Random Sampling–High Dimensional Model Representation (RS–HDMR) with Nonuniformly Distributed Variables: Application to an Integrated Multimedia/Multipathway Exposure and Dose Mode...

ARTICLE *in* THE JOURNAL OF PHYSICAL CHEMISTRY A · JUNE 2003

Impact Factor: 2.69 · DOI: 10.1021/jp022500f

Random Sampling−High Dimensional Model Representation (RS−HDMR) with Nonuniformly Distributed Variables: Application to an Integrated Multimedia/ Multipathway Exposure and Dose Model for Trichloroethylene

Sheng-Wei Wang and Panos G. Georgopoulos

Environmental and Occupational Health Sciences Institute, 170 Frelinghuysen Road, Piscataway, New Jersey 08854

Genyuan Li and Herschel Rabitz*

Department of Chemistry, Princeton University, Princeton, New Jersey 08544

Received: November 27, 2002

The high dimensional model representation (HDMR) technique is a procedure for representing high dimensional functions efficiently. A practical form of the technique, random sampling−high dimensional model representation (RS−HDMR), is based on randomly sampling the overall function. In reality, the samples are often obtained according to some probability density functions (pdfs). This paper extends our previous RS−HDMR work with uniformly distributed random samples to those with a nonuniform distribution and treats uniform sampling as a special case. Weighted orthonormal polynomial expansions are introduced to approximate the RS−HDMR component functions. Different pdfs give special formulas for the weighted orthonormal polynomials. However, the structure of the formulas for the RS−HDMR component functions represented by the Monte Carlo integration approximation are the same for all pdfs. The correlation method to reduce the variance of the Monte Carlo integration and the method to represent the high order terms by lower order terms in uniform RS−HDMR can also be used for nonuniform RS−HDMR. The theoretical basis of nonuniform RS−HDMR is provided, and an application is presented to an integrated environmental exposure and dose model for trichloroethylene.

1. Introduction

High dimensional model representation (HDMR) is a general set of quantitative model assessment and analysis tools for capturing high dimensional input−output system behavior.[1−6] Different forms of HDMR have been introduced. A practical form of the technique, random sampling−high dimensional model representation (RS−HDMR), is based on randomly sampling the overall function. In our previous work,[7−9] the random data are sampled over a uniform distribution; i.e., the probability density function (pdf) of the input variables is unity. However, in practice, the data are often nonuniformly distributed (for instance, a normal distribution, etc.). The formulation of RS−HDMR for a uniform distribution cannot be directly applied to data with a nonuniform distribution. In this paper, we extend our previous uniform RS−HDMR work to nonuniform distributions. After introducing weighted orthonormal polynomials related to the pdf, the general formulas of RS−HDMR can be constructed, including uniform RS−HDMR as a special case. The formulation of the Monte Carlo integration approximation for the RS−HDMR component functions is the same for all pdfs. The correlation method to reduce the variance of the Monte Carlo integration,[8] and the method to represent the high order terms by lower order terms in uniform RS−HDMR,[10] can be also applied within nonuniform RS−HDMR.

The paper is organized as follows. Section 2 provides the theoretical basis of nonuniform RS−HDMR. Section 3 presents an illustration of the method to an integrated environmental exposure and dose model for trichloroethylene. Finally, section 4 contains conclusions.

* Author to whom correspondence should be addressed. E-mail: hrabitz@princeton.edu.

2. Methodology

2.1. RS−HDMR. Because the impact of the multiple input variables on the output can be independent and cooperative, HDMR expresses the model output $f(\boldsymbol{x})$ as a finite hierarchical correlated function expansion in terms of the input variables:

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

where $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, the zeroth order (i.e., $l = 0$) component function f_0 is a constant representing the mean response of $f(\boldsymbol{x})$, and the first order (i.e., $l = 1$) component function $f_i(x_i)$ gives the independent contribution to $f(\boldsymbol{x})$ by the ith input variable acting alone, the second order (i.e., $l = 2$) component function $f_{ij}(x_i, x_j)$ gives the pair-correlated contribution to $f(\boldsymbol{x})$ by the input variables x_i and x_j, etc. The last term $f_{12\ldots n}(x_1, x_2, \ldots, x_n)$ contains any residual nth order correlated contribution of all input variables.

Distinct, but formally equivalent, HDMR expansions, all of which have the same structure as eq 1, may be constructed. When data are considered as randomly sampled, RS−HDMR can be obtained. For RS−HDMR, we first rescale the variables x_i by some suitable transformations such that $0 \le x_i \le 1$ for all i. The output function $f(\boldsymbol{x})$ is then defined in the unit hypercube $K^n = \{(x_1, x_2, \ldots, x_n) | 0 \le x_i \le 1, i = 1, 2, \ldots, n\}$. The independent input variable x_i ($i = 1, 2, \ldots, n$) has the pdf $w_i(x_i)$ satisfying

10.1021/jp022500f CCC: $25.00
Published on Web 05/17/2003

the conditions

$$\begin{cases} w_i(x_i) \geq 0 & (\text{for } 0 \leq x_i \leq 1) \\ \int_0^1 w_i(x_i)\, \mathrm{d}x_i = 1 & (i = 1, 2, ..., n) \end{cases} \tag{2}$$

The component functions of RS−HDMR with the input variables having the pdf $w_i(x_i)$ are defined as follows:

$$f_0 = \int_{K^n} \prod_{i=1}^{n} w_i(x_i) f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{3}$$

$$f_i(x_i) = \int_{K^{n-1}} \prod_{\substack{k=1 \\ k \neq i}}^{n} w_k(x_k) f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}^i - f_0 \tag{4}$$

$$f_{ij}(x_i, x_j) = \int_{K^{n-2}} \prod_{\substack{k=1 \\ k \neq i,j}}^{n} w_k(x_k) f(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}^{ij} - f_i(x_i) - f_j(x_j) - f_0 \tag{5}$$

...

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

The RS−HDMR component functions $f_i(x_i)$, $f_{ij}(x_i, x_j)$,... possess the property

$$\int_0^1 w_s(x_s) f_{i_1 i_2 ... i_l}(x_{i_1}, x_{i_2}, ..., x_{i_l})\, \mathrm{d}x_s = 0 \qquad (s \in \{i_1, i_2, ..., i_l\}) \tag{6}$$

which defines the mutual *weighted* orthogonality between two RS−HDMR component functions:

$$\int_{K^n} \prod_{i=1}^{n} w_i(x_i) f_{i_1 i_2 ... i_l}(x_{i_1}, x_{i_2}, ..., x_{i_l}) f_{j_1 j_2 ... j_k}(x_{j_1}, x_{j_2}, ..., x_{j_k})\, \mathrm{d}\boldsymbol{x} = 0 \tag{7}$$
$$(\{i_1, i_2, ..., i_l\} \neq \{j_1, j_2, ..., j_k\})$$

The component functions $f_i(x_i)$, $f_{ij}(x_i, x_j)$,... may be provided numerically, at discrete values of the input variables x_i, x_j,... produced from sampling the output function $f(\boldsymbol{x})$ for employment on the right-hand side of eqs 3−5. Thus, numerical data tables can be constructed for these component functions. A critical feature of the HDMR expansion is that its component functions are optimal choices tailored to a given $f(\boldsymbol{x})$ over the entire desired domain of $\boldsymbol{x}$.[3] Experience shows that the high order terms in the expansion often are negligible, and only the first few low order terms are needed to give a satisfactory approximation of $f(\boldsymbol{x})$. Thus, the approximate value of $f(\boldsymbol{x})$ for an arbitrary point $\boldsymbol{x}$ can be determined from these tables by performing only low dimensional interpolation over $f_i(x_i)$, $f_{ij}(x_i, x_j)$,...

2.1.1. Direct Monte Carlo Integration Approximation. To construct the numerical data tables for the RS−HDMR component functions, one needs to evaluate the previously mentioned integrals. Evaluation of the high dimensional integrals in the RS−HDMR expansion may be performed by Monte Carlo random sampling integration,[11,12] hence, the name RS−HDMR.

The theoretical foundation of Monte Carlo integration is the following.[11] Suppose $\boldsymbol{x} = (x_1, x_2, ..., x_n)$ are independent random variables with pdf $w_i(x_i)$ $(i = 1, 2, ..., n)$ and $F(\boldsymbol{x})$ is a function of $\boldsymbol{x}$. For N points of $\boldsymbol{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, ..., x_n^{(s)})$ $(s = 1, 2, ..., N)$ randomly generated in K^n, according to the pdf $w_i(x_i)$ $(i = 1, 2, ..., n)$,[12] the random variable

$$F_N = \frac{1}{N} \sum_{s=1}^{N} F(\boldsymbol{x}^{(s)}) \tag{8}$$

has the expectation

$$\langle F_N \rangle = \int_{K^n} \prod_{i=1}^{n} w_i(x_i) F(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x} \tag{9}$$

the variance

$$\mathrm{var}\{F_N\} = \frac{1}{N} \mathrm{var}\{F(\boldsymbol{x})\} \tag{10}$$

and the standard deviation (standard error)

$$\sigma\{F_N\} = (\mathrm{var}\{F_N\})^{1/2} = \frac{1}{\sqrt{N}} \sigma\{F(\boldsymbol{x})\} \tag{11}$$

Therefore, F_N can be used as an estimate of the integral $\int_{K^n} \prod_{i=1}^{n} w_i(x_i) F(\boldsymbol{x})\, \mathrm{d}\boldsymbol{x}$ with a standard error $\sigma\{F_N\}$ proportional to the standard error $\sigma\{F(\boldsymbol{x})\}$ of the integrand random variable $F(\boldsymbol{x})$.

For f_0, $f_i(x_i)$, and $f_{ij}(x_i, x_j)$, N sets of $\boldsymbol{x}^{(s)}$, $(x_i, \boldsymbol{x}^i)^{(s)} = (x_1^{(s)}, x_2^{(s)}, ..., x_{i-1}^{(s)}, x_i, x_{i+1}^{(s)}, ..., x_n^{(s)})$ with distinct fixed values of x_i and $(x_i, x_j, \boldsymbol{x}^{ij})^{(s)} = (x_1^{(s)}, x_2^{(s)}, ..., x_{i-1}^{(s)}, x_i, x_{i+1}^{(s)}, ..., x_{j-1}^{(s)}, x_j, x_{j+1}^{(s)}, ..., x_n^{(s)})$ with distinct fixed values of (x_i, x_j) are generated, according to the pdfs $\prod_{i=1}^{n} w_i(x_i^{(s)})$, $\prod_{k=1,k\neq l}^{n} w_k(x_k^{(s)})$, and $\prod_{k=1,k\neq i,j}^{n} w_k(x_k^{(s)})$, respectively, and then

$$f_0 \approx \frac{1}{N} \sum_{s=1}^{N} f(\boldsymbol{x}^{(s)}) \tag{12}$$

$$f_i(x_i) \approx \frac{1}{N} \sum_{s=1}^{N} f((x_i, \boldsymbol{x}^i)^{(s)}) - \frac{1}{N} \sum_{s=1}^{N} f(\boldsymbol{x}^{(s)}) \tag{13}$$

$$f_{ij}(x_i, x_j) \approx \frac{1}{N} \sum_{s=1}^{N} f((x_i, x_j, \boldsymbol{x}^{ij})^{(s)}) - \frac{1}{N} \sum_{s=1}^{N} f((x_i, \boldsymbol{x}^i)^{(s)}) - \frac{1}{N} \sum_{s=1}^{N} f((x_j, x^j)^{(s)}) + \frac{1}{N} \sum_{s=1}^{N} f(\boldsymbol{x}^{(s)}) \tag{14}$$

...

These formulas were obtained in our previous work for uniform RS−HDMR.[7−10] The pdfs $w_i(x_i)$ $(i = 1, 2, ..., n)$ are already involved in the sampling $f(\boldsymbol{x}^{(s)})$, $f((x_i, \boldsymbol{x}^i)^{(s)})$, and $f((x_i, x_j, \boldsymbol{x}^{ij})^{(s)})$ and do not explicitly appear in the previously described Monte Carlo integration formulas. Other high order component functions can be determined similarly.

2.1.2. Weighted Orthonormal Polynomial Expansion Approximation. For uniform RS−HDMR (i.e., all $w_i(x_i) = 1$), to reduce the sampling effort, the RS−HDMR component functions may be approximated by expansions, in terms of a suitable set of basis functions (for instance, orthonormal polynomials, spline functions, or simple polynomials).[7] When the points $\{x_i\}$ are not uniformly distributed, the RS−HDMR component functions can be approximated by weighted orthonormal polynomials $\{\varphi\}$ as

$$f_i(x_i) \approx \sum_{r=1}^{k} \alpha_r^i \varphi_r^i(x_i) \tag{15}$$

$$f_{ij}(x_i, x_j) \approx \sum_{p=1}^{l}\sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_p^i(x_i)\varphi_q^j(x_j) \tag{16}$$

$$f_{ijk}(x_i, x_j, x_k) \approx \sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''} \gamma_{pqr}^{ijk} \varphi_p^i(x_i)\varphi_q^j(x_j)\varphi_r^k(x_k) \tag{17}$$

$$\cdots$$

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

$$\int_0^1 w_i(x_i)\varphi_r^i(x_i)\,dx_i = 0 \qquad \text{(for all } r, i\text{)} \tag{18}$$

$$\int_0^1 w_i(x_i)[\varphi_r^i(x_i)]^2\,dx_i = 1 \qquad \text{(for all } r, i\text{)} \tag{19}$$

$$\int_0^1 w_i(x_i)\varphi_p^i(x_i)\varphi_q^i(x_i)\,dx_i = 0 \qquad (p \neq q) \tag{20}$$

i.e., they have a zero mean and a unit norm and are mutually orthogonal, with respect to the weight $w_i(x_i)$. In most cases, to achieve satisfactory accuracy using only $\varphi_1^i(x_i)$, the terms $\varphi_1^i(x_i)$, $\varphi_2^i(x_i)$, and $\varphi_3^i(x_i)$ are often sufficient (i.e., k, l, l', m, m', $m'' \leq 3$). Utilizing the conditions in eqs 18−20, the approximations for the RS−HDMR component functions given by eqs 15−17 will preserve the mutual weighted orthogonality in eq 7. In the following material, some examples of $\{\varphi\}$ are provided.

2.1.2.1. Uniform Distribution. For the uniform distribution, $w_i(x_i) = 1$, and

$$\varphi_1^i(x_i) = \sqrt{3}(2x_i - 1) \tag{21}$$

$$\varphi_2^i(x_i) = 6\sqrt{5}\left(x_i^2 - x_i + \frac{1}{6}\right) \tag{22}$$

$$\varphi_3^i(x_i) = 20\sqrt{7}\left(x_i^3 - \frac{3}{2}x_i^2 + \frac{3}{5}x_i - \frac{1}{20}\right) \tag{23}$$

$$\cdots$$

2.1.2.2. Triangular Distribution. For the triangular distribution

$$w_i(x_i) = \begin{cases} \dfrac{2}{\mu}x_i & (\text{for } 0 \leq x_i \leq \mu) \\ \dfrac{2}{1-\mu}(1 - x_i) & (\text{for } \mu < x_i \leq 1) \end{cases} \tag{24}$$

shown in Figure 1, we have

$$\varphi_1^i(x_i) = \sqrt{\frac{2(1+\mu)}{1+\mu^3}}[3x_i - (1+\mu)] \tag{25}$$

$$\varphi_2^i(x_i) = \alpha[-10(1 - \mu + \mu^2)x_i^2 + 2(4 - \mu - \mu^2 + 4\mu^3)x_i - (1 + 2\mu - 3\mu^2 + 2\mu^3 + \mu^4)] \tag{26}$$

$$\varphi_3^i(x_i) = \beta(c_3x_i^3 + c_2x_i^2 + c_1x_i + c_0) \tag{27}$$

$$\cdots$$

where

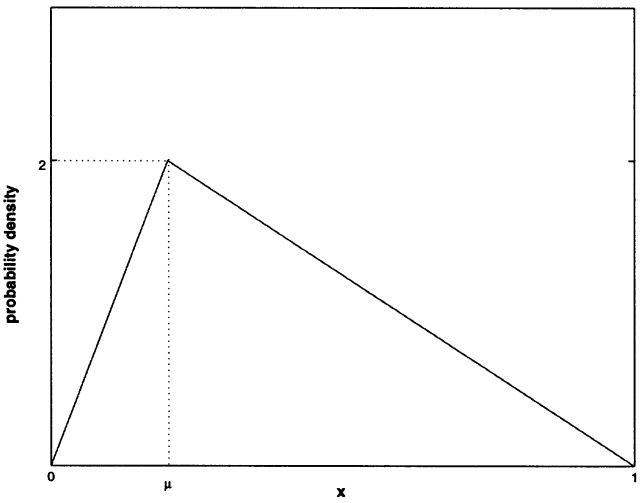

Figure 1. Triangular distribution.

$$\alpha = \sqrt{\frac{3}{(1 - \mu + \mu^2)(1 - 3\mu + 8\mu^2 - 11\mu^3 + 8\mu^4 - 3\mu^5 + \mu^6)}} \tag{28}$$

$$\beta = \sqrt{\frac{3}{p(\mu)q(\mu)}} \tag{29}$$

$$p(\mu) = 1 - 3\mu + 8\mu^2 - 11\mu^3 + 8\mu^4 - 3\mu^5 + \mu^6 \tag{30}$$

$$q(\mu) = 3 - 18\mu + 90\mu^2 - 285\mu^3 + 593\mu^4 - 860\mu^5 + 957\mu^6 - 860\mu^7 + 593\mu^8 - 285\mu^9 + 90\mu^{10} - 18\mu^{11} + 3\mu^{12} \tag{31}$$

$$c_0 = -2(1 + 3\mu - 14\mu^2 + 30\mu^3 - 18\mu^4 - 18\mu^5 + 30\mu^6 - 14\mu^7 + 3\mu^8 + \mu^9) \tag{32}$$

$$c_1 = 10(3 - 3\mu - 2\mu^2 + 29\mu^3 - 48\mu^4 + 29\mu^5 - 2\mu^6 - 3\mu^7 + 3\mu^8) \tag{33}$$

$$c_2 = -10(9 - 21\mu + 44\mu^2 - 26\mu^3 - 26\mu^4 + 44\mu^5 - 21\mu^6 + 9\mu^7) \tag{34}$$

$$c_3 = 70(1 - 3\mu + 8\mu^2 - 11\mu^3 + 8\mu^4 - 3\mu^5 + \mu^6) \tag{35}$$

Using the formulas in eqs 15−17, eq 1 can be expressed as

$$f(\boldsymbol{x}) \approx f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k} \alpha_r^i \varphi_r^i(x_i) + \sum_{1 \leq i < j \leq n}\sum_{p=1}^{l}\sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_p^i(x_i)\varphi_q^j(x_j) + \sum_{1 \leq i < j < k \leq n}\sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''} \gamma_{pqr}^{ijk} \varphi_p^i(x_i)\varphi_q^j(x_j)\varphi_r^k(x_k) + \cdots \tag{36}$$

The coefficients can be determined using the weighted orthonormality properties of $\{\varphi\}$ as follows:

$$\begin{aligned} \alpha_r^i &= \int_0^1 w_i(x_i) f_i(x_i) \varphi_r^i(x_i)\,dx_i \\ &= \int_{K^n} \prod_{k=1}^{n} w_k(x_k) f(\boldsymbol{x}) \varphi_r^i(x_i)\,d\boldsymbol{x} \\ &\approx \frac{1}{N}\sum_{s=1}^{N} f(\boldsymbol{x}^{(s)}) \varphi_r^i(x_i^{(s)}) \end{aligned} \tag{37}$$

$$\beta_{pq}^{ij} = \int_0^1 \int_0^1 w_i(x_i) w_j(x_j) f_{ij}(x_i, x_j) \varphi_p^i(x_i) \varphi_q^j(x_j)\, \mathrm{d}x_i\, \mathrm{d}x_j$$

$$= \int_{K^n} \prod_{k=1}^{n} w_k(x_k) f(\boldsymbol{x}) \varphi_r^i(x_i) \varphi_q^j(x_j)\, \mathrm{d}\boldsymbol{x}$$

$$\approx \frac{1}{N} \sum_{s=1}^{N} f(\boldsymbol{x}^{(s)}) \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)}) \tag{38}$$

$$\gamma_{pqr}^{ijk} = \int_0^1 \int_0^1 \int_0^1 w_i(x_i) w_j(x_j) w_k(x_k) f_{ijk}(x_i, x_j, x_k) \varphi_p^i(x_i) \varphi_q^j \times (x_j) \varphi_r^k(x_k)\, \mathrm{d}x_i\, \mathrm{d}x_j\, \mathrm{d}x_k$$

$$= \int_{K^n} \prod_{l=1}^{n} w_l(x_l) f(\boldsymbol{x}) \varphi_p^i(x_i) \varphi_q^j(x_j) \varphi_r^k(x_k)\, \mathrm{d}\boldsymbol{x}$$

$$\approx \frac{1}{N} \sum_{s=1}^{N} f(\boldsymbol{x}^{(s)}) \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)}) \varphi_q^k(x_k^{(s)}) \tag{39}$$

...

The formulas in eqs 37−39 are the same as those for uniform RS−HDMR,[7] because the sampling is drawn from the guidance by the weights $\{w_i(x_i)\}$.

When the Monte Carlo integration approximation is used, for a given set of N samples, the weighted orthogonality of the elements of $\{\varphi\}$ no longer holds; for example,

$$\int_0^1 w_i(x_i) \varphi_p^i(x_i) \varphi_q^i(x_i)\, \mathrm{d}x_i \approx \frac{1}{N} \sum_{s=1}^{N} \varphi_p^i(x_i^{(s)}) \varphi_q^i(x_i^{(s)}) \neq 0 \tag{40}$$

All the coefficients α_r^i, β_{pq}^{ij}, γ_{pqr}^{ijk},... in eq 36 then are coupled with each other and we may determine them simultaneously by solving a system of linear algebraic equations. However, for a high dimension n, the number of linear algebraic equations can be very large, and solving the equations is not computationally efficient.[8,10]

2.2. Correlation Method To Improve the Accuracy of Monte Carlo Integration in RS−HDMR. The error of Monte Carlo integration can be reduced either by increasing the sample size N or decreasing the variance of $F(\boldsymbol{x})$ in K^n. Monte Carlo integration error becomes troublesome when the random data of the integrand $F(\boldsymbol{x})$ have a large variance (i.e., $F(\boldsymbol{x})$ has rapid changes in the desired domain, especially in sign). This behavior is expected to arise when considering the integrands in the determination of β_{pq}^{ij}, γ_{pqr}^{ijk}, and other higher order coefficients in eqs 37−39 with products of the functions $\{\varphi\}$ such as $f(\boldsymbol{x}) \varphi_p^i(x_i) \varphi_q^j(x_j)$ and $f(\boldsymbol{x}) \varphi_p^i(x_i) \varphi_q^j(x_j) \varphi_r^k(x_k)$. The difficulty arises because the integrals exhibit rapid changes and the functions $\{\varphi\}$ are fixed, regardless of the form of $f(\boldsymbol{x})$. The determination of the expansion coefficients of high order RS−HDMR component functions by Monte Carlo integration generally requires additional samples. For example, to determine α_r^i by eq 37, a few hundred samples may give good accuracy; however, for β_{pq}^{ij} in eq 38, to achieve the same accuracy, thousands of samples may be needed, and for γ_{pqr}^{ijk} in eq 39, even more samples may be needed. However, the sample size is often restricted by the computational effort. These problems may be addressed by recognizing that the accuracy of Monte Carlo integration may be improved by reducing the variance of the integrand.

To improve the accuracy of Monte Carlo integration, the correlation method may be employed to reduce the variance of the integrand.[8,11] Consider an integral for any coefficient in eqs 37−39, for example,

$$\alpha_r^i = \int_{K^n} \prod_{k=1}^{n} w_k(x_k) f(\boldsymbol{x}) \varphi_r^i(x_i)\, \mathrm{d}\boldsymbol{x} \tag{41}$$

The variance of the integrand $\prod_{k=1}^{n} w_k(x_k) f(\boldsymbol{x}) \varphi_r^i(x_i)$ in K^n can be reduced if one can find a reference function $h(\boldsymbol{x})$ satisfying two conditions: (1) $f(\boldsymbol{x}) - h(\boldsymbol{x})$ is almost constant or zero in the entire domain, and (2) the integral

$$\int_{K^n} \prod_{k=1}^{n} w_k(x_k) h(\boldsymbol{x}) \varphi_r^i(x_i)\, \mathrm{d}\boldsymbol{x} = c_r^i \tag{42}$$

is known analytically. Then,

$$\alpha_r^i = \int_{K^n} \prod_{k=1}^{n} w_k(x_k) [f(\boldsymbol{x}) - h(\boldsymbol{x})] \varphi_r^i(x_i)\, \mathrm{d}\boldsymbol{x} + \int_{K^n} \prod_{k=1}^{n} w_k(x_k) h(\boldsymbol{x}) \varphi_r^i(x_i)\, \mathrm{d}\boldsymbol{x}$$

$$= \int_{K^n} \prod_{k=1}^{n} w_k(x_k) [f(\boldsymbol{x}) - h(\boldsymbol{x})] \varphi_r^i(x_i)\, \mathrm{d}\boldsymbol{x} + c_r^i \tag{43}$$

Now, the variance comes only from the first term in eq 43. Because $f(\boldsymbol{x}) - h(\boldsymbol{x})$ is almost constant or zero everywhere, we expect that

$$\mathrm{var}\{\prod_{k=1}^{n} w_k(x_k) [f(\boldsymbol{x}) - h(\boldsymbol{x})] \varphi_r^i(x_i)\} < \mathrm{var}\{\prod_{k=1}^{n} w_k(x_k) f(\boldsymbol{x}) \varphi_r^i(x_i)\} \tag{44}$$

α_r^i may be approximated by Monte Carlo integration,

$$\alpha_r^i \approx \frac{1}{N} \sum_{s=1}^{N} [f(\boldsymbol{x}^{(s)}) - h(\boldsymbol{x}^{(s)})] \varphi_r^i(x_i^{(s)}) + c_r^i \tag{45}$$

for data sampled according to $w_k(x_k)$ with better accuracy than that given by eq 37. Similarly, we also have

$$\beta_{pq}^{ij} \approx \frac{1}{N} \sum_{s=1}^{N} [f(\boldsymbol{x}^{(s)}) - h(\boldsymbol{x}^{(s)})] \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)}) + c_{pq}^{ij} \tag{46}$$

$$\gamma_{pqr}^{ijk} \approx \frac{1}{N} \sum_{s=1}^{N} [f(\boldsymbol{x}^{(s)}) - h(\boldsymbol{x}^{(s)})] \varphi_p^i(x_i^{(s)}) \varphi_q^j(x_j^{(s)}) \varphi_r^k(x_k^{(s)}) + c_{pqr}^{ijk} \tag{47}$$

where

$$c_{pq}^{ij} = \int_{K^n} \prod_{k=1}^{n} w_k(x_k) h(\boldsymbol{x}) \varphi_p^i(x_i) \varphi_q^j(x_j)\, \mathrm{d}\boldsymbol{x} \tag{48}$$

$$c_{pqr}^{ijk} = \int_{K^n} \prod_{k=1}^{n} w_k(x_k) h(\boldsymbol{x}) \varphi_p^i(x_i) \varphi_q^j(x_j) \varphi_r^k(x_k)\, \mathrm{d}\boldsymbol{x} \tag{49}$$

and formulas for other high order expansion coefficients.

A truncated RS−HDMR expansion of eq 36 satisfies the two previously stated conditions and can be used as $h(\boldsymbol{x})$. For example, we may choose the third order expansion,

$$h(\mathbf{x}) = f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k} \bar{\alpha}_r^i\, \varphi_r^i(x_i) + \sum_{1\le i<j\le n}\sum_{p=1}^{l}\sum_{q=1}^{l'} \bar{\beta}_{pq}^{ij}\, \varphi_p^i(x_i)\varphi_q^j(x_j) + \sum_{1\le i<j<k\le n}\sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''} \bar{\gamma}_{pqr}^{ijk}\, \varphi_p^i(x_i)\varphi_q^j(x_j)\varphi_r^k(x_k) \quad (50)$$

where the coefficients $\{\bar{\alpha}_r^i, \bar{\beta}_{pq}^{ij}, \bar{\gamma}_{pgr}^{ijk}\}$ are determined by direct Monte Carlo integration, given in eqs 37−39. The difference $f(\mathbf{x}) - h(\mathbf{x})$ should be small if the truncated RS−HDMR expansion is a good approximation of $f(\mathbf{x})$. Moreover, the second condition holds, using the weighted orthonormality property of $\{\varphi\}$:

$$\begin{aligned}\int_{K^n}\prod_{l=1}^{n} w_l(x_l) h(\mathbf{x})\varphi_r^i(x_i)\,\mathrm{d}\mathbf{x} &= \int_{K^n}\prod_{l=1}^{n} w_l(x_l)[f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k} \bar{\alpha}_r^i\, \varphi_r^i(x_i) + \sum_{1\le i<j\le n}\sum_{p=1}^{l}\sum_{q=1}^{l'} \bar{\beta}_{pq}^{ij}\, \varphi_p^i(x_i)\varphi_q^j(x_j) \\ &\quad + \sum_{1\le i<j<k\le n}\sum_{p=1}^{m}\sum_{q=1}^{m'}\sum_{r=1}^{m''} \bar{\gamma}_{pqr}^{ijk}\, \varphi_p^i(x_i)\varphi_q^j(x_j)\varphi_r^k(x_k)]\varphi_r^i(x_i)\,\mathrm{d}\mathbf{x} \\ &= \bar{\alpha}_r^i \end{aligned} \quad (51)$$

We then have

$$\alpha_r^i \approx \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})]\varphi_r^i(x_i^{(s)}) + \bar{\alpha}_r^i \quad (52)$$

Similarly, we also have

$$\beta_{pq}^{ij} \approx \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})]\varphi_p^i(x_i^{(s)})\varphi_q^j(x_j^{(s)}) + \bar{\beta}_{pq}^{ij} \quad (53)$$

$$\gamma_{pqr}^{ijk} \approx \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})]\varphi_p^i(x_i^{(s)})\varphi_q^j(x_j^{(s)})\varphi_r^k(x_k^{(s)}) + \bar{\gamma}_{pqr}^{ijk} \quad (54)$$

Equations 52−54 show that the first terms in these equations are corrections for the initial values $\bar{\alpha}_r^i$, $\bar{\beta}_{pq}^{ij}$, and $\bar{\gamma}_{pqr}^{ijk}$. The resultant α_r^i, β_{pq}^{ij}, and γ_{pqr}^{ijk} values may be reused as new initial values for the construction of a new $h(\mathbf{x})$ with even smaller values of $f(\mathbf{x}) - h(\mathbf{x})$ to repeat the calculation again, and more-accurate results may be obtained. Then, eqs 52−54 become an iteration procedure for a given set of random samples. The iteration should be convergent if the initial $h(\mathbf{x})$ value is similar to the $f(\mathbf{x})$ value and the sample size N is large enough.

If one chooses the second order expansion as a reference function,

$$h(\mathbf{x}) = f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k} \bar{\alpha}_r^i\, \varphi_r^i(x_i) + \sum_{1\le i<j\le n}\sum_{p=1}^{l}\sum_{q=1}^{l'} \bar{\beta}_{pq}^{ij}\, \varphi_p^i(x_i)\varphi_q^j(x_j) \quad (55)$$

then

$$\begin{aligned} c_{pqr}^{ijk} &= \int_{K^n}\prod_{l=1}^{n} w_l(x_l) h(\mathbf{x})\varphi_p^i(x_i)\varphi_q^j(x_j)\varphi_r^k(x_k)\,\mathrm{d}\mathbf{x} \\ &= \int_{K^n}\prod_{l=1}^{n} w_l(x_l)[f_0 + \sum_{i=1}^{n}\sum_{r=1}^{k} \bar{\alpha}_r^i\, \varphi_r^i(x_i) + \sum_{1\le i<j\le n}\sum_{p=1}^{l}\sum_{q=1}^{l'} \bar{\beta}_{pq}^{ij}\, \varphi_p^i(x_i)\varphi_q^j(x_j)]\varphi_p^i(x_i)\varphi_q^j(x_j)\varphi_r^k(x_k)\,\mathrm{d}\mathbf{x} \\ &= 0 \end{aligned} \quad (56)$$

and γ_{pqr}^{ijk} will be obtained without iteration as

$$\gamma_{pqr}^{ijk} \approx \frac{1}{N}\sum_{s=1}^{N}[f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})]\varphi_p^i(x_i^{(s)})\varphi_q^j(x_j^{(s)})\varphi_r^k(x_k^{(s)}) \quad (57)$$

The correlation method achieves improved accuracy through the enhanced quality of the Monte Carlo integration. Thus, for a particular choice of basis $\{\varphi\}$, reference function $h(\mathbf{x})$, and sample size N, the convergence of the iteration is expected to be good.

2.3. High Order Terms of RS−HDMR Expansion Approximated by Lower Order Ones. When n is large, eq 36 has a large number of terms, i.e., a large number of coefficients to be determined by approximate Monte Carlo integration. Each coefficient has its own Monte Carlo integration error, and the total error will be large when the number of terms is large. Even though the correlation method may be employed to improve the accuracy of the Monte Carlo integration, it is often inaccurate to determine high (greater than third) order RS−HDMR component functions by weighted orthonormal polynomial approximations. To further improve the accuracy, especially for high dimensional systems, an approach called low order term product−RS−HDMR (lp−RS−HDMR) has been developed for uniform RS−HDMR,[10] which can readily be extended to nonuniform RS−HDMR.

Define a set of new functions for $l = 1, 2,..., n$ and $p = 0, 1,..., l$:

$$g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, ..., x_{i_l}) = \begin{cases} \dfrac{\prod_{s=1}^{p} f_{i_s}(x_{i_s}) \prod_{r=1}^{(l-p)/2} f_{i_{p-1+2r} i_{p+2r}}(x_{i_{p-1+2r}}, x_{i_{p+2r}})}{\prod_{s=1}^{p} ||f_{i_s}(x_{i_s})|| \prod_{r=1}^{(l-p)/2} ||f_{i_{p-1+2r}\, i_{p+2r}}(x_{i_{p-1+2r}}, x_{i_{p+2r}})||} \\ 0 \text{ (if any } ||f_{i_s}(x_{i_s})|| \text{ and/or } ||f_{i_{p-1+2r} i_{p+2r}}(x_{i_{p-1+2r}}, x_{i_{p+2r}})|| = 0) \end{cases} \quad (58)$$

where $f_{i_s}(x_{i_s})$ and $f_{i_{p-1+2r} i_{p+2r}}(x_{i_{p-1+2r}}, x_{i_{p+2r}})$ are the first- and second-order RS−HDMR component functions, represented by weighted orthonormal polynomial expansions whose expansion coefficients are accurately determined by the correlation method, and

$$\begin{aligned} ||f_{i_s}(x_{i_s})|| &= \left[\int_0^1 w_{i_s}(x_{i_s}) f_{i_s}^2(x_{i_s})\,\mathrm{d}x_{i_s}\right]^{1/2} \\ &= \left[\sum_{r=1}^{k} (\alpha_r^{i_s})^2\right]^{1/2} \qquad (i_s \in \{1, 2,..., n\}) \end{aligned} \quad (59)$$

$$||f_{i_{p-1+2r}i_{p+2r}}(x_{i_{p-1+2r}}, x_{i_{p+2r}})|| = \left[\int_0^1\int_0^1 w_{i_{p-1+2r}}(x_{i_{p-1+2r}}) w_{i_{p+2r}}(x_{i_{p+2r}}) \times f^2_{i_{p-r+2r}i_{p+2r}}(x_{i_{p-1+2r}}, x_{i_{p+2r}})\, dx_{i_{p-1+2r}}\, dx_{i_{p+2r}}\right]^{1/2} = \left[\sum_{p=1}^{l}\sum_{q=1}^{l'}(\beta_{pq}^{i_{p-1+2r}i_{p+2r}})^2\right]^{1/2} \quad (60)$$

We call

$$g_i(x_i) = \frac{f_i(x_i)}{||f_i(x_i)||} \quad (61)$$

$$g_{ij}(x_i, x_j) = \frac{f_{ij}(x_i, x_j)}{||f_{ij}(x_i, x_j)||} \quad (62)$$

normalized first and second RS−HDMR component functions. For completeness, we set

$$g_0 = \frac{f_0}{||f_0||} \quad (63)$$

Considering that the functions $g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l})$ are a separable product, and using the property of RS−HDMR component functions given by eq 6, it can be readily proved that $g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l})$ ($l = 0, 1,..., n$) are orthonormal with the weights present, i.e.,

$$\int_{K^n}\prod_{i=1}^{n} w_i(x_i) g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l})\, dx_s = 0 \quad (s \in \{i_1, i_2,..., i_l\}) \quad (64)$$

$$\int_{K^n}\prod_{i=1}^{n} w_i(x_i) g^2_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l})\, d\mathbf{x} = 1 \quad (65)$$

$$\int_{K^n}\prod_{i=1}^{n} w_i(x_i) g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l}) g_{j_1j_2...j_k}(x_{j_1}, x_{j_2},..., x_{j_k})\, d\mathbf{x} = 0 \quad (\{i_1, i_2,..., i_l\} \neq \{j_1, j_2,..., j_k\}) \quad (66)$$

The only restriction is that any two functions in $\{g\}$ cannot have the same set of variables. For instance, if $\{g\}$ contains the function

$$g_{ijk}(x_i, x_j, x_k) = \frac{f_i(x_i) f_{jk}(x_j, x_k)}{||f_i(x_i)||||f_{jk}(x_j, x_k)||} \quad (67)$$

it cannot also have

$$g_{ijk}(x_i, x_j, x_k) = \frac{f_i(x_i) f_j(x_j) f_k(x_k)}{||f_i(x_i)||||f_j(x_j)||||f_k(x_k)||} \quad (68)$$

and vice versa, because the two functions are not weighted orthogonal functions.

The set of functions $\{g\}$ may be used as a basis to expand $f(\mathbf{x})$:

$$f(\mathbf{x}) \approx \alpha_0 g_0 + \sum_{i=1}^{n}\alpha_i g_i(x_i) + \sum_{1\le i<j\le n}\alpha_{ij} g_{ij}(x_i, x_j) + \cdots + \sum_{1\le i_1<\cdots<i_l\le n}\alpha_{i_1i_2...i_l} g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l}) + \cdots + \alpha_{12...n} g_{12...n}(x_1, x_2,..., x_n) \quad (69)$$

where $\{\alpha\}$ are constant coefficients, which can be obtained using the weighted orthonormality property of $\{g\}$, i.e.,

$$\alpha_{i_1i_2...i_l} = \int_{K^n}\prod_{i=1}^{n} w_i(x_i) f(\mathbf{x}) g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l})\, d\mathbf{x} \approx \frac{1}{N}\sum_{s=1}^{N} f(\mathbf{x}^{(s)}) g_{i_1i_2...i_l}(x_{i_1}^{(s)}, x_{i_2}^{(s)},..., x_{i_l}^{(s)}) \quad (\text{for } l = 0, 1,..., n) \quad (70)$$

Because $\{g\}$ is not a complete set, the expansion of an n-variate function $f(\mathbf{x})$ in this basis is likely only an approximation. Considering that

$$\alpha_0 = ||f_0|| \quad (71)$$

$$\alpha_i = ||f_i(x_i)|| \quad (\text{for } i = 1, 2,..., n) \quad (72)$$

$$\alpha_{ij} = ||f_{ij}(x_i, x_j)|| \quad (\text{for } 1 \le i < j \le n) \quad (73)$$

eq 70 becomes

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1\le i<j\le n} f_{ij}(x_i, x_j) + \cdots + \sum_{1\le i_1<\cdots<i_l\le n}\alpha_{i_1i_2...i_l} g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l}) + \cdots + \alpha_{12...n} g_{12...n}(x_1, x_2,..., x_n) \quad (74)$$

which implies the following approximations, upon comparison with eq 1:

$$f_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l}) \approx \alpha_{i_1i_2...i_l} g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l}) \quad (\text{for } l = 3, 4,..., n) \quad (75)$$

If only the normalized first order RS−HDMR component functions are used, we have

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1\le i<j\le n}\alpha_{ij} g_{ij}(x_i, x_j) + \cdots + \sum_{1\le i_1<\cdots<i_l\le n}\alpha_{i_1i_2...i_l} g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l}) + \cdots + \alpha_{12...n} g_{12...n}(x_1, x_2,..., x_n) \quad (76)$$

where

$$g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l}) = \prod_{s=1}^{l} g_{i_s} = \prod_{s=1}^{l}\frac{f_{i_s}(x_{i_s})}{||f_{i_s}(x_{i_s})||} \quad (\text{for } l = 2, 3,..., n) \quad (77)$$

Notice that all $g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l})$ ($l \ge 2$ or 3) are products of normalized first and second order RS−HDMR component functions. Hence, eqs 74 and 76 only require computation of the first and second order RS−HDMR component functions, which can be accurately represented by weighted orthonormal polynomial expansions, with the help of the correlation method. The high order component functions are no longer explicitly computed. The coefficient $\alpha_{i_1i_2...i_l}$ for $g_{i_1i_2...i_l}(x_{i_1}, x_{i_2},..., x_{i_l})$ is determined by eq 70, whose Monte Carlo integration approximation still involves the product of one and two variable functions. At first sight, the sampling effort and achieved accuracy might appear similar to the $\{\alpha_r^i, \beta_{pq}^{ij}, \gamma_{pqr}^{ijk},\cdots\}$ set for

the weighted orthonormal polynomial $\{\varphi\}$ expansions given by eqs 37−39. However, the $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ are obtained by averaging $f(\boldsymbol{x})$ over K^{n-1} and K^{n-2}, with respect to all x_k except x_i and x_i, x_j, respectively. This basis reflects the behavior of $f(\boldsymbol{x})$. In contrast, $\{\varphi\}$ is a fixed set of functions, without specific knowledge of the particular n-variate function $f(\boldsymbol{x})$ of interest; therefore, the variance of $\prod_{k=1}^{n} w_k(x_k) f(\boldsymbol{x}) \prod_l \varphi_l^i(x_i)$ should be larger than that for $\prod_{k=1}^{n} w_k(x_k) f(\boldsymbol{x}) g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})$. Therefore, $\{g\}$ is a natural basis for $f(\boldsymbol{x})$, and better accuracy should be attained, compared to $\{\varphi\}$, using the Monte Carlo integration approximation with a given sample size. Moreover, compared to the $\{\varphi\}$ expansion given by eq 36, the $\{g\}$ expansions given by eqs 74 and 76 have many fewer terms and, consequently, less unknown coefficients to be determined by Monte Carlo integration. For instance, the third order component functions $f_{ijk}(x_i, x_j, x_k)$ have 27 terms in eq 36 if the third order weighted orthonormal polynomial approximation is used. In contrast, eqs 74 and 76 have only one term: $g_{ijk}(x_i, x_j, x_k)$. The *reduction* in the number of terms to calculate with the $\{g\}$ expansion increases exponentially with the order l of RS−HDMR component functions. The reduction in the number of Monte Carlo integrations with the basis $\{g\}$ should also reduce the computational error.

Similarly, the correlation method can be used to reduce the variance of $\prod_{k=1}^{n} w_k(x_k) f(\boldsymbol{x}) g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})$ in the determination of $\alpha_{i_1 i_2 \ldots i_l}$, where the truncated expansions in eqs 74 and 76 are used as reference functions $h(\boldsymbol{x})$.

3. Application to an Integrated Exposure and Dose Model

3.1. Model Description. An integrated exposure and dose model has been developed to study multiroute residential human exposures to trichloroethylene (TCE) that is present in tap water. It incorporates dynamic microenvironmental and pharmacokinetic models, which consider the release of TCE from water into air within different rooms in the home, the activities of individuals, and the physiological uptake processes for three exposure routes (ingestion, inhalation, and dermal absorption).

Microenvironmental modeling[13] is used to quantify the levels of concentrations in different media (air, water, etc.) that are coming in contact with human receptors. Microenvironments have typically been defined as individual or aggregate locations where a homogeneous concentration of the pollutant is encountered, such as bedrooms and bathrooms in a residence. The daily exposure for an individual is the sum of the exposures in all microenvironments encountered within a day.

The governing mass-balance equation for a microenvironmental compartment is formulated as follows:

$$V_i \frac{\mathrm{d}C_i}{\mathrm{d}t} = \sum_{j=1}^{n} q_{j,i} C_j - \sum_{j=1}^{n} q_{i,j} C_i + \sum \text{sources} - \sum \text{sinks} \quad (78)$$

where C_i is the TCE concentration in air at compartment i (in units of μg/m^3), V_i is the volume of the compartment i (in cubic meters), t is the time (in minutes), $q_{i,j}$ is the volumetric air flow rate from compartment j to compartment i (in units of m^3/min), i and j are compartment numbers, and n is the total number of compartments. The term "sources" represents the contaminant generation mechanisms present in the compartment, whereas "sinks" represents the contaminant removal mechanisms that are present in the compartment.

The source strength for TCE through volatilization during showers is determined by the mass-transfer rate between water and air.[14−16] On the basis of a plug flow model of the water stream with the "pseudo-steady-state" assumption, the source strength of volatile TCE emissions from a shower is derived as

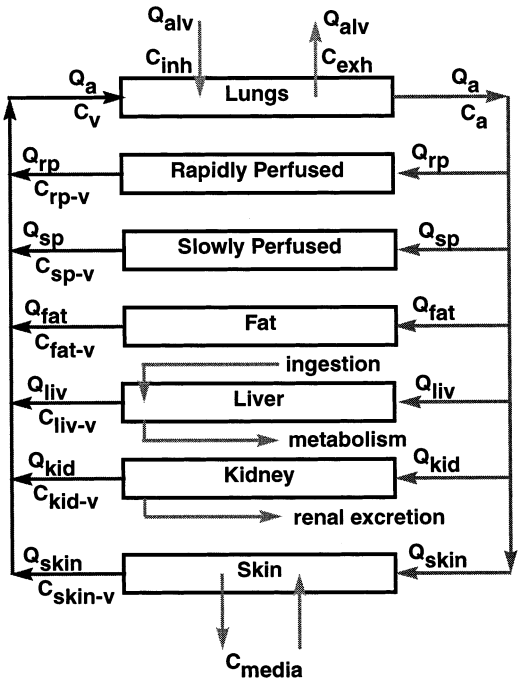

Figure 2. Schematic diagram of a general basic physiologically based pharmacokinetic (PBPK) model for volatile organic compounds.[13]

$$S = K_V\left(C_{\mathrm{in}} - \frac{y}{H}\right) \quad (79)$$

where S is the source strength (mass/time), K_V the volatilization coefficient (which depends on the mass-transfer coefficient and shower flow rate), C_{in} the concentration in water, y the concentration in air, and H the Henry's Law constant.

The calculated microenvironmental air concentration−time profiles of the pollutant are then used as inputs to pharmacokinetic models. Pharmacokinetic models are mathematical constructs that are used to calculate the concentrations or amounts of chemicals in body tissue and fluids as a function of time.[13,17] Physiologically based ("mechanistic") pharmacokinetic (PBPK) models typically represent the biological organism as a set of physiological compartments by lumping together similar tissues, and by describing transport between compartments on the basis of actual processes, such as blood circulation. The basic structure of a typical PBPK model describing the transport and metabolism of a volatile organic compound, such as TCE, in the body is shown schematically in Figure 2. There are potentially three major natural routes by which the chemical enters the body: (a) inhalation, (b) ingestion, and (c) dermal absorption.

The primary means of transport for xenobiotic chemicals that enter the body through one of more of these routes is via blood, which is the main vehicle for nutrient supply and waste removal from tissues. Xenobiotic chemicals in blood generally partition into free and protein-bound fractions, and the free fraction drives the transport of the chemical into tissue for the usual case of passive transport. In the basic PBPK model, transport of chemicals between blood and tissue is assumed to be flow-limited, which implies that the transport barriers between the free molecules in blood and tissue are negligible, and equilibration between free and bound fractions in blood and tissue is rapid. The chemical concentrations in venous blood exiting a tissue (tissue concentrations are assumed to be at equilibrium)

TABLE 1: Input Variable Ranges and Parameter μ

input	range: lower bound	range: upper bound	μ
age, x_1 (yr)	15	80	
TCE concentration in tap water, x_2 (ppm)	0.001	0.5	
bathroom volume, x_3 (m^3)	9	15	
drinking-water consumption rate, x_4 (L/day)	0.8	2.4	
shower flow rate, x_5 (L/min)	7.7	38.3	18.3
shower time, x_6 (min)	5	30	10
time after shower in bathroom, x_7 (min)	5	30	10

and in the tissue are assumed to be homogeneous, with respect to the concentration of the chemical. These assumptions enable tissue concentrations of the chemical to be described by ordinary differential equations (ODEs), similar to that of a continuously stirred tank reactor.

A mass balance around the equilibrium lung compartment results in

$$c_{\text{arterial}} = \frac{Q_{\text{cardiac}} c_{\text{venous}} + Q_{\text{alveolar}} C_{\text{air(inhaled)}}}{Q_{\text{cardiac}} + Q_{\text{alveolar}}/P_{\text{blood/air}}} \tag{80}$$

where

$$c_{\text{venous}} = \frac{1}{Q_{\text{cardiac}}} \sum_{j=1}^{n} Q_j C_j \tag{81}$$

c_{arterial} is the arterial blood concentration, and c_{venous} is the venous blood concentration. $C_{\text{air(inhaled)}}$ represents the inhaled air concentration, Q_{cardiac} is the cardiac blood flow rate, Q_{alveolar} is the alveolar blood flow rate, $P_{\text{blood/air}}$ is the partition coefficient between air and blood, Q_j is the blood flow rate in compartment j, and C_j is the concentration in compartment j.

The mass balance for any compartment j of volume V_j in the PBPK model, other than the viable skin and stratum corneum compartments, is given by

$$V_j \frac{dC_j}{dt} = Q_j \left(c_{\text{arterial}} - \frac{C_j}{P_{j/\text{blood}}} \right) - R_j \tag{82}$$

where V_j is the volume in compartment j, $P_{j/\text{blood}}$ is the partition coefficient between compartment j and blood, and R_j is the metabolism reaction rate in compartment j.

3.2. Nonuniform RS−HDMR Application. The nonuniform RS−HDMR methodology is used to construct the truncated RS−HDMR expansion as an efficient fully equivalent operational model (FEOM) for the previously described integrated exposure model, to relieve the computational burden of complex mechanistic modeling. The operations of the FEOM are very fast, because they only involve algebraic manipulations. The accuracy of the FEOM depends on the order of the truncated RS−HDMR expansion and the number of sampling points for constructing it, which is explored in the following.

Seven input variables are selected from the integrated exposure and dose model for TCE to construct the FEOM through the RS−HDMR with the nonuniform distribution method. The ranges of variation of these seven input variables are shown in Table 1.

The first four input variables—x_1, x_2, x_3, and x_4—have a uniform distribution. The last three input variables—x_5, x_6, and x_7—have a triangular distribution. The target model output is the total body burden of TCE accumulated after one month of continuous exposure via inhalation, ingestion, and dermal contact. The exposure pathways of inhalation and dermal contact

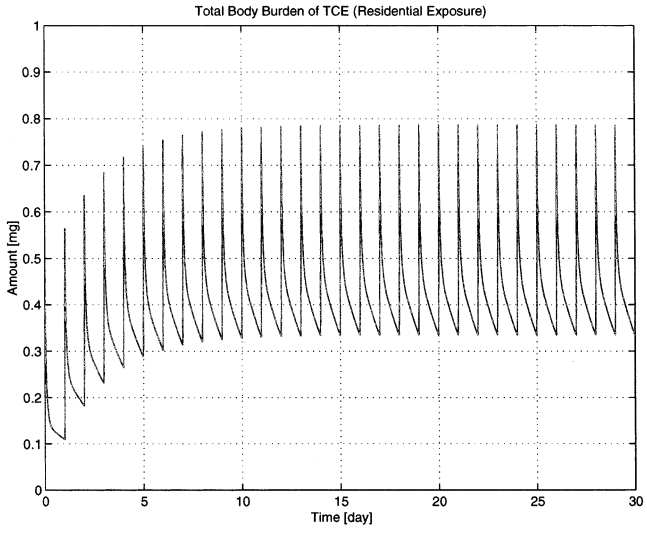

Figure 3. Total body burden predicted by the trichloroethylene (TCE) microenvironmental/PBPK model for one month of residential exposure.

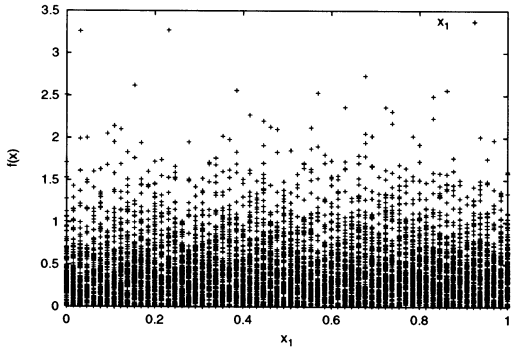

Figure 4. Data distribution with respect to the uniform distributed variable x_1 (normalized).

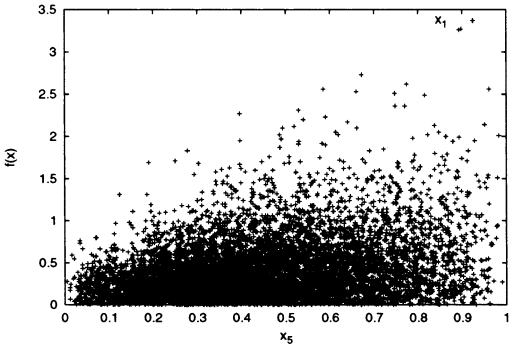

Figure 5. Data distribution with respect to the triangular distributed variable x_5 (normalized).

are mainly due to showering activities. The reason for choosing the total body burden of TCE after one month of exposure as the target model output ($f(\mathbf{x})$) for mapping with the seven inputs is that this amount reaches steady-state and it can be used to assess the health risk (see the output of the integrated exposure and dose model simulation in Figure 3 for a hypothetical case study).

Ten thousand random samples of $\mathbf{x}$, and their corresponding values of $f(\mathbf{x})$, were obtained from the model, according to the pdf $w_i(x_i)$. Figures 4 and 5 give examples of the data distribution, with respect to two different distributed input variables x_1 and x_5. Notice that x_1 is a discrete variable. The uniform and triangular distributions can be observed to have a distinct influence, and most data have values of $f(\mathbf{x}) < 0.5$.

The correlation method with bases $\{\varphi\}$ and $\{g\}$ was used to construct the RS−HDMR component functions for different sample sizes (500, 1000, 3000, 5000, and 10 000). The accuracy of different order RS−HDMR expansions whose component functions were obtained from $\{\varphi\}$ and $\{g\}$ with different sample sizes was then tested by comparison to the 10 000 exact data. The accuracy was represented by the portion of 10 000 data with relative errors not larger than 5%, 10%, and 20%.

3.3. Correlation Method with Basis $\{\varphi\}$. First, the weighted orthonormal polynomials $\{\varphi\}$ were used as a basis to approximate the RS−HDMR component functions given by eqs 15−17. The correlation method (eqs 52−54, 57) was used to determine the coefficients $\{\alpha_r^i, \beta_{pq}^{ij}, \gamma_{pqr}^{ijk}\}$. Different $h(\boldsymbol{x})$ and sample sizes were used to determine the coefficients. When $h(\boldsymbol{x})$ is the second order RS−HDMR expansion (eq 55), with the third order weighted orthonormal polynomial approximations (i.e., $k, l, l' = 3$), the iteration of the determination of first to third order RS−HDMR component functions was convergent only for the sample size $N = 10\,000$. When $h(\boldsymbol{x})$ is the second order RS−HDMR expansion with the second order weighted orthonormal polynomial approximations (i.e., $k, l, l' = 2$), the determination of first to third order RS−HDMR component functions was convergent only for the sample sizes 3000, 5000, and 10 000. For both choices of $h(\boldsymbol{x})$, the iteration-determining higher (greater than third) order RS−HDMR component functions were divergent for all the previously discussed sample sizes.

The accuracy of the resultant convergent first to third order RS−HDMR expansions are unsatisfactory for a special reason. For the third order RS−HDMR expansion, only ∼60% data have relative errors that are not larger than 5%. Observing Figures 4 and 5, one can see that many of the data have very small values of $f(\boldsymbol{x})$. For these points, even if the absolute errors of the RS−HDMR approximation are quite small, their relative errors can be very large. In this case, relative error does not give useful information. We set a threshold of 0.3 mg for $f(\boldsymbol{x})$. When the value of a datum is not larger than the threshold and the absolute value calculated by the RS−HDMR approximation is not larger than the threshold, we define the RS−HDMR approximation as giving the correct answer. The threshold dose value of 0.3 mg is chosen because the corresponding TCE concentration in water is less than the threshold concentration level used in the animal studies in the literature. Therefore, we added the portion of the data satisfying this condition to the data whose values are larger than the threshold and have relative errors of 5%, 10%, and 20% as a representation of the accuracy. The results are given in Table 2, which shows that the RS−HDMR approximation with the basis $\{\varphi\}$ is quite satisfactory. For the third order RS−HDMR expansion, ∼90% of the data have relative errors that are not larger than 5% or both their values and absolute calculated values given by RS−HDMR approximation are not larger than the threshold.

3.4. Correlation Method with Basis $\{g\}$. When the weighted orthonormal polynomials $\{\varphi\}$ were used as a basis, the iterative determination of the RS−HDMR component functions was divergent for sample sizes 500 and 1000. This is because the second order function $f_{ij}(x_i, x_j)$ cannot be accurately determined for small sample sizes with the basis $\{\varphi\}$. The first order function $f_i(x_i)$ can be accurately determined for a few hundred samples with the basis $\{\varphi\}$; therefore, we can use normalized $f_i(x_i)$, i.e., $g_i(x_i)$, as a basis to construct high order terms for small sample sizes. Different order truncation of eq 76 was used for the approximation, and the truncated second order expansion of eq 76 was used as $h(\boldsymbol{x})$. When the sample size satisfied $N \geq 3000$, the third order weighted orthonormal polynomial expansion was used to approximate $f_i(x_i)$. Otherwise, the second order weighted orthonormal polynomial expansion was used. The iterative determination of the coefficients $\alpha_{i_1 i_2 \ldots i_l}$ for all order terms was convergent for $N = 500$, 1000, 3000, 5000, and 10 000. Compared to the results given by the basis $\{\varphi\}$, the accuracy arising from $\{g\}$ composed of normalized $f_i(x_i)$ is worse for $N \geq 3000$; however, this basis can be used for $N < 3000$, whereas $\{\varphi\}$ cannot be used in this case.

TABLE 2: Relative Errors of the Different-Order RS−HDMR Expansions with the Basis $\{\varphi\}$ (Truncated eq 36) Obtained by the Correlation Method of Monte Carlo Integration with Different Second-Order $h(x)$ and Sample Sizes N (Threshold = 0.3)

	data portion (%)[a]		
relative error (%)	first order	second order	third order
	$k, l, l' = 3$; $N = 10\,000$		
5	46.2	83.5	86.7
10	57.1	94.6	94.6
20	73.3	97.9	97.1
	$k, l, l' = 2$; $N = 3000$		
5	46.5	82.8	86.3
10	57.5	94.0	95.3
20	73.6	97.8	97.8
	$k, l, l' = 2$; $N = 5000$		
5	46.5	83.1	87.9
10	57.3	94.1	95.9
20	73.5	97.7	97.8
	$k, l, l' = 2$; $N = 10\,000$		
5	46.2	83.9	90.7
10	57.2	94.5	96.8
20	73.4	97.8	98.1

[a] The percentage of 10 000 data whose values are >0.3 and relative errors are not larger than a given value, and those whose values and absolute calculated values of RS−HDMR approximation both are ≤0.3.

When the sample size satisfied $N \geq 3000$, the second order RS−HDMR component functions $f_{ij}(x_i, x_j)$ can be accurately determined. Therefore, normalized functions $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ can be used to compose the basis $\{g\}$ and truncation of eq 74 can be used for an approximation. After $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ were accurately determined by the correlation method (see Table 2), the coefficients $\alpha_{i_1 i_2 \ldots i_l}$ ($l \geq 3$) in eq 74 were determined by the correlation method, with $h(\boldsymbol{x})$ being the second order RS−HDMR expansion whose component functions are approximated by the second order weighted orthonormal polynomial expansion. All $g_{i_1 i_2 \ldots i_l}(x_{i_1}, x_{i_2}, \ldots, x_{i_l})$ ($l \geq 3$) are the products of the normalized $f_i(x_i)$ function given in eq 77. The results are given in Table 3. The high (greater than third) order terms are negligible; therefore, only the first to third order approximations are presented. Compared to Table 2, the results in Table 3 are quite satisfactory.

4. Conclusions

In the present paper, the uniform random sampling−high dimensional model representation (RS−HDMR) technique is extended to nonuniform distributions. After giving definitions of the RS−HDMR component functions and introducing weighted orthonormal polynomials related to the probability density function (pdf) of the input variables, all the formulas of the Monte Carlo integration approximation for uniform RS−HDMR can be used for nonuniform RS−HDMR. The pdf $w_i(x_i)$ reflects the random sampling of data and does not explicitly appear in the formulas. The correlation method, to reduce the variance of the Monte Carlo integration, and the low order term

TABLE 3: Relative Errors of the Different-Order RS−HDMR Expansions with the Basis $\{g\}$ (Truncated eq 74) Obtained by the Correlation Method of Monte Carlo Integration with Second-Order $h(x)$ and Different Sample Sizes N (Threshold = 0.3)

	data portion (%)[a]		
relative error (%)	first order	second order	third order
	$N = 3000$		
5	46.5	82.8	91.9
10	57.5	94.0	97.8
20	73.6	97.8	98.5
	$N = 5000$		
5	46.5	83.1	92.9
10	57.3	94.1	97.9
20	73.5	97.7	98.4
	$N = 10\,000$		
5	46.2	83.9	94.6
10	57.2	94.5	98.3
20	73.4	97.8	98.4

[a] The percentage of 10 000 data whose values are >0.3 and relative errors are not larger than a given value, and those whose values and absolute calculated values of RS−HDMR approximation both are ≤0.3.

product RS−HDMR (lp-RS−HDMR) method, to represent the high order terms by lower order terms in uniform RS−HDMR, can be also applied for nonuniform RS−HDMR. Thus, uniform RS−HDMR is only a special case of this general treatment. In practice, the data are often nonuniformly distributed, and the general treatment of RS−HDMR presented in this paper should be useful for realistic problems.

An integrated environmental exposure and dose model was used for illustrating the development of the nonuniform RS−HDMR expansions. To account for the physiological and demographic effects on the calculated dose output, seven input variables with pdfs of uniform and triangular shape were chosen to perform RS−HDMR mapping with the dose outputs. The generated RS−HDMR expansions have good accuracies for predicting the target dose outputs. Furthermore, the evaluations of the RS−HDMR expansions are very fast, because they only involve algebraic manipulations. Therefore, the RS−HDMR expansions can be used as a fully equivalent operational model (FEOM) to relieve the computational burden of the original mechanistic models. The creation of the FEOM is particularly useful in performing population exposure assessments, which are not computationally feasible by normal means, because large numbers of simulations employing the mechanistic models are often required to account for demographic and physiological variability within a population. The resultant FEOMs can then be used as accurate and efficient alternatives to the original complex mechanistic models. The RS−HDMR methodology provides an efficient route to treat this problem.

Acknowledgment. The authors acknowledge support from the Petroleum Research Fund of the American Chemical Society and the U.S. Environmental Protection Agency, under Cooperative Agreement No. EPAR-827033 to Environmental and Occupation Health Sciences Institute.

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. A; Ip, P. C.; Rabitz, H. An Efficient Chemical Kinetics Solver Using High Dimensional Model Representation. *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) Li, G.; Wang, S.-W.; Rabitz, H. Practical Approaches To Construct RS−HDMR Component Functions. *J. Phys. Chem. A* **2002**, *106*, 8721−8733.

(8) 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.

(9) Li, G.; Wang, S. W.; Rabitz, H.; Wang, S. K.; Jaffe, P. Global Uncertainty Assessments of a Mathematical Model for Trace Metal Bioremediation. *Chem. Eng. Sci.* **2002**, *57*, 4445−4460.

(10) 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.

(11) Kalos, M. H.; Whitlock, P. A. *Monte Carlo Methods. Vol. I: Basic*; John Wiley & Sons: New York, 1986.

(12) Press, W. H.; Teukolsky, S. A.; Vetterling, W. T.; Flannery, B. P. *Numerical Recipes in FORTRAN*; Cambridge University Press: New York, 1992; p 277, 299−319.

(13) Georgopoulos, P. G.; Walia, A.; Roy, A.; Lioy, P. J. An Integrated Exposure and Dose Modeling and Analysis System—Part I: Formulation and Testing of Microenvironmental and Pharmacokinetic Components. *Environ. Sci. Technol.* **1997**, *31*, 17−27.

(14) Little, J. C. Applying the Two-Resistance Theory to Contaminant Volatilization in Showers. *Environ. Sci. Technol.* **1992**, *26*, 1341−1349.

(15) Giardino, N. J.; Esmen, N. A.; Andelman, J. B.; Wilkes, C. R.; Davidson, C. I.; Small, M. J. Shower Volatilization Exposures in Homes Using Tap Water Contaminated with Trichloroethylene. *J. Exposure Anal. Environ. Epidemiol.* **1992**, *1 (Suppl.)*, 147−158.

(16) Wilkes, C. R.; Small, M. J.; Davidson, C. I.; Andelman, J. B. Modeling the Effects of Water Usage and Co-Behavior on Inhalation Exposures to Contaminants Volatilized from Household Water. *J. Exposure Anal. Environ. Epidemiol.* **1996**, *6*, 393−412.

(17) Menzel, D. B. Physiological Pharmacokinetic Modeling. *Environ. Sci. Technol.* **1997**, *21*, 944.