INTERNATIONAL JOURNAL OF
NUMERICAL ANALYSIS AND MODELING, *SERIES B*
Volume 2, Number 1, Pages 73–90

A NEW MULTIPLE SUB-DOMAIN RS-HDMR METHOD AND ITS APPLICATION TO TROPOSPHERIC ALKANE PHOTOCHEMISTRY MODEL

QIANG YUAN AND DONG LIANG

Abstract. The high dimensional model representation (HDMR) method was recently proposed as an efficient tool to capture the input-output relationships in high-dimensional systems for many problems in science and engineering. In this paper, we develop a new multiple sub-domain random sampling HDMR method (MSD-RS-HDMR) for general high dimensional input-output systems. The domain splitting technique is applied to divide the whole domain into multiple sub-domains. The RS-HDMR method is used to generate local approximations in sub-domains and a set of weight functions is introduced to obtain approximations near the sub-domain interfaces. Numerical experiments are carried out using the MSD-RS-HDMR method for given high dimensional functions and a real application of the Tropospheric Alkane Photochemistry model. The new method has been demonstrated to be very effective, and numerical results confirm the excellent performance of the method compared to the RS-HDMR and Cut-HMDR methods.

Key words. RS-HDMR, Multiple sub-domains, Weight functions, Random sequence.

1. Introduction

Many problems in science and engineering involve the interaction of a large number of input variables, which leads to complex high dimensional systems. For examples, when describing the condensing complex atmosphere chemistry mechanism in environmental study, the photochemical box-model in [27] is a high dimensional system with 58 input variables; the gas prices in a gas station normally depends on the crude oil market, the gasoline wholesale market, and the prices of nearby gas stations, etc ([29]); and the aerosol thermodynamic equilibrium prediction in the atmosphere involves aerosol components including aerosol water, aqueous sulfate ion, aerosol nitrate, $(NH_4)_2SO_4$, NH_4NO_3, and total sulfate, ammonium, and nitrate, etc, and environmental variables of temperature and relative humidity, which is a multi-phase and multi-component high dimensional system ([4]).

One important objective of high dimensional problems is to explore the relations between input variables and output variables. In a n-dimensional variable space, the computational complexity of a high dimensional problem scales exponentially ([19]). Many methods have been studied to establish the relations in high dimensions such as projection pursuit algorithms ([6, 9, 15]), multi-layer perceptions ([17]) and radial basis functions ([18]). However, these methods are not widely accepted due to their inefficient performance.

The high dimensional model representation (HDMR) method is a new technique of quantitative model assessment and analysis tool to effectively obtain the relationships of inputs and outputs in high dimensional systems ([1, 12, 13, 19, 20],

Received by the editors July 31, 2010 and, in revised form, February 15, 2011.
2000 *Mathematics Subject Classification.* 41A63, 65D99, 68W25.
This research was partly supported by the Natural Science & Engineering Research Council (NSERC) of Canada and by the Mathematics for Information Technology and Complex Systems (MITACS). The authors thank the editor and the referees for their suggestions which have helped to improve the paper.

etc). The HDMR method is developed to improve the efficiency of the deducing high dimensional behaviors. The method is formed by a particular organization of low dimensional component functions, in which each function is the contribution of one or more input variables to the output variables. Based on the definition of HDMR in [19], its approximations can be constructed in two common ways: the Cut-HDMR and the RS-HDMR ([14, 20]). The HDMR methods have a broad range of applications including fully equivalent operational models ([26]), global uncertainty assessments ([2, 8, 21, 28]), financial and econometrics applications, etc ([5, 22, 23, 27]). However, for many high dimensional systems, due to the complicated input-output relations, the standard HDMR methods, i.e. the Cut-HDMR and the RS-HDMR methods defined in the whole domain, can not achieve sufficient accuracy ([27]). Hence, there have been strong interests in developing efficient HDMR methods for high dimensional systems in large domains. In [13], the Multi-cut-HDMR method is proposed by using multi-cut points but the numerical approximations depend on the locations of the multi-cut points. More recently, Cheng et al. [4] developed the moving cut HDMR method by combining with the multiple local moving cut points to obtain accurate and efficient approximations for high dimensional systems. The method has been successfully applied in the aerosol thermodynamic equilibrium prediction in atmospheric environment ([4]).

In this paper, we develop a new multiple sub-domain random sample HDMR Method (MSD-RS-HDMR), in which the random sample technique is used in sub-domains instead of using the cut point technique. For high dimensional input-output systems in large domains, the domain splitting technique is first applied to divide the large domains into multiple sub-domains and the multiple output variables are then approximated by the random sample HDMR method in sub-domains. The proposed MSD-RS-HDMR method is a combination of the local RS-HDMR approximations built in sub-domains of multiple input variables. Depending on the properties of input-output systems, the sub-domains may be overlapping or non-overlapping. Meanwhile, a set of weight functions is presented to provide approximations near the sub-domain interfaces. The sample points of the high dimensional inputs are generated in a quasi-random sequence by a stochastic algorithm. Numerical experiments are reported using the MSD-RS-HDMR method. First, we consider examples of given multiple dimensional functions, and we focus on the accuracy and efficiency of approximations. Then, the method is applied to a real problem in Tropospheric Alkane Photochemistry model, which describes the species concentration in troposphere ([26, 27]). The computations are based on the reduced model using six input chemical species, i.e. O_3, NO_2, NO, and three lumped alkane species. The proposed MSD-RS-HDMR method improves the accuracy for high dimensional systems compared to the RS-HDMR and Cut-HDMR methods. Numerical tests demonstrate the excellent performance, and it can be regarded as an efficient tool for general high dimensional problems.

The paper is organized as follows. Section 2 gives an introduction of the HDMR method and presents the multiple sub-domain RS-HDMR method. Numerical experiments are reported in Section 3. Application to a real problem in the Tropospheric Alkane Photochemistry model is shown in Section 4. Finally, conclusions are presented in Section 5.

2. The Multiple Sub-domain RS-HDMR Method

In this section, we develop the multiple sub-domain RS-HDMR method for high dimensional systems.

2.1. HDMR method. For problems in high dimensional systems, it is very important to find the relations of the input variables and output variables. Even when the dimensions of input variables are high, we may have only one output. Let the input variables be $\mathbf{x} = (x_1, x_2, \ldots, x_n)$, where n ranges up to 10^2 or more, and the output be $f(\mathbf{x})$ ([14]). We call $(\mathbf{x}, f(\mathbf{x}))$ a point of the system, where $\mathbf{x}$ is the input and $f(\mathbf{x})$ is the output. To construct the structure of an output in an n-dimensional input space is not a easy work for large n. "Conventional logic implies that the computational complexity of sampling the input-output map scales exponentially as s^n where s is a parameter specific to the problem and n is the relevant dimension" ([16]). A recently developed technique to capture the inputs and outputs relations is the high dimensional model representation (HDMR) method.

The HDMR method is developed from the analysis of variance (ANOVA), which is a collection of statistical models and their associated procedures, in which the observed variance is partitioned into components due to different explanatory variables. The HDMR technique is capable of approximating high dimensional systems with low dimensional functions. The structure of the HDMR also makes it a good tool for many fields such as the construction of observation-based models directly from lab/field data, developing a fully equivalent operational model ([26]), global uncertainty assessment and identification of key variables and their interrelationship ([1, 19, 20, 23, 24, 28]), and efficient quantitative risk assessments, etc. The important feature of the HDMR method is that the input variables can independently and cooperatively contribute to the outputs. The general HDMR structure is written as:

$$
\begin{aligned}
f(\mathbf{x}) \equiv f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1\le i<j\le n} f_{ij}(x_i, x_j) + \sum_{1\le i<j<k\le n} f_{ijk}(x_i, x_j, x_k) + \cdots \\
+ \sum_{1\le i_1<\cdots<i_l\le n} f_{i_1 i_2 \cdots i_l}(x_{i_1}, x_{i_2}, \cdots, x_{i_l}) + \cdots + f_{12\cdots n}(x_1, x_2, \cdots, x_n)
\end{aligned} \tag{1}
$$

where the 0'th order component function f_0 is a constant representing the mean response to $f(\mathbf{x})$, and the first order component function $f_i(x_i)$ gives the independent contribution to $f(\mathbf{x})$ by the i'th input variable acting alone, and the second order 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 $f_{12\cdots n}(x_1, x_2, \cdots, x_n)$ contains any residual n'th order correlated contribution of all input variables. In real applications, we hope to find low l-order component functions which can provide good approximations to $f(\mathbf{x})$ in Eqn. (1), i.e. the significant terms in the HDMR expansion are expected to satisfy $l \ll n$ for $n \gg 1$.

There are two commonly used HDMR methods: the Cut-HDMR and RS-HDMR methods. The Cut-HDMR method expresses $f(\mathbf{x})$ in reference to a specified cut point $\bar{\mathbf{x}}$ in Ω, while the RS-HDMR method depends on the average value of $f(\mathbf{x})$ over the whole domain Ω. Here we only consider the RS-HDMR method. Generally speaking, we first re-scale variables x_i such that $0 \le x_i \le 1$ for all i. Then by a suitable transformation, the output function $f(\mathbf{x})$ is defined in the unit hypercube $\Omega = \{(x_1, x_2, \cdots, x_n) \mid 0 \le x_i \le 1, i = 1, 2, \cdots, n\}$. The component functions of the RS-HDMR approximation can be rewritten in the following forms ([14]):

$$
f_0 = \int_{\Omega} f(u)du \tag{2}
$$

$$
f_i(x_i) = \int_{\Omega_{n-1}} f(x_i, u^i)du^i - f_0 \tag{3}
$$

$$f_{ij}(x_i, x_j) = \int_{\Omega_{n-2}} f(x_i, x_j, u^{ij}) du^{ij} - f_i(x_i) - f_j(x_j) - f_0 \tag{4}$$

.....

where du^i is the product $du_1 du_2 \cdots du_n$ without du_i, and du^{ij} is the product $du_1 du_2 \cdots du_n$ without $du_i du_j$. The component functions can be achieved with the Monte Carlo integral, but it requires large sample points and the computation is expensive. In order to reduce the computing time, the component functions can be approximated by a set of orthonormal polynomial functions. We now write the component functions as

$$f_i(x_i) \approx \sum_{r=1}^{o_i} \alpha_r^i \varphi_r(x_i), \tag{5}$$

$$f_{ij}(x_i, x_j) \approx \sum_{p=1}^{o_i'} \sum_{q=1}^{o_j} \beta_{pq}^{ij} \varphi_{pq}(x_i, x_j), \tag{6}$$

.....

where o_i, o_i' and o_j are integers, α_r^i and β_{pq}^{ij} are constant coefficients to be determined, and $\varphi_r(x_i), \varphi_{pq}(x_i, x_j)$ are basis functions. Consequently, the sampling effort can be greatly reduced, because only the coefficients of component functions are unknown. With these formulas, (1) can be approximated by

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{n} \sum_{r=1}^{o_i} \alpha_r^i \varphi_r(x_i) + \sum_{1 \le i < j \le n} \sum_{p=1}^{o_i'} \sum_{q=1}^{o_j} \beta_{pq}^{ij} \varphi_{pq}(x_i, x_j) + \cdots . \tag{7}$$

The polynomials $\{\varphi_k(x)\}$ in domain $[a, b]$ are called orthonormal, where they satisfy

$$\int_a^b \varphi_k(x) dx = 0, \quad k = 1, 2, \cdots , \tag{8}$$

$$\int_a^b \varphi_k^2(x) dx = 1, \quad k = 1, 2, \cdots , \tag{9}$$

$$\int_a^b \varphi_k(x) \varphi_l(x) dx = 0, \quad k \neq l. \tag{10}$$

We can easily verify that the orthonormal polynomials constructed in domain $[0, 1]$ are:

$$\varphi_1(x) = \sqrt{3}(2x - 1), \tag{11}$$

$$\varphi_2(x) = 6\sqrt{5}(x^2 - x + \frac{1}{6}), \tag{12}$$

$$\varphi_3(x) = 20\sqrt{7}(x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20}). \tag{13}$$

Using the Monte Carlo integration, coefficients $\{\alpha_p^i, \beta_{pq}^{ij}\}$ under orthonormal polynomials can be further calculated by,

$$\alpha_r^i \approx \frac{1}{N} \sum_{s=1}^{N} f(\mathbf{x}^{(s)}) \varphi_r(x_i^{(s)}), \tag{14}$$

$$\beta_{pq}^{ij} \approx \frac{1}{N}\sum_{s=1}^{N} f(\mathbf{x}^{(s)})\varphi_p(x_i^{(s)})\varphi_q(x_j^{(s)}). \tag{15}$$

2.2. RS-HDMR over each sub-domain. In [12, 14], the input variable space is defined in or transferred to the n-dimensional space $[0,1]^n$. Now, we describe the RS-HDMR expansion in any sub-domain K=$[a_1,\ b_1]\times[a_2,\ b_2]\times\cdots\times[a_n,\ b_n]$ to construct multiple sub-domain RS-HDMR expansions later.

Definition 2.1. *Define an n-dimensional domain K=$[a_1,\ b_1]\times[a_2,\ b_2]\times\cdots\times[a_n,\ b_n]$ in R^n, where $b_i > a_i$, $i = 1, 2, \cdots, n$.*

In the multiple sub-domain RS-HDMR method, the orthonormal polynomials are defined in interval $[a_i, b_i]$. The following proposition gives the formulas of the orthogonal polynomials in any interval $[a, b]$.

Proposition 2.1. *Let $\{\varphi_i(x)\}$ be the orthonormal polynomials defined in interval $[0,1]$, then $\{\frac{1}{\sqrt{b-a}}\varphi_i(\frac{x-a}{b-a})\}$ is a set of orthonormal polynomials in interval $[a,b]$.*

The component functions of the RS-HDMR approximation on the sub-domain are given in the following proposition.

Proposition 2.2. *Let $\boldsymbol{x}$=$(x_1, x_2, \cdots, x_n) \in K$, then the component functions defined from (2)-(4) become*

$$f_0 = \prod_{k=1}^{n}\frac{1}{(b_k-a_k)}\int_{K_n} f(\boldsymbol{x})d\boldsymbol{x},$$

$$f_i(x_i) = \prod_{k=1,k\neq i}^{n}(\frac{1}{b_k-a_k})\int_{K_{n-1}} f(\boldsymbol{x})dx^i - f_0,$$

$$f_{ij}(x_i,x_j) = \prod_{k=1,k\neq i,j}^{n}(\frac{1}{b_k-a_k})\int_{K_{n-2}} f(\boldsymbol{x})dx^{ij} - f_i(x_i) - f_j(x_j) - f_0$$

$$\cdots\cdots$$

where $1 \le i,\ j \le n$, $i \neq j$, $dx^i = dx_1dx_2\cdots dx_{i-1}dx_{i+1}\cdots dx_n$, and $dx^{ij} = dx_1dx_2\cdots dx_{i-1}dx_{i+1}\cdots dx_{j-1}dx_{j+1}\cdots dx_n$ for $j > i$.

Proof. Integrating both sides of Eqn. (1) with respect to $x_1, x_2, \cdots, x_n$ leads to

$$\begin{aligned}\int_K f(\mathbf{x})d\mathbf{x} = \int_K \big[f_0 &+ \sum_{i=1}^{n} f_i(x_i) + \sum_{1\le i<j\le n} f_{ij}(x_i,x_j)\\ &+ \sum_{1\le i<j<k\le n} f_{ijk}(x_i,x_j,x_k) + \cdots + \sum_{1\le i_1<\cdots<i_l\le n} f_{i_1i_2\cdots i_l}(x_{i_1},x_{i_2},\cdots,x_{i_l})\\ &+ \cdots + f_{12\cdots n}(x_1,x_2,\cdots,x_n)\big]d\mathbf{x},\end{aligned}$$

we have

$$\int_K f(\mathbf{x})d\mathbf{x} = \int_K f_0 dx,$$

and noting that f_0 is a constant, then

$$f_0 = \prod_{k=1}^{n}\frac{1}{(b_k-a_k)}\int_K f(\mathbf{x})d\mathbf{x}.$$

Integrating both sides of (1) with respect to $x_1, x_2, \cdots, x_{i-1}, x_{i+1}, \cdots, x_n$, and using the fact that the integral of component function is zero, we have

$$\int_{K_{n-1}} f(\mathbf{x})dx^i = \int_{K_{n-1}} f_0 dx^i + \int_{K_{n-1}} f_i(x_i)dx^i,$$

thus,

$$f_i(x_i) = \prod_{k=1,k\neq i}^{n} (\frac{1}{b_k - a_k}) \int_{K_{n-1}} f(\mathbf{x})dx^i - f_0.$$

Following the same way, we obtain

$$f_{ij}(x_i, x_j) = \prod_{k=1,k\neq i,j}^{n} (\frac{1}{b_k - a_k}) \int_{K_{n-2}} f(\mathbf{x})dx^{ij},$$

as well as other terms. □

When a set of orthonormal polynomials is given in the sub-domain K, we have the following proposition to determine the coefficients $\{\alpha_r^i, \beta_{pq}^{ij}, \cdots\}$ by using the Monte Carlo integration.

Proposition 2.3. *If the sampling size N is large enough and the orthonormal polynomial functions are defined in K, then the coefficients* $\{\alpha_r^i, \beta_{pq}^{ij}\}$ *can be determined by the following formulas:*

$$\alpha_r^i \approx (b_i - a_i)\frac{1}{N}\sum_{s=1}^{N} f(x^{(s)})\varphi_r(x_i^{(s)}), \quad 1 \le i \le n$$

$$\beta_{pq}^{ij} \approx (b_i - a_i)(b_j - a_j)\frac{1}{N}\sum_{s=1}^{N} \mathrm{f}(x^{(s)})\varphi_p(x_i^{(s)})\varphi_q(x_j^{(s)}), \quad 1 \le i, j \le n, i \neq j.$$

2.3. Multiple Sub-domain RS-HDMR. We first divide the domain Ω into multiple sub-domains $\bigcup \Omega^{(k_1 k_2 \cdots k_n)}$.

Definition 2.2. *At the i'th direction, we divide the interval* $[a_i, b_i]$ *into* i' *subintervals* $[a_i^{(1)}, b_i^{(1)}]$, $[a_i^{(2)}, b_i^{(2)}]$, $\cdots$, $[a_i^{(i')}, b_i^{(i')}]$ *which satisfy the following properties*

$a_i^{(1)} = a_i$, $b_i^{(i')} = b_i$, $b_i^{(j)} > a_i^{(j)}$, $j = 1, 2, \cdots, i'$;

$a_i^{(i')} > a_i^{(i'-1)} > \cdots > a_i^{(2)} > a_i^{(1)}$, $b_i^{(i')} > b_i^{(i'-1)} > \cdots > b_i^{(2)} > b_i^{(1)}$.

Let $\Omega^{(k_1 k_2 \cdots k_n)}$ *be one sub-domain of* Ω *as*

$$\Omega^{(k_1 k_2 \cdots k_n)} = [a_1^{(k_1)}, b_1^{(k_1)}] \times [a_2^{(k_2)}, b_2^{(k_2)}] \times \cdots \times [a_n^{(k_n)}, b_n^{(k_n)}] \tag{16}$$

where $[a_j^{(k_j)}, b_j^{(k_j)}] \in [a_j, b_j]$, $1 \le k_j \le j'$, j' *is the number of sub-intervals in* $[a_j, b_j]$.

Remark 1. *There are two types of sub-domains: non-overlapping and overlapping. If the sub-domains only intersect in their boundaries, they are referred as non-overlapping. Otherwise, they are overlapping sub-domains. In* [4]*, a domain splitting method has been studied based on moving the multiple local cut points and building the corresponding sub-regions (sub-domains). On the other hand, the domain decompositions (splitting techniques) have been used in solving partial differential equations (see, for example,* [7, 25]*).*

Definition 2.3. *Let* $\boldsymbol{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \cdots, x_n^{(s)}) \in \Omega$, $s = 1, 2, \cdots, N$, *be the sampling points, the domain* center point *of* Ω *is defined as*

$$C = (\frac{1}{N}\sum_{s=1}^{N} x_1^{(s)}, \frac{1}{N}\sum_{s=1}^{N} x_2^{(s)}, \cdots, \frac{1}{N}\sum_{s=1}^{N} x_n^{(s)}).$$

Definition 2.4. *Let* $\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)} \in \Omega$, *the distance of these two points is defined as* $d(\boldsymbol{x}^{(i)}, \boldsymbol{x}^{(j)}) = \|\boldsymbol{x}^{(i)} - \boldsymbol{x}^{(j)}\|_2 = \sqrt{\sum_{k=1}^{n}(x_k^{(i)} - x_k^{(j)})^2}$, *where* $\boldsymbol{x}^{(i)} = (x_1^{(i)}, x_2^{(i)}, \cdots, x_n^{(i)})$ *and* $\boldsymbol{x}^{(j)} = (x_1^{(j)}, x_2^{(j)}, \cdots, x_n^{(j)})$.

To organize the local sub-domain RS-HDMR approximations near the interfaces, we consider the weight functions with the following properties:

1. If a given point $\mathbf{x}$ is not in subspace $\Omega^{(k_1 k_2 \cdots k_n)}$, then $\omega_{k_1 k_2 \cdots k_n}(\mathbf{x})$ should be zero.
2. If a given point $\mathbf{x}$ belongs to more than one subspaces, then their values $\omega_{k_1 k_2 \cdots k_n}(\mathbf{x})$ are not all zero.
3. It holds that $\sum_{k_1, k_2, \cdots, k_n} \omega_{k_1 k_2 \cdots k_n}(\mathbf{x}) = 1$.

Definition 2.5. *The weight function in sub-domain* $\Omega^{(k_1 k_2 \cdots k_n)}$ *is defined as*

$$\omega_{k_1 k_2 \cdots k_n}(\boldsymbol{x}) = \frac{\big(\sum_{i_1, i_2, \cdots, i_n} d(\boldsymbol{x}, C_{i_1, i_2, \cdots, i_n}) - d(\boldsymbol{x}, C_{k_1, k_2, \cdots, k_n})\big)\gamma_{k_1, k_2, \cdots, k_n}(\boldsymbol{x}, r)}{\sum_{j_1, j_2, \cdots, j_n}\big(\sum_{i_1, i_2, \cdots, i_n} d(\boldsymbol{x}, C_{i_1, i_2, \cdots, i_n}) - d(\boldsymbol{x}, C_{j_1, j_2, \cdots, j_n})\big)\gamma_{j_1, j_2, \cdots, j_n}(\boldsymbol{x}, r)} \tag{17}$$

where

$$\gamma_{k_1, k_2, \cdots, k_n}(\boldsymbol{x}, r) = \begin{cases} 1 & \boldsymbol{x} \in \Omega^{(k_1 k_2 \cdots k_n)} \cap B(\boldsymbol{x}, r) \\ 0 & else \end{cases} \tag{18}$$

where $r \geq 0$ *is a constant and* $B(\boldsymbol{x}, r)$ *is*

$$B(\boldsymbol{x}, r) = \{\boldsymbol{z} \mid \quad \|\boldsymbol{z} - \boldsymbol{x}\|_2 \leq r\} \tag{19}$$

With the above definition and propositions, we composite the local RS-HDMR approximations in sub-domains into a global MSD-RS-HDMR approximation in the whole domain. The multiple sub-domain RS-HDMR approximation (MSD-RS-HDMR) is defined as

$$f(\mathbf{x})^{(MSD)} = \sum_{k_1, k_2, \cdots, k_n} \omega_{k_1 k_2 \cdots k_n}(\mathbf{x}) f^{(k_1 k_2 \cdots k_n)}(\mathbf{x}), \quad \forall \mathbf{x} \in \Omega, \tag{20}$$

where $\omega_{k_1 k_2 \cdots k_n}(\mathbf{x})$ are the weight functions and $f^{(k_1 k_2 \cdots k_n)}$ are the local RS-HDMR approximations in sub-domains $\Omega^{(k_1 k_2 \cdots k_n)}$.

Remark 2. *The domain's center point is the mean of the inputs among the domain sample points. The sample points are chosen by a uniformly random method, thus the center point is actually the dense center of the domain. Experience shows that the accuracy is improved using the center point defined here.*

Remark 3. *The weight function* $\omega_{k_1 k_2 \cdots k_n}(\boldsymbol{x})$ *gives the weight of the local HDMR prediction at* $\boldsymbol{x}$. *The weight functions can automatically adjust different weight values of the sub-domain HDMR's contributions to the global output at* $\boldsymbol{x}$. *The function* $B(\boldsymbol{x}, r)$ *can extend the number of the related sub-domain HDMR functions used to form the output at* $\boldsymbol{x}$, *and resulting an improvement of the output accuracy near the boundaries.*

2.4. Generate sample data. The Monte Carlo integration (MCI) plays an important role in the RS-HDMR technique. The traditional MCI method converges to the integral with $1/\sqrt{N}$ because the pseudo-random number generator only produces uniformly distributed numbers with a large number of iterations ([3, 10]). When the sample points are quasi-random ([3, 10]), the convergence rate can be improved to $(\ln N)^n/N$, where n is the dimension of the integral function. In the sample points $(\mathbf{x}^{(s)}, f(\mathbf{x}^{(s)}))$, the sample inputs $\mathbf{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \cdots, x_i^{(s)}, \cdots, x_n^{(s)})$, $s = 1, 2, \cdots, N$, are randomly generated.

In computation, each $x_i^{(s)}$ is generated independently, and $x_1^{(s)}$, $x_2^{(s)}$, $\cdots$, $x_n^{(s)}$ are linked together to get a complete sample input $\mathbf{x}^{(s)}$. The random input set $\tilde{X}$ in $[0,1]^n$ is created in the following steps.

1. Input total number of records N, integer M and index j.
2. Divide the i'th directional interval $[0,1]$ uniformly into M sub-intervals: $D_i^{(1)}$, $D_i^{(2)}$, $\cdots$, $D_i^{(M)}$. The initial status symbols of the sub-intervals are set to zero, $i = 1, 2, \cdots, n$.
3. If the status symbols of all the i'th directional sub-intervals are 1, reset the status symbols to zero.
4. Apply the uniform random function to generate a number $x_i^{(s)}$ in interval $[0,1]$. For example, $x_i^{(s)}$ locates in one sub-interval $D_i^{(j)}$, $1 \le j \le M$.
5. If the status symbol of this $D_i^{(j)}$ is zero, set the status symbol of $D_i^{(j)}$ to 1; If the status symbol of $D_i^{(j)}$ is 1, goto Step 3.
6. Loop Steps 3 - 5 for $i = 1, 2, \cdots, n$ and generate random numbers $x_1^{(s)}$, $x_2^{(s)}$, $\cdots$, $x_n^{(s)}$.
7. Store $(x_1^{(s)}, x_2^{(s)}, \cdots, x_n^{(s)}, f(\mathbf{x}^{(s)}))$ as a record.
8. Let $s = s + 1$; if $s < N + 1$, goto Step 3; else goto Step 9.
9. Stop.

2.5. Efficient design for MSD-RS-HDMR. To design an efficient program for the multiple sub-domain RS-HDMR, we need to consider the key parts such as labeling sub-domains, sample points generation and organization, storage of frequently used values, and orthonormal polynomial dynamic generation, etc.

A. Labeling sub-domains. The domain Ω is defined in R^n. If the one-dimensional interval n is split into n_{s1}, n_{s2}, $\cdots$, n_{sn} sub-intervals, these sub-intervals can form $n_{s1} \times n_{s2} \times \cdots \times n_{sn}$ sub-domains. Thus, there will be $n_{s1} \times n_{s2} \times \cdots \times n_{sn}$ local RS-HDMR expansions.

We label sub-domains as follows. Let the number n_{l1}, n_{l2}, $\cdots$, and n_{ln} be the n_{l1}'th, n_{l2}'th, $\cdots$, and n_{ln}'th sub-intervals, respectively, where $1 \le n_{l1} \le n_{sn}$, $1 \le n_{l2} \le n_{s2}$, $\cdots$, and $1 \le n_{ln} \le n_{sn}$. The sub-domain is then labeled by a number as

$$n_{ln} + (n_{l(n-1)} - 1)n_{sn} + (n_{l(n-2)} - 1)n_{sn}n_{s(n-1)} + \cdots + (n_{l1} - 1)\prod_{j=2}^{n} n_{sj}.$$

This leads to each sub-domain with a unique number from 1 to $n_{s1} \times n_{s2} \times \cdots \times n_{sn}$, respectively.

B. Sample points generation and organization. The sample points $(\mathbf{x}^{(s)}, f(\mathbf{x}^{(s)}))$ are stored in the sample file line by line. In the multiple sub-domain RS-HDMR method, we prepare the sample points for each sub-domain, then looping every record in the sample file according to the sub-domain and saving the sample points

into its file named by the above defined number. The sample points stored in these files will be used to generate the local RS-HDMR expansions in sub-domains.

C. Storage of frequently used values. For each sub-domain, we construct its local RS-HDMR expansion in the MSD-RS-HDMR method. Some frequently used coefficients of these local RS-HDMR expansions are stored in their coefficient files.

D. Orthonormal polynomial dynamic generation. The orthonormal polynomials defined in $[0, 1]$ are given clearly. The polynomials in intervals $[a_i,\ b_i]$ will be dynamically generated from the transformation with the formulas in Proposition 2.1.

TABLE 1. Splitting sub-intervals of the input variables for $f(\mathbf{x})$.

Variable	Sub-interval (non-overlapping)	Sub-interval (overlapping)
x_1	[0, 0.6]; [0.6, 1]	[0, 0.61]; [0.59, 1]
x_2	[0, 1]	[0, 1]
x_3	[0, 1]	[0, 1]
x_4	[0, 1]	[0, 1]

3. Numerical Experiments

To validate the MSD-RS-HDMR method, we consider numerical experiments applying to high dimensional functions. The sample input points are generated by randomly with a uniform distribution and the output values are obtained by substituting the sample inputs into the given functions. Numerical results using the MSD-RS-HDMR method are compared with those based on the RS-HDMR method.

The following test cases of four-dimensional functions are investigated:

$$f(\mathbf{x}) = 10(x_1 - 0.2)^4 + x_1^2 + x_2 + x_3 + x_4 \tag{21}$$

$$h(\mathbf{x}) = 10sin(2\pi x_1 - 1.2\pi) + x_2 + x_3 + x_4 + x_1x_2 + x_3x_4 + x_1^3 + x_4^3 \tag{22}$$

where $\mathbf{x} = (x_1, x_2, x_3, x_4) \in [0, 1]^4$.

The splitting sub-intervals for the function $f(\mathbf{x})$ are listed in Table 1. Table 2 presents numerical results by the RS-HDMR method, the MSD-RS-HDMR methods with overlapping (MSD_o) and non-overlapping (MSD_{no}) for $f(\mathbf{x})$. The parameter r in Eqn. (18) is set to zero.

Data portion[1](%): the percents of the checked sample points with prediction errors within accuracies 5%, 10% or 20%. The number of total checked sample data is $10,000$. Sample size[2]: the number of sample points chosen for building the approximations.

From the simulations presented in Table 2, we observe that the MSD-RS-HDMR method with non-overlapping sub-domains (MSD_{no}) and overlapping sub-domains (MSD_o) produce comparable numerical results. However, it is note that they are more accurate compared to those using the RS-HDMR method (RS), in particularly for small order of polynomial order.

Another splitting sub-intervals for function $f(\mathbf{x})$ is given in Table 3. The simulation results are listed in Table 4. It is obvious that among the three approaches,

TABLE 2. Comparisons of first-order RS-HDMR method (RS) and MSD-RS-HDMR methods with MSD_o and MSD_{no}. The component functions are obtained using different sample sizes N and different orders of orthonormal polynomials.

Polynomial order	Sample size[2] (N)	Data portion[1](%)								
		Accuracy 5%			Accuracy 10%			Accuracy 20%		
		RS	MSD_{no}	MSD_o	RS	MSD_{no}	MSD_o	RS	MSD_{no}	MSD_o
1	1000	8.9	56.8	61.9	18.3	83.5	84.3	38.7	98.1	97.9
1	3000	8.7	61.4	58	18.8	89.5	89.5	39.8	99.1	98.8
1	5000	8.6	64.4	59.7	18.8	89.6	89.6	38.8	98.5	98.7
2	1000	30.1	61.4	63.1	57.8	87.1	87.5	85.7	97.9	97.7
2	3000	27.9	87.2	87	60.6	98.4	98.4	89.4	100	99.9
2	5000	27.3	93.3	93	60.6	98.7	98.6	89.7	99.9	99.9
3	1000	50.4	51.3	53.1	76.2	80	81.1	93.5	97.2	96.4
3	3000	75.9	85.7	85.6	95.3	96.2	97.3	99.7	99.8	99.9
3	5000	81.2	92.5	93.4	96.3	98.1	98.3	99.9	99.8	99.9

TABLE 3. Splitting sub-intervals of input variables for $f(\mathbf{x})$.

Variable	Sub-intervals (non-overlapping)	Sub-intervals (overlapping)
x_1	[0, 0.3]; [0.3, 1]	[0, 0.6]; [0.3, 1]
x_2	[0, 1]	[0, 1]
x_3	[0, 1]	[0, 1]
x_4	[0, 1]	[0, 1]

TABLE 4. Comparisons of first-order RS-HDMR method (RS) and the MSD-RS-HDMR methods with MSD_o and MSD_{no}. The component functions are obtained from different sample sizes N and different orders of orthonormal polynomials.

Polynomial order	Sample size (N)	Data portion(%)								
		Accuracy 5%			Accuracy 10%			Accuracy 20%		
		RS	MSD_{no}	MSD_o	RS	MSD_{no}	MSD_o	RS	MSD_{no}	MSD_o
1	1000	8.9	38.3	38.2	18.3	49	60.8	38.7	73.3	84.1
1	3000	8.7	39.1	40.2	18.8	49	60.7	39.8	73.4	84.9
1	5000	8.6	39.5	40.9	18.8	48.1	59.8	38.8	73.3	84.1
2	1000	30.1	59.7	63.8	57.8	85.3	92.4	85.7	95.4	99.3
2	3000	27.9	68.6	79.2	60.6	89.6	97.7	89.4	98.5	100
2	5000	27.3	75.9	87.1	60.6	92.8	98.5	89.7	99.3	99.9
3	1000	50.4	49.6	61.7	76.2	77.8	89.1	93.5	94.6	99
3	3000	75.9	75.3	90.3	95.3	93.5	99	99.7	99.6	100
3	5000	81.2	86.2	94.5	96.3	96.1	98.4	99.9	99.4	99.8

the MSD_o is the best method and the RS-HDMR is the worst. Thus, the MSD-RS-HDMR method with overlapping sub-domains is capable of producing more accurate results if a good overlapping domain decomposition is employed.

In the definition of the global approximation by the MSD-RS-HDMR method, the parameter r is introduced to extend the predictions on sub-domains. The effect of r is shown in Figure 1, where the x-axis is in r-direction and the y-axis gives the accuracy of the data portion. The test data is reported in Table 2. The figure shows that the accuracy could be improved using a appropriate choice of the value for r.

Moreover, for function $h(\mathbf{x})$, we split the domain as shown in Table 5. Table 6 gives numerical results by various methods.

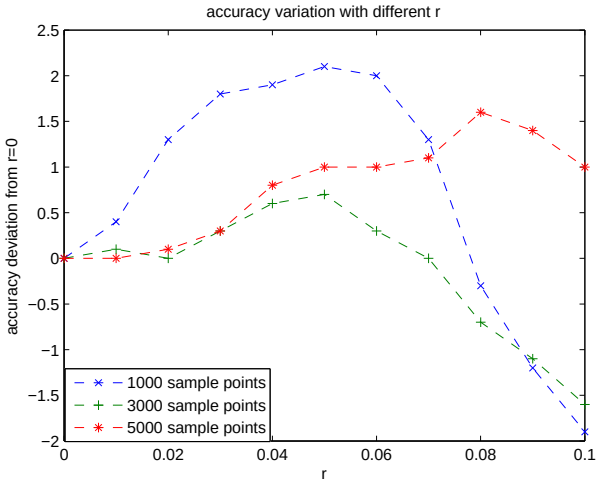


FIGURE 1. Comparison of accuracy with different r by the MSD-RS-HDMR method with non-overlapping sub-domains using 3-order orthonormal polynomials.

TABLE 5. Splitting sub-intervals of input variables for $h(\mathbf{x})$.

variable	Sub-intervals (non-overlapping)	Sub-intervals (overlapping)
x_1	[0, 0.4]; [0.4, 1]	[0, 0.45]; [0.4, 1]
x_2	[0, 1]	[0, 1]
x_3	[0, 1]	[0, 1]
x_4	[0, 1]	[0, 1]

TABLE 6. Comparison of second-order RS-HDMR (RS) and the MSD-RS-HDMR methods with MSD_o and MSD_{no}.

Polynomial order	Sample size (N)	Data portion(%)								
		Accuracy 5%			Accuracy 10%			Accuracy 20%		
		RS	MSD_{no}	MSD_o	RS	MSD_{no}	MSD_o	RS	MSD_{no}	MSD_o
1	1000	4	7.1	8.3	7.4	16	16.6	13.7	33.1	30.1
1	3000	3.5	7.5	7.7	7.4	15.1	14.7	12.8	31.2	28.5
1	5000	3.4	7.5	8.4	7.4	14.6	15.3	13	30.9	28.7
1	20000	3.2	6.6	7.1	6.9	13	13.8	13.5	29.7	28
2	1000	2.6	23.1	23.9	6.1	40.7	42.2	13.2	65.8	65.6
2	3000	2.8	23.2	25.3	5.1	45.3	47.3	10.8	69.4	70.3
2	5000	3.3	24.4	26.4	5.6	50.8	50.7	10.8	73.4	73.3
2	20000	3.1	26.6	25.9	5	55	52.9	10.8	76.9	76.8
3	1000	12.8	17.6	18.5	24.8	37.9	38.5	46.3	63.1	62.8
3	3000	18.6	30.7	32.3	34.7	53.9	55.7	55.5	76.3	76.6
3	5000	18.7	36.7	38.2	35.8	62.3	63.3	56.7	83	84
3	20000	15.6	60.1	61.5	34.8	82	83.4	60.6	92	91.6

Based on the numerical simulations for the four dimensional functions, it is clear that the MSD-RS-HDMR methods produce much more accurate results than those of the RS-HDMR method. On each sub-domain, it may requires appropriate sample points in order to keep the accuracy of the Monte Carlo integrations.

4. Tropospheric Alkane Photochemistry Model

In this section, we apply the MSD-RS-HDMR method for a real problem of the Tropospheric Alkane Photochemistry model. The model describes the species concentration in the troposphere of the atmosphere, and the problem has been studied in [26, 27]. Numerical simulations and comparisons with the RS-HDMR and Cut-HDMR methods will be reported.

4.1. Introduction of the Model. The troposphere is the lowest portion of the earth's atmosphere. It contains approximately three quarters of the atmosphere's mass and almost all of its water vapor and aerosols. The problem deals with alkane/NO_x/O_3 photochemistry in troposphere and involves 68 reactions and 52 species. In the tropospheric Alkane Photochemistry model, as the concentrations of some species keep unchanged or the action of some species is one variable, a reduced model by grouping variables for the actions of some species ([26, 27]) has been proposed. The Direct Constrained Approximate Lumping (DCAL) method has been applied to this Alkane model ([26, 27]), in which 30 nonmethane alkane species are lumped into three lumped species and the remaining 21 inorganic species and one methane are constrained treated as individual species. In the present study, we apply the photochemical box-model [27] to simulate the Alkane Photochemistry problem. This model has the advantage that it employs the Alkane/DCAL mechanism and the LSODE routine ([11]).

TABLE 7. Dynamic ranges (ppb) of the 33 input chemical species concentrations ([26]).

No.	Chemical Species	Range	No.	Chemical Species	Range
1	Ozone	1-200	18	Methylcyclopentane	4-18
2	Nitrogen dioxide	1-50	19	Heptane	60-275
3	Nitrogen oxide	1-100	20	3-Methylhexane	5-12
4	Ethane	5-24	21	2,4-Dimethylpentane	2-16
5	Propane	6-25	22	2,3-Dimethylpentane	3-17
6	Butane	7-38	23	Methylcyclohexane	2-16
7	Iso-butane	3-17	24	Octane	7-428
8	Pentane	2-20	25	4-Methylheptane	2-11
9	Iso-pentane	1-4	26	2,2,4-Trimethylpentane	2-11
10	Neo-pentane	1-3	27	Ethylcyclohexane	2-11
11	Cyclopentane	1-3	28	Nonane	3-37
12	Hexane	24-280	29	4-Ethylheptane	1-10
13	2-Methylpentane	3-11	30	Decane	1-9
14	3-Methylpentane	3-11	31	4-Propylheptane	1-8
15	2,2-Dimethylbutane	4-12	32	Undecane	1-7
16	2,3-Dimethylbutane	3-17	33	Dodecane	1-7
17	Cyclohexane	3-16			

For the photochemical box-model, the initial concentrations of 14 radical species are set to zero corresponding to the starting time at midnight, and the concentrations of 5 chemical species (N_2, O_2, H_2O, CO_2, and CH_4) remain constant during the simulation period. At the end, 6 chemical species (O_3, NO_2, NO, and

3 lumped alkane species) are left, that leads to a six input variables system. The appropriate dynamic ranges of the 33 species, including 30 explicit alkane species and 3 inorganic species (O_3, NO_2, and NO), are given in Table 7. The 30 explicit alkane species combine linearly into 3 lumped alkane species through the transformation of the DCAL lumping matrix ([27]), and the ranges are determined by the method given in [26]. The procedure is given as follows: (1) take ten thousand random sampling points from the 30-dimensional hypercube space consisting of 30 explicit alkane species ranges in Table 7; (2) transform above sampling points from the 30-dimensional hypercube space to the 3-dimensional lumped space through the DCAL lumping matrix; (3) determine the maximum and minimum values for each axis of the 3-dimensional lumped space among the ten thousand transformed points.

In the computational experiment, we compare the numerical results based on the MSD-RS-HDMR, RS-HDMR and Cut-HDMR methods. In the simulations, the HDMR function $f(\bar{x}_0, t_j)$ represents one of the chemical species concentrations after time t_j. Here, we suppose that the initial time is zero. For example, $f(\bar{\mathbf{x}}_0, 120)$ is the concentration of one species after 120 minutes, where the $\bar{\mathbf{x}}_0$ is the vector of the initial concentrations at time zero. For one output concentration of species, we need to generate 24 local HDMR expansions corresponding to $t_j = 60j$ minutes, $1 \leq j \leq 24$, over high dimensional domains. Thus, there are totally 6×24 local HDMR functions for 6 chemical species in 24 hour simulation.

The initial values for the input chemical species concentrations are given in Table 8. The ranges of the reduced model such as the ranges of O_3, NO_2, NO and three lumped variables are listed in Table 9, where the ranges of lumped variables is calculated by the above mentioned method. The values of the initial species of a reduced model are obtained through the transformation of the DCAL lumping matrix, which are shown in Table 10.

TABLE 8. Initial conditions (ppb) of 33 input chemical species concentrations ([26]).

No.	Chemical Species	Value	No.	Chemical Species	Value
1	Ozone	97.4	18	Methylcyclopentane	7.3
2	Nitrogen dioxide	35.7	19	Heptane	159
3	Nitrogen oxide	35.7	20	3-Methylhexane	6.4
4	Ethane	13.6	21	2,4-Dimethylpentane	8.0
5	Propane	18.9	22	2,3-Dimethylpentane	13.8
6	Butane	20.1	23	Methylcyclohexane	7.6
7	Iso-butane	12.1	24	Octane	320
8	Pentane	9.0	25	4-Methylheptane	5.3
9	Iso-pentane	2.9	26	2,2,4-Trimethylpentane	8.4
10	Neo-pentane	1.7	27	Ethylcyclohexane	5.0
11	Cyclopentane	2.2	28	Nonane	26.2
12	Hexane	108	29	4-Ethylheptane	3.8
13	2-Methylpentane	7.4	30	Decane	6.2
14	3-Methylpentane	5.4	31	4-Propylheptane	2.9
15	2,2-Dimethylbutane	8.2	32	Undecane	4.7
16	2,3-Dimethylbutane	6.7	33	Dodecane	2.5
17	Cyclohexane	9.4			

TABLE 9. Dynamic ranges (ppb) of the 6 input variables in Alkane/DCAL Mechanism ([26]).

No.	Chemical Species	Range
1	Ozone	(1,200)
2	Nitrogen dioxide	(1,50)
3	Nitrogen oxide	(1,100)
4	1st Lumped alkane species	(30,180)
5	2nd Lumped alkane species	(-5.5,41)
6	3rd Lumped alkane species	(-140,-36)

4.2. Simulation results. In this section, we compare the performance of five methods, the Cut-HDMR, the RS-HDMR, the MSD-RS-HDMR, the full model and the Dcal model applied to the Alkane Photochemistry model. The full model is the general model involving 68 reactions and 52 chemical species. The Dcal is a reduced model containing 6 species: O_3, NO_2, NO and 3 lumped species. The Cut-HDMR, the RS-HDMR and the MSD-RS-HDMR methods are constructed from data generated by the Dcal model ([27]). Simulation results of the full model, the Dcal model and the Cut-HDMR method are based on those reported in [27].

TABLE 10. Initial conditions (ppb) of 6 input species concentrations.

No.	Chemical Species	Value
1	Ozone	97.41499
2	Nitrogen dioxide	35.74295
3	Nitrogen oxide	35.74295
4	1st Lumped alkane species	124.7278
5	2nd Lumped alkane species	22.93971
6	3rd Lumped alkane species	-0.7262324

TABLE 11. Splitting sub-intervals (overlapping) of input variables.

Variable	Sub-intervals
x_1	[1, 100]; [90, 200]
x_2	[1 50]
x_3	[1, 50]; [50, 100]
x_4	[30, 180]
x_5	[-5.5, 41]
x_6	[-140, -36]

The sub-domain decomposition of the whole domain for the MSD-RS-HDMR method is given in Table 11. Results of the outputs O_3, NO_2 and NO by the RS-HDMR, the Cut-HDMR and the MSD-RS-HDMR methods are shown in Table 12. In all computations, the HDMR approximations contain only the first order component terms. The components functions are approximated by the third order orthonormal polynomials and with parameter $r = 7$. The predicted concentration

values are at time level 12×60 minutes. For the MSD-RS-HDMR computation, $10,000$ sample points are used in each sub-domain to construct the local HDMR approximations. To investigate the accuracy, $1,000$ checked sample points are chosen from the total random data.

From the simulation results presented in Table 12, it is clear that the MSD-RS-HDMR method produces the most accurate predictions compared to those using the Cut-HDMR and the RS-HDMR methods for all three species. The accuracy given by the Cut-HDMR and the RS-HDMR methods are almost the same.

TABLE 12. Comparison of RS-HDMR (RS), the MSD-RS-HDMR (MSD_o) and the Cut-HDMR (Cut) methods.

Item	Polynomial order	Data portion(%)								
		Accuracy 5%			Accuracy 10%			Accuracy 20%		
		RS	Cut	MSD_o	RS	Cut	MSD_o	RS	Cut	MSD_o
O_3	3	26.7	32.2	61.9	46.7	51.7	82.3	70.9	68.6	92.4
NO_2	3	36.3	40.75	71.7	57.2	61.1	89	74.8	76.84	95.5
NO	3	15.8	18.20	51.9	29.9	30.63	73.7	49.9	52.27	88.1

Figure 2 (a)(b) display the diurnal concentrations profiles of 24×60 minutes for O_3 and NO_2 by various methods. Figure 2 shows that the MSD-RS-HDMR expansion produces numerical predictions in excellent agreement with those obtained by the full model ([27]) and the DCAL reduced model, while the Cut-HDMR ([26]) method and the RS-HDMR method produce errors after time 15×60 minutes for O_3 and 20×60 minutes for NO_2, as well as before 5×60 minutes for NO_2.

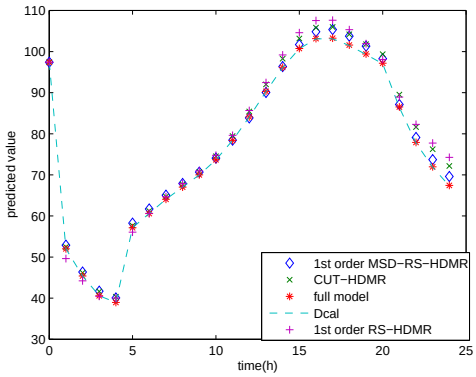
Figure 3(a)(b) illustrate the 5% accuracy of the Cut-HDMR ([26]), the RS-HDMR and the MSD-RS-HDMR methods for outputs O_3 and NO_2, respectively. It is clearly shown in Figure 3 that the MSD-RS-HDMR method gives the best result while the RS-HDMR method produces the worst prediction. For the Alkane Photochemistry model, numerical tests have verified that the MSD-RS-HDMR method is capable of producing more accurate results than the Cut-HDMR and the RS-HDMR methods.

5. Conclusions

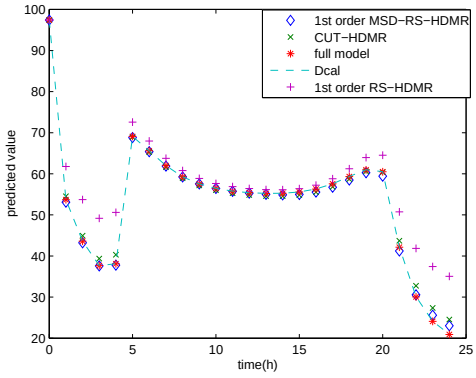
In this paper, we developed the multiple sub-domain random sampling HDMR methods (MSD-RS-HDMR) for general high dimensional input-output systems. We first divide the whole domain into multiple sub-domains by the domain splitting technique. Depending on the properties of systems, the sub-domains can be constructed as overlapping or non-overlapping. Secondly, the local RS-HDMR approximations are used on the sub-domains. The sample points of the high dimensional inputs are generated using a quasi-random sequence. The global MSD-RS-HDMR approximation is then formed with local RS-HDMR approximations. The weight functions are introduced to ensure good approximations near the interfaces. Numerical simulations for given multiple dimensional functions and a real application of the Tropospheric Alkane Photochemistry model confirmed that the MSD-RS-HDMR method is capable of producing accurate results and it can be applied efficiently for high dimensional systems in large domains.

References

[1] Alis, O. and Rabitz, H., Efficient Implementation of High Dimensional Model Representations. In Mathematical and Statistical Methods for sensitivity Analysis, Saltelli, A., Ed., John Wiley and Sons: New York, 2000.



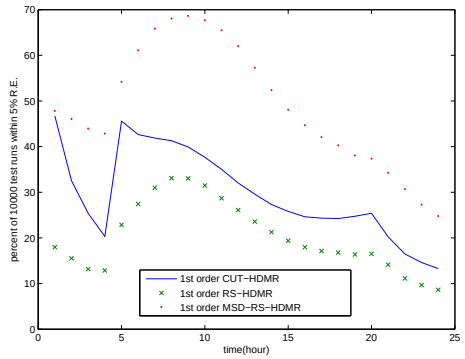
(a)



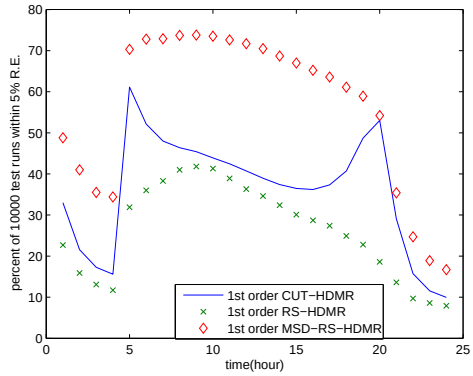
(b)

FIGURE 2. Diurnal concentration profiles of outputs O_3 (a) and NO_2 (b) predicted by different methods.

[2] Benson, J., Ziehn, T., Dixon, N.S. and Tomlin, A.S., Global sensitivity analysis of a 3D street canyon model-Part II: Application and physical insight using sensitivity analysis, Atmos. Environ., 42 (2008), 1874-1891.
[3] Carter, F. and Everett, D., Random Walks, Markov Chains and the Monte Carlo Method, Taygeta, Monterey, California, 1996.
[4] Cheng, Y., Liang, D., Wang, W., Gong, S. and Xue, M., An efficient approach of aerosol thermodynamic equilibrium predictions by the HDMR method, Atmos. Environ., 44 (2010), 1321-1330.
[5] Chowdhury, R., Rao, B.N., Hybrid High Dimensional Model Representation for reliability analysis, Comput. Methods Appl. Mech. Engrg., 198 (2009), 753-765
[6] Diaconis, P. and Shahshahani, M., On Nonlinear Functions of Linear Combinations, J. Sci. Statist. Comput., 5 (1984), 175-191.
[7] Du, C. and Liang, D., An Efficient S-DDM Iterative Approach for Compressible Contamination Fluid Flows in Porous Media, J. Comput. Phys., 229 (2010), 4501-4521.
[8] Feng, X.J., Hooshang, S., Chen, D., Li, G.Y., Weiss, R. and Rabitz, H., Optimizing Genetic Circuits by Global Sensitivity Analysis, Biophys. J., 87 (2004), 2195-2202.
[9] Friedman, J. and Stuetzle, W., Projection Pursuit Regression, J. Amer. Statist. Assoc., 76 (1981), 817-823.
[10] Guan, Y., A Monte Carlo Method of Integration, IB Diploma Programme, Nørre Gymnasium, Denmark, 2002.



(a)



(b)

FIGURE 3. Portion rates of percents within $10,000$ test runs predicted by different methods for outputs O_3 (a) and NO_2 (b).

[11] Hindmarsh, A.C. and ODEPACK, A., Systematized Collection of ODE Solvers, in Scientific Computation, Chapter, Livermore Solver for Ordinary Differential Equations, 5564, North-Holland, Amsterdam, 1983.
[12] Li, G.Y., Artamonov, M., Rabitz, H and Wang, S.W., High-Dimensional Model Representations Generated from Low Order Terms:lp-RS-HDMR, J. Comput. Chem., 24 (2003), 647-656.
[13] Li, G.Y., Schoendorf, J., Ho, T.S. and Rabitz, H., Multicut-HDMR with an Application to an Ionospheric Model, J. Comput. Chem., 25 (2004), 1149-1156.
[14] Li, G.Y., Rosenthal, C. and Rabitz, H., High dimensional model representations, J. Phys. Chem. A, 105 (2001), 7765-7777.
[15] Huber, P., Projection pursuit, Ann. Statist., 13 (1981), 435-525.
[16] Omer, F.A. and Rabitz, H., Efficient implementation of high dimensional model representations, J. Math. Chem., 29 (2001), 127-142.
[17] Parker, D., Learning logic, Working Paper No 47, Massachusetts Institute of Technology, Center for Computational Research in Economics and Management Science, 1985.
[18] Poggio, T. and Girosi, F., Networks for approximation and learning, Pro. IEEE, 78 (1990), 1481-1497.
[19] Rabitz, H. and Alis, O., General Foundations of High Dimensional Model Representations, J. Math. Chem., 25 (1999), 197-233.
[20] Rabitz, H., Alis, O., Shorter, J. and Shim, K., Efficient input-output Model Representations, Comput. Phys. Commun., 117 (1999), 11-20.

[21] Rao, B.N., Chowdhury, R., Prasad, A.M., Singh, R.K. and Kushwaha H.S., Probabilistic characterization of AHWR Inner Containment using High Dimensional Model Representation, Nucl. Eng. Des., 239 (2009), 1030-1041.
[22] Schoendorf, J., Rabitz, H. and Li, G.Y., A fast and accurate operational model of ionospheric electron density, Geophys. Res. Lett., 30 (2003), 1492.
[23] Shim, K. and Rabitz, H., Independent and Correlated Composition Behavior of Material Properties:Application to Energy Band Gaps for the $Ga_\alpha ln_{1-\alpha} PbAs_{1-\beta}$ and $Ga_\alpha ln_{1-\alpha} PbSbgAs_{1-\beta-\gamma}$ Alloys, Phys. Rev. B, 58 (1998), 1940-1946.
[24] Shorter, J., Precila, C.I. and Rabitz, H., An Efficient Chemical Kinetics Solver using High Dimensional Model Representations, J. Phys. Chem. A, 103 (1999), 7192-7198.
[25] Smith, B.F., Bjost, P.E. and Gropp, W.D., *Domain Decomposition Methods for Partial Differential Equations*, Cambridge University Press, 1996.
[26] Wang, S.W., Balakrishnan, S. and Georgopoulos, P., Fast Equivalent Operational Model of Tropospheric Alkane Photochemistry, AIChE J., 51 (2005), 11297-1303.
[27] Wang, S.W., Georgopoulos, P., Li, G.Y. and Rabitz, H., Condensing Complex Atmospheric Chemistry Mechanisms. 1. The Direct Constrained Approximate Lumping (DCAL) Method Applied to Alkane Photochemistry, Environ. Sci. Technol., 32 (1998), 2018-2024.
[28] Wang, S., Jaffe, P.R., Li, G.Y., Wang, S.W. and Rabitz, H., Simulating bioremediation of uranium-contaminated aquifers, uncertainty assessment of model parameters, Journal of Contaminant Hydrology, 64 (2003), 283-307.
[29] http://www.senate.gov/∼gov_affairs/042902gasreport/sectionv.pdf.

Department of Mathematics and Statistics, York University, Toronto, Ont., M3J 1P3, Canada
E-mail: qiangy@mathstat.yorku.ca and dliang@mathstat.yorku.ca