J Math Model Algor (2013) 12:265–276
DOI 10.1007/s10852-012-9211-7

Bound Analysis Through HDMR for Multivariate Data Modelling - CMMSE

M. Alper Tunga · Metin Demiralp

Received: 9 August 2012 / Accepted: 17 August 2012 / Published online: 7 September 2012

Abstract Multivariate data modelling problems consist of a number of nodes with associated function (class) values. The main purpose of these problems is to construct an analytical model to represent the characteristics of the problem under consideration. Because the devices, tools, and/or algorithms used to collect the data may have incapabilities or limited capabilities, the data set is likely to contain unavoidable errors. That is, each component of data is reliable only within an interval which contains the data value. To this end, when an analytical structure is needed for the given data, a band structure should be determined instead of a unique structure. As the multivariance of the given data set increases, divide–and–conquer methods become important in multivariate modelling problems. HDMR based methods allow us to partition the given multivariate data into less variate data sets to reduce the complexity of the given problem. This paper focuses on Interval Factorized HDMR method developed to determine an approximate band structure for a given multivariate data modelling problem having uncertainties on its nodes and function values.

Keywords High dimensional model representation · Multivariate functions · Interpolation · Multidimensional problems · Bound analysis · Approximation

M. A. Tunga (✉)
Faculty of Engineering, Software Engineering Department,
Bahçeşehir University, Beşiktaş 34349, İstanbul, Turkey
e-mail: alper.tunga@bahcesehir.edu.tr

M. Demiralp
Informatics Institute, Computational Science and Engineering Program,
İstanbul Technical University, Maslak 34469, İstanbul, Turkey
e-mail: metin.demiralp@gmail.com

1 Introduction

There may exist uncertainties on the location of the nodes and the function values at these nodes in some engineering and scientific problems. When the given problem is a multivariate data modelling problem and that problem has such uncertainties, a band structure construction is needed instead of a unique analytical structure. Hence, the methods used in modelling multivariate data should be reorganized due to this fact.

Dealing with many less variate data sets instead of a single multivariate date set as the training source, that is, learning system to construct an analytical model for the given problem is a good way in modelling purpose. High Dimensional Model Representation (HDMR) is a divide-and-conquer method and can be used as a representation technique of multivariate structures [6]. To this end, HDMR philosophy can be used as a tool for representing the characteristics of a given data modelling problem [9]. Generalized HDMR method is based on HDMR philosophy and is a method to partition the given multivariate data set into less variate data sets and to determine an analytical structure through these partitioned data sets as the model of the given problem [7].

In literature, we have Generalized HDMR to partition the given multivariate data set in which we assume that no construction errors appear [7]. However, when we need to take the uncertainties on data into consideration, that is, the given data has errors in its structure, a new HDMR based method is needed to obtain a band structure instead of a unique analytical model. The Interval Generalized HDMR method was developed for this purpose [10]. This method is the reorganized version of Generalized HDMR for the problems in which we need to generate band structure as the analytical model. The numerical results show us that this method works well for dominantly and purely additive natures. On the other hand, the representations obtained through the Interval Generalized HDMR method are insufficient for dominantly and purely multiplicative natures. The mentioned numerical results urge us to develop a new HDMR based method to overcome this disadvantage. This work aims to develop this new HDMR based method for obtaining acceptable band structures for these types of problems. We know from literature that the Factorized HDMR method [8] is a method to increase the performance of Generalized HDMR method in modelling problems having multiplicative nature. Taking this case into consideration, our new method is based on this Factorized HDMR philosophy. The name of this new method is Interval Factorized HDMR and includes the important features of the classical Factorized HDMR to build efficient band structure for the considered problems of this work.

HDMR method is also used by many scientists in various research areas and either the method is applied to several problems or new HDMR based methods are developed [1–3, 5, 11].

This paper is organized as follows. The second section covers the related mathematical background needed to develop our new method. The details of the new method are given in the third section. The fourth section includes a number of numerical implementations to examine the performance of Interval Factorized HDMR while the concluding remarks are discussed in the last section of this paper.

2 Mathematical Background

This section covers the multivariate data partitioning methods that were previously developed and are used in constructing the Interval Factorized HDMR method for bound analysis in multivariate data modelling problems. These methods are Plain HDMR [9] and Interval Generalized HDMR [10]. In addition, an interpolation technique including the bound analysis philosophy is given in the last sub-section.

2.1 The HDMR Method

The High Dimensional Model Representation (HDMR) method has the following finite expansion to express a given multivariate function in terms of less variate functions

$$f(x_1,\ldots,x_N) = f_0 + \sum_{i_1=1}^{N} f_{i_1}(x_{i_1}) + \sum_{\substack{i_1,i_2=1\\ i_1<i_2}}^{N} f_{i_1 i_2}(x_{i_1},x_{i_2}) + \cdots + f_{1\ldots N}(x_1,\ldots,x_N) \quad (1)$$

where N is the number of independent variables of the given function [6]. The main aim in this method is to uniquely determine the structure of each right hand side component of the expansion given in Eq. 1.

To determine the general structure of the constant HDMR component, f_0, the following operator is used

$$\mathcal{I}_0 F(x_1,\ldots,x_N) \equiv \int_{a_1}^{b_1} dx_1 \cdots \int_{a_N}^{b_N} dx_N W(x_1,\ldots,x_N) F(x_1,\ldots,x_N) \quad (2)$$

where $W(x_1,\ldots,x_N)$ is a product type weight and $F(x_1,\ldots,x_N)$ is a square integrable arbitrary function [9]. The operator, $\mathcal{I}_{i_1}$, is defined to obtain the structure of the univariate HDMR components, $f_{i_1}(x_{i_1})$ while $1 \le i_1 \le N$.

$$\begin{aligned}\mathcal{I}_{i_1} F(x_1,\ldots,x_N) \equiv{}& \int_{a_1}^{b_1} dx_1 W_1(x_1) \cdots \int_{a_{i_1-1}}^{b_{i_1-1}} dx_{i_1-1} W_{i_1-1}(x_{i_1-1}) \\ &\times \int_{a_{i_1+1}}^{b_{i_1+1}} dx_{i_1+1} W_{i_1+1}(x_{i_1+1}) \cdots \int_{a_N}^{b_N} dx_N W_N(x_N) F(x_1,\ldots,x_N)\end{aligned} \quad (3)$$

The weight function, $W(x_1,\ldots,x_N)$, appearing in the above conditions are assumed to be a product type weight and can be defined with normalization criteria as

$$W(x_1,\ldots,x_N) \equiv \prod_{j=1}^{N} W_j(x_j), \qquad \int_{a_j}^{b_j} dx_j W_j(x_j) = 1, \quad x_j \in [a_j, b_j] \quad (4)$$

The following vanishing conditions are used when we apply the above mentioned operators to both sides of the HDMR expansion to uniquely determine the HDMR components of the multivariate function under consideration.

$$\int_{a_1}^{b_1} dx_1 \cdots \int_{a_N}^{b_N} dx_N W(x_1, \ldots, x_N) f_i(x_i) = 0, \quad 1 \le i \le N \tag{5}$$

while these conditions correspond to the following orthogonality conditions

$$\int_{a_1}^{b_1} dx_1 W_1(x_1) \cdots \int_{a_N}^{b_N} dx_N W_N(x_N) f_{i_1 \ldots i_s}(x_{i_1}, \ldots, x_{i_s}) \\ \times f_{j_1 \ldots j_\ell}(x_{j_1}, \ldots, x_{j_\ell}) = 0, \quad (s \neq \ell) \vee \left[(i_1 \neq j_1) \vee \ldots \vee (i_s \neq j_s)\right] \tag{6}$$

When the operators, $\mathcal{I}_0$ and $\mathcal{I}_{i_1}$ are applied to the both sides of the Eq. 1 respectively, the constant and the univariate HDMR components are obtained as [9]

$$f_0 = \mathcal{I}_0 f(x_1, \ldots, x_N) \tag{7}$$

$$f_i(x_i) = \mathcal{I}_i f(x_1, \ldots, x_N) - f_0, \quad 1 \le i \le N \tag{8}$$

The higher variate HDMR components can be determined in the same manner. Because the aim of this study is to deal with at most the univariate approximation, we do not give the details of these components.

2.2 The Generalized HDMR Method

The HDMR method uses a product type weight in representing a given multivariate function by using the HDMR expansion. When we deal with multivariate data partitioning, it is obvious that we should know the function values at all possible nodes of the problem domain to obtain less variate data sets which allow us to represent the whole domain through the HDMR expansion because of product type weight need in the HDMR algorithm. In general, we can know the function values at only a small number of nodes of the problem domain. This urges us to use a non-product weight in HDMR. The Generalized HDMR method was developed for this purpose [7].

The first step in Generalized HDMR is to write the HDMR expansion of a general weight function as

$$W(x_1, \ldots, x_N) = W_0 + \sum_{i_1=1}^{N} W_{i_1}(x_{i_1}) + \cdots + W_{1\ldots N}(x_1, \ldots, x_N) \tag{9}$$

To determine the structure of each HDMR component of the general weight function, a product type auxiliary weight is needed

$$\Omega(x_1, \ldots, x_N) \equiv \prod_{j=1}^{N} \Omega_j(x_j), \qquad \int_{a_j}^{b_j} dx_j \Omega(x_j) = 1, \quad 1 \le j \le N \tag{10}$$

and the vanishing conditions under this auxiliary weight to determine the HDMR components of the general weight are defined as follows

$$\int_{a_{i_j}}^{b_{i_j}} dx_{i_j} \Omega_{i_j}(x_{i_j}) W_{i_1 \ldots i_k}\left(x_{i_1}, \ldots, x_{i_k}\right) = 0, \quad 1 \le i_j \le i_k \tag{11}$$

while the following relation is defined as the vanishing conditions for the determination of the multivariate function components [7]

$$\int_{a_1}^{b_1} dx_1 \cdots \int_{a_N}^{b_N} dx_N \Omega(x_1, \ldots, x_N) W(x_1, \ldots, x_N) f_i(x_i) = 0, \quad 1 \le i \le N \tag{12}$$

When we apply the operator given in Eq. 2 under the auxiliary weight function, $\Omega(x_1, \ldots, x_N)$ to the product of HDMR expansions of $W(x_1, \ldots, x_N)$ and $f(x_1, \ldots, x_N)$, the following relation is obtained as the constant Generalized HDMR component of the given multivariate function [7]

$$f_0 = \mathcal{I}_0 \left[W(x_1, \ldots, x_N) f(x_1, \ldots, x_N) \right] \tag{13}$$

where

$$W_0 = 1 \tag{14}$$

The general structure of the univariate Generalized HDMR components are obtained as follows in the same manner [7]

$$\begin{aligned}
&\mathcal{I}_i \left[W(x_1, \ldots, x_N) f(x_1, \ldots, x_N) \right] \\
&\quad = (1 + W_i(x_i))\, f_0 + (1 + W_i(x_i))\, f_i(x_i) \\
&\qquad + (1 + W_i(x_i)) \sum_{\substack{i_1=1 \\ i_1 \ne i}}^{N} \int_{a_{i_1}}^{b_{i_1}} dx_{i_1} \Omega_{i_1}(x_{i_1}) \left(1 + W_{i_1}(x_{i_1})\right) f_{i_1}(x_{i_1}) \\
&\qquad + \sum_{\substack{i_1, i_2 = 1, i_1 < i_2 \\ (i_1 = i) \vee (i_2 = i)}}^{N} \int_{(1-\delta_{i_1 i}) a_{i_1} + \delta_{i_1 i} a_{i_2}}^{(1-\delta_{i_1 i}) b_{i_1} + \delta_{i_1 i} b_{i_2}} \left[\left(1 - \delta_{i_1 i}\right) dx_{i_1} \Omega_{i_1}(x_{i_1}) + \delta_{i_1 i} dx_{i_2} \Omega_{i_2}(x_{i_2}) \right] \\
&\qquad \times \left[W_{i_1 i_2}(x_{i_1}, x_{i_2}) - W_{i_1}(x_{i_1}) W_{i_2}(x_{i_2}) \right] \left[\left(1 - \delta_{i_1 i}\right) f_{i_1}(x_{i_1}) + \delta_{i_1 i} f_{i_2}(x_{i_2}) \right]
\end{aligned} \tag{15}$$

where $\delta_{i_1 i}$ stands for Kronecker's Delta and $1 \le i \le N$. This relation is a set of integral equations whose unknowns are the univariate components of the Generalized HDMR expansion.

2.3 The Interval Generalized HDMR Method

The Interval Generalized HDMR method takes the uncertainties in the node locations and the function values into consideration to construct new relations through the Generalized HDMR method for obtaining an approximate band structure that models the given problem. In this sense, a weight function is defined as follows [7]

$$W(x_1, \ldots, x_N) \equiv \sum_{j=1}^{m} \alpha_j \delta\left(x_1 - x_1^{(j)}\right) \cdots \delta\left(x_N - x_N^{(j)}\right) \tag{16}$$

where α_j parameters are used for assigning a different importance to each individual datum and the multivariate data definition is

$$d_j \equiv \left(x_1^{(j)}, \ldots, x_N^{(j)}, \varphi_j\right), \quad \varphi_j \equiv f\left(x_1^{(j)}, \ldots, x_N^{(j)}\right), \quad 1 \le j \le m \tag{17}$$

When the weight function is inserted into the relations given in the previous subsection the Generalized HDMR method becomes to be applicable as a multivariate data partitioning technique. In addition, to obtain the error bands for the uncertainty analysis of those relations, we will approximate the differentiation operator with the corresponding order difference operator. In this sense, the relations will involve first order differences [10].

In this work, only the final relations of Interval Generalized HDMR are given for simplicity since the calculations were already given in the previous works of the authors [7, 10].

To this end, the structure of the constant Generalized HDMR component is obtained as [7]

$$f_0 = \sum_{j=1}^{m} \alpha_j \overline{\Omega}_j \varphi_j \tag{18}$$

while the error band for this constant component is as follows [10]

$$\Delta f_0 - \sum_{j=1}^{m} \left(\overline{\Omega}_j \varphi_j \Delta\alpha_j + \alpha_j \varphi_j \Delta\overline{\Omega}_j\right) = \sum_{j=1}^{m} \alpha_j \overline{\Omega}_j \Delta\varphi_j \tag{19}$$

where

$$\overline{\Omega}_j \equiv \prod_{k=1}^{N} \Omega_k\left(x_k^{(j)}\right), \quad \Delta\overline{\Omega}_j = \overline{\Omega}_j \left(\sum_{k=1}^{N} \frac{\Omega'_k\left(x_k^{(j)}\right)}{\Omega_k\left(x_k^{(j)}\right)} \Delta x_k^{(j)}\right), \quad 1 \le j \le m \tag{20}$$

Here, the uncertainties in node locations, function values and α coefficients of the weight function are assumed to be given.

Next step is to construct the general structure of the univariate Generalized HDMR components. The univariate components are the unknowns of the following linear equation system [7]

$$\beta_{i,k} = \overline{\alpha}_{i,k} \left(f_0 + f_i\left(\xi_i^{(k)}\right)\right) + \sum_{i_1=1}^{i-1} \sum_{\ell=1}^{m_{i_1}} \overline{\alpha}_{i_1,i;\ell,k} \Omega_{i_1} \left(\xi_{i_1}^{(l)}\right) f_{i_1} \left(\xi_{i_1}^{(l)}\right) + \\ + \sum_{i_1=i+1}^{N} \sum_{\ell=1}^{m_{i_1}} \overline{\alpha}_{i,i_1;k,\ell} \Omega_{i_1} \left(\xi_{i_1}^{(l)}\right) f_{i_1} \left(\xi_{i_1}^{(l)}\right) \tag{21}$$

where

$$\beta_{i,k} \equiv \sum_{j \in J_{i,k}} \frac{\alpha_j \overline{\Omega}_j}{\Omega_i\left(\xi_i^{(j)}\right)} \varphi_j, \quad \overline{\alpha}_{i,k} \equiv \sum_{j \in J_{i,k}} \frac{\alpha_j \overline{\Omega}_j}{\Omega_i\left(\xi_i^{(j)}\right)}, \quad 1 \le k \le m_i, \quad 1 \le i \le N \tag{22}$$

and

$$\overline{\alpha}_{i_1,i_2;k,\ell} \equiv \sum_{\substack{j\in J_{i_1,k}\\ j\in J_{i_2,\ell}}} \frac{\alpha_j \overline{\Omega}_j}{\Omega_{i_1}\left(\xi_{i_1}^{(k)}\right)\Omega_{i_2}\left(\xi_{i_2}^{(\ell)}\right)},$$

$$1 \le k \le m_{i_1}, \quad 1 \le \ell \le m_{i_2}, \quad 1 \le i_1 < i_2 \le N \tag{23}$$

The J sets are composed of identical x_i coordinate values while m_i values stand for the total number of identical values for each coordinate [7].

The error band for each univariate component is obtained by solving the following linear equation system [10]

$$\begin{aligned}
&\overline{\alpha}_{i,k}\left[\Delta f_0 + \Delta f_i\left(\xi_i^{(k)}\right)\right] + \Delta\overline{\alpha}_{i,k}\left[f_0 + f_i\left(\xi_i^{(k)}\right)\right] \\
&\quad - \sum_{j\in J_{i,k}}\left[\frac{\alpha_j\varphi_j\Delta\overline{\Omega}_j + \overline{\Omega}_j\varphi_j\Delta\alpha_j}{\Omega_i\left(\xi_i^{(j)}\right)}\right] \\
&\quad + \sum_{i_1=1}^{i-1}\sum_{\ell=1}^{m_{i_1}}\left[\Omega_{i_1}\left(\xi_{i_1}^{(l)}\right) f_{i_1}\left(\xi_{i_1}^{(l)}\right)\Delta\overline{\alpha}_{i_1,i;\ell,k} + \overline{\alpha}_{i_1,i;\ell,k}\Omega_{i_1}\left(\xi_{i_1}^{(l)}\right)\Delta f_{i_1}\left(\xi_{i_1}^{(l)}\right)\right] \\
&\quad + \sum_{i_1=i+1}^{N}\sum_{\ell=1}^{m_{i_1}}\left[\Omega_{i_1}\left(\xi_{i_1}^{(l)}\right) f_{i_1}\left(\xi_{i_1}^{(l)}\right)\Delta\overline{\alpha}_{i,i_1;k,\ell} + \overline{\alpha}_{i,i_1;k,\ell}\Omega_{i_1}\left(\xi_{i_1}^{(l)}\right)\Delta f_{i_1}\left(\xi_{i_1}^{(l)}\right)\right] \\
&\quad = \sum_{j\in J_{i,k}}\left[\frac{\Omega_i\left(\xi_i^{(j)}\right)\alpha_j\overline{\Omega}_j\Delta\varphi_j - \alpha_j\overline{\Omega}_j\varphi_j\Omega_i'\left(\xi_i^{(j)}\right)\Delta\xi_i^{(j)}}{\left(\Omega_i\left(\xi_i^{(j)}\right)\right)^2}\right] \\
&\qquad - \sum_{i_1=1}^{i-1}\sum_{\ell=1}^{m_{i_1}}\left[\overline{\alpha}_{i_1,i;\ell,k} f_{i_1}\left(\xi_{i_1}^{(l)}\right)\Omega_{i_1}'\left(\xi_{i_1}^{(l)}\right)\Delta\xi_{i_1}^{(l)}\right] \\
&\qquad - \sum_{i_1=i+1}^{N}\sum_{\ell=1}^{m_{i_1}}\left[\overline{\alpha}_{i,i_1;k,\ell} f_{i_1}\left(\xi_{i_1}^{(l)}\right)\Omega_{i_1}'\left(\xi_{i_1}^{(l)}\right)\Delta\xi_{i_1}^{(l)}\right]
\end{aligned} \tag{24}$$

where

$$\Delta\overline{\alpha}_{i,k} - \sum_{j\in J_{i,k}}\frac{\overline{\Omega}_j\Delta\alpha_j + \alpha_j\Delta\overline{\Omega}_j}{\Omega_i\left(\xi_i^{(j)}\right)} = \sum_{j\in J_{i,k}}\frac{\alpha_j\overline{\Omega}_j\Omega_i'\left(\xi_i^{(j)}\right)\Delta\xi_i^{(j)}}{\left(\Omega_i\left(\xi_i^{(j)}\right)\right)^2} \tag{25}$$

and

$$\Delta\overline{\alpha}_{i_1,i_2;k,\ell} - \sum_{\substack{j\in J_{i_1,k}\\ j\in J_{i_2,\ell}}} \frac{\alpha_j \Delta\overline{\Omega}_j + \overline{\Omega}_j \Delta\alpha_j}{\Omega_{i_1}\left(\xi_{i_1}^{(k)}\right)\Omega_{i_2}\left(\xi_{i_2}^{(\ell)}\right)} =$$

$$= - \sum_{\substack{j\in J_{i_1,k}\\ j\in J_{i_2,\ell}}} \frac{\alpha_j \overline{\Omega}_j \left[\Omega_{i_1}\left(\xi_{i_1}^{(k)}\right)\Omega'_{i_2}\left(\xi_{i_2}^{(\ell)}\right)\Delta\xi_{i_2}^{(\ell)} + \Omega_{i_2}\left(\xi_{i_2}^{(\ell)}\right)\Omega'_{i_1}\left(\xi_{i_1}^{(k)}\right)\Delta\xi_{i_1}^{(k)}\right]}{\left(\Omega_{i_1}\left(\xi_{i_1}^{(k)}\right)\right)^2 \left(\Omega_{i_2}\left(\xi_{i_2}^{(\ell)}\right)\right)^2} \tag{26}$$

The ξ values are the identical values identified in each coordinate, x_i. We also know the $\Delta\xi_i^{(j)}$ values.

2.4 Interpolation

The univariate components of both Generalized HDMR and Interval Generalized HDMR are partitioned data sets of the given multivariate data. To obtain an analytical structure for these components Langrange interpolation is used. In this sense, a multinomial should be built in terms of Lagrange polynomials

$$p_i(x_i) = \sum_{k_i=1}^{m_i} L_{k_i}(x_i) f_i\left(\xi_i^{(k_i)}\right),$$
$$L_{k_i}(x_i) = \mathbf{x}_i^T \mathbf{A}_i^{-1} \mathbf{e}_{k_i}, \quad 1 \le k_i \le m_i, \quad 1 \le i \le N \tag{27}$$

where

$$\mathbf{x}_i = \begin{bmatrix} 1 \\ x_i \\ x_i^2 \\ \vdots \\ x_i^{m_i-1} \end{bmatrix}, \ \mathbf{A}_i = \begin{bmatrix} 1 & \xi_i^{(1)} & \left(\xi_i^{(1)}\right)^2 & \dots & \left(\xi_i^{(1)}\right)^{m_i-1} \\ 1 & \xi_i^{(2)} & \left(\xi_i^{(2)}\right)^2 & \dots & \left(\xi_i^{(2)}\right)^{m_i-1} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 1 & \xi_i^{(m_i)} & \left(\xi_i^{(m_i)}\right)^2 & \dots & \left(\xi_i^{(m_i)}\right)^{m_i-1} \end{bmatrix} \tag{28}$$

Here, $1 \le i \le N$ and $\mathbf{A}$ is a Wandermonde matrix [10]. When we take into consideration the uncertainties mentioned before, the error band for the multinomial given in Eq. 27 is obtained as

$$\Delta p_i(x_i) = \sum_{k_i=1}^{m_i} \left[(\Delta L_{k_i}(x_i)) f_i\left(\xi_i^{(k_i)}\right) + L_{k_i}(x_i)\left(\Delta f_i\left(\xi_i^{(k_i)}\right)\right)\right] \tag{29}$$

where

$$\Delta L_{k_i}(x_i) = -\mathbf{x}_i^T \mathbf{A}_i^{-1}(\Delta\mathbf{A}_i)\mathbf{A}_i^{-1}\mathbf{e}_{k_i}, \quad 1 \le k_i \le m_i, \quad 1 \le i \le N \tag{30}$$

and

$$\Delta A_i = \begin{bmatrix} 0 & 1 & 2\xi_i^{(1)}\Delta\xi_i^{(1)} & \dots & (m_i-1)\left(\xi_i^{(1)}\right)^{m_i-2}\Delta\xi_i^{(1)} \\ 0 & 1 & 2\xi_i^{(2)}\Delta\xi_i^{(2)} & \dots & (m_i-1)\left(\xi_i^{(2)}\right)^{m_i-2}\Delta\xi_i^{(2)} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 1 & 2\xi_i^{(m_i)}\Delta\xi_i^{(m_i)} & \dots & (m_i-1)\left(\xi_i^{(m_i)}\right)^{m_i-2}\Delta\xi_i^{(m_i)} \end{bmatrix} \tag{31}$$

Finally, we can express the considered analytical structure in terms of Interval Generalized HDMR components and the following band structure is obtained as the model of the given problem

$$\bar{s}_1^{(l)}(x_1,\dots,x_N) \equiv (f_0 - |\Delta f_0|) + \sum_{i=1}^{N}(p_i(x_i) - |\Delta p_i(x_i)|)$$
$$\bar{s}_1^{(u)}(x_1,\dots,x_N) \equiv (f_0 + |\Delta f_0|) + \sum_{i=1}^{N}(p_i(x_i) + |\Delta p_i(x_i)|) \tag{32}$$

where $\bar{s}_1^{(l)}$ and $\bar{s}_1^{(u)}$ stand for the lower and upper bands of the model respectively [10].

3 Interval Factorized HDMR

Partitioning the given data including the errors in it and obtaining analytical structures for the univariate components of the expansion through Interval Generalized HDMR algorithm cannot be sufficient for the multivariate functions that do not have dominantly additive nature. A factorized form of this method is needed when the sought function has multiplicative nature. The expansion for this factorized form is defined as follows [8].

$$f(x_1,\dots,x_N) = r_0\left[\prod_{i_1=1}^{N}\left(1 + r_{i_1}(x_{i_1})\right)\right] \times \cdots \times \left[\left(1 + r_{1\dots N}(x_1,\dots,x_N)\right)\right] \tag{33}$$

The univariate Factorized HDMR approximation can be written as

$$\pi_1(x_1,\dots,x_N) = r_0\prod_{i_1=1}^{N}\left(1 + r_{i_1}(x_{i_1})\right) \tag{34}$$

where the Factorized HDMR components are written in terms of Generalized HDMR components as follows [8]

$$r_0 = f_0, \qquad r_{i_1}(x_{i_1}) = \frac{f_{i_1}(x_{i_1})}{f_0}, \quad 1 \le i_1 \le N \tag{35}$$

To obtain the error bands for the constant and the univariate terms of the Factorized HDMR expansion the first order differences of the relations of these components are evaluated. The error band for the constant term is obtained as

$$\Delta r_0 = \Delta f_0 \tag{36}$$

while the relation for the error band of the univariate terms are

$$\Delta r_{i_1}(x_{i_1}) = \frac{f_0 \Delta f_{i_1}(x_{i_1}) - f_{i_1}(x_{i_1})\Delta f_0}{f_0^2}, \quad 1 \le i_1 \le N \tag{37}$$

where the error band relations for the constant term and the univariate terms of the Interval Generalized HDMR algorithm are given in Eqs. 19 and 24 respectively.

The first order difference of the univariate Factorized HDMR approximant given in Eq. 34 can be obtained as follows

$$\Delta \pi_1(x_1, \ldots, x_N) = (\Delta r_0) \prod_{i_1=1}^{N} \left(1 + r_{i_1}(x_{i_1})\right) + r_0 \Delta \left(\prod_{i_1=1}^{N} \left(1 + r_{i_1}(x_{i_1})\right) \right) \tag{38}$$

If the calculations for the above relation are done the following relation is obtained

$$\Delta \pi_1(x_1, \ldots, x_N) = \left[\frac{\Delta r_0}{r_0} + \sum_{k=1}^{N} \frac{\Delta r_k(x_k)}{1 + r_k(x_k)} \right] \pi_1(x_1, \ldots, x_N) \tag{39}$$

The univariate Interval Factorized HDMR approximants for the lower curve and the upper curve structures can be written respectively as

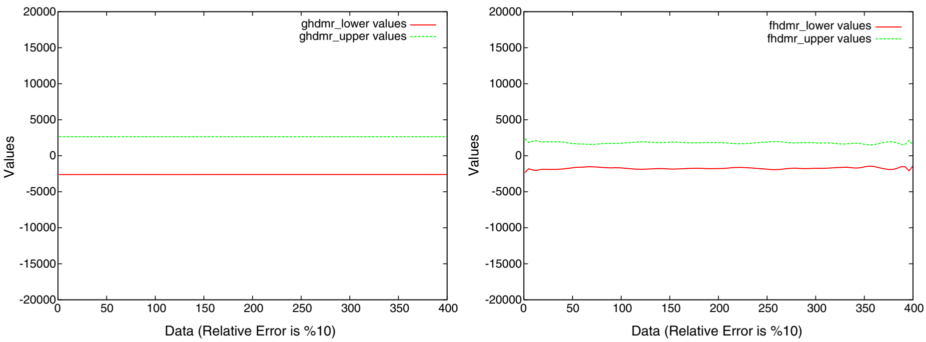
$$\pi_1^{(l)}(x_1, \ldots, x_N) = \pi_1(x_1, \ldots, x_N) - |\Delta \pi_1(x_1, \ldots, x_N)| \tag{40}$$

$$\pi_1^{(u)}(x_1, \ldots, x_N) = \pi_1(x_1, \ldots, x_N) + |\Delta \pi_1(x_1, \ldots, x_N)| \tag{41}$$

where $\pi_1^{(l)}$ and $\pi_1^{(u)}$ stand for the lower and upper bands of the model respectively. Since we deal with at most univariate approximation in this work, the relations of higher variate approximations are not included here.

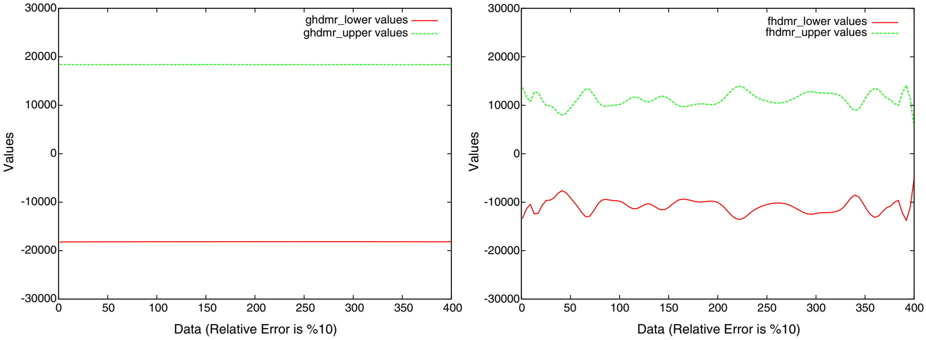
4 Numerical Implementations

This section covers a number of examples to examine the bound analysis performance of our new method, Interval Factorized HDMR. The calculations are done in MuPAD [4] within 20-digits precision. Because we offer a new method that



(a) Interval Generalized HDMR (b) Interval Factorized HDMR

Fig. 1 Band structures for $f_1(x_1, \ldots, x_5)$



(a) Interval Generalized HDMR (b) Interval Factorized HDMR

Fig. 2 Band structures for $f_2(x_1, \ldots, x_5)$

works well for cases that are dominantly and purely multiplicative nature, the testing functions are selected as

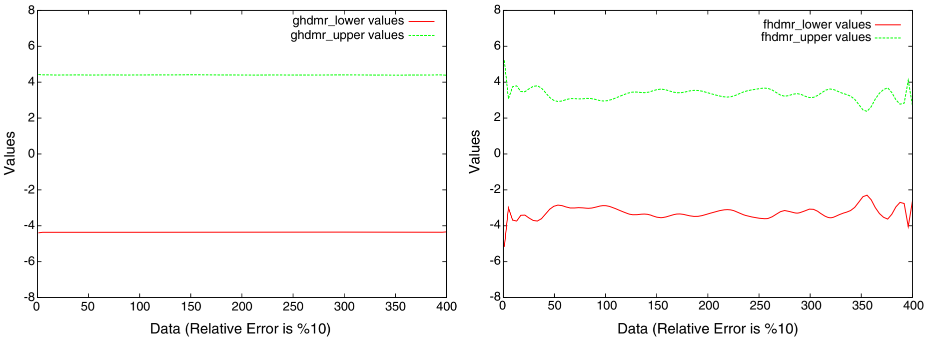
$$f_1(x_1, \ldots, x_5) = \left[\sum_{i=1}^{5} x_i\right]^3, \quad f_2(x_1, \ldots, x_5) = \left[\sum_{i=1}^{5} x_i\right]^5,$$

$$f_3(x_1, \ldots, x_5) = \prod_{i=1}^{5} x_i \tag{42}$$

where each has 5 independent variables. The modelling process will be executed on 400 nodes with an uncertainty of 10 % on the locations of the nodes and the function values. The domains of the independent variables are selected as follows

$$1 \le x_1 \le 4, \quad 3 \le x_2 \le 7, \quad 2 \le x_3 \le 5, \quad 4 \le x_4 \le 8, \quad 3 \le x_5 \le 9 \tag{43}$$

Figures 1, 2, and 3 show the performance of both Interval Generalized HDMR and Interval Factorized HDMR methods in randomly constructed multivariate data modelling problems by using the testing functions and the domain of each independent variable of the related function given in Eqs. 42 and 43 respectively.



(a) Interval Generalized HDMR (b) Interval Factorized HDMR

Fig. 3 Band structures for $f_3(x_1, \ldots, x_5)$

It is clear that the Interval Factorized HDMR method works better than the other method. Our proposed method constructs a tighter band structure.

5 Concluding Remarks

Generalized HDMR is a multivariate data partitioning method and can be used to obtain an analytical structure as the model for the multivariate data modelling problems. When uncertainties on data that describe such problems occur, Generalized HDMR method should be reconstructed to determine error bands for the components of the method. This results in a band structure as the model instead of a unique analytical structure. The Interval Generalized HDMR method is the method that has the ability to determine that type of a band structure. However, it is known that this method works well for cases of additive nature. When the nature of the given problem becomes dominantly or purely multiplicative then a new method is needed to have an acceptable band structure. This new method is named as Interval Factorized HDMR and is based on the standard Factorized HDMR method. The numerical results and the corresponding figures show us that this new method gives better band structures than the Interval Generalized HDMR method for cases of dominantly and purely multiplicative nature. This means that our new method closes the gap that the Interval Generalized HDMR method has in modelling cases of non-additive nature.

References

1. Banerjee, I., Ierapetritou, M.G.: Model independent parametric decision making. Ann. Oper. Res. **132**(1–4), 135–155 (2004)
2. Chowdhury, R., Rao, B.N.: Hybrid high dimensional model representation for reliability analysis. Comput. Methods Appl. Mech. Eng. **198**, 753–765 (2009)
3. Gomez, M.C., Tchijov, V., Leon, F., Aguilar, A.: A tool to improve the execution time of air quality models. Environ. Model. Softw. **23**, 27–34 (2008)
4. Oevel, W., Postel, F., Wehmeier, S., Gerhard, J.: The MuPAD Tutorial. Springer, New York (2000)
5. Rabitz, H., Alış, Ö.F.: General foundations of high dimensional model representations. J. Math. Chem. **25**, 197–233 (1999)
6. Sobol, I.M.: Sensitivity estimates for nonlinear mathematical models. Math. Model. Comput. Exper. **1**, 407–414 (1993)
7. Tunga, M.A., Demiralp, M.: Data partitioning via generalized high dimensional model representation (ghdmr) and multivariate interpolative applications. Math. Res. **9**, 447–462 (2003)
8. Tunga, M.A., Demiralp, M.: A factorized high dimensional model representation on the nodes of a finite hyperprismatic regular grid. Appl. Math. Comput. **164**, 865–883 (2005)
9. Tunga, M.A., Demiralp, M.: A new approach for data partitioning through high dimensional model representation. Int. J. Comput. Math. **85**(12), 1779–1792 (2008)
10. Tunga, M.A., Demiralp, M.: Bound analysis in univariately truncated generalized high dimensional model representation for random-data partitioning: interval ghdmr. Appl. Numer. Math. **59**, 1431–1448 (2009)
11. Ziehn, T., Tomlin, A.S.: Gui-hdmr—a software tool for global sensitivity analysis of complex models. Environ. Model. Softw. **24**, 775–785 (2009)