by orthonormal polynomials and applications in global uncertainty assessments are provided. Sections 3 and 4 illustrate the applications of RS-HDMR for global uncertainty assessments to an atmospheric photochemistry model and a trace metal bioremediation model. Finally, Section 5 contains conclusions and a discussion.

2. Principles of HDMR with applications in global uncertainty assessments

Many important problems in science and engineering reduce to finding an efficiently constructed map of the relationship between sets of high dimensional system input and output variables. The system may be described by a mathematical model (e.g., typically a set of ordinary or partial differential equations), where the input variables might be specified initial and/or boundary conditions, parameters or functions in the model, and the output variables would be the model solutions or a functional of them. The IO behavior may also be based on observations in the laboratory or field where a mathematical model cannot readily be constructed for an observed system. In this case the system is simply considered as a black box where the input consists of the measured laboratory or field (control) variables and the output(s) is the observed system response. Regardless of the circumstances, the input is often very high dimensional with many variables even if the output is only a single quantity. We refer to the input variables collectively as $\mathbf{x} = \{x_1, x_2, \ldots, x_n\}$ with n ranging up to $\sim 10^2 - 10^3$ or more, and the output as $f(\mathbf{x})$. For simplicity in the remainder of the paper and without loss of generality, we shall refer to the system as a model regardless of whether it involves modeling, laboratory experiments or field studies.

2.1. Theoretical foundations of HDMR

As the impact of the multiple input variables on the output can be independent and cooperative, HDMR expresses the model output $f(\mathbf{x})$ as a finite hierarchical correlated function expansion in terms of the input variables (Alis & Rabitz, 1999):

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \leqslant i < j \leqslant n} f_{ij}(x_i, x_j) + \cdots + \sum_{1 \leqslant i_1 < \cdots < i_l \leqslant n} f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) + \cdots + f_{12\ldots n}(x_1, x_2, \ldots, x_n), \quad (1)$$

where the zeroth order (i.e., $l = 0$) component function f_0 is a constant representing the mean response to $f(\mathbf{x})$, and the first order (i.e., $l = 1$) component function $f_i(x_i)$ gives the independent contribution to $f(\mathbf{x})$ by the ith input variable acting alone, the second order (i.e., $l = 2$) component function $f_{ij}(x_i, x_j)$ gives the pair correlated contribution to $f(\mathbf{x})$ by the input variables x_i and x_j, etc. The last term $f_{12\ldots n}(x_1, x_2, \ldots, x_n)$ contains any residual nth order correlated contribution of all input variables. The above HDMR expansion has a finite number of terms and is always exact. Other expansions have been suggested (Ghanem & Spanos, 1991), but they commonly have an infinite number of terms with some specified functions (e.g., Hermite polynomials).

The basic conjecture underlying HDMR is that the component functions in Eq. (1) arising in typical real problems are not likely to exhibit high order cooperatives among the input variables such that the significant terms in the HDMR expansion are expected to satisfy the relation: $l \ll n$ for $n \gg 1$. Experience shows that an HDMR expansion to second order

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \leqslant i < j \leqslant n} f_{ij}(x_i, x_j) \quad (2)$$

often provides a satisfactory description of $f(\mathbf{x})$ for many high dimensional systems when the input variables are properly chosen. HDMR attempts to exploit this observation to efficiently determine high dimensional IO system mapping. The presence of only low order variable cooperativity does not necessarily imply a small set of significant variables nor does it limit the nonlinear nature of the IO relationship. Fig. 1 gives an example of typical first and second order HDMR component functions which reveal the nonlinear relationships between model inputs and outputs.

Exploiting the expected low order variable cooperativity in high dimensional systems can only be done if practical formulations of the HDMR component functions can be found. To achieve this goal, the HDMR expansion

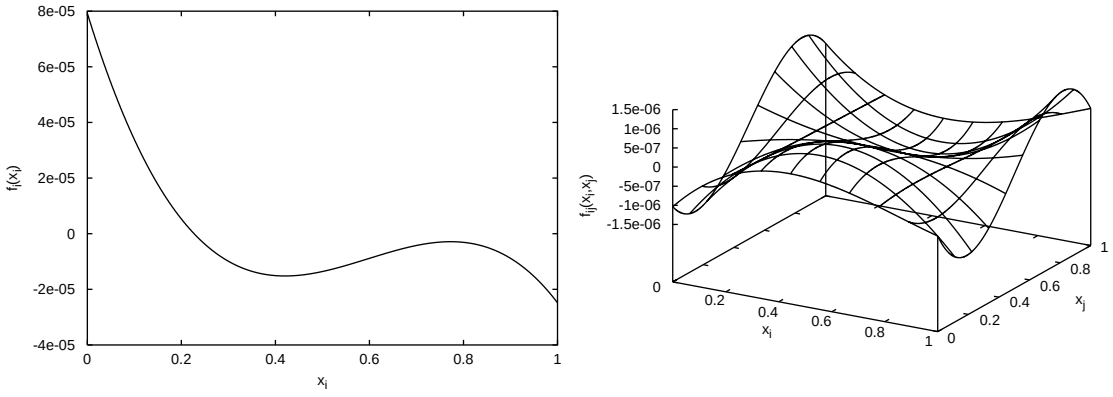

Fig. 1. The functional behavior of typical first (left) and second (right) order HDMR component functions from a reactive transport bioremediation model for uranium contaminated groundwater (Wang, Jaffé, Li, Wang, & Rabitz, 2002). Here the input variables x_i and x_j are rate constants in the model, and the output f is the cumulative flux of uranium (U^{+4}) passing through a given cross-section of the aquifer. This nonlinear behavior is typical of many chemical/physical models.

component functions f_0, $f_i(x_i), f_{ij}(x_i,x_j),\ldots$ are *optimally tailored* to a given $f(\mathbf{x})$ over the entire desired domain Ω of $\mathbf{x}$. A particular component function $f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})$ $(l = 0, 1, \ldots, n-1)$ is obtained by minimizing the functional

$$\min_{f_{i_1 i_2 \ldots i_l}} \int_{\Omega} w_{i_1 i_2 \ldots i_l}(\hat{\mathbf{x}}, \mathbf{u}) \Bigg[f(\mathbf{u}) - f_0 - \sum_{i=1}^{n} f_i(u_i) - \sum_{1 \leqslant i < j \leqslant n} f_{ij}(u_i, u_j) - \cdots - \sum_{1 \leqslant i_1 < \cdots < i_l \leqslant n} f_{i_1 i_2 \ldots i_l}(u_{i_1}, u_{i_2}, \ldots, u_{i_l}) \Bigg]^2 \mathrm{d}\mathbf{u} \tag{3}$$

under a suitable specified orthogonality condition which guarantees that all the component functions are determined step-by-step. Here, $\hat{\mathbf{x}} = (x_{i_1}, x_{i_2}, \ldots, x_{i_l})$, $\mathrm{d}\mathbf{u} = \mathrm{d}u_1\, \mathrm{d}u_2 \cdots \mathrm{d}u_n$, and $w_{i_1 i_2 \ldots i_l}(\hat{\mathbf{x}}, \mathbf{u})$ is a weight function.

Different weight functions will produce distinct, but formally equivalent HDMR expansions, all of the same structure as Eq. (1). There are two commonly used HDMR expansions: Cut- and RS-HDMR corresponding to ordered and random sampling, respectively. Cut-HDMR expresses $f(\mathbf{x})$ in reference to a specified cut point $\bar{\mathbf{x}}$ in Ω while RS-HDMR depends on the average value of $f(\mathbf{x})$ over the whole domain Ω.

2.1.1. Ordered sampling: Cut-HDMR

For Cut-HDMR a reference point $\bar{\mathbf{x}}$ is first chosen in Ω of the n-dimensional input variable $\mathbf{x}$ space. When Cut-HDMR is taken to convergence, the representation of $f(\mathbf{x})$ is invariant to the choice of $\bar{\mathbf{x}}$. In practical circumstances it can be wise to choose $\bar{\mathbf{x}}$ in the neighborhood of a local of interest in Ω.

The Cut-HDMR component functions with respect to reference point $\bar{\mathbf{x}}$ have the following forms:

$$f_0 = f(\bar{\mathbf{x}}), \tag{4}$$

$$f_i(x_i) = f(x_i, \bar{\mathbf{x}}^i) - f_0, \tag{5}$$

$$f_{ij}(x_i, x_j) = f(x_i, x_j, \bar{\mathbf{x}}^{ij}) - f_i(x_i) - f_j(x_j) - f_0,$$

$$\vdots \tag{6}$$

where

$$(x_i, \bar{\mathbf{x}}^i) = (\bar{x}_1, \ldots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \ldots, \bar{x}_n), \tag{7}$$

$$(x_i, x_j, \bar{\mathbf{x}}^{ij}) = (\bar{x}_1, \ldots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \ldots, \bar{x}_{j-1}, x_j, \bar{x}_{j+1}, \ldots, \bar{x}_n). \tag{8}$$

The last term $f_{12\ldots n}(x_1, x_2, \ldots, x_n)$ is determined by the difference between $f(\mathbf{x})$ and all other component functions in Eq. (1).

The above formulas can be readily obtained from Eq. (3) or simply by substituting $(x_{i_1}, x_{i_2}, \ldots, x_{i_l}, \bar{\mathbf{x}}^{i_1 i_2 \ldots i_l})$ with different sets of $\{i_1, i_2, \ldots, i_l\} \subset \{1, 2, \ldots, n\}$ for $\mathbf{x}$ on the both sides of Eq. (1) and using the following specified condition: a component function of Cut-HDMR vanishes when any of its own variables takes the value of the corresponding element in $\bar{\mathbf{x}}$, i.e.,

$$f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})|_{x_s = \bar{x}_s} = 0, \quad s \in \{i_1, i_2, \ldots, i_l\}. \tag{9}$$

Eq. (9) serves to define an orthogonal relation between two different component functions of Cut-HDMR as

$$f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l}) f_{j_1 j_2 \ldots j_k}(x_{j_1}, x_{j_2}, \ldots, x_{j_k})|_{x_s = \bar{x}_s} = 0, \tag{10}$$

$s \in \{i_1, i_2, \ldots, i_l\} \cup \{j_1, j_2, \ldots, j_k\}$.

The Cut-HDMR component functions $f_i(x_i), f_{ij}(x_i, x_j), \ldots$ are typically attained numerically at discrete values of the input variables $x_i, x_j, \ldots$ produced from ordered sampling the output function $f(\mathbf{x})$ for employment on the RHS of Eqs. (4)–(6). Note that the Cut-HDMR component functions are defined along *cut* lines, planes, subvolumes, etc. through the reference point $\bar{\mathbf{x}}$ in Ω.

Since all the Cut-HDMR component functions satisfy a minimization condition in Eq. (3), they are optimal choices for a given output $f(\mathbf{x})$, and often only low order terms of Cut-HDMR are needed to give a satisfactory approximation for $f(\mathbf{x})$. Numerical data tables can be constructed for these component functions, and the values of $f(\mathbf{x})$ for an arbitrary point $\mathbf{x}$ in Ω can be determined from these tables by performing only low dimensional interpolation over $f_i(x_i), f_{ij}(x_i,x_j),\ldots$. If each input variable is sampled at s different values, the required number of model runs to construct the $f_i(x_i)$, $f_{ij}(x_i,x_j),\ldots$ tables is

$$1 + n(s-1) + \frac{n(n-1)(s-1)^2}{2} + \cdots,$$

which grows only polynomically with n and s. As only low order component functions of Cut-HDMR are used, the sample savings for large n are significant compared to traditional s^n exponential sampling. Thus, Cut-HDMR renders the original exponential difficulty to a problem of only polynomic complexity.

2.1.2. Random sampling: RS-HDMR

For RS-HDMR, we first rescale variables x_i such that $0 \leqslant x_i \leqslant 1$ for all i by performing suitable transformations of the input variables. The output function $f(\mathbf{x})$ is then defined in the unit hypercube $K^n = \{(x_1,x_2,\ldots,x_n)|0 \leqslant x_i \leqslant 1, i = 1,2,\ldots,n\}$. The component functions of RS-HDMR possess the following forms:

$$f_0 = \int_{K^n} f(\mathbf{x})\,\mathrm{d}\mathbf{x}, \tag{11}$$

$$f_i(x_i) = \int_{K^{n-1}} f(x_i,\mathbf{x}^i)\,\mathrm{d}\mathbf{x}^i - f_0, \tag{12}$$

$$\begin{aligned} f_{ij}(x_i,x_j) &= \int_{K^{n-2}} f(x_i,x_j,\mathbf{x}^{ij})\,\mathrm{d}\mathbf{x}^{ij} - f_i(x_i) - f_j(x_j) - f_0, \\ &\vdots \end{aligned} \tag{13}$$

where $\mathrm{d}\mathbf{x}^i$ and $\mathrm{d}\mathbf{x}^{ij}$ are just the product $\mathrm{d}x_1\,\mathrm{d}x_2\cdots\mathrm{d}x_n$ without $\mathrm{d}x_i$ and $\mathrm{d}x_i\,\mathrm{d}x_j$, respectively. Finally, the last term $f_{12\ldots n}(x_1,x_2,\ldots,x_n)$ is determined from the difference between $f(\mathbf{x})$ and all the other lower order component functions in Eq. (1).

Considering that the domain Ω is a unit hypercube, f_0 is just the mean value $\bar{f}$ of $f(\mathbf{x})$ over the whole domain in contrast with f_0 of Cut-HDMR which is the value of $f(\mathbf{x})$ at the specified single reference point $\bar{\mathbf{x}}$.

All the above formulas can be readily obtained from Eq. (3) or simply by integrating both sides of Eq. (1) with respect to different sets of input variables $\{x_{i_1},x_{i_2},\ldots,x_{i_l}\}$ $(l = n,n-1,\ldots,1)$, and using the following specified condition: the integral of a component function of RS-HDMR with respect to any of its own variables is zero, i.e.,

$$\int_0^1 f_{i_1 i_2\ldots i_l}(x_{i_1},x_{i_2},\ldots,x_{i_l})\,\mathrm{d}x_s = 0, \quad s \in \{i_1,i_2,\ldots,i_l\}, \tag{14}$$

which defines the orthogonality relation between two different RS-HDMR component functions as

$$\begin{aligned} &\int_{K^n} f_{i_1 i_2\ldots i_l}(x_{i_1},x_{i_2},\ldots,x_{i_l}) \\ &\quad\times f_{j_1 j_2\ldots j_k}(x_{j_1},x_{j_2},\ldots,x_{j_k})\,\mathrm{d}\mathbf{x} = 0. \\ &\{i_1,i_2,\ldots,i_l\} \neq \{j_1,j_2,\ldots,j_k\}. \end{aligned} \tag{15}$$

Evaluation of the high dimensional integrals in the RS-HDMR expansion may be carried out by Monte Carlo random sampling integration (Press, Teukolsky, Vetterling, & Flannery, 1992) and hence the name RS (Random Sampling)-HDMR.

According to the above formulas one can see that all the component functions of the Cut- and RS-HDMR expansions can be directly constructed from the values of output $f(\mathbf{x})$ either at ordered or randomly sampled points of $\mathbf{x}$, which makes the determination of f_0, $f_i(x_i)$, $f_{ij}(x_i,x_j),\ldots$ straight forward.

2.2. Global uncertainty assessments by RS-HDMR

The orthogonality property of the RS-HDMR component functions implies that *the component functions in the RS-HDMR expansion are independent and give unique physical information on the relationships amongst the model input variables for their actions on the output properties*, i.e., the individual component functions of the RS-HDMR expansion have a direct statistical correlation interpretation, that permits the model output variance σ_f^2 to be decomposed into its input variable contributions. Using the orthogonality of the component functions we have

$$\begin{aligned} \sigma_f^2 &= \int_{K^n} [f(\mathbf{x}) - \bar{f}]^2\,\mathrm{d}\mathbf{x} = \int_{K^n} [f(\mathbf{x}) - f_0]^2\,\mathrm{d}\mathbf{x} \\ &= \int_{K^n} \left[\sum_{i=1}^n f_i(x_i) + \sum_{1\leqslant i<j\leqslant n} f_{ij}(x_i,x_j) + \cdots\right]^2 \mathrm{d}\mathbf{x} \\ &= \sum_{i=1}^n \int_{K^n} f_i^2(x_i)\,\mathrm{d}\mathbf{x} + \sum_{1\leqslant i<j\leqslant n} \int_{K^n} f_{ij}^2(x_i,x_j)\,\mathrm{d}\mathbf{x} + \cdots \\ &= \sum_{i=1}^n \int_0^1 f_i^2(x_i)\,\mathrm{d}x_i \\ &\quad + \sum_{1\leqslant i<j\leqslant n} \int_0^1\int_0^1 f_{ij}^2(x_i,x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j + \cdots \\ &= \sum_{i=1}^n \sigma_i^2 + \sum_{1\leqslant i<j\leqslant n} \sigma_{ij}^2 + \cdots. \end{aligned} \tag{16}$$

Analogous expressions were given by Sobol (1993) and Cukier et al. (1978). Then the determination of $\sigma_i^2, \sigma_{ij}^2, \ldots$ is the evaluation of the integrals $\int f_i^2(x_i)\,dx_i$, $\int\int f_{ij}^2(x_i,x_j)\,dx_i\,dx_j$, etc. Monte Carlo integration would likely be the best means for these evaluations, and one may view these operations as a re-utilization of traditional Monte Carlo sampling within the model (Press et al., 1992). For instance, a set of N n-dimensional vector $\mathbf{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \ldots, x_n^{(s)})$ $(s = 1, 2, \ldots, N)$ are randomly generated, and then

$$f_0 = \int_{K^n} f(\mathbf{x})\,d\mathbf{x} \approx \frac{1}{N}\sum_{s=1}^{N} f(\mathbf{x}^{(s)}), \tag{17}$$

$$\sigma_f^2 = \int_{K^n} [f(\mathbf{x}) - f_0]^2\,d\mathbf{x} \approx \frac{1}{N}\sum_{s=1}^{N} f^2(\mathbf{x}^{(s)}) - f_0^2. \tag{18}$$

When $N \to \infty$, accurate values of f_0, σ_f can be obtained. Very often they converge quite fast, and a modest value of N may give very good results. Moreover, the quality of Monte Carlo integration does not depend on the dimension n. This property is extremely beneficial for treating high dimensional systems.

The direct determination of the component functions of RS-HDMR at different values of $x_i, x_j, \ldots$ by Monte Carlo integration requires a large number of random samples. For instance, distinct sets of Monte Carlo random samples of $f(x_i, \mathbf{x}^i)$ at different fixed values of x_i are needed to determine $f_i(x_i)$ (Sobol, 1993). When n is large, the sampling effort for the determination of the RS-HDMR component functions can be prohibitive. For instance, if the mesh for x_i takes m distinct values, then mN random samples are necessary to construct the $f_i(x_i)$ numerical table; if the meshes for both x_i and x_j take m distinct values, then m^2N random samples are necessary to construct the $f_{ij}(x_i, x_j)$ table. The required number of random samples increases exponentially with the order of the RS-HDMR component functions. Thus, often it is prohibitively expensive to construct high order RS-HDMR component function tables.

To reduce the sampling effort, the RS-HDMR component functions $f_i(x_i)$, $f_{ij}(x_i, x_j), \ldots$ may be approximately represented by some suitable functions, such as orthonormal polynomials, spline functions, or even simple polynomial functions (Li, Wang, & Rabitz, 2002). Here we will discuss orthonormal polynomials. The polynomials $\varphi_k(x)$ in the domain $[a, b]$ are referred to as orthonormal when they satisfy

$$\int_a^b \varphi_k(x)\,dx = 0 \quad (k = 1, 2, \ldots), \tag{19}$$

$$\int_a^b \varphi_k^2(x)\,dx = 1 \quad (k = 1, 2, \ldots), \tag{20}$$

$$\int_a^b \varphi_k(x)\varphi_l(x)\,dx = 0 \quad (k \neq l), \tag{21}$$

i.e., they have a zero mean, unit norm and are mutually orthogonal. For the domain $[0, 1]$, the orthonormal polynomials can be readily constructed from the above conditions:

$$\varphi_1(x) = \sqrt{3}\,(2x - 1), \tag{22}$$

$$\varphi_2(x) = 6\sqrt{5}\left(x^2 - x + \frac{1}{6}\right), \tag{23}$$

$$\varphi_3(x) = 20\sqrt{7}\left(x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20}\right), \tag{24}$$

$$\vdots$$

The orthonormal polynomials in Eqs. (22)–(24) can be used as a basis to approximate $f_i(x_i), f_{ij}(x_i, x_j), \ldots$ as follows:

$$f_i(x_i) = \sum_{k=1}^{\infty} \alpha_k^i \varphi_k(x_i), \tag{25}$$

$$f_{ij}(x_i, x_j) = \sum_{k,l=1}^{\infty} \beta_{kl}^{ij} \varphi_k(x_i)\varphi_l(x_j), \tag{26}$$

$$\vdots$$

In most cases, to achieve a desired accuracy using $\varphi_1(x)$, $\varphi_2(x)$ and $\varphi_3(x)$ is sufficient (i.e., $k, l \leqslant 3$).

Using these expressions, Eq. (1) can be expressed as

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{n}\sum_{k=1}^{s_i} \alpha_k^i \varphi_k(x_i) + \sum_{1 \leqslant i < j \leqslant n}\sum_{k=1}^{s_i'}\sum_{l=1}^{s_j} \beta_{kl}^{ij} \varphi_k(x_i)\varphi_l(x_j) + \ldots, \tag{27}$$

where s_i and s_i' may be different.

The coefficient α_k^i and β_{kl}^{ij} can be determined by minimization of the functionals (Li et al., 2002)

$$\min_{\alpha_k^i} \int_0^1 \left[f_i(x_i) - \sum_{k=1}^{s_i} \alpha_k^i \varphi_k(x_i)\right]^2 dx_i + \lambda_i \int_0^1 \left[\partial^2\left(\sum_{k=1}^{s_i} \alpha_k^i \varphi_k(x_i)\right)/\partial x_i^2\right]^2 dx_i, \tag{28}$$

$$\min_{\beta_{kl}^{ij}} \int_0^1\int_0^1 \left[f_{ij}(x_i, x_j) - \sum_{k=1}^{s_i'}\sum_{l=1}^{s_j} \beta_{kl}^{ij} \varphi_k(x_i)\varphi_l(x_j)\right]^2 dx_i\,dx_j$$

$$+\lambda_{ij}\sum_{s,t\in\{i,j\}}\int_0^1\int_0^1\left[\partial^2\left(\sum_{k=1}^{s_i'}\sum_{l=1}^{s_j}\beta_{kl}^{ij}\varphi_k(x_i)\varphi_l(x_j)\right)\Big/\partial x_s\partial x_t\right]^2 \mathrm{d}x_i\,\mathrm{d}x_j, \tag{29}$$

where the first terms in Eqs. (28) and (29) minimize the square of the difference between $f_i(x_i)$, $f_{ij}(x_i,x_j)$ and their orthonormal polynomial approximations, while the second terms minimize the square of the second order derivatives of the orthonormal polynomial approximations. Simultaneous minimization of the second order derivatives will make the approximation functions smooth and diminish the oscillations of the functions around the exact solutions caused by Monte Carlo integration error, especially for small numbers of random samples. λ_i and λ_{ij} are parameters to control the relative weight of the first and second terms.

For Eq. (28) and $s_i = 3$, the minimization yields

$$\begin{pmatrix} 1 & 0 & 0 \\ 0 & 1+720\lambda_i & 0 \\ 0 & 0 & 1+8400\lambda_i \end{pmatrix}\begin{pmatrix}\alpha_1^i\\ \alpha_2^i\\ \alpha_3^i\end{pmatrix} = \begin{pmatrix}\int_0^1 f_i(x_i)\varphi_1(x_i)\,\mathrm{d}x_i\\ \int_0^1 f_i(x_i)\varphi_2(x_i)\,\mathrm{d}x_i\\ \int_0^1 f_i(x_i)\varphi_3(x_i)\,\mathrm{d}x_i\end{pmatrix}$$

$$= \begin{pmatrix}\int_{K^n} f(\mathbf{x})\varphi_1(x_i)\,\mathrm{d}\mathbf{x}\\ \int_{K^n} f(\mathbf{x})\varphi_2(x_i)\,\mathrm{d}\mathbf{x}\\ \int_{K^n} f(\mathbf{x})\varphi_3(x_i)\,\mathrm{d}\mathbf{x}\end{pmatrix} \approx \frac{1}{N}\sum_{s=1}^{N}\begin{pmatrix} f(\mathbf{x}^{(s)})\varphi_1(x_i^{(s)})\\ f(\mathbf{x}^{(s)})\varphi_2(x_i^{(s)})\\ f(\mathbf{x}^{(s)})\varphi_3(x_i^{(s)})\end{pmatrix}. \tag{30}$$

Eq. (30) shows that one can choose a λ_i to diminish α_2^i and especially α_3^i.

Similarly for Eq. (29) and $s_i', s_j = 3$, the minimization yields

$$A\beta = \mathbf{b}, \tag{31}$$

where

$$A = \begin{Bmatrix} 1+144\lambda_{ij} & 0 & 48\sqrt{21}\lambda_{ij} & 0 & 0\\ 0 & 1+1440\lambda_{ij} & 0 & 0 & 0\\ 48\sqrt{21}\lambda_{ij} & 0 & 1+10416\lambda_{ij} & 0 & 0\\ 0 & 0 & 0 & 1+1440\lambda_{ij} & 0\\ 0 & 0 & 0 & 0 & 1+5040\lambda_{ij}\\ 0 & 0 & 0 & 240\sqrt{21}\lambda_{ij} & 0\\ 48\sqrt{21}\lambda_{ij} & 0 & 336\lambda_{ij} & 0 & 0\\ 0 & 240\sqrt{21}\lambda_{ij} & 0 & 0 & 0\\ 336\lambda_{ij} & 0 & 672\sqrt{21}\lambda_{ij} & 0 & 0 \end{Bmatrix}$$

$$\begin{Bmatrix} 0 & 48\sqrt{21}\lambda_{ij} & 0 & 336\lambda_{ij}\\ 0 & 0 & 240\sqrt{21}\lambda_{ij} & 0\\ 0 & 336\lambda_{ij} & 0 & 672\sqrt{21}\lambda_{ij}\\ 240\sqrt{21}\lambda_{ij} & 0 & 0 & 0\\ 0 & 0 & 0 & 0\\ 1+19200\lambda_{ij} & 0 & 0 & 0\\ 0 & 1+10416\lambda_{ij} & 0 & 672\sqrt{21}\lambda_{ij}\\ 0 & 0 & 1+19200\lambda_{ij} & 0\\ 0 & 672\sqrt{21}\lambda_{ij} & 0 & 1+45024\lambda_{ij} \end{Bmatrix}, \tag{32}$$

$$\beta = (\beta_{11}^{ij}\ \beta_{12}^{ij} \cdots \beta_{32}^{ij}\ \beta_{33}^{ij})^{\mathrm{T}}, \tag{33}$$

$$\mathbf{b} = \begin{pmatrix} \int_0^1 \int_0^1 f_{ij}(x_i,x_j)\varphi_1(x_i)\varphi_1(x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j \\ \int_0^1 \int_0^1 f_{ij}(x_i,x_j)\varphi_1(x_i)\varphi_2(x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j \\ \vdots \\ \int_0^1 \int_0^1 f_{ij}(x_i,x_j)\varphi_3(x_i)\varphi_2(x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j \\ \int_0^1 \int_0^1 f_{ij}(x_i,x_j)\varphi_3(x_i)\varphi_3(x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j \end{pmatrix}.$$

$$\approx \frac{1}{N}\sum_{s=1}^{N} \begin{pmatrix} f(\mathbf{x}^{(s)})\varphi_1(x_i^{(s)})\varphi_1(x_j^{(s)}) \\ f(\mathbf{x}^{(s)})\varphi_1(x_i^{(s)})\varphi_2(x_j^{(s)}) \\ \vdots \\ f(\mathbf{x}^{(s)})\varphi_3(x_i^{(s)})\varphi_2(x_j^{(s)}) \\ f(\mathbf{x}^{(s)})\varphi_3(x_i^{(s)})\varphi_3(x_j^{(s)}) \end{pmatrix}. \tag{34}$$

Eq. (34) shows that for a given λ_{ij} the terms related to high order orthonormal polynomials $\varphi_k(x_i)$ and $\varphi_l(x_j)$ diminish.

After all the coefficients α_k^i and β_{kl}^{ij} are determined, σ_i^2 and σ_{ij}^2 can be readily obtained by using the orthonormal property of $\varphi_k(x_i)$:

$$\sigma_i^2 = \int_0^1 f_i^2(x_i)\,\mathrm{d}x_i \approx \int_0^1 \left[\sum_{k=1}^{s_i} \alpha_k^i \varphi_k(x_i)\right]^2 \mathrm{d}x_i$$

$$= \sum_{k=1}^{s_i} (\alpha_k^i)^2, \tag{35}$$

$$\sigma_{ij}^2 = \int_0^1 \int_0^1 f_{ij}^2(x_i,x_j)\,\mathrm{d}x_i\,\mathrm{d}x_j$$

$$\approx \int_0^1 \int_0^1 \left[\sum_{k=1}^{s_i'} \sum_{l=1}^{s_j} \beta_{kl}^{ij}\varphi_k(x_i)\varphi_l(x_j)\right]^2 \mathrm{d}x_i\,\mathrm{d}x_j$$

$$= \sum_{k=1}^{s_i'} \sum_{l=1}^{s_j} (\beta_{kl}^{ij})^2. \tag{36}$$

Note that only one random sampling set of $f(\mathbf{x})$ is needed to determine f_0 and all the expansion coefficients α_k^i, β_{kl}^{ij}, ... for f_0, $f_i(x_i)$, $f_{ij}(x_i,x_j)$,... and consequently, σ_f^2, σ_i^2, σ_{ij}^2, etc. This dramatically reduces the sampling effort and makes global uncertainty assessments of a model very efficient. Moreover, RS-HDMR provides not only quantitative assessments of σ_f^2, σ_i^2, σ_{ij}^2,..., but the qualitative behaviors of the independent and collective actions by $f_i(x_i)$, $f_{ij}(x_i,x_j)$,... of input variables on the output as well. Hence, the uncertainty assessment given by RS-HDMR is also valuable for attaining a physical understanding of the origins of output uncertainty as well as providing a basis to suggest additional laboratory experiments, field studies or parameter refinements to best improve the quality of the model. Furthermore, the resultant f_0, $f_i(x_i)$ and $f_{ij}(x_i,x_j)$ may be used to compose a FEOM. Using the FEOM, approximate values of $f(\mathbf{x})$ for given $\mathbf{x}$'s can be obtained. When a mathematical model for a system cannot be constructed, the FEOM plays a role like a mathematical model. When a system has a mathematical model, approximate values of $f(\mathbf{x})$ can be efficiently obtained without solving the original time-consuming mathematical model. The saving of computational effort can be significant especially the mathematical model is a set of stiff differential equations.

3. Example 1: a photochemical box model

The first example for the illustration of global uncertainty assessments given by RS-HDMR is a zero-dimensional photochemical box model designed to treat the ozone chemistry in the background troposphere used for the study of three-dimensional global chemical-transport (Wang, Levy, Li, & Rabitz, 1999). This box model consists of 63 reactions and 28 chemical species. Using this box model the rates of ozone production P and destruction D are calculated and incorporated into the overall three-dimensional model. The details of this process are not relevant here, but the box model provides a good testing ground for global uncertainty assessments given by RS-HDMR. The rates of ozone production P and destruction D are chosen as two output variables controlled by seven input variables (month, latitude, altitude and the concentrations of 4 precursors: H_2O, CO, NO_x and O_3). Here the four precursors: H_2O, CO, NO_x and O_3 are considered as independent random variables $\mathbf{x} = (x_1,x_2,x_3,x_4)$ with a uniform distribution, and their uncertainty effects on the uncertainty of outputs P and D will be determined. The ranges of the four inputs are shown in Table 1.

A set of 5000 random samples of the 4 input variables $\mathbf{x}^{(s)} = (x_1^{(s)},x_2^{(s)},x_3^{(s)},x_4^{(s)})$ (s = 1–5000) within the above ranges were generated by quasi-random sampling method (Press et al., 1992). The corresponding outputs P and D at $\mathbf{x}^{(s)}$ were obtained by solving the differential equations of the box model. Using these data, the average values $(\bar{P},\bar{D})$, the total variances $(\sigma_{\bar{P}}^2,\sigma_{\bar{D}}^2)$, the first and second order

Table 1
The ranges of input variables

Input	Lower bound	Upper bound
Relative humidity, x_1 (%)	5	100
CO, x_2 (ppb)	10	200
NO_x, x_3 (ppt)	50	950
O_3, x_4 (ppb)	10	150

Table 2
The convergence of f_0 and $\sigma^2_{\bar{f}}$ with respect to sample size N

Sample size	Production, P		Destruction, D	
(N)	f_0	$\sigma^2_{\bar{P}}$	f_0	$\sigma^2_{\bar{D}}$
500	18.43	96.09	24.72	288.05
1000	18.42	96.24	24.73	287.30
3000	18.40	96.03	24.76	288.21
5000	18.40	96.19	24.77	288.47

variances ($\sigma_i^2, \sigma_{ij}^2$) for P and D are calculated with the formulas given in Section 2. The $\sigma_i^2, \sigma_{ij}^2$ are obtained from Eqs. (35) and (36), and the coefficients α_k^i and β_{kl}^{ij} for the orthonormal polynomial approximations of $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ are determined by Eqs. (30) and (31).

As the global uncertainty assessments given by RS-HDMR utilize Monte Carlo integration whose error decreases as $\sim 1/\sqrt{N}$ (Press et al., 1992), the accuracy of the assessments depends on the sample size (N). We will first discuss the effect of sample size.

3.1. Sample size

In order to find the influence of sample size, the f_0, $\sigma^2_{\bar{f}}$ for P and D were calculated. The results for f_0 are shown in Table 2.

This result shows that the convergence is quite fast. There is no significant difference between the values of f_0 and $\sigma^2_{\bar{f}}$ obtained from 500 to 5000 points.

3.2. Accuracy of the second order RS-HDMR expansion

In the overwhelming body of statistical analysis data of many systems, it is rarely found that more than covariances (i.e., cooperativity order $l = 2$) are necessary to describe the input multivariate contributions. Moreover, HDMR expansions are constructed optimally for a given $f(\mathbf{x})$ which guarantees their fast convergence. This implies that the total variance may be well approximated as

$$\sigma_{\bar{f}} \approx \sum_{i=1}^{n} \sigma_i^2 + \sum_{1 \leqslant i < j \leqslant n} \sigma_{ij}^2. \tag{37}$$

This is true only if the second order RS-HDMR expansion

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{n} \sum_{k=1}^{s_i} \alpha_k^i \varphi_k(x_i) + \sum_{1 \leqslant i < j \leqslant n} \sum_{k=1}^{s_i'} \sum_{l=1}^{s_j} \beta_{kl}^{ij} \varphi_k(x_i) \varphi_l(x_j) \tag{38}$$

is a good approximation for $f(\mathbf{x})$, and therefore can be used as a FEOM. As an illustration, second order RS-HDMR expansions were constructed from different sample sizes, both with and without simultaneously minimizing the second order derivatives. Since the Monte Carlo integration error depends on sample size N, λ_i and λ_{ij} were chosen as

$$\lambda_i = c_i / N^3, \tag{39}$$

$$\lambda_{ij} = c_{ij} / N^3. \tag{40}$$

The coefficients c_i and c_{ij} were determined by comparison of the accuracy of second order FEOMs for different choices of their values.

Fig. 2 gives $f_2(x_2)$ for P, with and without simultaneously minimizing second order derivatives. Fig. 2 shows that without minimization of the second order derivatives there exist oscillations for the curves obtained from small sample sizes around the curve obtained from 5000 samples. This oscillation comes from Monte Carlo integration approximation error. However, with minimization of second order derivatives this error can be diminished. The resultant curves of $f_2(x_2)$ are very close to one another for different sample sizes, and thus 500 samples can give similar accurate result to that obtained with 5000 samples. Other $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ are similar to $f_2(x_2)$.

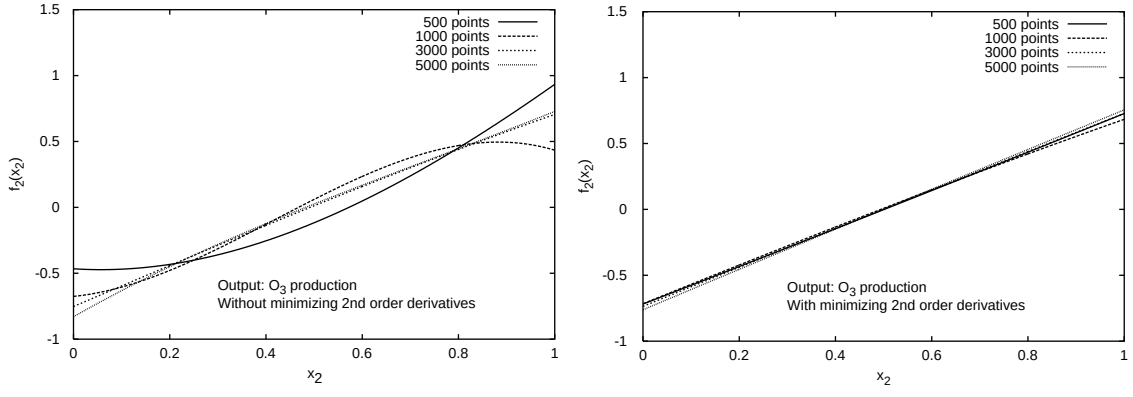

Fig. 2. $f_2(x_2)$ for P constructed from different sample sizes and with or without minimizing the second order derivatives.

Table 3
The accuracy of second order FEOMs obtained from different sample sizes (N), with and without minimizing second order derivatives (the percentage of N data with relative error not larger than a given value)[a]

Minimization of second order derivatives	Sample size (N)	Relative error					
		5%		10%		20%	
		P	D	P	D	P	D
Without	500	0.570	0.338	0.772	0.566	0.900	0.786
	1000	0.687	0.614	0.854	0.815	0.931	0.928
	3000	0.813	0.831	0.923	0.939	0.963	0.979
	5000	0.857	0.914	0.927	0.966	0.963	0.989
With	500	0.796	0.850	0.922	0.944	0.974	0.984
	1000	0.851	0.893	0.952	0.974	0.978	0.999
	3000	0.871	0.933	0.958	0.984	0.979	0.998
	5000	0.880	0.946	0.958	0.986	0.982	0.995

[a]linear, quadratic and cubic orthonormal polynomials were used to approximate $f_i(x_i)$ and $f_{ij}(x_i,x_j)$ (i.e., s_i, s_i' and s_j are all 3).

Table 4
Grid resolution of the exact data table

Input variable	Grid points	Number of grids
Relative humidity, x_1 (%)	$20,25,30,\ldots,100$	17
CO, x_2 (ppb)	$50,60,70,\ldots,200$	16
NO_x, x_3 (ppt)	$50,100,150,\ldots,700$	14
O_3, x_4 (ppb)	$20,30,40,\ldots,150$	14

Table 3 gives the accuracy of second order FEOMs constructed from different samples sizes. The data employed to test the accuracy are just the N samples used to construct the FEOM. The accuracy of the resultant second order FEOM predictions was further tested against a table of exact solutions obtained from box-model runs for the complete set of uniformly sampled points in the desired region of the four-dimensional input variable space. By multiplying the numbers of grid points along each input variable axis specified in Table 4 ($17(H_2O)\times 16(CO)\times 14(NO_x)\times 14(O_3)$), the table of exact solutions is composed of 53,312 box model runs for a given latitude ($-1.5°$), altitude (surface level), month (July) and surface albedo type (land). The results are shown in Table 5.

The results in Table 5 show that with minimization of the second order derivatives the FEOMs even constructed from only 500 random samples have more than 90% of the tested points with relative errors less than 5%. But, without minimization of the second order derivatives only the FEOM constructed from 5000 samples reaches this accuracy. Notice that the results in Table 5 were obtained from 53,312 data which uniformly cover the whole desired domain of $\mathbf{x}$, but the results in Table 3 were tested only for N data. Therefore, Table 5 tests the accuracy of the FEOMs on the whole domain. Since the accuracy given in Table 5 is better than that in Table 3, it implies that the choices of the values of c_i, c_{ij} and consequently λ_i and λ_{ij} from small samples can guarantee the accuracy of the FEOM in the whole domain of $\mathbf{x}$.

Table 5
The comparison between second order FEOMs obtained from different sample sizes (N), with and without minimizing second order derivatives (the percentage of 53,312 data with relative error not larger than a given value)[a]

Minimization of second order derivatives	Sample size (N)	Relative error					
		5%		10%		20%	
		P	D	P	D	P	D
Without	500	0.571	0.346	0.783	0.596	0.917	0.802
	1000	0.726	0.586	0.883	0.814	0.966	0.936
	3000	0.859	0.864	0.957	0.955	0.991	0.992
	5000	0.904	0.918	0.968	0.976	0.994	0.996
With	500	0.928	0.930	0.995	0.983	1.000	0.995
	1000	0.939	0.943	0.995	0.996	1.000	1.000
	3000	0.950	0.967	0.997	0.998	1.000	1.000
	5000	0.968	0.971	1.000	0.998	1.000	1.000

[a]linear, quadratic and cubic orthonormal polynomials were used to approximate $f_i(x_i)$ and $f_{ij}(x_i,x_j)$ (i.e., s_i, s_i' and s_j are all 3).

Table 6
The $\sigma^2_{\bar{f}}$, $\sum\sigma_i^2+\sum\sigma_{ij}^2$ and $\sum\sigma_i^2$ for P and D resulted from different sample size N

	Sample size	$\bar{f}$	$\sigma^2_{\bar{f}}$	$\sum\sigma_i^2+\sum\sigma_{ij}^2$, $(/\sigma^2_{\bar{f}})$	$\sum\sigma_i^2$, $(/\sigma^2_{\bar{f}})$
P	500	18.43	96.09	98.42 (102.4%)	92.51 (96.3%)
	1000	18.42	96.24	97.09 (100.9%)	91.90 (95.5%)
	3000	18.40	96.03	95.73 (99.7%)	91.50 (95.3%)
	5000	18.40	96.19	95.82 (99.6%)	91.55 (95.2%)
D	500	24.72	288.1	296.0 (102.8%)	265.1 (92.0%)
	1000	24.73	287.3	285.8 (99.5%)	265.1 (92.3%)
	3000	24.76	288.2	287.7 (99.8%)	267.7 (92.9%)
	5000	24.77	288.5	288.2 (99.9%)	268.5 (93.1%)

As the accuracy of second order FEOMs constructed from 500 to 5000 samples with minimization of second order derivatives are all good, the variance analysis was performed from them.

3.3. Variance analysis

In Table 6, the $\sum\sigma_i^2+\sum\sigma_{ij}^2$ and $\sum\sigma_i^2$ given by RS-HDMR are compared to the total variance $\sigma^2_{\bar{f}}$ for different sample sizes.

The mean value of $\bar{P}$ and standard deviation $\sigma_{\bar{P}}$ are about 18.4 and 9.8; the mean value of $\bar{D}$ and standard deviation $\sigma_{\bar{D}}$ are about 24.8 and 17.0, respectively. These values show that significant variations are occurring over the domain of $\mathbf{x}$. From Eq. (16) we know that

$$\sigma^2_{\bar{f}} \geqslant \sum_{i=1}^{n}\sigma_i^2 + \sum_{1\leqslant i<j\leqslant n}\sigma_{ij}^2. \tag{41}$$

Table 7
The standard derivations σ_i for output P obtained from different sample sizes

Sample size							
500		1000		3000		5000	
x_i	σ_i	x_i	σ_i	x_i	σ_i	x_i	σ_i
x_3	8.95	x_3	8.92	x_3	8.90	x_3	8.90
x_1	3.16	x_1	3.15	x_1	3.17	x_1	3.17
x_4	1.49	x_4	1.45	x_4	1.42	x_4	1.40
x_2	0.42	x_2	0.41	x_2	0.42	x_2	0.44

Table 8
The standard derivations σ_i for output D obtained from different sample sizes

Sample size							
500		1000		3000		5000	
x_i	σ_i	x_i	σ_i	x_i	σ_i	x_i	σ_i
x_4	14.27	x_4	14.30	x_4	14.37	x_4	14.40
x_1	7.70	x_1	7.69	x_1	7.74	x_1	7.73
x_3	1.20	x_3	1.17	x_3	1.15	x_3	1.14
x_2	0.41	x_2	0.39	x_2	0.23	x_2	0.22

Most of the results in Table 6 satisfy this relation, except of those with sample sizes 500 and 1000. This is obviously caused by Monte Carlo integration error with small sample sizes, but the error is not very large because the values of $\sum \sigma_i^2 + \sum \sigma_{ij}^2$ are very close to one another and very close to σ_f^2 for all sample sizes. This implies that the independent variable action $\sum \sigma_i^2$ and the pair correlation action $\sum \sigma_{ij}^2$ are dominant for outputs P and D in this photochemical box model. Other high order correlation actions are negligible. Moreover, the values of $\sum \sigma_i^2$ in the last column for P and D show that the independent variable action $\sum \sigma_i^2$ much larger than the pair correlation action $\sum \sigma_{ij}^2$.

The magnitudes of σ_i and σ_{ij} will identify which inputs and their pair cooperativity have the most influence on output uncertainties. Tables 7 and 8 list the standard deviations σ_i for P and D obtained from the 3rd order orthonormal polynomial approximations.

Tables 7 and 8 show that the significance of the input variables obtained from different sample sizes are essentially the same. Therefore, in this model 500 samples are sufficient to provide reliable information about the independent variable action σ_i upon the output uncertainty. From the two tables, one can see that for output P , NO_x is the most important input variable, followed by H_2O, O_3 down to CO having the least influence; for output D, O_3 is the most important input variable, followed by H_2O, NO_x down to CO having the least influence.

Similarly, Tables 9 and 10 give the pair correlations σ_{ij} for outputs P and D.

Table 9
The standard derivations σ_{ij} for output P obtained from different sample sizes

Sample size							
500		1000		3000		5000	
(x_i,x_j)	σ_{ij}	(x_i,x_j)	σ_{ij}	(x_i,x_j)	σ_{ij}	(x_i,x_j)	σ_{ij}
(x_1,x_3)	1.803	(x_1,x_3)	1.819	(x_1,x_3)	1.772	(x_1,x_3)	1.756
(x_3,x_4)	0.609	(x_3,x_4)	0.866	(x_3,x_4)	0.716	(x_3,x_4)	0.742
(x_2,x_3)	0.399	(x_2,x_3)	0.513	(x_2,x_3)	0.464	(x_2,x_3)	0.446
(x_1,x_2)	0.064	(x_2,x_4)	0.139	(x_1,x_2)	0.115	(x_1,x_2)	0.090
(x_2,x_4)	0.021	(x_1,x_2)	0.112	(x_2,x_4)	0.105	(x_2,x_4)	0.050
(x_1,x_4)	0.002	(x_1,x_4)	0.039	(x_1,x_4)	0.038	(x_1,x_4)	0.015

Table 10
The standard derivations σ_{ij} for output D obtained from different sample sizes

Sample size							
500		1000		3000		5000	
(x_i,x_j)	σ_{ij}	(x_i,x_j)	σ_{ij}	(x_i,x_j)	σ_{ij}	(x_i,x_j)	σ_{ij}
(x_1,x_4)	4.358	(x_1,x_4)	4.300	(x_1,x_4)	4.376	(x_1,x_4)	4.364
(x_3,x_4)	0.641	(x_3,x_4)	0.628	(x_3,x_4)	0.536	(x_3,x_4)	0.447
(x_2,x_4)	0.248	(x_2,x_4)	0.318	(x_1,x_3)	0.209	(x_1,x_3)	0.145
(x_1,x_3)	0.158	(x_1,x_3)	0.179	(x_2,x_4)	0.135	(x_2,x_4)	0.072
(x_2,x_3)	0.005	(x_2,x_3)	0.143	(x_2,x_3)	0.082	(x_2,x_3)	0.052
(x_1,x_2)	0.004	(x_1,x_2)	0.037	(x_1,x_2)	0.031	(x_1,x_2)	0.015

For different sample sizes, Tables 9 and 10 give almost the same ordering for all pairs, especially the same order for two largest pairs. Considering that the σ_{ij} magnitudes of the remainder pairs are quite small, their ordering is not significant. Combing the results for the standard deviations σ_i in Tables 7 and 8, one can see that for this photochemical model, a few hundred random samples are sufficient to provide reliable uncertainty assessments over the input region. From the results of Tables 9 and 10, we see that the pair cooperativity of H_2O and NO_x has the largest influence on output P, and the pair cooperativity of H_2O and O_3 has the largest influence on output D among all pairs; the pair cooperativity of NO_x and O_3 is the second important one for both P and D. Moreover, the calculation is very fast. Using an SGI O2 workstation with 225 MHz CPU, to calculate the total variance, σ_i and σ_{ij} for an output, the CPU time for 500, 1000 and 5000 samples are 0.06, 0.11 and 0.53 s, respectively.

3.4. Optimal orthonormal polynomials

All the above results were obtained by using orthonormal polynomials whose formulas are given in Eqs. (22)–(24). However, the coefficients α_k^i and β_{kl}^{ij} are determined by Eqs. (30) and (31) where the Monte Carlo integration approximation is employed. This is equivalent to the

approximations

$$\int_0^1 \varphi_k(x)\,\mathrm{d}x \approx \frac{1}{N}\sum_{s=1}^{N}\varphi_k(x^{(s)}) = \xi_k \neq 0 \quad (k=1,2,\ldots), \tag{42}$$

$$\int_0^1 \varphi_k^2(x)\,\mathrm{d}x \approx \frac{1}{N}\sum_{s=1}^{N}\varphi_k^2(x^{(s)}) = \zeta_k \neq 1 \quad (k=1,2,\ldots), \tag{43}$$

$$\int_0^1 \varphi_k(x)\varphi_l(x)\,\mathrm{d}x \approx \frac{1}{N}\sum_{s=1}^{N}\varphi_k(x^{(s)})\varphi_l(x^{(s)}) = \eta_{kl} \neq 0 \quad (k \neq l), \tag{44}$$

where ξ_k and η_{kl} are small numbers, but not exactly equal to zero, ζ_k are close to but not equal to unity; and, their values depend on the sample used. This implies that the orthonormal properties are not exactly fulfilled when the Monte Carlo integration approximation is employed. This may causes some new error. In order to reduce this error, we define optimal orthonormal polynomials for different samples as follows:

$$\varphi_1(x) = a_1 x + a_0, \tag{45}$$

$$\varphi_2(x) = b_2 x^2 + b_1 x + b_0, \tag{46}$$

$$\varphi_3(x) = c_3 x^3 + c_2 x^2 + c_1 x + c_0, \tag{47}$$

$$\vdots \tag{48}$$

where $a_0, a_1, b_0, \ldots, c_3$ are undetermined coefficients which depend on the sample used. To determine the coefficients, optimization procedures (e.g., Simplex method (Press, Teukolsky, Vetterling, Flannery, 1992a)) were used to minimize the sum of ξ_k^2, $(\zeta_k - 1)^2$ and η_{kl}^2 for a given sample. We refer to the resultant polynomials as *optimal* orthonormal polynomials, and Eqs. (22)–(24) as *theoretical* ones. Table 11 gives the comparison of these two kinds of orthonormal polynomials for input x_1 with 500 and 5000 sample sizes of the box model. The results for other input variables and other sample sizes are similar.

Table 11 shows that optimal orthonormal polynomials satisfy the orthonormal properties much better than theoretical ones from different sample sizes. Using the resultant optimal orthonormal polynomials we repeated all the above calculations. Tables 12 and 13 give the magnitude orders of standard deviations σ_i for P and D. Compared to Tables 7 and 8, the input orders are the same, and there is no significant difference between the values of σ_i, especially for the sample sizes larger than 500. Therefore, optimal orthonormal polynomials are only necessary for very small samples.

Table 11
The comparison between optimal and theoretical orthonormal polynomials for input x_1

	Sample size			
	500		5000	
	Theoretical	Optimal	Theoretical	Optimal
ξ_1	0.115×10^{-2}	-0.188×10^{-6}	-0.365×10^{-3}	-0.100×10^{-6}
ξ_2	-0.700×10^{-2}	-0.858×10^{-8}	-0.391×10^{-3}	-0.126×10^{-6}
ξ_3	0.364×10^{-2}	0.657×10^{-8}	-0.455×10^{-4}	0.506×10^{-6}
ζ_1	0.99374	1.00000	0.99965	1.00000
ζ_2	0.99058	1.00000	0.99915	1.00000
ζ_3	0.98192	1.00000	0.99916	1.00000
η_{12}	0.422×10^{-2}	-0.123×10^{-6}	-0.367×10^{-3}	0.981×10^{-7}
η_{13}	-0.112×10^{-1}	0.215×10^{-6}	-0.957×10^{-3}	0.462×10^{-6}
η_{23}	0.664×10^{-2}	0.660×10^{-6}	-0.470×10^{-3}	-0.346×10^{-6}

Table 12
The standard derivations σ_i for output P given by optimal orthonormal polynomials

Sample size							
500		1000		3000		5000	
x_i	σ_i	x_i	σ_i	x_i	σ_i	x_i	σ_i
x_3	8.92	x_3	8.91	x_3	8.90	x_3	8.91
x_1	3.10	x_1	3.14	x_1	3.16	x_1	3.17
x_4	1.47	x_4	1.45	x_4	1.45	x_4	1.45
x_2	0.42	x_2	0.43	x_2	0.43	x_2	0.43

Table 13
The standard derivations σ_i for output D given by optimal orthonormal polynomials

Sample size							
500		1000		3000		5000	
x_i	σ_i	x_i	σ_i	x_i	σ_i	x_i	σ_i
x_4	14.38	x_4	14.37	x_4	14.38	x_4	14.39
x_1	7.69	x_1	7.70	x_1	7.72	x_1	7.74
x_3	1.14	x_3	1.12	x_3	1.14	x_3	1.15
x_2	0.46	x_2	0.36	x_2	0.23	x_2	0.23

4. Example 2: a trace metal bioremediation model

A mathematical model to simulate the bioremediation of uranium-contaminated groundwater system (Wang, Jaffé, Li, Wang, Rabitz, 2002) was used as another example for global uncertainty assessments by RS-HDMR.

The mobility of trace metals and radioisotopes in the groundwater can be controlled to a significant extend via in situ manipulation of key biogeochemical reactions that directly or indirectly alter their solubility. For this

Table 14
The 20 biotic and abiotic redox reactions

	Reaction No.	Oxidant	Reductant
	1	O_2	NH_3
	2	O_2	CH_4
	3	O_2	Organic carbon
Biotic	4	NO_3^-	Organic carbon
	5	Mn^{+4}	Organic carbon
	6	Fe^{+3}	Organic carbon
	7	SO_4^{-2}	Organic carbon
	8	CO_2	Organic carbon
	9	NO_3^-	Fe^{+2}
	10	O_2	Fe^{+2}
	11	O_2	Mn^{+2}
	12	O_2	HS^-
	13	SO_4^{-2}	CH_4
Abiotic	14	O_2	Mn^{+2} (ads)
	15	O_2	Fe^{+2} (ads)
	16	Mn^{+4}	HS^-
	17	Mn^{+4}	Fe^{+2}
	18	Fe^{+3}	HS^-
	19	O_2	U^{+4}
	20	U^{+6}	Fe^{+2} (ads)

ads: adsorbed.

Table 15
The ranges of input variables used in the simulations

	Rate constant	Lower bound	Upper bound
		(mole/year)	
	$k_1(x_1)$	15	60
	$k_2(x_2)$	15	60
	$k_3(x_3)$	0.5	2.0
Biotic	$k_4(x_4)$	0.02	0.08
	$k_5(x_5)$	0.005	0.02
	$k_6(x_6)$	0.0025	0.01
	$k_7(x_7)$	0.085	0.34
	$k_8(x_8)$	0.025	0.10
		(/mole/year)	
	$k_9(x_9)$	0.8×10^3	3.2×10^3
	$k_{10}(x_{10})$	1.05×10^7	4.20×10^7
	$k_{11}(x_{11})$	2.3	9.2
	$k_{12}(x_{12})$	1.0×10^5	4.0×10^5
	$k_{13}(x_{13})$	0.5×10^4	2.0×10^4
Abiotic	$k_{14}(x_{14})$	1.05×10^7	4.20×10^7
	$k_{15}(x_{15})$	2.5×10^6	1.0×10^7
	$k_{16}(x_{16})$	4.0×10^5	1.6×10^6
	$k_{17}(x_{17})$	5.0×10^3	2.0×10^4
	$k_{18}(x_{18})$	5.0×10^2	2.0×10^3
	$k_{19}(x_{19})$	3.0×10^6	1.2×10^7
	$k_{20}(x_{20})$	0.65×10^2	2.6×10^2

purpose, specified redox conditions need to be established in the ground-water. The required redox conditions can be developed by stimulating in-situ bacterial growth via the injection of a carbon source. As the carbon is biodegraded, bacteria will use different electron acceptors in the sequence of decreasing energy yield. Once the most favorable electron acceptors (i.e., oxygen and nitrate) have been depleted, some bacteria can use compounds such as soluble U^{+6} as an electron acceptor and reduce it to insoluble U^{+4}. A time-dependent one-dimensional reaction transport model has been developed to simulate such a process (Wang et al., 2002). The model consists of a set of coupled mass balance equations, accounting for advection, diffusion, dispersion, and a kinetic formulation of the biotic and abiotic transformations affecting an organic substrate, electron acceptors corresponding reduced species and contaminant trace metals/radionuclides of interest.

The numerical studies were performed to evaluate the effects of the uncertainties of the various rate constants used by the formulation on four chemical species simulated by the model including uranium. The global output uncertainty for uranium in the entire desired domain of the 20 input variables was quantified by RS-HDMR using up to 1000 model runs with randomly sampled input variables. In the following treatment global uncertainty assessments are presented with respect to the 20 inputs for two outputs: (1) the total precipitation of U^{+4} and (2) the cumulative amount of Fe^{+3} transformed to Fe^{+2}. The 20 biotic and abiotic reactions are given in Table 14. The ranges over which their respective rate constants were evaluated in this analysis are shown in Table 15.

A set of 1000 random data points over the 20 input variables $\mathbf{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \ldots, x_{20}^{(s)})$ $(s = 1 - 1000)$ within the specified ranges was generated by a quasi-random sampling method (Press et al., 1992). The corresponding outputs: T_U (total precipitation of U^{+4}) and C_{Fe} (cumulative amount of Fe^{+3} transformed to Fe^{+2}) for $\mathbf{x}^{(s)}$ were obtained using the bioremediation model. Using the data generated from the model simulations, the average values ($\bar{T}_U$, $\bar{C}_{Fe}$), the total variances ($\sigma^2_{\bar{T}_U}$, $\sigma^2_{\bar{C}_{Fe}}$), and the coefficients α_k^i, β_{kl}^{ij}, and consequently $f_i(x_i)$, $f_{ij}(x_i, x_j)$ and σ_i, σ_{ij} of specified model outputs were determined for different sample sizes (200, 400, 600, 800 and 1000). It was found that most of the first order and all the second order RS-HDMR component functions for the two outputs are negligible. Some RS-HDMR component functions oscillate around zero and vanish after minimizing the second order derivatives. These oscillations come from the error associated with the Monte Carlo integration approximation, especially for small samples. For nonzero RS-HDMR component functions minimization of second order derivatives make the functions smooth.

4.1. Output 1: total precipitation of U^{+4} (T_U)

For T_U, most of the first order and all of the second order RS-HDMR component functions are negligible, only $f_i(x_i)$ $(i = 3, 4, 5, 19, 20)$ have contributions for the output. The FEOM given by the first order RS-HDMR approximation has quite good accuracy. The results are shown in Table 16.

Table 16
The accuracy of FEOMs for T_U obtained from different sample sizes (N), both with and without minimizing the second order derivatives (the percentage of 1000 test points with relative error not larger than a given value)[a]

Minimization of second order derivatives	Sample size (N)	Relative error		
		5%	10%	20%
Without	200	0.385	0.680	0.915
	400	0.685	0.943	0.995
	600	0.850	0.980	0.998
	800	0.861	0.986	0.998
	1000	0.968	0.995	0.997
With	200	0.967	0.996	0.997
	400	0.976	0.997	0.997
	600	0.983	0.997	0.997
	800	0.987	0.997	0.997
	1000	0.985	0.997	0.997

[a]linear, quadratic and cubic orthonormal polynomials were used to approximate $f_i(x_i)$ (i.e., s_i is 3).

Table 17
The difference between $\sigma^2_{\bar{T}_U}$ and $\sum \sigma_i^2$ for output T_U

Sample size	$\sigma^2_{\bar{T}_U}$	$\sum \sigma_i^2$	$\sum \sigma_i^2/\sigma^2_{\bar{T}_U}$
200	0.276×10^{-12}	0.311×10^{-12}	112.7%
400	0.282×10^{-12}	0.280×10^{-12}	99.7%
600	0.283×10^{-12}	0.277×10^{-12}	97.8%
800	0.285×10^{-12}	0.291×10^{-12}	102.3%
1000	0.283×10^{-12}	0.286×10^{-12}	100.8%

Table 18
Ordering of the largest standard deviations σ_i of the output T_U obtained from different sample sizes

Sample size					
200		600		1000	
x_i	σ_i	x_i	σ_i	x_i	σ_i
x_{20}	0.488×10^{-6}	x_{20}	0.462×10^{-6}	x_{20}	0.471×10^{-6}
x_4	0.257×10^{-6}	x_4	0.237×10^{-6}	x_4	0.242×10^{-6}
x_5	0.607×10^{-7}	x_5	0.712×10^{-7}	x_5	0.622×10^{-7}
x_3	0.519×10^{-7}	x_3	0.355×10^{-7}	x_3	0.318×10^{-7}
x_{19}	0.109×10^{-7}	x_{19}	0.270×10^{-7}	x_{19}	0.192×10^{-7}

For all sample sizes more than 95% of the 1000 data points have relative error less than 5%, even at a sample size of 200 when second derivative smoothing is included. The difference between $\sigma^2_{\bar{T}_U}$ and $\sum \sigma_i^2$ for output T_U and the magnitude order of σ_i are given in Tables 17 and 18.

The mean value $\bar{T}_U$ and the total standard deviation $\sigma_{\bar{T}_U}$ for different sample sizes are about 0.35×10^{-5} and 0.53×10^{-6}, respectively. Therefore, the total standard deviation is only about 15% of the mean value for $\bar{T}_U$. The results given in Tables 17 and 18 show that $\sigma^2_{\bar{T}_U}$ and $\sum \sigma_i^2$ for different sample sizes do not differ very much if the sample size is larger than 200. The convergence is good even if the sample sizes are quite small. Moreover, $\sum \sigma_i^2/\sigma^2_{\bar{T}_U}$ are close to 100%. This implies that the higher order correlation actions are very small. The ordering of the σ_i in Table 18 are the same for all sample sizes. There is no significant influence on T_U of the other inputs within the ranges given in Table 15.

Table 19
The accuracy of FEOMs for C_{Fe} obtained from different sample sizes (N), both with and without minimizing the second order derivatives (the percentage of 1000 test points with relative error not larger than a given value)[a]

Minimization of second order derivatives	Sample size (N)	Relative error		
		5%	10%	20%
Without	200	0.385	0.640	0.925
	400	0.600	0.880	0.980
	600	0.810	0.972	0.995
	800	0.833	0.973	0.996
	1000	0.915	0.989	0.995
With	200	0.924	0.976	0.993
	400	0.958	0.992	0.996
	600	0.961	0.991	0.996
	800	0.960	0.996	0.996
	1000	0.965	0.993	0.996

[a]linear, quadratic and cubic orthonormal polynomials were used to approximate $f_i(x_i)$ (i.e., s_i is 3).

4.2. *Output 2: cumulative amount of* Fe^{+3} *transformed to* Fe^{+2} (C_{Fe})

The uncertainty assessment for C_{Fe} was performed for different sample sizes. Similar to output 1, only $f_i(x_i)$ ($i = 3, 4, 5, 6, 7$) have significant contributions to the output, and the other $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ are negligible. The FEOM given by the first order RS-HDMR approximation also has good accuracy. The results are shown in Table 19.

Table 19 shows that the accuracy is slightly less than that for T_U. For all sample sizes more than 92% of 1000 data have relative error less than 5% when second order derivative smoothing is used. This implies that in this model a few hundred samples are sufficient to give a reliable variance analysis. The difference between $\sigma^2_{\bar{C}_{Fe}}$ and $\sum \sigma_i^2$ for output C_{Fe} and the magnitude order of σ_i are given in Tables 20 and 21, respectively.

The mean value $\bar{C}_{Fe}$ and the total standard deviation $\sigma_{\bar{C}_{Fe}}$ for different sample sizes are about 0.51×10^{-2} and 0.12×10^{-2}, respectively. Therefore, the total standard deviation is about 23% of the mean value for $\bar{C}_{Fe}$. The results given in Table 20 show that the $\sum \sigma_i^2$ for different sample sizes do not differ much and are close to $\sigma^2_{\bar{C}_{Fe}}$. This implies

Table 20
The difference between $\sigma^2_{\bar{C}_{Fe}}$ and $\sum \sigma_i^2$ for output C_{Fe}

Sample size	$\sigma^2_{\bar{C}_{Fe}}$	$\sum \sigma_i^2$	$\sum \sigma_i^2/\sigma^2_{\bar{C}_{Fe}}$
200	0.141×10^{-5}	0.151×10^{-5}	107.6%
400	0.142×10^{-5}	0.142×10^{-5}	100.1%
600	0.142×10^{-5}	0.140×10^{-5}	98.7%
800	0.142×10^{-5}	0.135×10^{-5}	95.3%
1000	0.141×10^{-5}	0.140×10^{-5}	99.3%

Table 21
Ordering of the largest standard deviations σ_i of the output C_{Fe} obtained from different sample sizes

Sample size					
200		600		1000	
x_i	σ_i	x_i	σ_i	x_i	σ_i
x_6	0.115×10^{-2}	x_6	0.112×10^{-2}	x_6	0.112×10^{-2}
x_4	0.378×10^{-3}	x_4	0.320×10^{-3}	x_4	0.330×10^{-3}
x_5	0.232×10^{-3}	x_5	0.183×10^{-3}	x_5	0.192×10^{-3}
x_7	0.143×10^{-3}	x_3	0.550×10^{-4}	x_3	0.520×10^{-4}
x_3	0.637×10^{-4}	x_7	0.387×10^{-4}	x_7	0.440×10^{-4}

that the higher order correlation contributions for C_{Fe} are negligible.

The ordering of the σ_i in Table 21 are the same for all sample sizes except of inputs 3 and 7 for 200 data.

5. Conclusions and a discussion

This paper presented an efficient means to attain global uncertainty assessments by RS-HDMR. With a modest number of random samples, RS-HDMR can efficiently provide reliable global uncertainty assessments for a model. Its advantage comes from the properties of RS-HDMR: high dimensional input–output systems often can be well approximated by low dimensional components, and the individual RS-HDMR component functions have a direct statistical correlation interpretation. The first property of RS-HDMR dramatically reduces the sampling effort. The second one permits the model output variance σ^2 to be decomposed into its input contributions $\sigma^2 = \sum_i \sigma_i^2 + \sum_{i<j} \sigma_{ij}^2 + \cdots$ due to the independent variable action σ_i^2, the pair correlation action σ_{ij}^2, etc. The information gained from this decomposition can be most valuable for attaining a physical understanding of the origins of output uncertainty as well as providing suggestions for additional laboratory/field studies or parameter refinements to best improve the quality of the model. The methodology was illustrated for applications to an atmospheric photochemistry model and a trace metal bioremediation model.

Orthonormal polynomials were used to approximate the RS-HDMR component functions. Simultaneously minimizing the second order derivatives of the component functions provides for better accuracy. The illustrations show that a fully global uncertainty assessment of reasonable models only needed a modest set of data at some randomly sampled inputs. In the two illustrations, first and second order RS-HDMR expansions used as FEOMs were shown to have good accuracy with more than 90% of test data having a relative error less than 5%. The variance analysis is reliable because $(\sum_i \sigma_i^2 + \sum_{i<j} \sigma_{ij}^2)/\sigma_f^2$ or $\sum_i \sigma_i^2/\sigma_f^2 > 90\%$ with very stable behavior for different sample sizes. Furthermore, the calculations are quite straight forward and simple. The important input variables, the influence patterns of the inputs on the output (independent and/or cooperative) can be readily identified. There is no other easy way to determine σ_{ij} and other high order variable cooperativities, and RS-HDMR forms an efficient and practical approach for global uncertainty assessments.

Notation

Scalars

a_i	linear polynomial coefficients, $i = 0, 1$
b_i	quadratic polynomial coefficients, $i = 0, 1, 2$
c_i	coefficients or cubic polynomial coefficients, $i = 0, 1, 2, 3$
c_{ij}	coefficient
C_{Fe}	accumulative amount of Fe^{+3} transformed to Fe^{+2}
D	rate of ozone destruction
$f(\mathbf{x})$	output
$f(\bar{\mathbf{x}})$	output at $\bar{\mathbf{x}}$
$f(x_i, \bar{\mathbf{x}}^i)$	output at $(x_i, \bar{\mathbf{x}}^i)$
$f(x_i, x_j, \bar{\mathbf{x}}^{ij})$	output at $(x_i, x_j, \bar{\mathbf{x}}^{ij})$
f_0	zeroth order component function of HDMR expansion
$f_i(x_i)$	first order component function of HDMR expansion
$f_{ij}(x_i, x_j)$	second order component function of HDMR expansion
$f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})$	lth order component function of HDMR expansion
$f_{12\ldots n}(x_1, x_2, \ldots, x_n)$	last term of HDMR expansion
$\bar{f}$	mean value of $f(\mathbf{x})$
K^n	n-dimensional hypercube
n	integer, the number of inputs
N	integer, sample size
P	rate of ozone production
s	integer
s_i	integer
s_i'	integer
s_j	integer

T_U	total precipitation of U^{+4}
$w_{i_1 i_2 \ldots i_l}(\hat{\mathbf{x}}, \mathbf{u})$	weight function
x_i	input i
x_j	input j
x_k	input k

Vectors and matrices

Capital letters represent matrices; bold-face lower case letters represent vectors.

A	(9×9)-matrix
$\mathbf{b}$	constant vector
$\mathbf{x}$	n-dimensional input variable vector
$\mathbf{x}^{(s)}$	sth sample of $\mathbf{x}$
$\bar{\mathbf{x}}$	reference point of Cut-HDMR
$\hat{\mathbf{x}}$	defined as $\hat{\mathbf{x}} = (x_{i_1}, x_{i_2}, \ldots, x_{i_l})$
$(x_i, \bar{\mathbf{x}}^i)$	defined as $(\bar{x}_1, \ldots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \ldots, \bar{x}_n)$
$(x_i, x_j, \bar{\mathbf{x}}^{ij})$	defined as $(\bar{x}_1, \ldots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \ldots, \bar{x}_{j-1}, x_j, \bar{x}_{j+1}, \ldots, \bar{x}_n)$

Greek letters

α_k^i	coefficient of orthonormal polynomial expansion for $f_i(x_i)$
β_{kl}^{ij}	coefficient of orthonormal polynomial expansion for $f_{ij}(x_i, x_j)$
β	vector
ξ_k	defined as $1/N \sum_{s=1}^{N} \varphi_k(x^{(s)})$
ζ_k	defined as $1/N \sum_{s=1}^{N} \varphi_k^2(x^{(s)})$
η_{kl}	defined as $1/N \sum_{s=1}^{N} \varphi_k(x^{(s)}) \varphi_l(x^{(s)})$
λ_i	weight coefficients
λ_{ij}	weight coefficients
$\sigma_{\tilde{f}}^2$	total variance
σ_i	standard deviation caused by input i
σ_{ij}	standard deviation caused by inputs i and j
$\varphi_k(x)$	kth order orthonormal polynomial
Ω	desired region of input variable vector $\mathbf{x}$

Acknowledgements

The authors acknowledge support from Department of Defense, the Environmental Protection Agency, the Hercules Incorporated, and the Natural and Accelerated Bioremediation Research (NABIR) program, Office of Biological and Environmental Research (OBER), Department of Energy (grant DE-FG02-98ER62705).

References

Alis, O., & Rabitz, H. (1999). General foundations of high dimensional model representations. *Journal of Mathematical Chemistry*, *25*, 197–233.

Alis, O., & Rabitz, H. (2000). Efficient input–output model representations. In A. Saltelli (Ed.), *Mathematical and statistical methods for sensitivity analysis*. New York: Wiley.

Cukier, R. I., Levine, H. B., & Schuler, K. E. (1978). Nonlinear sensitivity analysis of multiparameter model systems. *Journal of Computational Physics*, *26*, 1–42.

Doll, J. D., & Freeman, D. L. (1986). Randomly exact method. *Science*, *234*, 1356–1360.

Gardner, R. H. (1983). *Error analysis and sensitivity analysis in ecology. Encyclopedia of systems and control*. New York: Pergamon Press.

Ghanem, R. G., & Spanos, P. D. (1991). *Stochastic finite elements: A spectral approach*. New York: Springer.

Iman, R. L., & Conover, W. J. (1980). Small sample sensitivity analysis techniques for computer models, with an application to risk assessment. *Communication in Statistics, Part A. Theory and Methods*, *17*, 1749–1842.

Isukapalli, S. S., Roy, A., & Georgopoulos, P. G. (2000). Efficient sensitivity/uncertainty analysis using the combined stochastic response surface method and automated differentiation: Application to environmental and biological systems. *Risk Analysis*, *20*, 591–602.

Li, G., Rosenthal, C., & Rabitz, H. (2001a). High dimensional model representations. *Journal of Physical Chemistry*, *105*, 7765–7777.

Li, G., Wang, S. W., & Rabitz, H. (2002). Practical approaches to construct RS-HDMR component functions. *Journal of Physical Chemistry*, 106, 8721–8733.

McRae, G. J., Goodin, W. R., & Seinfeld, J. H. (1982). *Mathematical modeling of photochemical air pollution*. Technical Report (EQL Report No. 18), Environmental Quality Laboratory, California Institute of Technology, Pasadena, CA.

McRae, G. J., Tilden, J. W., & Seinfeld, J. H. (1981). Global sensitivity analysis: A computational implementation of the Fourier amplitude sensitivity test (FAST). *Computer and Chemical Engineering*, *6*, 15–25.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992). *Numerical recipes in* FORTRAN (pp. 299–319). New York: Cambridge University Press.

Press, W. H., Teukolsky, S. A., Vetterling, W. T., & Flannery, B. P. (1992a). *Numerical recipes in* FORTRAN (pp. 402–406). New York: Cambridge University Press.

Rabitz, H. (1989). System analysis at the molecular scale. *Science*, *246*, 221–226.

Rabitz, H., Alis, O. F., Shorter, J., & Shim, K. (1999). Efficient input–output model representations. *Computer Physics Communications*, *117*, 11–20.

Saltelli, A., & Bolado, R. (1998). An alternative way to compute Fourier amplitude sensitivity test. *Computational Statistics and Data Analysis*, *26*, 445–460.

Saltelli, A., Tarantola, S., & Chan, P. S. (1999). A quantitative model-independent method for global sensitivity analysis of model output. *Technometrics*, *41*, 39–56.

Shorter, J., Precila, C. Ip., & Rabitz, H. (1999). An efficient chemical kinetics solver using high dimensional model representations. *Journal of Physical Chemistry A*, *103*(36), 7192–7198.

Shorter, J., & Rabitz, H. (2000). Radiation transport simulation by means of a fully equivalent operational model. *Geophysical Research Letters*, *27*, 3485–3488.

Sobol, I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiments*, *1*, 407–414.

Stolarski, R., Butler, D., & Rundle, R. (1978). Uncertainty propagation in a stratospheric model 2. Monte-Carlo analysis of imprecisions due to reaction rates. *Journal of Geophysical Research*, *83C*, 3074–3078.

Tatang, M. A., Pan, W. W., Prinn, R. G., & McRae, G. J. (1997). An efficient method for parametric uncertainty analysis of a numerical model. *Journal of Geophysical Research—Atmospheres*, *102*(D18), 21925–21932.

Wang, S. K., Jaffé, P., Li, G., Wang, S. W., & Rabitz, H. (2002). Simulating bioremediation of uranium-contaminated aquifers. *Journal of Contaminant Hydrology*, in press.

Wang, S. W., Levy, II. H., Li, G., & Rabitz, H. (1999). Fully equivalent operational models for atmospheric chemical kinetics within global chemistry-transport models. *Journal of Geophysical Research*, *104*(D23), 30417–30426.

Yetter, R., Eslava, L. A., Dryer, F. L., & Rabitz, H. (1984). Elementary and derived sensitivity information in chemical kinetics. *Journal of Physical Chemistry*, *88*, 1497–1507.