University of Leeds

School of Process, Environmental and Materials Engineering
Energy and Resources Research Institute

User Documentation
(Draft)

GUI-HDMR

Version 1.1

Developer: Tilo Ziehn

Contact:

Professor Alison Tomlin: a.s.tomlin@leeds.ac.uk

Tilo Ziehn: pretz@leeds.ac.uk

Contents

1 Introduction

Global sensitivity analysis is a very important part of the model evaluation process and should be applied in order to assess and increase the reliability of computer models. However, many models used in environmental engineering and other fields have a large number of uncertain parameters and are computationally expensive to run. This restricts the usage of traditional methods of global sensitivity analysis such as Monte Carlo analysis due to the computational expense of the model and the difficulty in interpreting the results (e.g. scatter plots) for large numbers of parameters.

The random sampling high dimensional model representation (RS-HDMR) method (Rabitz et al., 1999), (Li et al., 2002) provides a straightforward approach to explore the input output mapping of a model without requiring a large number of model runs. Furthermore, variance based sensitivity indices can be determined in an automatic way in order to rank the importance of input parameters and to explore the influence of parameter interactions. However, various case studies have shown that the existing set of RS-HDMR tools do not always provide the most efficient and most accurate way to construct a metamodel and to calculate sensitivity indices. This has lead to the development of an optimisation algorithm as an extension to existing HDMR tools in Ziehn and Tomlin (2008a), which automatically determines the optimal order of the polynomials used in the HDMR expansion for the approximation of the component functions. Further, a threshold has been introduced in Ziehn and Tomlin (2008b) in order to exclude unimportant component functions from the HDMR expansion. This is particularly useful if the number of input parameters is very large (greater than 20) and if only a few parameters contribute to the overall output variance. Commonly a screening method such as the Morris method (Morris, 1991) is applied first for cases with a large input space dimension. The extension of the RS-HDMR method by the idea of excluding unimportant component functions from the HDMR expansion via a threshold generally prevents the need for a prior screening. In some cases, the combination of a screening approach and the RS-HDMR method might be computationally more efficient (Ziehn and Tomlin, 2008b), (Ziehn et al., 2008a).

Although the RS-HDMR method is conceptually simple, the application is not always easy since many HDMR tools and extensions exist. The GUI-HMDR software (Ziehn and Tomlin, 2008c) described here combines RS-HDMR tools and their extensions in one Matlab package equipped with a graphical user interface. This makes the HDMR method and its extensions easily available for all interested users in all fields. The user has to supply only one set of input values (random or experimental data) and the corresponding output values. Therefore, the application of the GUI-HDMR software is not restricted to a certain type of model and can be used independently from the type of platform used as long as the standard Matlab package is provided.

2 High Dimensional Model Representation (HDMR)

The high dimensional model representation (HDMR) method is a set of tools explored by Rabitz et al. (1999) in order to express the input-output relationship of complex models with a large number of input variables. The mapping between the input variables $x_1, \ldots, x_n$ and the output variables $f(\mathbf{x}) = f(x_1, \ldots, x_n)$ in the domain R^n can be written in the following form:

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^{n} f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j) + \ldots + f_{12\ldots n}(x_1, x_2, \ldots, x_n) \tag{1}$$

Here f_0 denotes the mean effect (zeroth order), which is a constant. The function $f_i(x_i)$ is a first order term giving the effect of variable x_i acting independently (although generally non-linearly) upon the output $f(\mathbf{x})$. The function $f_{ij}(x_i, x_j)$ is a second order term describing the cooperative effects of the variables x_i and x_j upon the output $f(\mathbf{x})$. The higher order terms reflect the cooperative effects of increasing numbers of input variables acting together to influence the output $f(\mathbf{x})$. If there is no interaction between the input variables, then only the zeroth order term f_0 and the first order terms $f_i(x_i)$ will appear in the HDMR expansion.

The HDMR expansion is computationally very efficient if higher order input variable correlations are weak and can therefore be neglected. For many systems a HDMR expression up to second order already provides satisfactory results and a good approximation of $f(\mathbf{x})$ (Li et al., 2001).

The developed GUI-HDMR software is based on the RS-HDMR approach (Li et al., 2002) where a set of random sample points N over the entire domain R^n is used. The zeroth order term f_0 can be approximated by the average value of $f(\mathbf{x})$. The determination of the higher order component functions is based on the approximation of the component functions by orthonormal basis functions:

$$f_i(x_i) \approx \sum_{r=1}^{k} \alpha_r^i \varphi_r(x_i) \tag{2}$$

$$f_{ij}(x_i, x_j) \approx \sum_{p=1}^{l} \sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_p(x_i) \varphi_q(x_j) \tag{3}$$
$$\ldots$$

where k, l, l' represent the order of the polynomial expansion, α_r^i and β_{pq}^{ij} are constant coefficients to be determined, and $\varphi_r(x_i)$, $\varphi_p(x_i)$ and $\varphi_q(x_j)$ are the orthonormal basis functions.

Note that only one set of random samples N is necessary in order to determine all RS-HDMR component functions (Li et al., 2002).

2.1 Variance Reduction Methods

The determination of the expansion coefficients for the orthonormal polynomials is based on Monte Carlo (MC) integration. The error of the MC integration controls the accuracy of the RS-HDMR expansion. Variance reduction methods can be applied to improve the accuracy of the MC integration. Two methods have been successfully applied in connection with the RS-HDMR method and are implemented in the GUI-HDMR software: The correlation method described in Kalos and Whitlock (1986) and used by Li at el. (2003) and the ratio control variate method as given by Li and Rabitz (2006). In both cases the determination of the expansion coefficients becomes an iterative process and requires an analytical reference function $h(\mathbf{x})$. This function has to be similar to $f(\mathbf{x})$ and as shown in Li at el. (2003) and Li and Rabitz (2006) a truncated RS-HDMR expansion can be used as a reference function whose expansion coefficients were calculated by direct MC integration.

The application of a variance reduction method allows improvements in the accuracy of the metamodel without increasing the sample size N. This is very beneficial for practical applications of RS-HDMR especially for high dimensional systems.

2.2 Optimal Polynomial Order and Threshold

The set of RS-HDMR tools as described so far uses the same polynomial order for all first order and second order component functions respectively. However, because the HDMR component functions are independent, the order of the polynomial approximation should be chosen separately for each component function. Consequently, an optimisation method has been developed in Ziehn and Tomlin (2008a), which automatically chooses the best polynomial order for each of the component functions. The computational effort to calculate the optimal order for all polynomials is low and therefore it is worthwhile to calculate it in order to improve the accuracy of the final metamodel.

Further, a threshold has been introduced in Ziehn and Tomlin (2008b) in order to exclude unimportant component functions from the HDMR expansion. The idea of this approach is to reduce the number of component functions to be approximated by polynomials and therefore to reduce the number of parameters so that screening methods can be avoided. This is particularly useful if the model has a high number of input parameters (greater than 20) and only low higher order effects. If significant parameter interactions exist, then it might be computationally more efficient to apply a screening method such as the Morris method (Morris, 1991) beforehand in order to reduce the dimensionality of the problem. This can result in a more accurate estimation of the higher order component functions for smaller samples sizes N.

2.3 Global Sensitivity Analysis using HDMR

A commonly used method in global SA is the method of Sobol' (Sobol, 2001), which is conceptually the same as the RS-HDMR approach. In statistics the decomposition of $f(\mathbf{x})$ into summands of increasing dimensionality (see equation (1)) is called ANOVA decomposition which is also a member of the high dimensional model representations known as ANOVA-HDMR (Rabitz et al., 1999), (Rabitz and Ali\c{s}, 1999). However, the calculation of the partial variances on the basis of the RS-HDMR function expansion provides a much more efficient approach (Ali\c{s} and Rabitz, 2001).

The total variance D can be obtained by

$$D = \int_{K^n} f^2(\mathbf{x}) \mathrm{d}\mathbf{x} - f_0^2 \tag{4}$$

and the partial variances $D_{i_1,\dots,i_s}$ can be calculated from each of the terms in equation (1):

$$D_i \;\;=\;\; \int_0^1 f_i^2(x_i) \mathrm{d}x_i \tag{5}$$

$$D_{ij} \;\;=\;\; \int_0^1 \int_0^1 f_{ij}^2(x_i, x_j) \mathrm{d}x_i \mathrm{d}x_j \tag{6}$$

$$\dots$$

Once the partial variances are determined the sensitivity indices can be calculated as follows:

$$S_{i_1,\dots,i_s} = \frac{D_{i_1,\dots,i_s}}{D}, \qquad 1 \le i_1 < \cdots < i_s \le n \tag{7}$$

so that all its terms add up to 1:

$$\sum_{i=1}^n S_i + \sum_{1 \le i < j \le n} S_{ij} + \dots + S_{1,2,\dots,n} = 1 \tag{8}$$

The first order sensitivity index S_i measures the main effect of the input variable x_i on the output, or in other words the fractional contribution of x_i to the variance of $f(\mathbf{x})$. The second order sensitivity index S_{ij} measures the interaction effect of x_i and x_j on the output and so on. More detailed information about the calculation of the sensitivity indices on basis of the RS-HDMR component functions can be found in Li at el. (2002a).

3 Software Overview

The software GUI-HDMR was developed to combine existing RS-HDMR tools and developed RS-HDMR extensions and to make them easily available for all interested users. The software is written in Matlab and requires the basic Matlab package. A second order HDMR expansion is used which allows the investigation of parameter interactions.

The software comes with a graphical user interface (GUI), but can also be used without the GUI as a script based approach. In both cases the user has to supply only two files. The first one contains the rescaled input values ($0 \leq x_i \leq 1$) which were used to run the model and the second one contains the corresponding output values. Both files have to be provided as ASCII files within a matrix format. The rows represent the sample number ($1 \ldots N$) and the columns stand for the different inputs ($i = 1 \ldots n$) or the number of considered outputs respectively. If only one output is considered then the output file is simply a column vector.

The set of input values can be any Monte Carlo sample (if the input parameters can be controlled) or measured values (if using experimental data). If the input parameters can be controlled, then a quasi-random sampling method such as the Sobol' sequence (Sobol, 1976) is preferable. This guarantees that the input space is covered more uniformly than by using random values and it provides a better convergence rate (Kucherenko, 2007).

Additionally a second set of input and output values may be provided in the same format as the first set. In order to show the accuracy of the constructed metamodel the relative error (RE) and the coefficient of determination (R^2) are calculated for a number of samples. If a different set of input and output values is provided, then the metamodel is tested against values which were not used to construct the HDMR expansion. If no second set is provided, then the same set of input and output values is used for the accuracy test. Optionally one can also provide the original input parameter ranges in order to produce the plots of the component functions with the correct ranges instead of the rescaled ones.

A fully functional metamodel can be constructed using only the one set of samples provided. Additionally, first and second order sensitivity indices can be calculated. Component functions of first and second order can also be plotted and compared with scatter plots.

3.1 Graphical User Interface (GUI)

After uncompressing the file `gui_hdmr_software_v1p1.zip` the graphical user interface of the GUI-HDMR software can be started by typing `gui_hdmr` into the matlab command window [1]. The main window as shown in figure 1 will open. From the main window calculations can be started and progress is shown. Three other windows can be accessed

---

[1]Please ensure that you are in the right folder or that the path is set within Matlab.

from the main window: The *Setup* window, the *Results* window showing the optimal polynomial orders and the accuracy of the metamodel and the *Global Sensitivity Analysis* window.
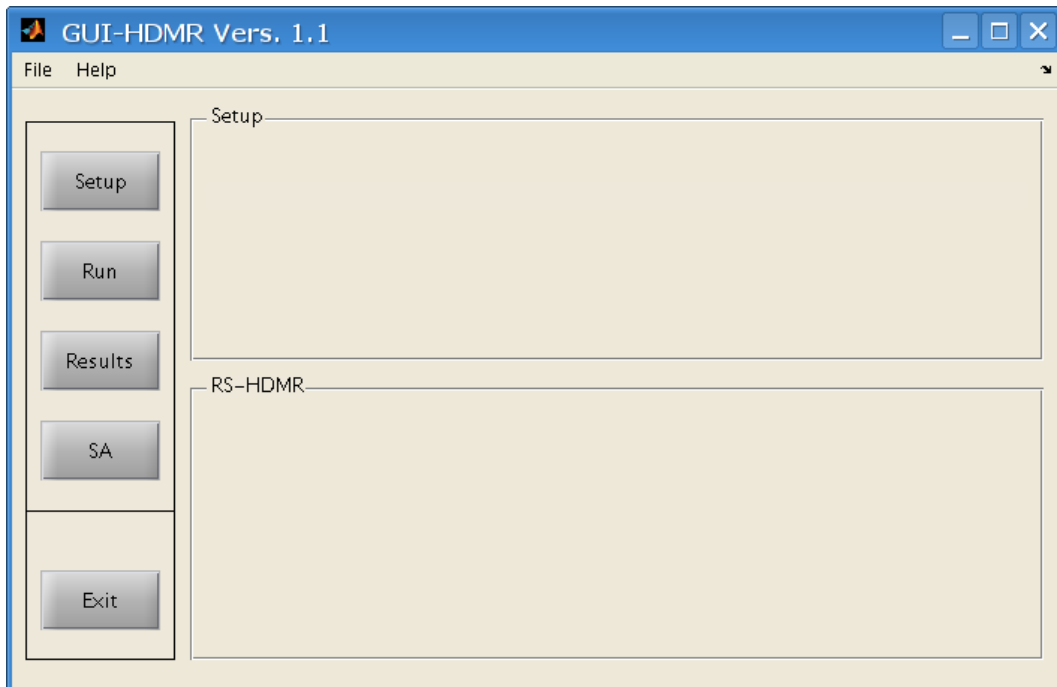

Figure 1: GUI-HDMR main window.

In case a previous HDMR analysis has been performed using the GUI-HDMR software and results have been saved, the result file can be loaded by choosing **File - Load results**. The **Open** dialog will appear and a result file with the ending `.mat` can be selected. If the selection is confirmed by clicking **Open** the result file will be loaded in the matlab workspace and all data are available within the GUI-HDMR software.

3.1.1 Setup

From the main window the *Setup* window (figure 2) can be accessed by clicking the **Setup** button on the left. In the *Setup* window, the sample input and output files which are based on the performed model runs can be loaded and different settings for the HDMR analysis can be applied. The sample input file is loaded by choosing **File - Open - Input-File (for HDMR analysis)** and the sample output file by choosing **File - Open - Output-File (for HDMR analysis)**. The set of samples provided by these two files is used to construct the HDMR expansion (metamodel) up to second order and to calculate the variance based sensitivity indices. Optionally, a second set of input and output samples can be loaded in order to perform an accuracy test based on the relative

error and coefficient of determination R^2. In this case the metamodel is tested against values which were not used for the construction of the HDMR expansion. The additional set of input and output samples is loaded by choosing **File - Open - Input-File (for accuracy test)** and **File - Open - Output-File (for accuracy test)** respectively. It is also possible to load a file containing the original input parameter ranges by choosing **File - Open - Input Ranges (for plots)** in order to plot component functions with the real ranges instead of the rescaled ones. In each of the cases the *Open* dialog will appear and the corresponding file can be selected and confirmed by clicking **Open**. In order to reset the file settings choose **File - Reset file settings**.
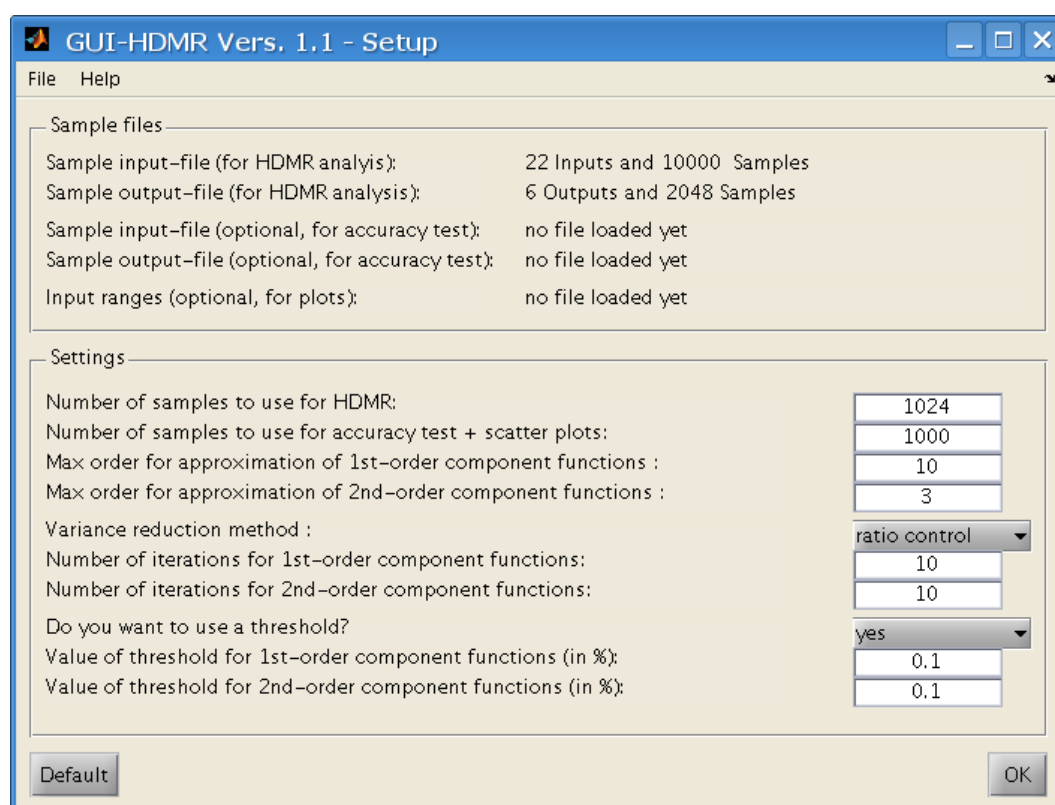

Figure 2: GUI-HDMR *Setup* window. Input and output files can be loaded and settings can be applied for the HDMR analysis.

After loading the input and output files, a summary will be shown in the "Sample files" section of the *Setup* window stating the number of inputs and outputs and the number of samples provided in each file.

Various settings can now be defined within the *Setup* window in the "Settings" sections. The sample size N can be set which is used for the construction of the HDMR expansion. The maximum sample size N corresponds with the number of samples supplied in the input and output files. Additionally, the number of samples can be defined which are used

for the accuracy test (calculation of the relative error and coefficient of determination, see also section 3.1.3) and the scatter plots (see also section 3.1.4).

The optimisation method (see section 2.2) which automatically calculates the optimal order for each of the polynomials is included in the GUI-HDMR software and one can define a maximum polynomial order for the approximation of the first and second order component functions. Up to tenth order polynomials are supported by the Software.

A variance reduction method (see section 2.1) to control the Monte Carlo integration error can also be applied. The user has the choice between two approaches: the correlation method and the ratio control variate method. In both cases the number of iterations can be set for the first and second order HDMR component functions.

Finally a threshold (see section 2.2) can be applied, which is used to exclude unimportant component functions from the HDMR expansion. The threshold is given in % and is particularly useful if the model has a high number of input parameters. Suggestions on how to apply the threshold in connection with the HDMR method are given in Ziehn and Tomlin (2008b).

By pressing the **Default** button at the lower left corner of the Setup window the standard settings will be applied. In this case no variance reduction method and no threshold is used. By pressing the **OK** button the settings will be saved and the *Setup* window is closed.

3.1.2 Construction of the HDMR Expansion

After loading the input and output files and applying all the settings, a summary is shown in the "Setup" section of the main window. The HDMR analysis is started by pressing the **Run** button. The progress of the calculation is shown in the "RS-HDMR" section of the main window (figure 3).

If a variance reduction method has been applied, then the current iteration number and the total number of iterations (figure 3) are also shown for the calculation of the optimal order for the first and second order component functions.

After the calculations have been completed the results can be saved by choosing **File - Save results as...**. This will create a MAT-file with the ending `.mat` in the current folder. This file can be loaded into the software again for further analysis (i.e. in order to create plots of the component functions).

Please note, that if a higher maximum polynomial order (≥ 5) has been assigned for the approximation of the second order component functions, then the whole calculation process might be more time consuming, especially in connection with a large input space dimension and if using a variance reduction method with a large number of iterations.
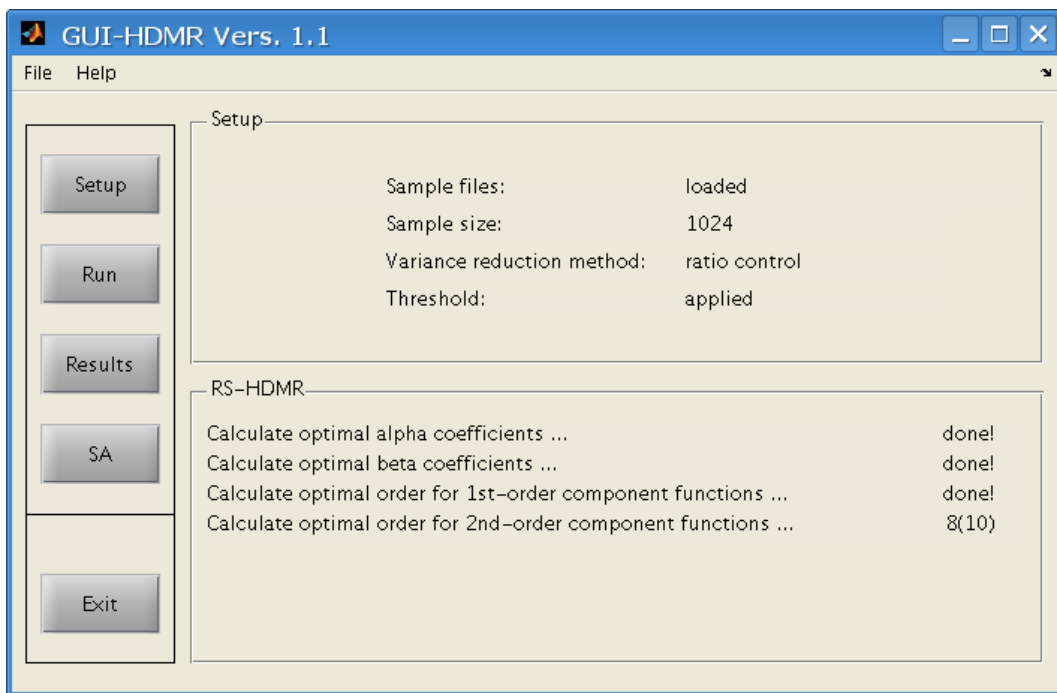
Figure 3: GUI-HDMR main window. The summary of the settings is shown and the progress of the calculation.

3.1.3 Results

The results of the constructed HDMR expansion are accessable by pressing the **Results** button in the main window after the calculations have been finished or after a results file has been loaded. The *Results* window (figure 4) states the number of component functions which have been approximated by polynomials and by what order for first order and second order (by pressing **1st-order** and **2nd-order** respectively from the "Results" menu on the left hand side).

It is also possible to verify the accuracy of the constructed first and second order HDMR metamodel based on various methods. If an additional set of sample inputs and outputs has been provided (see section 3.1.1), then the relative error and coefficient of determination R^2 are calculated using a set of samples which has not been used for the construction of the HDMR expansion. Otherwise, the metamodel is tested against the same set of samples which has been used for its construction. Two graphical methods are also provided: a scatter plot showing the relationship between the original model and the HDMR metamodel and a comparison of the empirical probability distribution functions (pdf) for both, the original model and the HDMR metamodel. The plot of the empirical pdf requires the Matlab statistics toolbox. However, the GUI-HDMR software automatically checks whether or not the statistics toolbox is available and if not then a comparison of
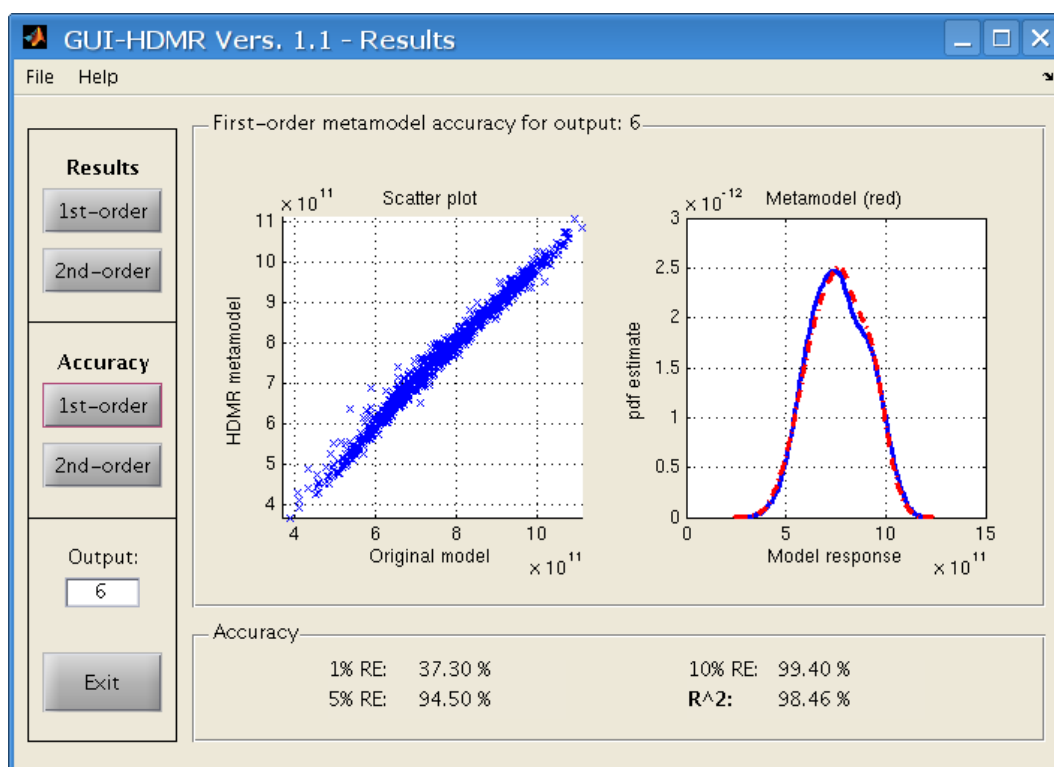
Figure 4: GUI-HDMR *Results* window. The accuracy of the HDMR metamodel is presented in terms of the relative error (1 %, 5 % and 10 %), the coefficient of determination (R^2), the scatter plot and the plot of the empirical pdf.

the histograms is presented instead of the pdfs. By pressing **1st-order** in the "Accuracy" menu on the right hand side the relative error (absolute error divided by the magnitude of the exact value) and the coefficient of determination are calculated for the first order HDMR metamodel using the number of samples as defined in the setup (see section 3.1.1). Additionally, the scatter plot and the plot of the pdfs (or histograms) are also constructed. Clicking on **2nd-order** in the "Accuracy" menu provides the relative error, coefficient of determination and plots for the second order HDMR metamodel. The relative error and the coefficient of determination are given in the "Accuracy" section. It is stated how many of the tested samples (in %) are within an a range of 1 %, 5 % and 10 % of the relative error. The coefficient of determination R^2 shows the amount of variation (in %) in the original model output that can be explained by the HDMR metamodel.

If more than one output has been considered (sample output file has more than one columns), then the output of interest can be chosen by typing in the corresponding number in the "Output:" field on the left corner. The results and the accuracy are then presented for the selected output.

The scatter plot and the plot of the empirical pdf (or histogram) can be saved in four

different formats: Portable Network Graphics (png), Portable Document Format (pdf), Encapsulated PostScript (eps) and Matlab format. This can be done by choosing **File - Save plot a (left)** in order to save the current plot in the left figure or **File - Save plot b (right)** to save the current plot in the right figure. The **Save** dialog will appear and a file name can be chosen. If the selection is confirmed by clicking **Save** the plot will be saved in four different versions with the endings `.png`, `.pdf`, `.eps` and `.fig`.

The *Results* window can be closed by clicking the **Exit** button and one gets back to the main window.

3.1.4 Sensitivity Analysis

The *Global Sensitivity Analysis* window is opened by pressing the **SA** button in the main window. If clicking the **1st-order** button in the "Results" menu on the right a ranking of the five most important input parameters (first order effects) is constructed and displayed in the "Sensitivity Indices section" (figure 5). Each of the five parameters is identified by its input number (number of the column in the input sample file) and the first order sensitivity index S_i is given. The sum of all first order sensitivity indices is also calculated and displayed.

By clicking the **2nd-order** button in the "Results" menu on the right a ranking of the five most important input parameter interactions (the second order effects) is constructed and displayed in the "Sensitivity Indices" section (figure 6). Each of the five parameter interactions is identified by its input number combination (number of the column in the input sample file) and the second order sensitivity index S_{ij} is given. Additionally, the sum of all second order sensitivity indices and the sum of all first and second order sensitivity indices is calculated and displayed.

Each of the component functions can also be plotted in order to illustrate the relationship between a certain input parameter (or combination) and the considered output. The plot of the first order effects can be accessed by clicking **1st-order** in the "Plots" menu on the left hand side. Then, one has three options:

1. By pressing the **Scatter** button next to the right or left diagram a scatter plot is created.

2. By pressing the **HDMR** button the first order RS-HDMR component function is plotted showing the effect of one parameter acting independently upon the output.

3. By pressing the **Both** button the scatter plot and the plot of the first order component function (in this case the mean f_0 is added to component function $f_i(x_i)$) are combined in one plot for comparison.

In all three cases the input of interest can be selected in the "Input:" field next to the plot.
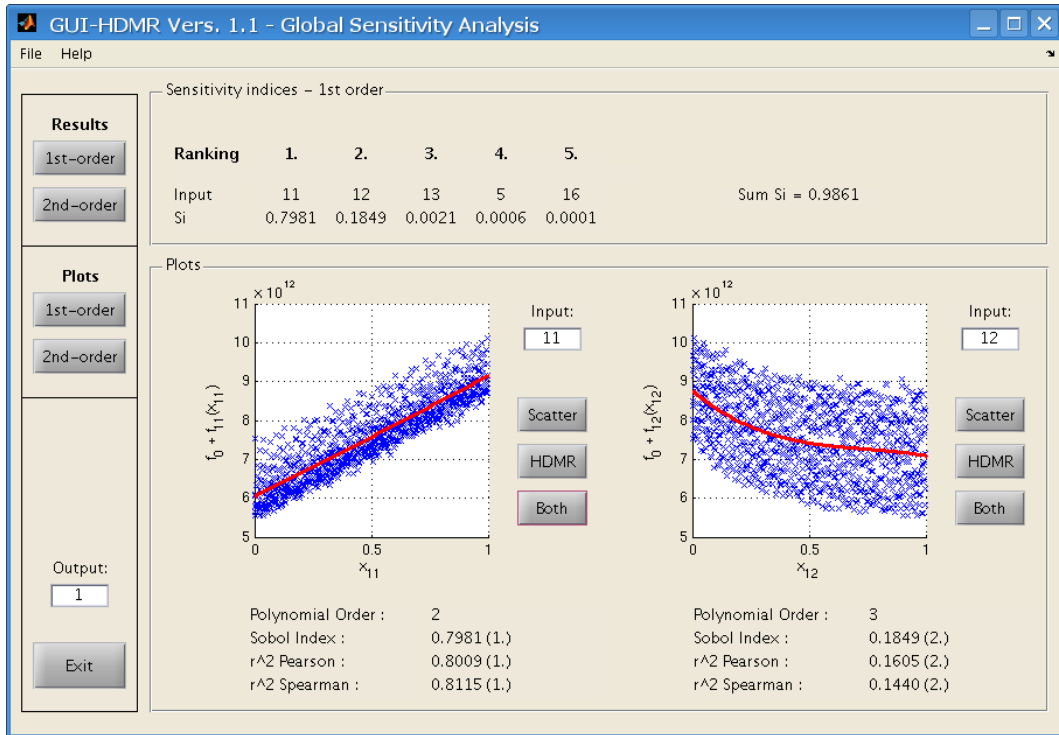
Figure 5: GUI-HDMR *Global Sensitivity Analysis* window: The ranking of the five most important parameters based on the first order sensitivity indices is shown. Additionally all component functions can be plotted.

Additionally, for each of the plots the following information is given:

1. The polynomial order which has been used to approximate the corresponding component function.

2. The first order sensitivity index and its rank (displayed in brackets).

3. The r^2 value using the Pearson correlation coefficient and its rank (displayed in brackets).

4. The r^2 value using the Spearman rank correlation coefficient and its rank (displayed in brackets).

The Pearson correlation coefficient measures the correlation between the input and the output (a correlation of zero means there is no linear relationship between the two variables). The square of the correlation coefficient can be used as a measure of the association between the input and the output. For example, if r^2 is 0.70, then 70 % of the variance of the output can be explained by changes in the input assuming a linear relationship
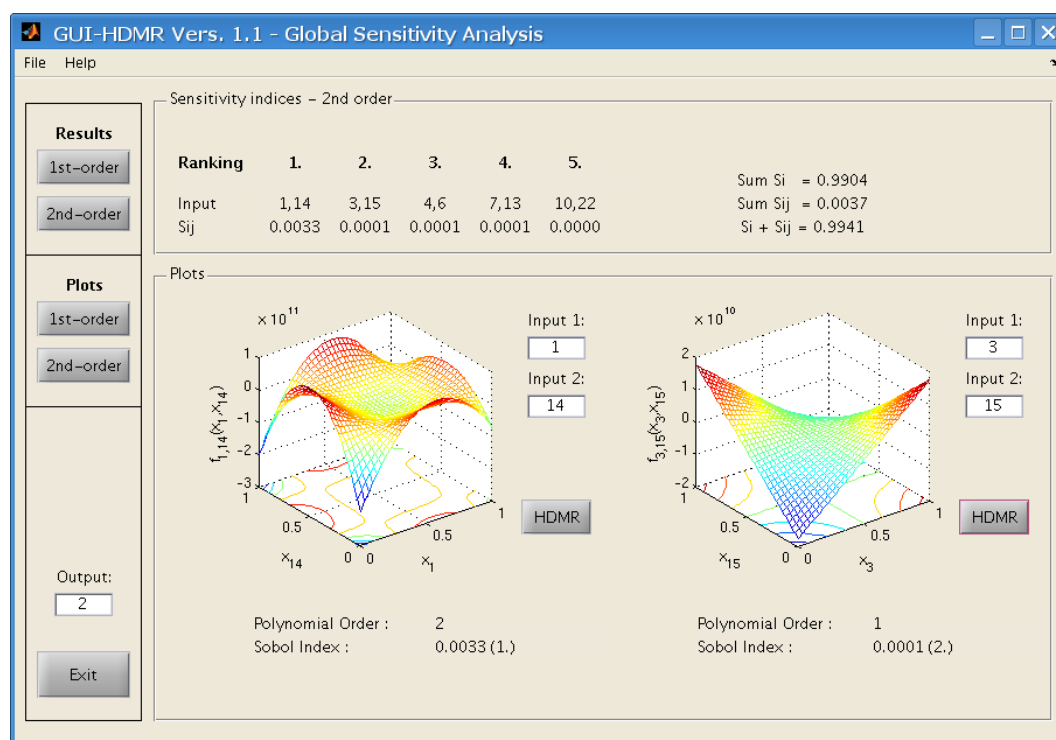
Figure 6: GUI-HDMR *Global Sensitivity Analysis* window: The ranking of the five most important parameter interactions based on the second order sensitivity indices is shown. Additionally all component functions can be plotted.

between the input and the output. The use of the Spearman rank correlation coefficient is similar, however it assesses how well an arbitrary monotonic function could describe the relationship between the input and the output. More detailed information about the correlation coefficients and their application can be found in Moore (2007). The variance based sensitivity indices do not require any assumptions regarding the relationship between the input and the output and are therefore more reliable than the correlation coefficients.

The plots of the second order component functions can be accessed by clicking **2nd-order** in the "Plots" menu on the left hand side. The input combination of interest can be chosen be editing the "Input 1:" and "Input 2:" field next to the plots. By pressing the **HDMR** button the plots can be created individually.

If a file for the input ranges has been provided in the *Setup* window, then all plots are created using the original input parameter ranges. Otherwise, all plots are given for the rescaled input values between 0 and 1.

Again, if more than one output has been considered (sample output file has more than one column), then the output of interest can be chosen by typing in the corresponding

number in the "Output:" field on the left corner. The sensitivity indices and plots are then shown for the selected output.

All constructed plots of the first and second order component functions (and/or scatter plots) can be saved in four different formats. To do so one has to choose **File - Save plot a (left)** in order to save the current plot in the left figure or **File - Save plot b (right)** to save the current plot in the right figure.

The **Save** dialog will appear and a file name can be chosen. If the selection is confirmed by clicking **Save** the plot will be saved in four different versions with the endings `.png`, `.pdf`, `.eps` and `.fig`.

3.1.5 GUI-HDMR Results File

All data, settings and results of the HDMR analysis can be saved in one external MAT-file by choosing **File - Save results as...** in the main window, *Results* window or *Global Sensitivity analysis window*. Additionally, if closing the GUI-HDMR software with the **Exit** button in the main window all results will be automatically saved in an external file with the allocated name `workspace.mat`. This file will be overwritten each time the **Exit** button is used and guarantees that all results will be saved in case the user forgets to use the **File - Save results as...** option. The results file can be loaded into the GUI-HDMR software again in order to access the sensitivity indices for the five most important parameters and parameter interactions or to plot component functions. However, the results file can also be loaded into the Matlab workspace (`load <filename>.mat`) e.g. in order to access all sensitivity indices and their ranks. It consists of four structure arrays:

- `sample_1` containing the input and output samples used for the HDMR analysis loaded in the *Setup* window.

- `sample_2` containing the additional input and output samples used for the accuracy test loaded in the *Setup* window (they will be equal to the input and output values in `sample_1` if no additional set has been loaded).

- `settings` containing the HDMR settings from the *Setup* window.

- `rs_hdmr` containing all results from the HDMR analysis.

The `rs_hdmr` structure array consits of the following fields:

```
rs_hdmr =
               f0: [double]
            alpha: [double]
             beta: [double]
```

```
            fac_fi: [double]
           fac_fij: [double]
    fac_fi_nonzero: [struct]
   fac_fij_nonzero: [struct]
  matrix_1st_order: [double]
  matrix_2nd_order: [double]
                Si: [double]
        Si_ranking: [double]
               Sij: [double]
       Sij_ranking: [double]
          R2_pears: [double]
  R2_pears_ranking: [double]
          R2_spear: [double]
  R2_spear_ranking: [double]
       beta_number: [double]
```

`f0`
Contains the mean values for all considered outputs in a row vector.

`alpha`
Contains the α coefficients for the orthonormal polynomials for the approximation of the first order HDMR component functions in a 3D matrix.

`beta`
Contains the β coefficients for the orthonormal polynomials for the approximation of the second order HDMR component functions in a 3D matrix.

`fac_fi`
Contains the optimal polynomial order for the approximation of the first order HDMR component functions in a 2D matrix.

`fac_fij`
Contains the optimal polynomial order for the approximation of the second order HDMR component functions in a 2D matrix.

`fac_fi_nonzero`
Contains a set of structure arrays. Each structure array corresponds to one output and contains the optimal polynomial order for the approximation of each of the first order HDMR component functions which are non-zero in a 2D array.

`fac_fij_nonzero`
Contains a set of structure arrays. Each structure array corresponds to one output and contains the optimal polynomial order for the approximation of each of the second order HDMR component functions which are non-zero in a 2D array.

`matrix_1st_order`
Contains the number of the first order component functions which have been approximated by orthonormal polynomials and by what order in a 2D matrix.

`matrix_2nd_order`
Contains the number of the second order component functions which have been approximated by orthonormal polynomials and by what order in a 2D matrix.

`Si`
Contains the first order variance based sensitivity indices for all input parameters in a 2D matrix.

`Si_ranking`
Contains the ranking based on the first order variance based sensitivity indices for the input parameters in a 2D matrix.

`Sij`
Contains the second order variance based sensitivity indices for all input parameter combinations in a 2D matrix.

`Sij_ranking`
Contains the ranking based on the second order variance based sensitivity indices for the input parameter combinations in a 2D matrix.

`R2_pears`
Contains the Pearson correlation coefficient for each of the input parameters in a 2D matrix.

`R2_pears_ranking`
Contains the ranking based on the Pearson correlation coefficient for the input parameters in a 2D matrix.

`R2_spear`
Contains the Spearman rank correlation coefficient for each of the input parameters in a 2D matrix.

`R2_spear_ranking`
Contains the ranking based on the Spearman rank correlation coefficient for the input parameters in a 2D matrix.

`beta_number`
Contains the possible combinations of orthonormal polynomials up to tenth order in a 2D matrix (used only internally).

In order to access a field in a structure array, one has to specify the name of the field. For example to access the first order sensitivity indices S_i one has to type `rs_hdmr.Si` into the matlab command window after loading the results file.

If using the **File - Save results as...** option in the main window, *Results* window or *Global Sensitivity analysis window* then ASCII text files are also created in the current folder using the given file name as specified:

`<filename>_Si.txt`
Contains the first order sensitivity indices for all input parameters and the ranking.

`<filename>_Sij.txt`
Contains the second order sensitivity indices for all input parameter combinations and the ranking.

`<filename>_R2_Pears.txt`
Contains the Pearson correlation coefficient for each of the input parameters and the ranking.

`<filename>_R2_Spear.txt`
Contains the Spearman rank correlation coefficient for each of the input parameters and the ranking.

`<filename>_fac_fi.txt`
Contains the optimal polynomial order for the approximation of the first order HDMR component functions.

`<filename>_fac_fij.txt`
Contains the optimal polynomial order for the approximation of the second order HDMR component functions.

3.2 Script Based Approach

The proposed GUI-HDMR software can also be used without the graphical user interface by calling the corresponding functions in a Matlab m-file. Various template files are provided.

3.2.1 Simple Example

A simple example in order to calculate the sensitivity indices without using variance reduction methods and without applying a threshold would look like this:

```
clear all;

% load sample input and output
sample_input=load('input_values.txt');
sample_output=load('output_values.txt');
```

```
% Number of samples to be used
N = 1024;
sample_input=sample_input(1:N,:);
sample_output=sample_output(1:N,:);

% set maximum polynomial order
max_1st = 10;
max_2nd = 5;

% calculate f0, alpha and beta
[alpha,f0]=sub_alpha_1st(sample_input,sample_output,max_1st);
[beta_numbers] = sub_beta_numbers;
[beta]=sub_beta_2nd(sample_input,sample_output,beta_numbers,max_2nd);

% calculate optimal polynomial order
[opt_1st]=sub_opt_order_1st(f0,alpha,sample_input,sample_output,max_1st);
[approx1st] = sub_comp_opt_1st(alpha,f0,opt_1st,sample_input,sample_output);
[opt_2nd]=sub_opt_order_2nd(beta,sample_input,sample_output,...
          approx1st,beta_numbers,max_2nd);

% calculate first and second order sensitivity indices
[D,Di,Dij,Si,Sij] = sub_sensitivity_indices(alpha,beta,opt_1st,...
                    opt_2nd,sample_input,sample_output);
```

The sample size N has to be smaller or equal to the number of samples provided by the two files. At first the mean (f_0) and the expansion coefficients (α and β) for the orthonormal polynomials are calculated up to the maximum order provided by `max_1st` and `max_2nd`. Then the optimal order for all of the component functions is calculated and finally the overall variance (D), the partial variances (D_i, D_{ij}) and the sensitivity indices (S_i, S_{ij}) are computed.

3.2.2 Template Files

Four template files are provided in order to calculate the first and second order variance based sensitivity indices:

1. `template_simple.m`
   Similar to the code shown in the latter section. No variance reduction methods are used and no threshold is applied.

2. `template_th.m`
   A threshold is applied and can be defined separately for first and second order to exclude unimportant component functions.

3. `template_varred.m`

   A variance reduction method is applied. The user can choose between the correlation method and the ratio control variate method. The number of iterations can be set separately for first and second order.

4. `template_varred_th.m`

   Both, a variance reduction method and a threshold, are applied.

4 Analytical Test Models

The HDMR software has been extensively tested using analytical functions and has been successfully applied to a number of case studies (see for example Ziehn and Tomlin (2008a) or Ziehn and Tomlin (2008b)).

Sample input and output files for two analytical models which are widely used in the literature are supplied with the GUI-HMDR software. Both models are highly non-linear, non-monotonic and have significant second order effects (parameter interactions). The Sobol' sequence (Sobol, 1976) which typically contains $N = 2^m$ points (m integer) has been used to generate the sample input files for the two models.

4.1 Ishigami Function

The first analytical model has three input parameters and one output and is known as the Ishigami function (Ishigami and Homma, 1990):

$$f(\mathbf{x}) = \sin(x_1) + a \sin^2(x_2) + b x_3^4 \sin(x_1) \tag{9}$$

where x_i is uniformly distributed within $(-\pi, \pi)$ and the constants $a = 7, b = 0.1$. The sensitivity indices can be calculated analytically and are given in table 1 for comparison. The interesting thing about the model is the dependence on x_3, which has no additive effect (first order) on the model output but contributes via interaction with x_1.

GUI-HDMR setup:

- Sample input file: `ishigami_inputs.txt`

- Sample output file: `ishigami_output.txt`

- Input ranges (optional for plots): `ishigami_ranges.txt`

- Sample size for HDMR analysis: $N = 512$

- Maximum order for approximation of first order component functions: 10

- Maximum order for approximation of second order component functions: 5

- Variance Reduction method: Correlation method

- Iterations for the first order component functions: 10

- Iterations for the second order component functions: 10

- No threshold has been used.

Method	S_1	S_2	S_3	$\sum S_i$	S_{12}	S_{13}	S_{23}	$\sum S_i + \sum S_{ij}$
Analytical	0.3139	0.4424	0	0.7563	0	0.2437	0	1.0000
GUI-HDMR	0.3171	0.4391	0	0.7562	0	0.2418	0	0.9980

Table 1: First and second order sensitivity indices for the Ishigami function.

As shown in table 1 a small sample size of $N = 512$ is already sufficient in order to accurately estimate the first and second order sensitivity indices. Parameter 3 and the interactions between parameter 1,2 and 2,3 have been correctly excluded from the HDMR expansion since the polynomial order has been automatically set to zero. Obviously they have no contribution at all towards the overall output variance.

4.2 Sobol' g-function

The second analytical test model is the Sobol' g-function (Saltelli and Sobol, 1995):

$$f(\mathbf{x}) = \prod_{i=1}^{n} g_i(x_i), \tag{10}$$

where

$$g_i(x_i) = \frac{\mid 4x_i - 2 \mid + a_i}{1 + a_i} \tag{11}$$

with x_i uniformly distributed within $(0, 1)$. The value of a_i determines the role for the corresponding input parameter x_i, the higher the value of a_i the lower the sensitivity of the input parameter x_i. In this case we are considering eight input parameters and one output. The spectrum of a_i has been chosen according to Ratto et al. (2007) as $a_i = \{4.5, 4.5, 1, 0, 1, 9, 0, 9\}$.

GUI-HDMR setup:

- Sample input file: `gfunction_inputs.txt`

- Sample output file: `gfunction_output.txt`

- Sample size for HDMR analysis: $N = 4\,096$

- Maximum order for approximation of first order component functions: 10

- Maximum order for approximation of second order component functions: 5

- Variance Reduction method: Correlation method

- Iterations for the first order component functions: 10

- Iterations for the second order component functions: 10

- No threshold has been used.

A comparison of the analytical results and the results obtained by using the GUI-HDMR software is shown in table 2 for the first order sensitivity indices and in table 3 for the second order sensitivity indices. The second order sensitivity indices are only shown for the five most important parameter interactions. A smaller sample size of $N = 1\,024$ would be sufficient in order to get the correct importance ranking for the parameters and parameter interactions. However, a sample size of $N = 4\,096$ is necessary in order to obtain more accurate values for each of the sensitivity indices.

Method	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	$\sum S_i$
Analytical	0.0096	0.0096	0.0727	0.2906	0.0727	0.0029	0.2906	0.0029	0.7516
GUI-HDMR	0.0098	0.0127	0.0707	0.2834	0.0731	0.0029	0.2828	0.0033	0.7386

Table 2: First order Sensitivity indices for the Sobol' g-function.

Method	S_{47}	S_{34}	S_{37}	S_{45}	S_{57}	$\sum S_{ij}$	$\sum S_i + \sum S_{ij}$
Analytical	0.0969	0.0242	0.0242	0.0242	0.0242	0.2209	0.9725
GUI-HDMR	0.0922	0.0240	0.0244	0.0244	0.0226	0.2246	0.9633

Table 3: Second order Sensitivity indices for the Sobol' g-function.

A more detailed comparison of the analytical results and the results calculated by the GUI-HDMR software using different sample sizes N is given in Ziehn and Tomlin (2008c). The application of the threshold for the Sobol' g-function using 50 input parameters is also given and plots of selected first and second order component functions are shown. The sample input file and the corresponding output file for this test case are provided with the GUI-HDMR software (`gfunction_inputs_50inp.txt` and `gfunction_output_50inp.txt`).

5 Quick Guide

This quick guide describes the function of all buttons and menu options for all windows of the GUI-HDMR software in an overview.

5.1 Main Window

Starts HDMR analysis and shows progress.

- Menu:
  - **File - Save results as...** : Save all results in external file.
  - **File - Load results** : Load results from external file.
  - **Help - Main Help** : Help to the main window.
  - **Help - About** : Version number and contact details.

- Buttons:
  - **Setup** : Open the Setup window.
  - **Run** : Start HDMR analysis.
  - **Results** : Open the Results window.
  - **SA** : Open the Global Sensitivity Analysis window.
  - **Exit** : Exit the GUI-HDMR program and save all results in external file `workspace.mat`.

5.2 Setup Window

Load external files and apply settings for HDMR analysis.

- Menu:
  - **File - Open - Input-File (for HDMR analysis)** : Load sample input values from external file.
  - **File - Open - Output-File (for HDMR analysis)** : Load sample output values from external file.
  - **File - Open - Input-File (for accuracy test)** : Load additional sample input values from external file.
  - **File - Open - Output-File (for accuracy test)** : Load additional sample output values from external file.

– **File - Open - Input Ranges (for plots)** : Load input parameter ranges from external file.

– **Help - Setup Help** : Help to the Setup window.

– **Help - About** : Version number and contact details.

- Buttons:

  – **Default** : Apply the default setting.

  – **Ok** : Save settings and exit Setup window.

5.3 Results Window

Show optimal order of the applied polynomials and calculate relative error to verify accuracy of the metamodel.

- Menu:

  – **File - Save plot a (left)** : Opens the **Save** dialog in order to save the current plot on the left hand side as `.png`, `.pdf`, `.eps` and `.fig`.

  – **File - Save plot b (right)** : Opens the **Save** dialog in order to save the current plot on the right hand side as `.png`, `.pdf`, `.eps` and `.fig`.

  – **File - Save results as...** : Save all results in external file.

  – **File - Load results** : Load results from external file.

  – **Help - Results Help** : Help to the Results window.

  – **Help - About** : Version number and contact details.

- Buttons - Results:

  – **1st-order** : Show optimal order of the polynomials used for the approximation of the first order component functions.

  – **2nd-order** : Show optimal order of the polynomials used for the approximation of the second order component functions.

- Buttons - Accuracy:

  – **1st-order** : Show the accuracy of the first order HDMR metamodel in terms of the relative error, coefficient of determination, scatter plot (original model vs. HDMR metamodel) and empirical pdf plot (or histogram if statistics toolbox is not available).

– **2nd-order** : Show the accuracy of the second order HDMR metamodel in terms of the relative error, coefficient of determination, scatter plot (original model vs. HDMR metamodel) and empirical pdf plot (or histogram if statistics toolbox is not available).

- Buttons:

    – **Exit** : Exit Results window.

5.4 Global Sensitivity Analysis Window

Show first and second order sensitivity indices and plot component functions.

- Menu:

    – **File - Save plot a (left)** : Opens the **Save** dialog in order to save the current plot on the left hand side as `.png`, `.pdf`, `.eps` and `.fig`.

    – **File - Save plot b (right)** : Opens the **Save** dialog in order to save the current plot on the right hand side as `.png`, `.pdf`, `.eps` and `.fig`.

    – **File - Save results as...** : Save all results in external file.

    – **File - Load results** : Load results from external file.

    – **Help - SA Help** : Help to the Global Sensitivity Analysis window.

    – **Help - About** : Version number and contact details.

- Buttons - Results:

    – **1st-order** : Show the first order sensitivity indices ranked by their importance for the five most important parameters.

    – **2nd-order** : Show the second order sensitivity indices ranked by their importance for the five most important parameter interactions.

- Buttons - Plots:

    – **1st-order** : Plot the first order component functions.

    – **Scatter** : Show the scatter plot for chosen input-output combination.

    – **HDMR** : Show the plot of the first order component function for chosen input-output combination.

    – **Both** : Show the scatter plot and the plot of the first order component function for chosen input-output combination in one figure for comparison.

    – **2nd-order** : Plot the second order component functions (pre-view with only 11 grid points).

- **HDMR** : Show the plot of the second order component function for chosen inputs-output combination (combined with contour plot and higher resolution using 33 grid points).

- Buttons:

  - **Exit** : Exit Global Sensitivity Analysis window.

References

Ali�, Ö.F., Rabitz, H., 2001. Efficient Implementation of High Dimensional Model Representations. Journal of Mathematical Chemistry 29, 127 - 142.

Ishigami, T., Homma, T., 1990. An importance quantification technique in uncertainty analysis for computer models. In: Proceedings of the ISUMA'90, First International Symposium on Uncertainty Modelling and Analysis, 398-403.

Kalos, M. H., Whitlock, P. A., 1986. Monte Carlo Methods. Volume I: Basics. John Wiley & Sons.

Kucherenko, S., 2007. Application of global sensitivity indices for measuring the effectiveness of quasi-monte carlo methods and parameter estimation. In: Proceedings of the fifth International Conference on Sensitivity Analysis of Model Output (SAMO 2007), 35-36.

Li, G., Rosenthal, C., Rabitz, H., 2001. High Dimensional Model Representations. Journal of Physical Chemistry A 105, 7765 -7777.

Li, G., Wang, S.-W., Rabitz, H., 2002. Practical Approaches To Construct RS-HDMR Component Functions. Journal of Physical Chemistry A 106, 8721-8733.

Li, G., Wang, S.-W., Rabitz, H., Wang, S., Jaffé, P., 2002. Global uncertainty assessments by high dimensional model representations (HDMR). Chemical Engineering Science 57, 4445-4460.

Li, G., Rabitz, H., Wang, S.-W., Georgopoulos, P. S., 2003. Correlation Method for Variance Reduction of Monte Carlo Integration in RS-HDMR. Journal of Computational Chemistry 24, 277-283.

Li, G., Rabitz, H., 2006. Ratio control variate method for efficiently determining high-dimensional model representations. Journal of Computational Chemistry 27, 1112 - 1118.

Moore, D. S., 2007. The Basic Practice of Statistics. New York, NY : W.H. Freeman and Company.

Morris, M. D., 1991. Factorial Sampling Plans for Preliminary Computational Experiments. Technometrics 33, 161-174.

Rabitz, H., Ali�, Ö. F., Shorter, J., Shim, K., 1999. Efficient input–output model representations. Computer Physics Communications 117, 11-20.

Rabitz, H., Ali�, Ö. F., 1999. General foundations of high-dimensional model representations. Journal of Mathematical Chemistry 25, 197-233.

Ratto, M., Pagano, A., Young, P., 2007. State Dependent Parameter metamodelling and sensitivity analysis. Computer Physics and Communications 177, 863-876.

Saltelli, A., Sobol',I. M., 1995. About the use of rank transformation in sensitivity analysis of model output. Reliability Engineering & System Safety 50, 225-239.

Sobol, I. M., 1976. On the distribution of points in a cube and the approximate evaluation of integrals. USSR Computational Mathematics and Mathematical Physics 7, 86-112.

Sobol, I. M., 2001. Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates, Mathematics and Computers in Simulation 55, 271 - 280.

Ziehn, T., Tomlin, A. S., 2008a. Global sensitivity analysis of a 3D street canyon model - Part I: The developemnt of high dimensional model representations. Atmospheric Environment 42, 1857-1873.

Ziehn, T., Tomlin, A. S., 2008b. A global sensitivity study of sulphur chemistry in a pre-mixed methane flame model using HDMR. International Journal of Chemical Kinetics 40, 742-753.

Ziehn, T., Tomlin, A. S., 2008c. GUI-HDMR - A Software Tool for Global Sensitivity Analysis of Complex Models. Environmental Modelling & Software, accepted for publication.

Ziehn, T., Hughes, K. J., Griffiths, J. F., Porter, R., Tomlin, A. S., 2008a. A global sensitivity study of cyclohexane oxidation under low temperature fuel rich conditions using HDMR methods. Combustion Theory and Modelling, under review.