Random Sampling-High Dimensional Model Representation (RS-HDMR) and Orthogonality of Its Different Order Component Functions

Genyuan Li,[†] Jishan Hu,[‡] Sheng-Wei Wang,[§] Panos G. Georgopoulos,[§] Jacqueline Schoendorf,[#] and Herschel Rabitz*,[†]

Department of Chemistry, Princeton University, Princeton, New Jersey 08544, Department of Mathematics, Hong Kong University of Science and Technology, Clear Water Bay, Kowloon, Hong Kong, Environmental and Occupational Health Sciences Institute, 170 Frelinghuysen Road, Piscataway, New Jersey 08854, and ATK Mission Research, 589 West Hollis Street, Suite 201, Nashua, New Hampshire 03062-1323

Received: July 27, 2005; In Final Form: November 11, 2005

High dimensional model representation is under active development as a set of quantitative model assessment and analysis tools for capturing high-dimensional input−output system behavior based on a hierarchy of functions of increasing dimensions. The HDMR component functions are optimally constructed from zeroth order to higher orders step-by-step. This paper extends the definitions of HDMR component functions to systems whose input variables may not be independent. The orthogonality of the higher order terms with respect to the lower order ones guarantees the best improvement in accuracy for the higher order approximations. Therefore, the HDMR component functions are constructed to be mutually orthogonal. The RS-HDMR component functions are efficiently constructed from randomly sampled input−output data. The previous introduction of polynomial approximations for the component functions violates the strictly desirable orthogonality properties. In this paper, new orthonormal polynomial approximation formulas for the RS-HDMR component functions are presented that preserve the orthogonality property. An integrated exposure and dose model as well as ionospheric electron density determined from measured ionosonde data are used as test cases, which show that the new method has better accuracy than the prior one.

1. Introduction

In the chemical sciences, many laboratory experiments, environmental and industrial processes, as well as modeling exercises, are characterized by large numbers of variables. For example, a petrochemical system or a protein can be composed of thousands species or atoms viewed as variables. Process simulations are an increasingly important part of industrial design, and a major difficulty arises from the computational burden of the chemical kinetics calculations involved, especially for petrochemical systems. The kinetics can consume as much as 90% of the total CPU time in simulations employing detailed chemical mechanisms. The full modeling of these processes for design and control purposes becomes computationally prohibitive on even the largest supercomputers. Another class of problems with high dimensional inputs occurs in molecular modeling. Any molecular simulation, regardless of its nature, can be viewed as establishing a functional relationship between the physical observables (e.g., cross sections, rate constants, etc.) and the underlying potential energy surfaces. This relationship can be very complex and nonlinear even for few-body systems, and the relationship generally becomes more involved for many-body systems. Producing a computationally efficient and accurate global map of the potential → observable relationships would enable a deeper physical understanding of the dynamical processes involved and also provide the means to accelerate molecular simulations.

An important concern when undertaking these explorations is the number of experiments or modeling excursions necessary to effectively learn the system input → output behavior. Although simple logic suggests that the number of runs could grow exponentially with the number of input variables, broad-scale evidence indicates that the required effort often scales far more comfortably.

High dimensional model representation (HDMR)[1−15] is under active development as a set of quantitative model assessment and analysis tools for capturing high-dimensional input−output system behavior. As the impact of the multiple input variables on the output can be independent and cooperative, HDMR expresses the output $f(\mathbf{x})$ as a finite hierarchical correlated function expansion in terms of the input variables $\mathbf{x} = (x_1, x_2, ..., x_n)$:

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

The above expansion was introduced as a statistical ANOVA decomposition,[16,17] and HDMR has developed systematic optimal procedures to construct the distinct component functions in the above expansion[1,3,5] to meet different requirements in a variety of scientific problems.[2,4,7,8,10,11,18] As the HDMR component functions are optimally constructed to suit each application, for many realistic systems only low order component functions have significant contributions for the output. Previous work[1,2,4,7,8,18] indicated that HDMR expansions truncated up to

* To whom correspondence should be addressed.
[†] Princeton University.
[‡] Hong Kong University of Science and Technology.
[§] Environmental and Occupational Health Sciences Institute.
[#] ATK Mission Research.

10.1021/jp054148m CCC: $33.50
Published on Web 01/28/2006

second order often provide a satisfactory description of the output for many high dimensional systems when the input variables are properly chosen. Thus, HDMR can render the originally perceived exponential difficulty of creating an input → output map down to a problem of only low order polynomic sampling complexity, which makes the treatment of many high dimensional input−output problems feasible. These advantages are retained even when the number of input variables n is in the thousands.[18] Once the map is determined for a particular application, typically with modest computational or experimental overhead, it can be used to replace the original system model and dynamical equations; the high efficiency of evaluating the map permits a thorough exploration of the system physical behavior.

Recently, groundwork was laid for improving the capabilities of the HDMR technique through the introduction of random sampling (RS)-HDMR, which is a practical procedure based on randomly sampling the input variables. RS-HDMR is very efficient for treating high dimensional input−output mapping problems and has been successfully utilized in several scientific modeling applications including atmospheric chemistry,[9,13,14] environmental metal bioremediation,[10] integrated exposure and dose studies,[11,19] and bio-kinetics modeling.[20]

In the prior formulations of the HDMR component functions, all input variables were considered to be independent. In this paper, we extend the definitions of the HDMR component functions to treat systems whose input variables may not be independent. The resultant formulas for the HDMR component functions are the same as the original ones except that the probability density functions are replaced by conditional probability density functions, and the formulas after using the orthonormal polynomial and Monte Carlo approximations are exactly the same. The formulas for the HDMR component functions with independent input variables are only a special case of the general treatment given in this paper.

The previous introduction of polynomial approximations to the RS-HDMR component functions violates their strictly desirable orthogonality properties. In this paper, new orthonormal polynomial approximation formulas for the RS-HDMR component functions are presented that preserve the orthogonality property. These modifications of HDMR enlarge its application to different scientific problems and further improve its accuracy. The treatment of ionospheric electron dynamics given in this paper is an example of the new application to real measured field data. An H_2/air combustion model with 8 species and a nitrate−ester propellant explosion model with 59 species were also successfully treated by the new modified RS-HDMR technique and will be reported later. Additional applications in physical, chemical and other scientific systems are now open as well.

The paper is organized as follows. Section 2 presents the extension of the definitions of the HDMR component functions to treat systems whose input variables may not be independent. Section 3 proves that the new formulas for the RS-HDMR component functions are the same as the prior ones, regardless of whether the input variables are independent or not. Section 4 provides new orthonormal polynomial approximation formulas for the RS-HDMR component functions that preserve the orthogonality property. The mathematical proof of the orthogonality for the new method is given in the Appendix. Section 5 presents illustrations of the method to an integrated environmental exposure and dose model for trichloroethylene, as well as for ionospheric electron density determined from measured ionosonde data. Finally, section 6 contains conclusions.

2. HDMR with Dependent Variables

In practice, the chosen input variables $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ may not be independent. For convenience, the input variables are normalized, i.e., $\mathbf{x} \in K^n = \{(x_1, x_2, ..., x_n)|0 \le x_i \le 1, i = 1, 2, ..., n\}$ where K^n is an n-dimensional unit hypercube. Let $w(\mathbf{x})$ be the probability density function (pdf) for $\mathbf{x}$ satisfying the conditions

$$\begin{cases} w(\mathbf{x}) \ge 0 & (\mathbf{x} \in K^n) \\ \int_{K^n} w(\mathbf{x})\, d\mathbf{x} = 1 \end{cases} \tag{2}$$

Here, $w(\mathbf{x})$ may not be separable when the variables x_i's are dependent. The HDMR component functions are defined as

$$f_0 = \int_{K^n} w(\mathbf{x})\, f(\mathbf{x})\, d\mathbf{x} \tag{3}$$

$$f_i(x_i) = \int_{K^{n-1}} w_{\mathbf{x}^i|x_i}(\mathbf{x}^i)\, f(\mathbf{x})\, d\mathbf{x}^i - f_0 \tag{4}$$

$$f_{ij}(x_i,x_j) = \int_{K^{n-2}} w_{\mathbf{x}^{ij}|x_i,x_j}(\mathbf{x}^{ij})\, f(\mathbf{x})\, d\mathbf{x}^{ij} - f_i(x_i) - f_j(x_j) - f_0 \tag{5}$$

etc., where

$$w_i(x_i) = \int_{K^{n-1}} w(x_i,\mathbf{x}^i)\, d\mathbf{x}^i \tag{6}$$

$$w_{ij}(x_i,x_j) = \int_{K^{n-2}} w(x_i,x_j,\mathbf{x}^{ij})\, d\mathbf{x}^{ij} \tag{7}$$

$$w_{\mathbf{x}^i|xi}(\mathbf{x}^i) = w(x_i,\mathbf{x}^i)/w_i(x_i) \tag{8}$$

$$w_{\mathbf{x}^{ij}|x_i,x_j}(\mathbf{x}^{ij}) = w(x_i,x_j,\mathbf{x}^{ij})/w_{ij}(x_i,x_j) \tag{9}$$

and $\mathbf{x}^i$, $\mathbf{x}^{ij}$ are $\mathbf{x}$ without the element x_i and x_i, x_j with range K^{n-1} and K^{n-2}, respectively. $w_{\mathbf{x}^i|x_i}(\mathbf{x}^i)$, $w_{\mathbf{x}^{ij}|x_i,x_j}(\mathbf{x}^{ij})$ are the conditional pdf's of $\mathbf{x}$ for a fixed value of x_i, and x_i, x_j.[21] $w(x_1, \mathbf{x}^1)$, $w(x_i, x_j, \mathbf{x}^{ij}) \equiv w(\mathbf{x})$ with $\mathbf{x}$ written as $(x_1, \mathbf{x}^1)$ and $(x_i, x_j, \mathbf{x}^{ij})$ to emphasize that the integration in eqs 6 and 7 are with respect to $\mathbf{x}^i$ and $\mathbf{x}^{ij}$ only. When all of the input variables x_i are sampled independently, i.e.,

$$w(\mathbf{x}) = \prod_{i=1}^{n} w_i(x_i) \tag{10}$$

then

$$w_{\mathbf{x}^i|x_i}(\mathbf{x}^i) = \prod_{\substack{k=1 \\ k\neq i}}^{n} w_k(x_k) \tag{11}$$

$$w_{\mathbf{x}^{ij}|x_i,x_j}(\mathbf{x}^{ij}) = \prod_{\substack{k=1 \\ k\neq i,j}}^{n} w_k(x_k) \tag{12}$$

and the formulas given in eqs 3−5 become

$$f_0 = \int_{K^n} \prod_{i=1}^{n} w_i(x_i)\, f(\mathbf{x})\, d\mathbf{x} \tag{13}$$

$$f_i(x_i) = \int_{K^{n-1}} \prod_{\substack{k=1 \\ k\neq i}}^{n} w_k(x_k)\, f(\mathbf{x})\, d\mathbf{x}^i - f_0 \tag{14}$$

$$f_{ij}(x_i,x_j) = \int_{K^{n-2}} \prod_{\substack{k=1 \\ k\neq i,j}}^{n} w_k(x_k)\, f(\mathbf{x})\, d\mathbf{x}^{ij} - f_i(x_i) - f_j(x_j) - f_0 \tag{15}$$

etc., which were obtained before for HDMR with independent variables.[9−11] The formulas given in eqs 13−15 are only a special case of those given in eqs 3−5. However, the mutual orthogonality between two distinct HDMR component functions

$$\int_{K^n} w(\mathbf{x})\, f_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})\, f_{j_1 j_2 \ldots j_k}(x_{j_1}, x_{j_2}, \ldots, x_{j_k})\, \mathrm{d}\mathbf{x} = 0 \qquad (16)$$
$$\{i_1, i_2, \ldots, i_l\} \neq \{j_1, j_2, \ldots, j_k\}$$

cannot be preserved for the new definitions of HDMR component functions given in eqs 3−5.

3. Polynomial Approximations of the RS−HDMR Component Functions

The HDMR component functions $f_i(x_i)$, $f_{ij}(x_i, x_j)$, ... can be generated numerically at discrete values of the input variables x_i, x_j, ... produced from random sampling the output function $f(\mathbf{x})$ according to the weight $w(\mathbf{x})$ and employing the right-hand side of eqs 3−5. Thus, numerical data tables can be constructed for these component functions, and the approximate value of $f(\mathbf{x})$ for an arbitrary point $\mathbf{x}$ can be determined from these tables by performing only low dimensional interpolation over $f_i(x_i)$, $f_{ij}(x_i, x_j)$, ...

Constructing the numerical data tables for the HDMR component functions requires evaluating the multidimensional integrals. Evaluation of the high dimensional integrals may be carried out by Monte Carlo random sampling,[22] and this method was referred to as RS-HDMR. However, the direct determination of all RS-HDMR component functions at different values of x_i, x_j, ... by direct Monte Carlo integration requires a large number of random samples.[17]

To reduce the sampling effort, the RS-HDMR component functions were previously approximated by expansions in terms of a suitable set of basis functions, for instance, weighted orthonormal polynomials $\{\varphi\}$ as[9,11]

$$f_i(x_i) \approx \sum_{r=1}^{k} \alpha_r^i\, \varphi_r^i(x_i) \qquad (17)$$

$$f_{ij}(x_i, x_j) \approx \sum_{p=1}^{l}\sum_{q=1}^{l'} \beta_{pq}^{ij}\, \varphi_p^i(x_i)\, \varphi_q^j(x_j) \qquad (18)$$

$$f_{ijk}(x_i, x_j, x_k) \approx \sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''} \gamma_{pqr}^{ijk}\, \varphi_p^i(x_i)\, \varphi_q^j(x_j)\, \varphi_r^k(x_k) \qquad (19)$$

where k, l, l', m, m', and m'' are integers, α_r^i, β_{pq}^{ij}, and γ_{pqr}^{ijk} are constant coefficients to be determined, and the polynomials $\{\varphi\}$ possess the weighted orthonormality properties:

$$\int_0^1 w_i(x_i)\, \varphi_r^i(x_i)\, \mathrm{d}x_i = 0 \qquad \text{for all } r, i \qquad (20)$$

$$\int_0^1 w_i(x_i)[\varphi_r^i(x_i)]^2\, \mathrm{d}x_i = 1 \qquad \text{for all } r, i \qquad (21)$$

$$\int_0^1 w_i(x_i)\, \varphi_p^i(x_i)\, \varphi_q^i(x_i)\, \mathrm{d}x_i = 0 \qquad p \neq q \qquad (22)$$

i.e., they have a zero mean and unit norm and are mutually orthogonal with respect to the weight $w_i(x_i)$. In most cases, to achieve satisfactory accuracy, using only $\varphi_r^i(x_i)$, $r \leq 3$, is sufficient.

By using the formulas in eqs 17−19, eq 1 can be expressed as

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k} \alpha_r^i\, \varphi_r^i(x_i) + \sum_{1 \leq i < j \leq n}\sum_{p=1}^{l}\sum_{q=1}^{l'} \beta_{pq}^{ij}\, \varphi_p^i(x_i)\, \varphi_q^j(x_j) + \sum_{1 \leq i < j < k \leq n}\sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''} \gamma_{pqr}^{ijk}\, \varphi_p^i(x_i)\, \varphi_q^j(x_j)\, \varphi_r^k(x_k) + \ldots \qquad (23)$$

The coefficients α_r^i, β_{pq}^{ij}, γ_{pqr}^{ijk}, ... can be determined by using the weighted orthonormality properties of $\{\varphi\}$ combined with the Monte Carlo integration approximation as follows:

$$\alpha_r^i = \int_0^1 w_i(x_i)\, f_i(x_i)\, \varphi_r^i(x_i)\, \mathrm{d}x_i = \int_{K^n} w_i(x_i)\, w_{\mathbf{x}^{i}|x_i}(x_i)\, f(\mathbf{x})\, \varphi_r^i(x_i)\, \mathrm{d}\mathbf{x}$$
$$= \int_{K^n} w(\mathbf{x})\, f(\mathbf{x})\, \varphi_r^i(x_i)\, \mathrm{d}\mathbf{x} \approx \frac{1}{N}\sum_{s=1}^{N} f(\mathbf{x}^{(s)})\, \varphi_r^i(x_i^{(s)}) \qquad (24)$$

$$\beta_{pq}^{ij} = \int_0^1\int_0^1 w_{ij}(x_i, x_j)\, f_{ij}(x_i, x_j)\, \varphi_p^i(x_i)\, \varphi_q^j(x_j)\, \mathrm{d}x_i\, \mathrm{d}x_j$$
$$= \int_{K^n} w_{ij}(x_i, x_j)\, w_{\mathbf{x}^{ij}|x_i,x_j} f(\mathbf{x})\, \varphi_p^i(x_i)\, \varphi_q^i(x_j)\, \mathrm{d}\mathbf{x}$$
$$= \int_{K^n} w(\mathbf{x})\, f(\mathbf{x})\, \varphi_p^i(x_i)\, \varphi_q^j(x_j)\, \mathrm{d}\mathbf{x}$$
$$\approx \frac{1}{N}\sum_{s=1}^{N} f(\mathbf{x}^{(s)})\, \varphi_p^i(x_i^{(s)})\, \varphi_q^j(x_j^{(s)}) \qquad (25)$$

$$\gamma_{pqr}^{ijk} = \int_0^1\int_0^1\int_0^1 w_{ijk}(x_i, x_j, x_k)\, f_{ijk}(x_i, x_j, x_k)\, \varphi_p^i(x_i)\, \varphi_q^j(x_j)\, \varphi_r^k(x_k)\, \mathrm{d}x_i\, \mathrm{d}x_j\, \mathrm{d}x_k$$
$$= \int_{K^n} w_{ijk}(x_i, x_j, x_k)\, w_{\mathbf{x}^{ijk}|x_i,x_j,x_k}(\mathbf{x}^{ijk})\, f(\mathbf{x})\, \varphi_p^i(x_i)\, \varphi_q^j(x_j)\, \varphi_r^k(x_k)\, \mathrm{d}\mathbf{x}$$
$$= \int_{K^n} w(\mathbf{x})\, f(\mathbf{x})\, \varphi_p^i(x_i)\, \varphi_q^j(x_j)\, \varphi_r^k(x_k)\, \mathrm{d}\mathbf{x}$$
$$\approx \frac{1}{N}\sum_{s=1}^{N} f(\mathbf{x}^{(s)})\, \varphi_p^i(x_i^{(s)})\, \varphi_q^j(x_j^{(s)})\, \varphi_r^k(x_k^{(s)}) \qquad (26)$$

where N is the sample size and $\mathbf{x}^{(s)}$ is the sth sample. These formulas are exactly the same as those for the systems whose input variables are independent.[9,11] As the sample was taken according to the weight $w(\mathbf{x})$, the advantage of eqs 24−26 is that $w(\mathbf{x})$ does not need to be known explicitly. This feature is especially useful for cases when the data is drawn from industrial, biological, environmental, etc. circumstances where the $w(\mathbf{x})$ is likely not known.

Equations 24−26 were obtained by using the orthonormality property of $\{\varphi\}$. However, when Monte Carlo integration is employed, a small, but significant error is introduced:

$$\int_0^1 w_i(x_i)\, \varphi_r^i(x_i)\, \mathrm{d}x_i \approx \frac{1}{N}\sum_{s=1}^{N} \varphi_r^i(x_i^{(s)}) = \xi_r^i \neq 0 \qquad (r = 1, 2, \ldots) \qquad (27)$$

$$\int_0^1 w_i(x_i)[\varphi_r^i(x_i)]^2\, \mathrm{d}x_i \approx \frac{1}{N}\sum_{s=1}^{N} [\varphi_r^i(x_i^{(s)})]^2 = \zeta_r^i \neq 1 \qquad (r = 1, 2, \ldots) \qquad (28)$$

$$\int_0^1 w_i(x_i)\, \varphi_p^i(x_i)\, \varphi_q^i(x_i)\, \mathrm{d}x_i \approx \frac{1}{N}\sum_{s=1}^{N} \varphi_p^i(x_i^{(s)})\, \varphi_q^i(x_i^{(s)}) = \eta_{pq}^i \neq 0 \qquad (p \neq q) \qquad (29)$$

here ξ_r^i and η_{pq}^i are small numbers, but not equal to zero, and ζ_r^i are close to, but not equal to, unity; the values of these errors depend on the sample used. The smaller the sample size, the larger the violation of orthonormality, which may cause errors in the HDMR expansion. To reduce this error, *optimal* weighted orthonormal polynomials for different samples were defined as follows:[11]

$$\varphi_1^i(x_i) = a_1 x_i + a_0 \tag{30}$$

$$\varphi_2^i(x_i) = b_2 x_i^2 + b_1 x_i + b_0 \tag{31}$$

$$\varphi_3^i(x_i) = c_3 x_i^3 + c_2 x_i^2 + c_1 x_i + c_0 \tag{32}$$

etc., where the coefficients a_0, a_1, b_0, ..., c_3 are determined in such a way that for a given set of random samples (generated by $w_i(x_i)$) $\sum_r[(\xi_r^i)^2 + (\zeta_r^i - 1)^2] + \sum_{p<q}(\eta_{pq}^i)^2$ is minimized. This implies that the weighted orthonormality property is forced to be best satisfied for a given set of data.

Optimal weighted orthonormal polynomials for the single variable x_i best satisfy the weighted orthonormality property, but the Monte Carlo integration approximation of the integrals with more variables (e.g., the products $\varphi_p^i(x_i)\,\varphi_q^j(x_j)$ $(i \neq j)$) cannot guarantee the weighted orthonormality property for any sample size. Therefore, even if optimal weighted orthonormal polynomials are used, the RS-HDMR component functions with different variables will generally still not be strictly orthogonal, especially for small sample sizes. The material below addresses this problem.

4. Orthogonality of RS−HDMR Component Functions

The violation of orthogonality pointed out above diminishes the accuracy of the RS-HDMR expansion. In this section, new orthonormal polynomial approximation formulas for the RS-HDMR component functions are presented. The basis functions $\{\varphi\}$ used for the low order RS-HDMR component functions are always subsets of those for the higher order ones. The new orthonormal polynomial approximations for the RS-HDMR component functions guarantee that the different order approximated RS-HDMR component functions are mutually orthogonal when the Monte Carlo integration approximation is used for any sample size.

The HDMR component functions are optimally determined from zeroth to higher orders in a step-by-step fashion

$$f(\mathbf{x}) \approx f_0 + f^{(1)} + f^{(2)} + f^{(2)} + ... \tag{33}$$

where $f^{(i)}$ denotes the collection of ith order HDMR terms.

The *sufficient* condition for $f^{(i)}$ to be orthogonal to $f^{(j)}(j < i)$ is that the subspace spanned by the basis functions of $f^{(j)}$ is a normal subspace of the subspace spanned by the basis functions of $f^{(i)}$. To satisfy the condition, the RS-HDMR component functions can be approximated by orthonormal polynomials $\{\varphi\}$ as follows

$$f_i(x_i) \approx \sum_{r=1}^{k} \alpha_r^{(0)i}\,\varphi_r^i(x_i) \tag{34}$$

$$f_{ij}(x_i,x_j) \approx \sum_{r=1}^{k}[\alpha_r^{(ij)i}\,\varphi_r^i(x_i) + \alpha_r^{(ij)j}\,\varphi_r^j(x_j)] + \sum_{p=1}^{l}\sum_{q=1}^{l'}\beta_{pq}^{(0)ij}\,\varphi_p^i(x_i)\,\varphi_q^j(x_j) \tag{35}$$

$$f_{ijk}(x_i,x_j,x_k) \approx \sum_{r=1}^{k}[\alpha_r^{(ijk)i}\,\varphi_r^i(x_i) + \alpha_r^{(ijk)j}\,\varphi_r^j(x_j) + \alpha_r^{(ijk)k}\,\varphi_r^k(x_k)] + \sum_{p=1}^{l}\sum_{q=1}^{l'}[\beta_{pq}^{(ijk)ij}\,\varphi_p^i(x_i)\,\varphi_q^j(x_j) + \beta_{pq}^{(ijk)ik}\,\varphi_p^i(x_i)\,\varphi_q^k(x_k) + \beta_{pq}^{(ijk)jk}\,\varphi_p^j(x_j)\,\varphi_q^k(x_k)] + \sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''}\gamma_{pqr}^{(0)ijk}\varphi_p^i(x_i)\,\varphi_q^j(x_j)\,\varphi_r^k(x_k) \tag{36}$$

etc., where the basis $\{\varphi\}$ for approximating the lower order RS-HDMR component functions is always a subset of those for the higher order ones. Notice that the basis $\{\varphi\}$ given in eqs 34−36 contains all the corresponding basis $\{\varphi\}$ given in eqs 17−19. Therefore, the bases $\{\varphi\}$ given in eqs 34−36 are referred to as *extended bases*, and those in eqs 17−19 are called *nonextended bases*.

Using eqs 34−36 the mutual orthogonality of the different order approximate RS-HDMR component functions can be guaranteed when the coefficients $\{\alpha, \beta, \gamma\}$ are obtained by stepwise least squares regression. The details of this technique and its mathematical proof are given in the Appendix.

Using the formulas in eqs 34−36, the third order RS-HDMR expansion for an n-variate function $f(\mathbf{x})$ can be expressed as

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k}(\alpha_r^{(0)i} + \sum_{\substack{j=1\\ j\neq i}}^{n}\alpha_r^{(ij)i} + \sum_{\substack{j<k=1\\ j,k\neq i}}^{n}\alpha_r^{(ijk)i})\varphi_r^i(x_i) + \sum_{1\le i<j\le n}\sum_{p=1}^{l}\sum_{q=1}^{l'}(\beta_{pq}^{(0)ij} + \sum_{\substack{k=1\\ k\neq i,j}}^{n}\beta_{pq}^{(ijk)ij})\varphi_p^i(x_i)\varphi_q^j(x_j) + \sum_{1\le i<j<k\le n}\sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''}\gamma_{pqr}^{(0)ijk}\varphi_p^i(x_i)\,\varphi_q^j(x_j)\,\varphi_r^k(x_k) \tag{37}$$

After combining all the coefficients $\{\alpha\}$, $\{\beta\}$, respectively, the final number of constant coefficients is the same as that for the truncated third order RS-HDMR expansion given in eq 23.

5. Illustrations

Two examples will be used to illustrate the new method described in section 4 and the Appendix. The first example is an integrated exposure and dose model where the input variables are sampled independently; i.e., the weight $w(\mathbf{x})$ is known and separable. For the second example, ionospheric critical frequencies are determined from ground-based ionosonde measurements, which may not be independent as input variables. In this case the distribution $w(\mathbf{x})$ of the inputs was not explicitly known.

5.1. Application to an Integrated Exposure and Dose Model. An integrated exposure and dose model was previously applied to test the correlation method to determine the RS-HDMR expansion coefficients given in eq 23.[11] Here the same model will be used to test the difference between employing eqs 23 and 37. The model was developed for the study of multiroute residential human exposures to trichloroethylene (TCE) present in tap water. It incorporates dynamic microenvironmental and pharmacokinetic models for the release of TCE from water into air considering different rooms in the home, the activities of individuals, and the physiological uptake processes for three exposure routes (ingestion, inhalation, and dermal absorption). The details of the model can be found in ref 11.

TABLE 1: Input Variable Ranges and Parameter

input	range: lower bound	range: upper bound
age (year), x_1	15	80
TCE concentration in tap water (ppb), x_2	0.001	0.5
shower stall volume (m^3), x_3	9	15
drinking water consumption rate (L/day), x_4	0.8	2.4
shower flow rate (L/min), x_5	7.7	38.3
shower time (min), x_6	5	30
time after shower in bathroom (min), x_7	5	30

The model output of interest, $f(\mathbf{x})$, is the total body burden of TCE accumulated after one month of continuous exposure via inhalation, ingestion and dermal contact. Seven input variables are selected from the integrated exposure and dose model. The ranges spanned by these seven input variables are shown in Table 1. The first four input variables x_1, x_2, x_3, x_4 have a uniform distribution, and the last three input variables x_5, x_6, x_7 have an asymmetric triangular distribution.[11]

Ten thousand random samples of $\mathbf{x}$ and the corresponding values of the output $f(\mathbf{x})$ were obtained from the model according to the specified pdf $w(\mathbf{x}) = \prod_{i=1}^{n} w_i(x_i)$. Figure 1 gives examples of the data distribution of $f(\mathbf{x})$ with respect to the two distributed input variables x_1 and x_5. Note that x_1 is a discrete variable, and the results for $f(\mathbf{x})$ correspond to samples over all of the seven input variables. The uniform and triangular distributions can be observed to have a distinct influence, and most of the data $f(\mathbf{x})$ have values less than 0.5.

The nonuniform RS-HDMR methodology was used to construct a third order RS-HDMR expansion (whose component functions were approximated by second order weighted orthonormal polynomials, i.e., $k, l, l', m, m', m'' = 2$) to form an efficient fully equivalent operational model (FEOM) for the above integrated exposure and dose model to relieve the computational burden of complex mechanistic assessment. Once the RS-HDMR component functions are determined, the resultant approximation for $f(\mathbf{x})$ serves as a FEOM.

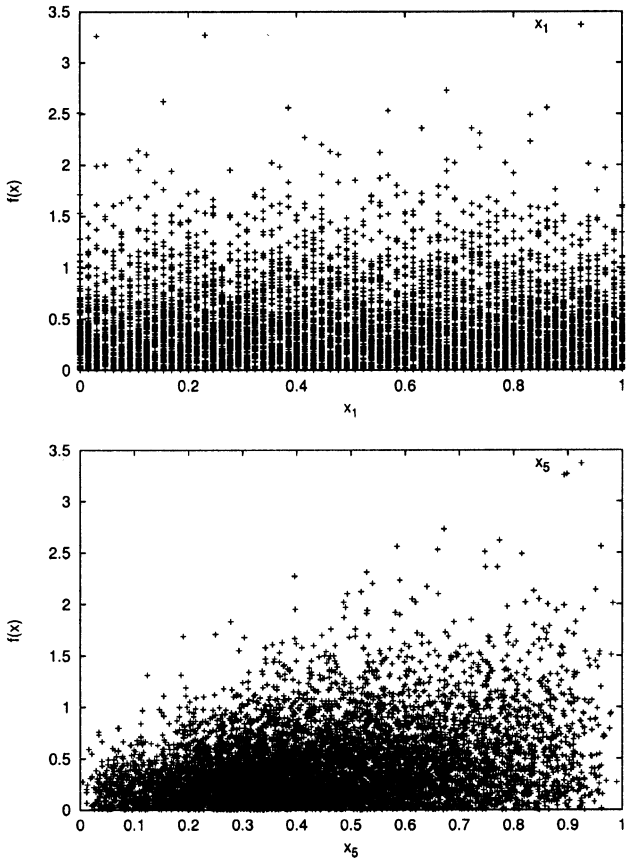

Figure 1. Data distributions for the integrated exposure and dose model with respect to the normalized uniformly distributed variable x_1 and the triangularly distributed variable x_5.

TABLE 2: Relative Errors of Different Order RS-HDMR Expansions Given by Eq 23 Whose Expansion Coefficients Were Obtained by Stepwise Least Squares Regression with Different Sample Sizes and a Second Order Orthonormal Polynomial Approximation for the RS-HDMR Component Functions[a]

		data portion (%)					
		used data			test data		
sample size N ($10000 - N$)	rel error (%)	first order	second order	third order	first order	second order	third order
500	5	14.4	39.6	59.6	15.7	36.0	32.2
(9500)	10	31.2	61.0	78.6	30.1	54.4	50.9
	20	50.2	77.0	90.2	49.2	72.1	68.9
1000	5	16.3	45.5	60.8	15.9	42.5	49.0
(9000)	10	30.0	65.6	78.6	30.0	62.5	69.3
	20	51.1	82.0	89.0	50.1	78.6	83.5
1500	5	15.7	47.7	65.9	15.9	46.7	58.6
(8500)	10	30.8	68.9	82.9	29.9	66.9	77.0
	20	50.6	83.7	91.5	50.2	81.0	87.3
2000	5	16.5	50.1	65.9	15.3	48.8	61.5
(8000)	10	30.9	69.9	82.7	29.7	69.6	79.3
	20	50.5	82.4	90.3	50.4	82.1	88.4
3000	5	16.4	55.9	72.8	15.2	53.8	70.0
(7000)	10	30.8	72.6	84.9	29.9	71.9	83.4
	20	50.8	82.8	91.4	50.6	82.4	90.4
5000	5	16.4	55.0	75.7	15.9	54.9	74.0
(5000)	10	30.1	72.5	85.7	29.9	71.7	85.5
	20	50.8	82.9	91.5	50.2	82.7	91.3

[a] Orthogonality of the HDMR component functions is not assured by this procedure.

Stepwise least squares regression was used to determine the expansion coefficients given in eqs 23 and 37 for different sample sizes (500, 1000, 1500, 2000, 3000, 5000). The accuracy of the FEOM was calculated for the data used to construct the RS-HDMR expansion (referred as *used data*) and the remaining data of the 10000 points (i.e., 9500, 9000, 8500, 8000, 7000, 5000, denoted as *test data*). The accuracy is represented by the portion of the data whose output values given by the FEOM have relative errors not larger than 5, 10 and 20%. The results are given in Tables 2 and 3 where the "data portion" gives the percentage of the data whose relative errors are not larger than a given value.

Tables 2 and 3 show that the third order RS-HDMR given by eq 37 has better accuracy than that given by eq 23 for all sample sizes. When the sample size is larger than 2000 points, the accuracy of eq 37 for the used data and test data is almost the same. This implies that from using only ~2000 data a reliable FEOM given by the RS-HDMR expansion can be constructed, which is valid over the entire input domain. This demonstrates that orthogonality of the different order RS-HDMR component functions is important in the construction of the RS-HDMR expansion, and using the extended bases given in eqs 34−36 and stepwise least squares regression can guarantee the orthogonality of the different order RS-HDMR component functions.

From Figure 1 one can see that many data have very small internal dose values. For these points, even if the absolute errors of the FEOMs are quite small, their relative errors can be very large. In this case, relative error does not give useful information. Therefore, we set a threshold value of 0.3 mg for the internal

TABLE 3: Relative Errors of Different Order RS-HDMR Expansions Given by Eq 37 Whose Expansion Coefficients Were Obtained by Stepwise Least Squares Regression with Different Sample Sizes and a Second Order Orthonormal Polynomial Approximation for the RS-HDMR Component Functions[a]

		data portion (%)					
		used data			test data		
sample size N (10000 − N)	rel error (%)	first order	second order	third order	first order	second order	third order
500	5	14.4	53.2	94.8	15.7	47.1	60.2
(9500)	10	31.2	71.4	97.2	30.1	65.0	76.4
	20	50.2	81.6	98.4	46.6	77.9	86.5
1000	5	16.3	53.9	85.7	15.9	50.5	78.3
(9000)	10	30.0	72.8	92.6	29.9	68.8	85.8
	20	51.1	82.9	96.4	50.1	80.2	92.2
1500	5	15.7	56.7	83.2	15.5	53.3	77.6
(8500)	10	30.8	72.7	90.9	29.9	70.6	87.6
	20	50.6	82.3	95.1	50.2	81.3	93.2
2000	5	16.5	55.1	80.8	15.3	52.9	76.8
(8000)	10	30.9	72.1	89.5	29.7	71.3	88.1
	20	50.5	81.9	94.8	50.4	81.9	93.4
3000	5	16.4	56.5	82.7	15.2	54.7	80.5
(7000)	10	30.8	72.5	90.8	29.9	71.5	89.7
	20	50.8	82.5	95.2	50.7	81.9	94.3
5000	5	16.3	56.1	81.5	15.9	55.4	80.2
(5000)	10	30.1	72.3	89.9	29.9	71.4	89.5
	20	50.8	82.4	94.4	50.2	81.9	94.1

[a] Orthogonality of the HDMR component functions is assured in this formulation.

dose. When the value of a datum is not larger than the threshold and the absolute value calculated by the FEOMs is not larger than the threshold as well, then we define the FEOMs as giving the correct answer. The threshold value of 0.3 mg is chosen because it is smaller than the 30 day reference dose (RfD) of a standard 70 kg male for assessing health risks.[23] Therefore, we added the portion of the data satisfying this condition to the data whose values are larger than the threshold and with 5, 10 and 20% relative errors as a representation of the accuracy. The resultant accuracy for eq 37 is given in Table 4 and is very satisfactory.

5.2. Application to Measured Ionosonde Data. RS-HDMR has been tested using ionosonde data measured at Huancayo, Peru, between the years 1957 and 2000. An ionosonde transmits radio wave signals that are reflected when the transmitted frequency is equal to the local plasma frequency in the ionosphere. Electron densities as a function of altitude at a given time are calculated from these returned frequencies. The ionospheric electron density is characterized by the "critical frequencies" returned from the peak density in the E-region (*foE*), and the peak density in the F-region (*foF2*) of the ionosphere. These critical frequencies are the RS-HDMR outputs, and there are six measured geophysical input parameters: year, day-of-year, time-of-day, $F_{10.7}$, *Kp*, and *Dst*. Here $F_{10.7}$ represents the 10.7 cm solar flux index, which is a surrogate index for solar output: high values of $F_{10.7}$ occur during a solar maximum, and low values occur during a solar minimum. *Kp* is a 3-hourly index of the solar particle radiation derived from geomagnetic field variations measured at 13 subauroral locations. *Dst* is also an index based on the geomagnetic field, which is derived from mid- and low-latitude sites and monitors occurrences of magnetic storms.

The critical frequencies *foE* and *foF*2 follow the 11 year solar cycle variation seen in the $F_{10.7}$ solar flux measurements, so the input variable "year" is transformed to "year = (year − 1957 mod 11)". The ionosphere exhibits much greater day-to-night variations within a day than it does at the same hour from day-to-day, so separate RS-HDMR expansions were constructed at each hour of the day, eliminating the need to use time-of-day as an RS-HDMR input. The values of the outputs at anytime within a day are obtained by simple interpolation. The sixth input, the previous day's measured *foE* or *foF*2, is a lagged data value that is necessary to eliminate large errors in the RS-HDMR expansion coefficients caused by autocorrelation due to the time series nature of the data. Traditionally, output is treated as linearly dependent on its lagged value, but RS-HDMR treats it as an arbitrary nonlinear function.

The results for the data at one time-of-day, 12 UT (universal time), are given below. The weight function $w(\mathbf{x})$ is not known, but data analysis shows that the input variables are dependent; i.e., $w(\mathbf{x})$ is not separable. The relationship between the input year and $F_{10.7}$ is illustrated in the upper panel of Figure 2, and the relationship between the inputs *Kp* and *Dst* is plotted on the lower panel.

TABLE 4: Relative Errors of Different Order RS-HDMR Expansions Given by Eq 37 Whose Expansion Coefficients Were Obtained by Stepwise Least Squares Regression with Different Sample Sizes and a Second Order Orthonormal Polynomial Approximation for the RS-HDMR Component Functions (Threshold = 0.3 mg)[a]

		data portion (%)					
		used data			test data		
sample size N (10000 − N)	rel error (%)	first order	second order	third order	first order	second order	third order
500	5	47.4	85.2	100.0	49.9	82.6	87.6
(9500)	10	61.2	97.0	100.0	61.6	93.0	95.4
	20	75.4	98.4	100.0	75.8	97.2	98.0
1000	5	48.1	84.2	99.6	49.9	83.3	95.4
(9000)	10	58.3	95.0	100.0	60.8	94.3	98.3
	20	74.3	98.6	100.0	75.4	97.7	98.9
1500	5	48.1	86.1	99.5	49.0	85.2	97.0
(8500)	10	59.7	95.0	99.7	60.2	94.8	98.9
	20	74.1	98.5	99.7	75.2	98.0	99.1
2000	5	49.9	85.5	99.2	48.8	84.1	96.8
(8000)	10	60.7	95.2	99.6	59.8	94.3	98.8
	20	75.0	98.2	99.6	74.8	97.9	99.0
3000	5	49.6	86.6	99.0	48.8	85.5	97.5
(7000)	10	60.8	95.2	99.6	60.2	94.7	99.0
	20	75.3	98.4	99.7	75.5	98.0	99.1
5000	5	49.7	85.9	98.7	49.3	86.1	97.9
(5000)	10	60.5	94.5	99.3	60.4	94.4	99.0
	20	75.4	97.9	99.3	75.4	97.9	99.1

[a] Orthogonality of the HDMR component functions is assured in this formulation.

TABLE 5: Relative Errors of Different Order RS-HDMR Expansions Given by Eq 23 Whose Expansion Coefficients Were Obtained by Stepwise Least Squares Regression with a Second Order Orthonormal Polynomial Approximation for the RS-HDMR Component Functions[a]

			data portion (%)					
			used data			test data		
output	sample size N used data (test data)	rel error (%)	first order	second order	third order	first order	second order	third order
	3000	5	68.4	77.8	81.3	56.3	62.0	63.9
foE	(5711)	10	94.9	97.4	98.3	88.6	91.0	91.4
		20	99.7	99.7	99.8	99.0	98.9	98.7
	3000	5	36.4	41.8	45.3	29.0	32.6	31.0
foF2	(6521)	10	63.8	73.1	77.0	54.2	57.7	57.9
		20	93.6	95.1	97.0	86.1	86.7	86.7

[a] Orthogonality of the HDMR component functions is not assured by this procedure.

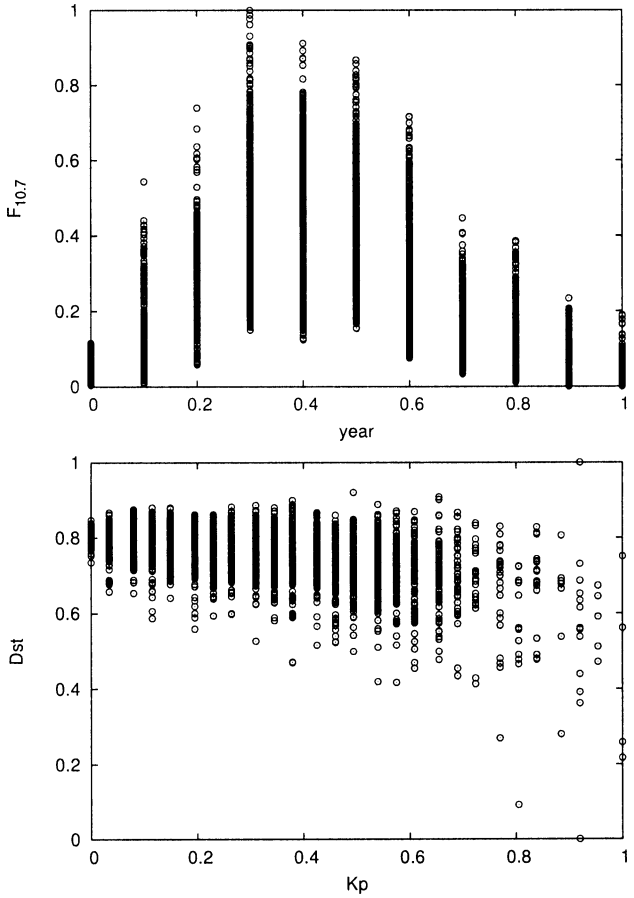

Figure 2. Ionosonde data distribution. The dependences between normalized input variables: year and $F_{10.7}$, *Kp* and *Dst* for the data at 12 UT.

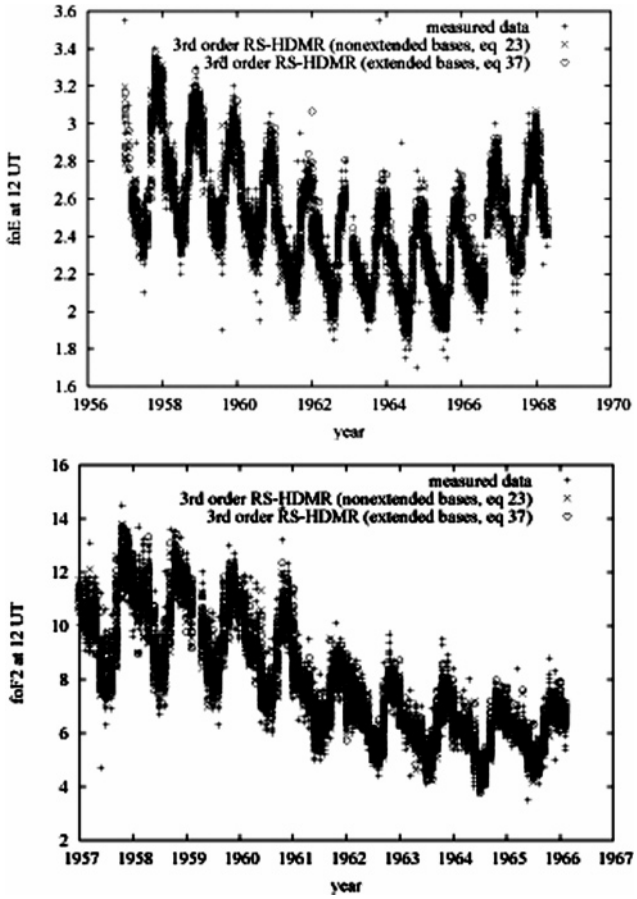

Figure 3. Comparison between the measured used data and the second order RS-HDMR approximations for *foE* and *foF*2 in the ionosonde illustration.

TABLE 6: Relative Errors of Different Order RS-HDMR Expansions Given by Eq 37 Whose Expansion Coefficients Were Obtained by Stepwise Least Squares Regression with a Second Order Orthonormal Polynomial Approximation for the RS-HDMR Component Functions[a]

			data portion (%)					
			used data			test data		
output	sample size *N* used data (test data)	rel error (%)	first order	second order	third order	first order	second order	third order
foE	3000 (5711)	5	68.4	83.3	86.7	56.3	72.1	70.0
		10	94.9	97.8	98.5	88.6	93.5	92.1
		20	99.7	99.7	99.7	99.0	98.8	98.1
foF2	3000 (6521)	5	36.4	45.4	48.2	29.0	36.1	35.0
		10	63.8	76.2	79.4	54.2	65.2	62.5
		20	93.6	96.1	97.5	86.1	90.9	89.5

[a] Orthogonality of the HDMR component functions is assured by this procedure.

Despite the unknown distribution $w(\mathbf{x})$, these data can be treated by eqs 23 and 37 because the sample is taken with respect to $w(\mathbf{x})$, and the Monte Carlo approximations do not explicitly need $w(\mathbf{x})$. The third order RS-HDMR is constructed with the first 3000 data points (between the years 1957 and 1968), referred to as *used data*. The RS-HDMR approximations are tested on the subsequent ~6000 data points (between the years 1968 and 2000), referred to as *test data*. The accuracy of the RS-HDMR expansions acting as FEOMs for *foE* and *foF*2 (the total numbers of the data for *foE* and *foF*2 are 8711 and 9521, respectively) are given in Tables 5 and 6 for eqs 23 and 37.

Tables 5 and 6 show that second and third order RS-HDMR expansion accuracies given by eq 37 are better than those given by eq 23 for both used data and test data. This implies that eq 37 has better interpolation and extrapolation accuracy. The first-order RS-HDMR expansions for both equations are the same, so they have the same accuracy. These results further establish that a system with a nonseparable weight $w(\mathbf{x})$ can be treated in the same way as a system with independent input variables, and orthogonality of the different order RS-HDMR component functions is important in the construction of the RS-HDMR expansion.

Comparisons between the measured data and the RS-HDMR approximations of *foE* and *foF*2 for used data and test data are given in Figures 3 and 4. The accuracy is satisfactory, indicating that a FEOM built with data from one solar cycle (approximately 11 years of data) can be constructed to reliably predict the values of *foE* and *foF*2 for the following two solar cycles (about 20 years).

Finally, a comparison of the results in Tables 4 and 6 shows that the exposure and dose model produced a more accurate FEOM than that arising from the ionosonde data. Comparing different physical examples is difficult, but a basic distinction in these two examples is that the exposure and dose model is built around inputs that are error free whereas the ionosonde data have errors which are not explicitly characterized. The latter errors are likely a major contribution to the quality difference between the FEOMs for the two examples. Nevertheless, the ionosonde example shows that an effective FEOM can be generated from real data including where there is correlation between the input variables.

6. Conclusion

The prior definitions of the HDMR component functions are extended in this paper to treat systems whose input variables

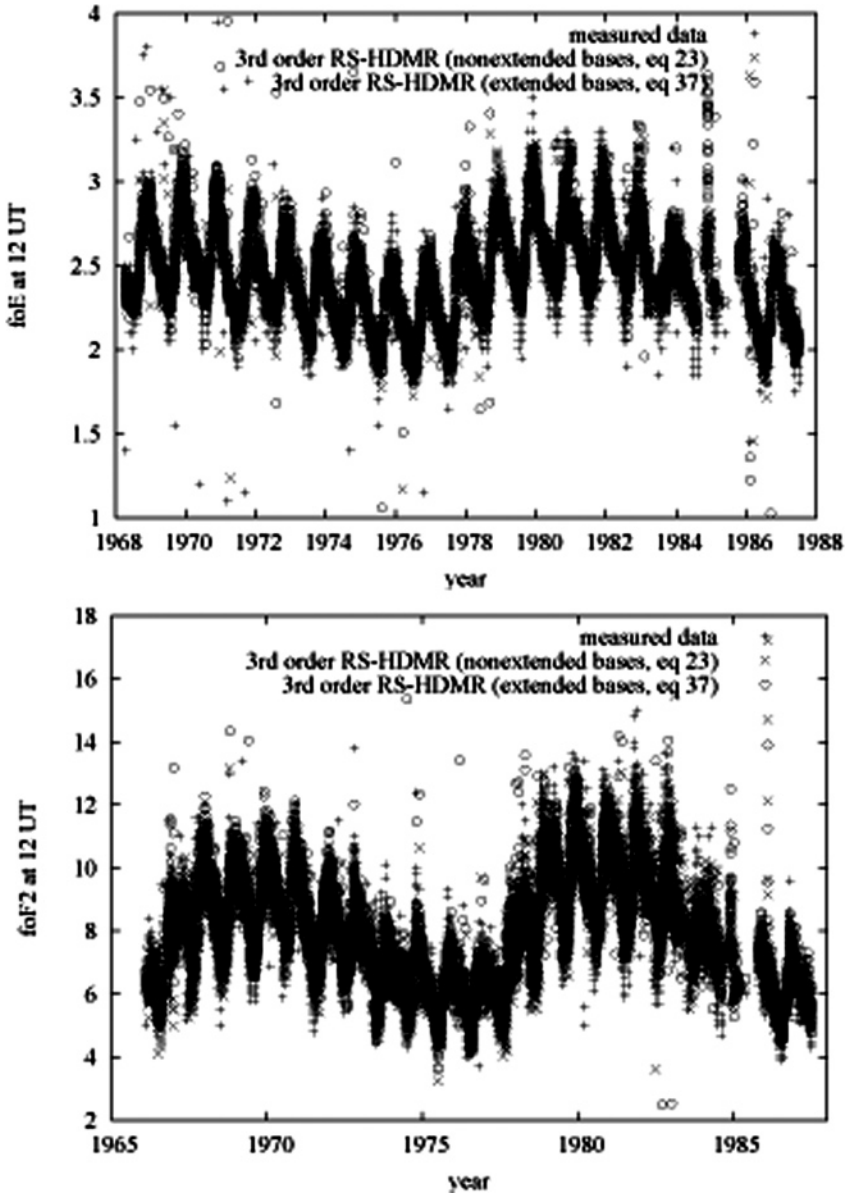

Figure 4. Comparison between the measured test data and the second order RS-HDMR approximations for *foE* and *foF*2 in the ionosonde illustration. The prediction accuracy for years 1968−1988 data is satisfactory.

may not be independent. The resultant formulas for the HDMR component functions are the same as the original ones except that the probability density functions are replaced by conditional probability density functions, and the formulas after using the orthonormal polynomial and Monte Carlo approximations are exactly the same. The formulas for the HDMR component functions with independent input variables are only a special case of the general treatment. The HDMR component functions are originally defined to be mutually orthogonal. The orthogonality of the different order HDMR component functions guarantees that a higher order truncated HDMR expansion always has better accuracy than any lower order one. RS-HDMR is a practical approach within the family of HDMR formulations. The component functions of RS-HDMR are constructed from randomly sampled input−output data by the Monte Carlo integration approximation. For this reason, with a modest sample size, the different component functions of RS-HDMR, as originally formulated, are not strictly orthogonal after they are approximated by orthonormal polynomial expansions. In this paper, new orthonormal polynomial approximation formulas for the RS-HDMR component functions are presented. The basis functions used for the low order RS-HDMR component functions are always subsets of those for higher order ones. By utilizing stepwise least squares regression, the new orthonormal polynomial approximations of the RS-HDMR component functions guarantee that the different order approximated RS-HDMR component functions are mutually orthogonal for any sample size when the Monte Carlo integration approximation is used. The tests for an integrated exposure and dose model and measured ionosonde data demonstrate that the new method has better accuracy than the prior one.

Acknowledgment. We acknowledge support from the Army Research Office and the U.S. Environmental Protection Agency, under Cooperative Agreement No. EPA CR-827033 to the Environmental and Occupational Health Sciences Institute. This work was also supported by Air Force AFRL/VS contract number F19628-02-C-0051. Ionosonde data was obtained from the Space Physics Interactive Data Resource (SPIDR) web page, National Geophysical Data Center (NGDC), NOAA Satellite and Information Service (http://spidr.ngdc.noaa.gov/spidr).

Appendix

A. New Basis Functions. When the basis functions $\{\varphi\}$ used for the low order RS-HDMR component functions are subsets of those for the higher order ones, and the expansion coefficients $\{\alpha, \beta, \gamma\}$ are determined by stepwise least squares regression, then the new orthonormal polynomial approximations for the RS-HDMR component functions guarantee that the different order approximated RS-HDMR component functions are mutually orthogonal upon using the Monte Carlo integration approximation for any sample size. The proof is given below.

The HDMR component functions are optimally determined from zeroth to higher orders in a step-by-step fashion

$$f(\mathbf{x}) \approx f_0 + f^{(1)} + f^{(2)} + f^{(2)} + ... \tag{38}$$

where $f^{(i)}$ denotes the collection of ith order HDMR terms. Such systematic approaches to function approximation are quite common. Consider the Hilbert space H of continuous functions where a set of linearly independent elements $\{v_1, v_2, ..., v_{n_1}\} \in H$ are selected to form the basis for a subspace $V \subset H$. Suppose that the first order approximation v for a model output $f(\mathbf{x})$ is of the form

$$v = \sum_{i=1}^{n_1} b_i v_i \tag{39}$$

where b_i are constant coefficients. To get better accuracy, we wish to find the second-order approximation $g \in H$ which is consistent with a given set of data and is as close as possible to v. Let

$$g = v + r \tag{40}$$

where r is in a subspace $R = \{r_1, r_2, ..., r_{n_r}\}(n_r > n_1)$ of H. The best g is provided by r that is orthogonal to v.[24] The *sufficient* condition for r to be orthogonal to v is that V is a normal subspace of R, i.e.,

$$V \subset R \tag{41}$$

In this case, R can be decomposed as

$$R = V \oplus V^{\perp} \tag{42}$$

where $V^{\perp} \neq \{0\}$ is the orthogonal complementary subspace of V in R, and one can always find a $r \in V^{\perp}$ orthogonal to v. Similarly, we can consider the third-order approximation

$$g = u + v + r \tag{43}$$

and the sufficient condition to find the best v and r for a given u is

$$U \subset V \subset R \tag{44}$$

where U and V are subspaces spanned by the bases $\{u_1, u_2, ..., u_{n_1}\}$ and $\{v_1, v_2, ..., v_{n_2}\}$ $(n_r > n_2 > n_1)$. In this case, u, v, r can be mutually orthogonal.

To satisfy the condition given in eq 44, the RS-HDMR component functions can be approximated by orthonormal polynomials $\{\varphi\}$ as follows

$$f_i(x_i) \approx \sum_{r=1}^{k} \alpha_r^{(0)i} \varphi_r^i(x_i) \tag{45}$$

$$f_{ij}(x_i,x_j) \approx \sum_{r=1}^{k} [\alpha_r^{(ij)i} \varphi_r^i(x_i) + \alpha_r^{(ij)j} \varphi_r^j(x_j)] + \sum_{p=1}^{l}\sum_{q=1}^{l'} \beta_{pq}^{(0)ij} \varphi_p^i(x_i) \varphi_q^j(x_j) \tag{46}$$

$$f_{ijk}(x_i,x_j,x_k) \approx \sum_{r=1}^{k} [\alpha_r^{(ijk)i} \varphi_r^i(x_i) + \alpha_r^{(ijk)j} \varphi_r^j(x_j) + \alpha_r^{(ijk)k} \varphi_r^k(x_k)] + \sum_{p=1}^{l}\sum_{q=1}^{l'} [\beta_{pq}^{(ijk)ij} \varphi_p^i(x_i) \varphi_q^j(x_j) + \beta_{pq}^{(ijk)ik} \varphi_p^i(x_i) \varphi_q^k(x_k) + \beta_{pq}^{(ijk)jk} \varphi_p^j(x_j) \varphi_q^k(x_k)] + \sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''} \gamma_{pqr}^{(0)ijk} \varphi_p^i(x_i) \varphi_q^j(x_j) \varphi_r^k(x_k) \tag{47}$$

where the basis $\{\varphi\}$ for approximating the lower order RS-HDMR component functions is always a subset of those for the higher order ones. Then the mutual orthogonality of the different order approximated RS-HDMR component functions can be guaranteed with properly chosen coefficients $\{\alpha, \beta, \gamma\}$.

B. Second-Order Expansion. The remaining task is to determine all the coefficients in eqs 45−47 so that the different order orthonormal polynomial terms, and consequently, different order approximated RS-HDMR component functions are mutually orthogonal.

Suppose N random samples $\{\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(N)}\}$ for $\mathbf{x}$ are generated according to a given pdf $w(\mathbf{x})$. The inner product of two functions $v(\mathbf{x})$ and $r(\mathbf{x})$ is defined as

$$(v(\mathbf{x}), r(\mathbf{x})) = \frac{1}{N}\sum_{s=1}^{N} v(\mathbf{x}^{(s)})\, r(\mathbf{x}^{(s)}) \tag{48}$$

Let $v(\mathbf{x})$, $r(\mathbf{x})$ represent the collections of the first and second-order RS-HDMR component functions, respectively. Then, $g(\mathbf{x})$, the second-order RS-HDMR approximation of $f(\mathbf{x}) - f_0$, is

$$g(\mathbf{x}) = v(\mathbf{x}) + r(\mathbf{x}) = f(\mathbf{x}) - f_0 \quad \forall\, \mathbf{x} \in (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(N)}) \tag{49}$$

with

$$v(\mathbf{x}) = \sum_{i=1}^{n_1} b_i v_i(\mathbf{x}) \tag{50}$$

$$r(\mathbf{x}) = \sum_{i=1}^{n_r} a_i r_i(\mathbf{x}) \tag{51}$$

where $v_i(\mathbf{x})$, $r_i(\mathbf{x})$ are first and second order RS-HDMR terms approximated by the orthonormal polynomials $\{\varphi\}$ given in eqs 45−46. The basis $\{r_1(\mathbf{x}), r_2(\mathbf{x}), ..., r_{n_r}(\mathbf{x})\}$ is composed of all the elements and their possible products of $\{\varphi\}$ with one and two variables. As $\{v_1, v_2, ..., v_{n_1}\}$ is a subset of $\{r_1(\mathbf{x}), r_2(\mathbf{x}), ..., r_{n_r}(\mathbf{x})\}$, we arrange $\{r_i\}$ in the order $\{v_1, v_2, ..., v_{n_1}, r_{n_1+1}, r_{n_1+2}, ..., r_{n_r}\}$.

Consider the inner products

$$(r_i(\mathbf{x}), g(\mathbf{x})) = (r_i(\mathbf{x}), f(\mathbf{x}) - f_0) = d_i \qquad (i = 1, 2, ..., n_r) \tag{52}$$

Substituting eqs 49−51 into eq 52 yields

$$(r_i(\mathbf{x}), \sum_{j=1}^{n_1} b_j v_j(\mathbf{x}) + \sum_{j=1}^{n_r} a_j r_j(\mathbf{x})) = (r_i(\mathbf{x}), \sum_{j=1}^{n_1} b_j v_j(\mathbf{x})) + (r_i(\mathbf{x}), \sum_{j=1}^{n_r} a_j r_j(\mathbf{x})) = d_i \quad (53)$$

Equation 53 may be represented in matrix form as

$$B\mathbf{b} + A\mathbf{a} = \mathbf{d} \quad (54)$$

where A is an $(n_r \times n_r)$ symmetric nonsingular matrix with the (i, j)-entry

$$A_{ij} = (r_i(\mathbf{x}), r_j(\mathbf{x})) \quad (55)$$

B is an $(n_r \times n_1)$ rectangular matrix with with the (i, j)-entry

$$B_{ij} = (r_i(\mathbf{x}), v_j(\mathbf{x})) \quad (56)$$

and

$$\mathbf{a} = (a_1, a_2, ..., a_{n_r})^{\mathrm{T}} \quad (57)$$

$$\mathbf{b} = (b_1, b_2, ..., b_{n_1})^{\mathrm{T}} \quad (58)$$

$$\mathbf{d} = (d_1, d_1, ..., d_{n_r})^{\mathrm{T}} \quad (59)$$

Because the first n_1 elements of $\{r_i\}$ are $\{v_i\}$, the matrix B is just the first n_1 columns of the matrix A.

If $r(\mathbf{x})$ is chosen from $V^{\perp}$, then $r(\mathbf{x})$ is orthogonal to V, i.e., $r(\mathbf{x})$ is orthogonal to the basis $\{v_i\}$:

$$(v_i(\mathbf{x}), r(\mathbf{x})) = (v_i(\mathbf{x}), \sum_{j=1}^{n_r} a_j r_j(\mathbf{x})) = 0 \quad (i = 1, 2, ..., n_1) \quad (60)$$

The above equation may also be represented in matrix form as

$$B^{\mathrm{T}}\mathbf{a} = \mathbf{0} \quad (61)$$

where B^{T} is the transpose of B.

Combining eqs 54 and 61 gives

$$\begin{bmatrix} A & B \\ B^{\mathrm{T}} & 0 \end{bmatrix}\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \mathbf{0} \end{bmatrix} \quad (62)$$

To attain a better understanding of eq 62, we consider eqs 54 and 61 separately. Because A is nonsingular, A^{-1} exists. As B^{T} is the first n_1 rows of A, then

$$B^{\mathrm{T}}A^{-1}B = [I_{n_1}|\mathbf{0}]B = \hat{B} \quad (63)$$

where I_{n_1} is the n_1-dimensional identity matrix, $\mathbf{0}$ is the $(n_1 \times (n_r - n_1))$ null matrix, and $\hat{B}$ is an $(n_1 \times n_1)$ matrix composed of the first n_1 rows of B and it is also the submatrix of A composed of the elements of A in the first n_1 rows and first n_1 columns.

Multiplying both sides of eq 54 by $B^{\mathrm{T}}A^{-1}$ from the left and using eqs 61 and 63 gives

$$B^{\mathrm{T}}A^{-1}B\mathbf{b} + B^{\mathrm{T}}A^{-1}A\mathbf{a} = B^{\mathrm{T}}A^{-1}\mathbf{d}$$

$$\hat{B}\mathbf{b} + B^{\mathrm{T}}\mathbf{a} = [I_{n_1}|\mathbf{0}]\mathbf{d}$$

$$\hat{B}\mathbf{b} = \hat{\mathbf{d}} \quad (64)$$

where $\hat{\mathbf{d}}$ is the first n_1 elements of $\mathbf{d}$. Note that eq 64 is the equation to determine the coefficients $\mathbf{b}$ of the first-order approximated RS-HDMR component functions by least squares regression.

The coefficients $\mathbf{a}$ for the second order approximated RS-HDMR component functions can be obtained by solving the following equation

$$A\mathbf{a} = \mathbf{d} - B\mathbf{b} = \mathbf{d} - B\hat{B}^{-1}\hat{\mathbf{d}} \quad (65)$$

which gives

$$\mathbf{a} = A^{-1}\mathbf{d} - A^{-1}B\hat{B}^{-1}\hat{\mathbf{d}} = A^{-1}\mathbf{d} - \begin{bmatrix} I_{n_1} \\ \mathbf{0} \end{bmatrix}\hat{B}^{-1}\hat{\mathbf{d}} = A^{-1}\mathbf{d} - \begin{bmatrix} \hat{B}^{-1}\hat{\mathbf{d}} \\ \mathbf{0} \end{bmatrix} \quad (66)$$

Equation 66 is the least squares solution for the coefficients $\mathbf{a}$ of the second order approximated RS-HDMR component functions after substituting in the resultant coefficients $\mathbf{b}$ of the first-order functions.

Combining all these results, we can rewrite eq 62 as

$$\begin{bmatrix} A & B \\ & \hat{B} \end{bmatrix}\begin{bmatrix} \mathbf{a} \\ \mathbf{b} \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \hat{\mathbf{d}} \end{bmatrix} \quad (67)$$

Equation 67 implies that to construct orthogonal first and second order approximate RS-HDMR component functions, the coefficients $\mathbf{b}$ and $\mathbf{a}$ should be sequentially determined by least squares regression.

The orthogonality between $v(\mathbf{x})$ and $r(\mathbf{x})$ given by eq 61 can be proved using eqs 63 and 66 as follows:

$$B^{\mathrm{T}}\mathbf{a} = B^{\mathrm{T}}A^{-1}\mathbf{d} - B^{\mathrm{T}}A^{-1}B\hat{B}^{-1}\hat{\mathbf{d}} = [I_{n_1}|\mathbf{0}]\boldsymbol{d} - \hat{B}\hat{B}^{-1}\hat{\mathbf{d}} = \hat{\mathbf{d}} - \hat{\mathbf{d}} = \mathbf{0} \quad (68)$$

C. Third-Order Expansion. Now let $g(\mathbf{x})$ represent the third order approximated RS-HDMR expansion of $f(\mathbf{x}) - f_0$, i.e.,

$$g(\mathbf{x}) = u(\mathbf{x}) + v(\mathbf{x}) + r(\mathbf{x}) = f(\mathbf{x}) - f_0 \quad \forall\, \mathbf{x} \in (\mathbf{x}^{(1)}, \mathbf{x}^{(2)}, ..., \mathbf{x}^{(N)}) \quad (69)$$

with

$$u(\mathbf{x}) = \sum_{i=1}^{n_1} c_i u_i(\mathbf{x}) \quad (70)$$

$$v(\mathbf{x}) = \sum_{i=1}^{n_2} b_i v_i(\mathbf{x}) \quad (71)$$

$$r(\mathbf{x}) = \sum_{i=1}^{n_r} a_i r_i(\mathbf{x}) \quad (72)$$

where $u(\mathbf{x})$, $v(\mathbf{x})$, and $r(\mathbf{x})$ are the collections of first, second and third order RS-HDMR component functions approximated by the orthonormal polynomials $\{\varphi\}$ given in eqs 45−47. The basis $\{r_1, r_2, ..., r_{n_r}\}$ is all the elements and their possible products of $\{\varphi\}$ with one, two and three variables, and the elements of $\{u_i\}$, $\{v_i\}$, $\{r_i\}$ are arranged as

$$\{u_1, u_2, ..., u_{n_1}\}$$

$$\{u_1, u_2, ..., u_{n_1}, v_{n_1+1}, v_{n_1+2}, ..., v_{n_2}\}$$

$$\{u_1, u_2, ..., u_{n_1}, v_{n_1+1}, v_{n_1+2}, ..., v_{n_2}, r_{n_2+1}, r_{n_2+2}, ..., r_{n_r}\}$$

Consider the inner products

$$(r_i(\mathbf{x}), g(\mathbf{x})) = (r_i(\mathbf{x}), f(\mathbf{x}) - f_0) = d_i \quad (i = 1, 2, ..., n_r) \quad (73)$$

and substituting eqs 70−72 into eq 73 yields

$$(r_i(\mathbf{x}), f(\mathbf{x}) - f_0) = (r_i(\mathbf{x}), \sum_{j=1}^{n_1} c_j u_j(\mathbf{x}) + \sum_{i=1}^{n_2} b_j v_j(\mathbf{x}) + \sum_{j=1}^{n_r} a_j r_j(\mathbf{x}))$$

$$= (r_i(\mathbf{x}), \sum_{j=1}^{n_1} c_j u_j(\mathbf{x})) + (r_i(\mathbf{x}), \sum_{j=1}^{n_2} b_j v_j(\mathbf{x})) + (r_i(\mathbf{x}), \sum_{j=1}^{n_r} a_j r_j(\mathbf{x})) = d_i \quad (74)$$

Equation 74 can be represented in matrix form as

$$C\mathbf{c} + B\mathbf{b} + A\mathbf{a} = \mathbf{d} \quad (75)$$

where A is an $(n_r \times n_r)$ symmetric matrix; B and C are $(n_r \times n_2)$ and $(n_r \times n_1)$ rectangular matrixes, respectively; and

$$\mathbf{a} = (a_1, a_2, ..., a_{n_r})^{\mathrm{T}} \quad (76)$$

$$\mathbf{b} = (b_1, b_2, ..., b_{n_2})^{\mathrm{T}} \quad (77)$$

$$\mathbf{c} = (c_1, c_2, ..., c_{n_1})^{\mathrm{T}} \quad (78)$$

$$\mathbf{d} = (d_1, d_1, ..., d_{n_r})^{\mathrm{T}} \quad (79)$$

The orthogonality of the different order terms is given by the following equations. Because $r(\mathbf{x})$ is orthogonal to U and V, $r(\mathbf{x})$ must be orthogonal to their bases, i.e.,

$$(u_i(\mathbf{x}), r(\mathbf{x})) = (u_i(\mathbf{x}), \sum_{j=1}^{n_r} a_j r_j(\mathbf{x})) = 0 \quad (i = 1, 2, ..., n_1) \quad (80)$$

$$(v_i(\mathbf{x}), r(\mathbf{x})) = (v_i(\mathbf{x}), \sum_{j=1}^{n_r} a_j r_j(\mathbf{x})) = 0 \quad (i = 1, 2, ..., n_2) \quad (81)$$

Similarly, if $v(\mathbf{x})$ is chosen from $U^{\perp}$, then v must be orthogonal to the basis of U, i.e.,

$$(u_i(\mathbf{x}), v(\mathbf{x})) = (u_i(\mathbf{x}), \sum_{j=1}^{n_2} b_j v_j(\mathbf{x})) = 0 \quad (i = 1, 2, ..., n_1) \quad (82)$$

The above equations may also be represented in matrix form as

$$C^{\mathrm{T}}\mathbf{a} = 0 \quad (83)$$

$$B^{\mathrm{T}}\mathbf{a} = 0 \quad (84)$$

$$D^{\mathrm{T}}\mathbf{b} = 0 \quad (85)$$

where D is an $(n_2 \times n_1)$ rectangular matrix. Note that all the matrixes B, C, D are submatrixes of A because they are obtained from the inner products of the orthonormal polynomial basis functions which are all contained in $\{r_i\}$.

Multiplying both sides of eq 75 from the left by C^TA^{-1} and $B^{\mathrm{T}}A^{-1}$, respectively, and using eqs 83 and 84 yields

$$C^{\mathrm{T}}A^{-1}C\mathbf{c} + C^{\mathrm{T}}A^{-1}B\mathbf{b} = C^{\mathrm{T}}A^{-1}\mathbf{d} \quad (86)$$

$$B^{\mathrm{T}}A^{-1}C\mathbf{c} + B^{\mathrm{T}}A^{-1}B\mathbf{b} = B^{\mathrm{T}}A^{-1}\mathbf{d} \quad (87)$$

Let

$$\hat{B} = B^{\mathrm{T}}A^{-1}B = \begin{bmatrix} A_{11} & \cdots & A_{1n_2} \\ \vdots & \ddots & \vdots \\ A_{n_2 1} & \cdots & A_{n_2 n_2} \end{bmatrix} \quad (88)$$

$$\hat{C} = C^{\mathrm{T}}A^{-1}C = \begin{bmatrix} A_{11} & \cdots & A_{1n_1} \\ \vdots & \ddots & \vdots \\ A_{n_1 1} & \cdots & A_{n_1 n_1} \end{bmatrix} \quad (89)$$

$$D = B^{\mathrm{T}}A^{-1}C = \begin{bmatrix} A_{11} & \cdots & A_{1n_1} \\ \vdots & \ddots & \vdots \\ A_{n_2 1} & \cdots & A_{n_2 n_1} \end{bmatrix} \quad (90)$$

$$\hat{\mathbf{d}}_{n_1} = C^{\mathrm{T}}A^{-1}\mathbf{d} = [I_{n_1}|\mathbf{0}]\mathbf{d} = (d_1, ..., d_{n_1})^{\mathrm{T}} \quad (91)$$

$$\hat{\mathbf{d}}_{n_2} = B^{\mathrm{T}}A^{-1}\mathbf{d} = [I_{n_2}|\mathbf{0}]\mathbf{d} = (d_1, ..., d_{n_2})^{\mathrm{T}} \quad (92)$$

By using eq 85, it follows that eqs 86 and 87 reduce to

$$\hat{C}\mathbf{c} = \hat{\mathbf{d}}_{n_1} \quad (93)$$

$$D^{\mathrm{T}}\mathbf{c} + \hat{B}\mathbf{b} = \hat{\mathbf{d}}_{n_2} \quad (94)$$

Then,

$$\mathbf{c} = \hat{C}^{-1}\hat{\mathbf{d}}_{n_1} \quad (95)$$

which is the least-squares solution for the coefficients $\mathbf{c}$ of the first order approximated RS-HDMR component functions. Then

$$\mathbf{b} = \hat{B}^{-1}\hat{\mathbf{d}}_{n_2} - \hat{B}^{-1}D\hat{C}^{-1}\hat{\mathbf{d}}_{n_1} = \hat{B}^{-1}\hat{\mathbf{d}}_{n_2} - \begin{bmatrix} I_{n_1} \\ \mathbf{0} \end{bmatrix}\hat{C}^{-1}\hat{\mathbf{d}}_{n_1} = \hat{B}^{-1}\hat{\mathbf{d}}_{n_2} - \begin{bmatrix} \hat{C}^{-1}\hat{\mathbf{d}}_{n_1} \\ \mathbf{0} \end{bmatrix} \quad (96)$$

This is the solution for $\mathbf{b}$ obtained by least squares after substituting the resultant values of $\mathbf{c}$ from eq 95.

The orthogonality between $u(\mathbf{x})$ and $v(\mathbf{x})$ given by eq 85 can be proved by using eq 96

$$D^{\mathrm{T}}\mathbf{b} = D^{\mathrm{T}}\hat{B}^{-1}\hat{\mathbf{d}}_{n_2} - D^{\mathrm{T}}\hat{B}^{-1}D\hat{C}^{-1}\hat{\mathbf{d}}_{n_1} = [I_{n_1}|\mathbf{0}]\hat{\mathbf{d}}_{n_2} - \hat{C}\hat{C}^{-1}\hat{\mathbf{d}}_{n_1} = \hat{\mathbf{d}}_{n_1} - \hat{\mathbf{d}}_{n_1} = \mathbf{0} \quad (97)$$

The coefficient vector $\mathbf{a}$ is obtained by

$$\mathbf{a} = A^{-1}(\mathbf{d} - B\mathbf{b} - C\mathbf{c}) \quad (98)$$

which is also the solution of **a** obtained by least squares after substituting in the resultant values of **b** and **c**. Thus, eqs 75, 86, and 87 can be written in matrix form as

$$\begin{bmatrix} A & B & C \\ & \hat{B} & D \\ & & \hat{C} \end{bmatrix} \begin{bmatrix} \mathbf{a} \\ \mathbf{b} \\ \mathbf{c} \end{bmatrix} = \begin{bmatrix} \mathbf{d} \\ \mathbf{d}_{n_2} \\ \mathbf{d}_{n_1} \end{bmatrix} \tag{99}$$

Equation 99 implies that to construct mutually orthogonal first, second and third order approximated RS-HDMR component functions, the coefficients **c**, **b** and **a** should be determined step-by-step through least squares regression. The orthogonality between $r(\mathbf{x})$ and $u(\mathbf{x})$, $v(\mathbf{x})$ given in eqs 83 and 84 can be proved as follows:

$$\begin{aligned} C^{\mathrm{T}}\mathbf{a} &= C^{\mathrm{T}}A^{-1}(\mathbf{d} - B\mathbf{b} - C\mathbf{c}) \\ &= [I_{n_1}|\mathbf{0}](\mathbf{d} - B\mathbf{b} - C\mathbf{c}) \\ &= \mathbf{d}_{n_1} - D^{\mathrm{T}}\mathbf{b} - \hat{C}\mathbf{c} \\ &= \mathbf{d}_{n_1} - \mathbf{d}_{n_1} = \mathbf{0} \end{aligned} \tag{100}$$

$$\begin{aligned} B^{\mathrm{T}}\mathbf{a} &= B^{\mathrm{T}}A^{-1}(\mathbf{d} - B\mathbf{b} - C\mathbf{c}) \\ &= [I_{n_2}|\mathbf{0}](\mathbf{d} - B\mathbf{b} - C\mathbf{c}) \\ &= \mathbf{d}_{n_2} - \hat{B}\mathbf{b} - D^{\mathrm{T}}\mathbf{c} \\ &= \mathbf{d}_{n_2} - \mathbf{d}_{n_2} = \mathbf{0} \end{aligned} \tag{101}$$

The above results show that using the bases given in eqs 45−47 and stepwise least squares regression to determine the coefficients $\{\alpha, \beta, \gamma, ...\}$, the resultant different order approximated RS-HDMR component functions are mutually orthogonal for a given set of data.

References and Notes

(1) Rabitz, H.; Alis, O. F.; Shorter, J.; Shim, K. Efficient input-output model representations. *Comput. Phys. Commun.* **1999**, *117*, 11−20.

(2) Shim, K.; Rabitz, H. Independent and correlated composition behavior of material properties: application to energy band gaps for the $Ga_\alpha In_{1-\alpha}P_\beta As_{1-\beta}$ and $Ga_\alpha In_{1-\alpha}P_\beta Sb_\gamma As_{1-\beta-\gamma}$ alloys. *Phys. Rev. B* **1998**, *58*, 1940−1946.

(3) Alis, O. F.; Rabitz, H. General foundations of high dimensional model representations. *J. Math. Chem.* **1999**, *25*, 197−233.

(4) Shorter, J.; Rabitz, H. An efficient chemical kinetics solver using high dimensional model representations. *J. Phys. Chem. A* **1999**, *103*, 7192−7198.

(5) Alis, O. F.; Rabitz, H. Efficient implementation of high dimensional model representations. *J. Math. Chem.* **2001**, *29*, 127−142.

(6) Li, G.; Rosenthal, C.; Rabitz, H. High dimensional model representations. *J. Phys. Chem. A* **2001**, *105*, 7765−7777.

(7) Wang, S. W.; Levy, H., II; Li, G.; Rabitz, H. Fully equivalent operational models for atmospheric chemical kinetics within global chemistry-transport models. *J. Geophys. Res.* **1999**, *104* (No. D23), 30417−30426.

(8) Shorter, J.; Rabitz, H. Radiation transport simulation by means of a fully equivalent operational model. *Geophys. Res. Lett.* **2000**, *27*, 3485−3488.

(9) Li, G.; Wang, S. W.; Rabitz, H. Practical approaches to construct RS-HDMR component functions. *J. Phys. Chem.* **2002**, *106*, 8721−8733.

(10) Li, G.; Wang, S. W.; Rabitz, H.; Wang, S. K.; Jáffe, P. Global uncertainty assessments of a mathematical model for trace metal bioremediation. *Chem. Eng. Sci.* **2002**, *57*, 4445−4460.

(11) Wang, S. W.; Georgopoulos, P. G.; Li, G.; Rabitz, H. Random sampling-high dimensional model representation (RS-HDMR) with nonuniformly distributed variables: application to an integrated multimedia/multipathway exposure and dose model for trichloroethylene. *J. Phys. Chem. A* **2003**, *107*, 4707−4718.

(12) Li, G.; Wang, S. W.; Rosenthal, C.; Rabitz, H. High dimensional model representations generated from low dimensional data samples, I. mp-Cut-HDMR. *J. Math. Chem.* **2001**, *30*, 1−30.

(13) Li, G.; Artamonov, M.; Rabitz, H.; Wang, S. W.; Georgopoulos, P. G.; Demiralp, M. High dimensional model representations generated from low order terms − lp-RS-HDMR. *J. Comput. Chem.* **2003**, *24*, 647−656.

(14) Li, G.; Rabitz, H.; Wang, S. W.; Georgopoulos, P. G. Correlation method for variance reduction of Monte Carlo integration in RS-HDMR. *J. Comput. Chem.* **2003**, *24*, 277−283.

(15) Li, G.; Rabitz, H. Ratio control variate method for variance reduction of Monte Carlo integration in RS-HDMR. *J. Comput. Chem.*, in press.

(16) Takemura, A., Tensor analysis of ANOVA decomposition. *J. Am. Statist. Assoc.* **1983**, *78*, 894−900.

(17) Sobol, I. M. Sensitivity Estimates for Nonlinear Mathematical Models. *Mathematical Modeling and Computational Experiments* **1993**, *1*, 407−414.

(18) Geremia, J. M.; Rabitz, H.; Rosenthal, C. Constructing global functional maps between molecular potentials and quantum observables. *J. Chem. Phys.* **2001**, *114*, 9325−9336.

(19) Li, G.; Rabitz, H.; Wang, S. W.; Georgopoulos, P. G. Characterizing uncertainties in human exposure modeling through the random sampling-high dimensional model representation (RS-HDMR) Methodology. *IJRAM* **2005**, *5*, 387−406.

(20) Feng, X. J.; Hooshangi, S.; Chen, D.; Li, G.; Weiss, R.; Rabitz, H. Optimizing genetic Circuits by global sensitivity analysis. *Biophys. J.* **2004**, *87*, 2195−2202.

(21) Larsen, R. J.; Marx, M. L. *An Introduction to Probability and its Application*; Prentice-Hall: Englewood Cliffs, New Jersey, 1985.

(22) Evan, M.; Swartz, T. *Approximating Integrals via Monte Carlo and Deterministic Methods*; Oxford University Press: Oxford, NY, 2000.

(23) Risk assessment toxicity values in the Risk Assessment Information System (RAIS) database (http://rais.ornl.gov/tox/profiles/trichloroethene_r-agsa.shtml). Oak Ridge National Laboratory: Oak Ridge, TN, 2004.

(24) Deutsch, F. *Best Approximation in Inner Product Space*; Springer: New York, 2001.