

Development and Application of Global Sensitivity Analysis Methods in Environmental and Safety Engineering

Tilo Ziehn

Submitted in accordance with the requirements of the degree of PhD

THE UNIVERSITY OF LEEDS

School of Process, Environmental and Materials Engineering

November 2008

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others. This copy has been supplied on the understanding that it is copyright material and that no quotation from the thesis may be published without proper acknowledgement.

Acknowledgements

Above all, I would like to express my sincere gratitude to Professor Alison Tomlin, who has been my supervisor. She provided me with many helpful suggestions, important advice and constant encouragement during this work.

I would also like to thank Dr. Nick Dixon for useful discussions and constructive advice. Furthermore I am indebted to my former and present colleagues Dr. Emily Agus, Dr. David Young, Dr. Ahmed Balogun and Mr. Alex Charlton for providing a great work environment and for their encouragement and conversations. Also, I would like to thank Mr. Christian Groh for helpful discussions, support and close friendship.

I am very grateful to my girlfriend Amy for her encouragement, understanding and support.

Finally I would like to thank my parents for their never ending support over all these years and for always being there when needed.

Abstract

As computing power increases and data relating to elementary chemical and physical processes improves, the use of computational modelling as a design tool in environmental and safety engineering is becoming increasingly important. Sensitivity analysis (SA) and uncertainty analysis (UA) can help to gain a better physical insight into the model and they can highlight possible discrepancies between experimental results and model predictions. Global methods are the best approach for this purpose, however using current methodologies their calculation consumes large amounts of computational effort. Therefore, efficient methods for global UA and SA based on an approach called high dimensional model representation (HDMR) are developed in this work. Two commonly used HDMR methods, cut-HDMR and random sampling (RS)-HMDR, are introduced and compared against each other by applying them to an analytical test function and a case study investigating the flow field in a 3D street canyon using a micro-scale Reynolds averaged computational fluid dynamics (CFD) model. It is shown that discrepancies between experimental velocity and turbulence profiles and model predictions exist and that this is not only due to the uncertainties in the four considered input parameters. It is also demonstrated that RS-HDMR has various advantages over cut-HDMR and therefore it is used for all further studies exclusively. Extensions to the existing set of RS-HDMR tools are developed in order to make this approach even more suitable for the requirements that arise in environmental and safety engineering. The extended RS-HDMR approach is then applied to other case studies. The focus is on a turbulent atmospheric plume of nitrogen oxides reacting with background ozone using a Lagrangian stochastic model. Uncertainties in 22 physical and chemical parameters are investigated. It is shown that the mixing model used within the Lagrangian scheme has a significant influence on the predicted concentrations. Furthermore, a

model of sulphur-nitrogen chemistry interactions in a one dimensional low pressure premixed methane flame is studied considering uncertainties in 176 parameters (rate constants and enthalpies of formation). Here, it is illustrated that only a few input parameters affect the predicted final nitrogen oxide mole fraction under fuel rich conditions and that the ranking of the important parameters critically depends on the uncertainty ranges chosen. Low temperature isothermal cyclohexane oxidation under fuel rich conditions is also investigated considering uncertainties in 238 rate constants. This case study has been a challenge for the extended HDMR method because significant higher order effects exist. The experimental form of reactant consumption (quadratic autocatalysis) could not be recovered taking into account the uncertainties in the rate constants only and therefore wall losses of important radicals were included in the chemical scheme. In this way it was possible to recover the experimental form of reactant consumption. The application of the extended RS-HDMR method has also led to the development of the software package GUI-HDMR which is freely available and can be easily applied by interested users. A summary of software features is provided.

Contents

Acknowledgements	i
Abstract	ii
Table of Contents	iv
Nomenclature	x
List of Figures	xiv
List of Tables	xxiii
1 Introduction	1
1.1 Motivation	2
1.2 Aim and Objectives	5
1.3 Outline of the Thesis	6
2 Uncertainty and Sensitivity Analysis	10
2.1 Overview of Sensitivity Analysis Methods	11
2.2 Local Methods	13
2.3 Screening Methods	14
2.3.1 One at a Time (OAT) Design	15
2.3.2 Morris Method	16

2.4	Global Methods	19
2.4.1	Overview	19
2.4.2	Sampling	20
2.4.3	Monte Carlo Analysis	28
2.4.4	Response Surface Method (RSM)	36
2.4.5	Method of Sobol'	40
2.4.6	Fourier Amplitude Sensitivity Test (FAST)	45
2.5	Summary	49
3	High Dimensional Model Representation (HDMR)	52
3.1	Cut-HDMR	54
3.1.1	Reference Point	57
3.1.2	Multicut-HDMR	58
3.2	RS-HDMR	58
3.2.1	Computation of the Higher Order Component Functions	60
3.2.2	Variance Reduction Methods	66
3.3	Accuracy of the HDMR Metamodel	71
3.3.1	Probability Density Function Estimate and Empirical Cumulative Distribution Function	73
3.3.2	Residual Variance	75
3.3.3	Relative Error	75
3.3.4	Scatter Plots and Coefficient of Determination	76
3.3.5	Sample Plots	78
3.4	Sensitivity Analysis Using HDMR	79
3.4.1	Sensitivity Indices based on cut-HDMR	80
3.4.2	Sensitivity Indices based on RS-HDMR	83
3.4.3	Plot of the HDMR Component Functions	85
3.5	Summary	87

3.5.1	HDMR Approaches	87
3.5.2	HDMR in Comparison to Other Global SA Methods	89
4	Comparison of cut-HDMR and RS-HDMR in Practice	92
4.1	Optimisation Method for RS-HDMR	92
4.2	Test Model: Ishigami Function	95
4.2.1	Properties of the Ishigami Function	96
4.2.2	Application of cut-HDMR	97
4.2.3	Application of RS-HDMR	99
4.2.4	Discussion and Conclusions	102
4.3	Case Study: Simulation of the Turbulent Flow Field in a 3D Street Canyon	103
4.3.1	Introduction	103
4.3.2	Model Description	106
4.3.3	Results	114
4.3.4	Discussion	126
4.3.5	Comparison of cut-HDMR and RS-HDMR	127
4.3.6	Conclusions	128
4.4	Summary	129
5	Application of RS-HDMR for Systems with a Large Number of Input Parameters	131
5.1	Threshold to Exclude Component Functions	132
5.2	Test Model: Sobol' g-function	134
5.2.1	Properties of the Sobol' g-function	134
5.2.2	Application of RS-HDMR	135
5.2.3	Discussion	138
5.3	Introduction to Chemical Kinetics	138

5.3.1	Rate of Reaction	140
5.3.2	Rate Law	140
5.3.3	Pressure Dependence of Reactions	142
5.3.4	Temperature Effect - Arrhenius Equation	143
5.3.5	Photolysis	143
5.3.6	Chemical Thermodynamics	144
5.3.7	Summary	145
5.4	Case Study: Reactive Plume Model	145
5.4.1	Introduction	147
5.4.2	Model Validation and Uncertainty Analysis	149
5.4.3	Extended Chemical Mechanism	152
5.4.4	Results	155
5.4.5	Conclusions	163
5.5	Case Study: Premixed Methane Flame Model	164
5.5.1	Introduction	165
5.5.2	Model Description	167
5.5.3	Investigated Scenarios and Experimental Setup	169
5.5.4	Results and Discussions	173
5.5.5	Conclusions	185
5.6	Summary	187
6	Application of RS-HDMR for Systems with Significant Parameter Interactions	189
6.1	Need for Screening	189
6.2	Test Model: Sobol' g-function	190
6.2.1	Properties of the Sobol' g-function	191
6.2.2	Application of RS-HDMR	192
6.2.3	Combination of Screening and RS-HDMR	194

6.2.4	Discussion	198
6.3	Case Study: Cyclohexane Oxidation	199
6.3.1	Introduction	200
6.3.2	Evaluation of the Mechanism and Initial Comparison with Ex- periment	204
6.3.3	Selected Inputs and Outputs for Sensitivity Analysis	206
6.3.4	Results	207
6.3.5	Conclusions	225
6.4	Summary	227
7	GUI-HDMR Software	230
7.1	Introduction	230
7.1.1	Graphical User Interface (GUI)	231
7.1.2	Script Based Approach	233
7.2	Summary	233
8	Limitations and Extensions of the RS-HDMR Approach and Future Work	235
8.1	Investigation of Higher Order Effects	235
8.1.1	Transformation of the Model Output	236
8.1.2	Inclusion of Third-Order Effects	239
8.1.3	Conclusions	241
8.2	Non-Uniform Input Distributions	241
8.3	Non-Independent Input Parameters	243
8.4	Extensions and Improvements of the GUI-HDMR Software	244
8.5	Summary	245
9	Summary and Conclusions	247

A Mathematical Background and Additional Information	258
A.1 Example for the Conversion of cut-HDMR Component Functions	258
A.2 Shifted Legendre Polynomials	259
B GUI-HDMR Software Documentation	261
B.1 Graphical User Interface (GUI)	262
B.1.1 Setup	263
B.1.2 Construction of the HDMR Expansion	266
B.1.3 Results	267
B.1.4 Sensitivity Analysis	269
B.2 Script Based Approach	272
B.2.1 Simple Example	272
B.2.2 Template Files	274
Bibliography	275

Nomenclature

Terminology

Metamodel Approximation of the real model (e.g. HDMR expansion).

Parameter Is defined here as the constants and independent variables which define a mathematical equation or model, e.g. $y = ax_1 + bx_2$ the independent variables x_1, x_2 and the constants a, b are all parameters.

Real model Refers to the model (computer code) which is used for the simulations, also referred to as original model.

Surrogate Approximation of the real model, which is used as a fast equivalent instead of the real model in order to perform simulations.

Abbreviations

AE Absolute Error

ANOVA Analysis of Variance

ccdf Complementary Cumulative Distribution Function

cdf Cumulative Distribution Function

CFD Computational Fluid Dynamics

Defra Department for Environment, Food and Rural Affairs

DNS Direct Numerical Simulation

EU European Union

FAST Fourier Amplitude Sensitivity Test

GUI Graphical User Interface

HDMR High Dimensional Model Representation

LES Large Eddy Simulation

MC Monte Carlo

OAT One At A Time

pdf Probability Density Function

RANS Reynolds Averaged Navier Stokes

RE Relative Error

RS Random Sampling

RSM Response Surface Method

SA Sensitivity Analysis

TKE Turbulent Kinetic Energy

UA Uncertainty Analysis

WHO World Health Organization

Latin Symbols

A Arrhenius factor

c_0	Structure function coefficient
D	Overall variance
D_i, D_{ij}	Partial variances
E	Activation energy
$f(x)$	Model output, also y
f_0	Mean effect (RS-HDMR)
$f_i(x_i)$	First-order HDMR component function
$f_{ij}(x_i, x_j)$	Second-order HDMR component function
h	Planck's constant
k	Turbulent kinetic energy
K^n	Unit hypercube
L	Order of the HDMR expansion
m	Number of grid points (cut-HDMR)
N	Sample size
n	Input space dimension (number of input parameters)
R	Gas constant
r	Pearson correlation coefficient
r^2	Coefficient of determination
R^n	Domain
r_s	Spearman rank correlation coefficient

S_i	First-order sensitivity index
S_{ij}	Second-order sensitivity index
T	Temperature
t_m	Mixing time scale
TS	Total sensitivity index
u, v, w	Wind components
u_*	Friction velocity
U_m	Mean wind speed measurement
\bar{x}	Reference point (cut-HDMR)
x	Parameter
y	Model output, also $f(x)$
z	Solar zenith angle
z_0	Surface roughness length

Greek Symbols

α	Mixing time scale coefficient
α_r^i	Expansion coefficient for the RS-HDMR basis functions (e.g. orthonormal polynomials) in order to approximate the first-order component functions.
β_{pq}^{ij}	Expansion coefficient for the RS-HDMR basis functions (e.g. orthonormal polynomials) in order to approximate the second-order component functions.

δ_{ij}	Kronecker delta
ϵ	Turbulent dissipation rate
κ	Von Karman's constant
μ	Mean
ν	Photons frequency
ω	Frequency
$\varphi(x)$	RS-HDMR basis function (e.g orthonormal polynomial)
σ	Standard Deviation
θ	Background wind direction

List of Figures

2.1	Possible classification of sensitivity analysis methods.	12
2.2	Example of a trajectory for two parameters ($n = 2$) and a four-level grid ($p = 4$).	17
2.3	Overview of various classes of SA methods according to two properties: the amount of information produced in terms of the model sensitivity and the computational cost of the method. The computational cost is measured in terms of the number of required model evaluations and depends on the number of input parameters n and the complexity of the model.	18
2.4	Example of a Latin hypercube sequence for two parameters ($n = 2$) and a samples size of $N = 4$	22
2.5	(a) 1024 random sampling points, (b) 1024 Latin hypercube sampling points, (c) 1024 points of the Halton sequence, (d) 1024 points of the Sobol' sequence. All sampling procedures are based on a uniform distribution in the interval $[0, 1]$	25
2.6	Plot of (a) the mean and (b) variance for the simple test model using different sampling procedures.	27
2.7	(a) Uniform distribution with $a = 2$ and $b = 4$, (b) Normal distribution with $\mu = 0$ and $\sigma = 1$	30

2.8	Scatter plots showing (a) a strong positive linear relationship, (b) a strong negative linear relationship.	32
2.9	Scatter plots showing (a) a strong nonlinear (monotonic) relationship, (b) a strong nonlinear (non-monotonic) relationship.	32
3.1	(a) Calculation of $f_1(x_1)$ along the x_1 -axis through the reference point, (b) calculation of $f_2(x_2)$ along the x_2 -axis through the reference point, (c) calculation of $f_{12}(x_1, x_2)$ along the surface of x_1 and x_2 through the reference point.	56
3.2	Example probability density function (pdf) estimate using a (a) first- and (b) second-order HDMR metamodel.	74
3.3	Example empirical cumulative distribution function (cdf) using a (a) first- and (b) second-order HDMR metamodel.	74
3.4	Example scatter plot for the model output and (a) the first-order HDMR metamodel and (b) the second-order HDMR metamodel. . . .	77
3.5	Comparison of the response from the original model and (a) the first- order HDMR metamodel and (b) the second-order HDMR metamodel.	79
3.6	(a) Example of a first-order cut-HDMR component function and (b) a second-order cut-HDMR component function with linear interpola- tion between the exact solutions (indicated by the red circles) using $m = 11$ grid points.	85
3.7	(a) Example of a first-order RS-HDMR component function and (b) a second-order RS-HDMR component function using second-order or- thonormal polynomials.	86
4.1	First-order component function (a) for input parameter x_1 and (b) for input parameter x_2 using a grid size of $m = 21$	98

4.2	Second-order component function for input parameters x_1 and x_3 using a grid size of $m = 21$	99
4.3	First-order component function (a) for input parameter x_1 (approximated by a 7th order polynomial) and (b) for input parameter x_2 (approximated by a 8th order polynomial) using $N = 512$	101
4.4	Second-order component function for input parameters x_1 and x_3 approximated by 4th order polynomials using $N = 512$	102
4.5	Flow regimes associated with perpendicular background winds in street canyons with different aspect ratios.	105
4.6	Schematic of the recirculating flow showing the transport of pollutants to the leeward side of the street canyon.	107
4.7	Grid and building configuration around Gillygate.	108
4.8	Normalised velocity components against background wind direction from experimental data and model data for G4.	110
4.9	Normalised TKE against background wind direction θ from experimental data and model data for G3.	111
4.10	Normalised TKE against background wind direction θ from experimental data and model data for G4.	111
4.11	Street canyon un-normalised TKE and wind vector cross-section for a background wind direction of 90°	113
4.12	Comparison of the output response from the MISKAM model and (a) the first-order and (b) the second-order cut-HDMR metamodel for the normalised wind component v/U_m at the measurement point G4.	116
4.13	Comparison of different order approximations for first-order component functions for input parameter (a) x_4 (background wind direction) and (b) x_1 (inflow roughness length) and model output G4 w/U_m . . .	117

4.14 Comparison of the statistics between the MISKAM flow model and the second-order cut-HDMR metamodel.	119
4.15 Scatter plot of x_4 (background wind direction) and (a) output G4 w/U_m (normalised vertical wind velocity) and (b) Mast TKE/ U_m^2 . . .	120
4.16 First-order component function (a) for input parameter x_4 (background wind direction) and output G4 w/U_m and (b) for input parameter x_4 (background wind direction) and output Mast TKE/ U_m^2 . . .	121
4.17 Second-order component function for input parameters x_1 (inflow roughness length) and x_4 (background wind direction) and output G4 TKE/ U_m^2	122
5.1 First order component function (a) for input parameter x_1 (approximated by a 10th order polynomial) and (b) for input parameter x_2 (approximated by a 6th order polynomial) using a sample size of $N = 1024$	137
5.2 Radial profile of mean $[NO_x]$ at $x = 2.2\text{ m}$, $x = 3.8\text{ m}$ and $x = 4.8\text{ m}$ from the point source obtained form simulations using the nominal values for σ , α and c_0 and from wind tunnel experiments. Error bars are based on 400 simulations varying σ , α and c_0 in their assigned ranges.	150
5.3 Accuracy of the RS-HDMR metamodel based on $N = 512$ model runs for $[NO_x]$ at the plume centre ($x = 4.8\text{ m}$) showing (a) the scatter plot between the original model response and the first-order RS-HDMR metamodel ($r^2 = 98.93\%$) and (b) the probability density function estimate for both the original model and the first-order RS-HDMR metamodel.	156

5.4	First-order component functions and scatter plots for (a) the structure function coefficient c_0 and (b) the mixing time scale coefficient α with respect to $[NO_x]$ at the plume centre for $x = 2.2\text{ m}$. The mean f_0 is added to f_i for comparison with the scatter plot.	158
5.5	Error bars for peak $[NO_x]$ at 2.2 m from the point source using a: $c_0 = [3 \dots 6]$ and b: $c_0 = [4 \dots 6]$	159
5.6	First-order component functions and scatter plots for (a) E/R for $NO + O_3 \longrightarrow NO_2$ and (b) the mixing time scale coefficient α with respect to $[O_3]$ at the plume centre for $x = 2.2\text{ m}$. The mean f_0 is added to f_i for comparison with the scatter plot.	161
5.7	Comparison between model simulations and experimental measurements for the relative change in NO concentration on the addition of varying amounts of SO_2 to the fuel.	168
5.8	Accuracy of the RS-HDMR metamodel based on $N = 1024$ model runs for NO concentration (mole fraction) showing (a) the scatter plot between the original model response and the first-order RS-HDMR metamodel ($r^2 = 98.50\%$) and (b) the probability density function estimate for both the original model and the first-order RS-HDMR metamodel.	175
5.9	First-order component functions and scatter plots for (a) $SO + NH = NO + SH$, (b) $SO + N = NO + S$, (c) $SO + OH = SO_2 + H$, (d) SO (scenario 3).	178
5.10	Second-order component function $f_{ij}(x_i, x_j)$ for x_i : $SO + OH = SO_2 + H$ and x_j : $SO + NH = NO + SH$ (scenario 3).	180
5.11	(a) First-order component function f_i and (b) scatter plots for SN (scenario 2). The mean f_0 is added to f_i to make the plot comparable with the scatter plot.	184

6.1	Results of the Morris analysis using $r = 10$ trajectories ($N = 510$). Only the first eight parameters $x_1 - x_8$ are potentially important. Parameters $x_9 - x_{50}$ are identified as unimportant.	195
6.2	Results of the Morris analysis using 10, 50 and 100 trajectories.	196
6.3	(a) First-order component function and scatter plot for parameter x_4 , (b) second-order component function for the parameter interaction between x_4 and x_7 for the Sobol' g-function using a sample size of $N = 4096$	198
6.4	Theoretical curve of isothermal quadratic autocatalysis in a closed vessel.	201
6.5	Experimental (dashed lines) and numerically simulated fractional O ₂ consumption (a) EXGAS mechanism (b) Silke mechanism (c) Sample simulation using quasi-random inputs for A-factors (d) Sample sim- ulation using quasi-random inputs for A-factors and wall losses. All simulations at $T = 503$ K.	202
6.6	Morris plot for output y_1 (time at point of inflexion) using 100 tra- jectories. The ten most important parameters ranked by their mean are labelled.	208
6.7	Morris plot for output y_2 (O ₂ ratio) using 100 trajectories. The ten most important parameters ranked by their mean are labelled.	209
6.8	Accuracy of the RS-HDMR metamodel for $N = 131\,072$ based on the scatter plot (1000 random points) for (a) the time at point of inflexion using a first-order HDMR metamodel (b) the O ₂ ratio using a first-order HDMR metamodel, (c) the time at point of inflexion using a second-order HDMR metamodel and (d) the O ₂ ratio using a second-order HDMR metamodel.	214

6.9 Accuracy of the RS-HDMR metamodel for $N = 131\,072$ based on the probability density function (pdf) using 1 000 random points for (a) the time at point of inflection using a first-order HDMR metamodel (b) the O ₂ ratio using a first-order HDMR metamodel, (c) the time at point of inflection using a second-order HDMR metamodel and (d) the O ₂ ratio using a second-order HDMR metamodel.	215
6.10 Carbon element flux analysis for isothermal oxidation using the 60 species 238 reactions mechanism: 1:2 cyclohexane in air, 470 K, 1 atm. Arrow thickness scaled to magnitude of element flux.	217
6.11 First-order component function f_i for (a) 2C ₆ H ₁₁ OO → 2C ₆ H ₁₁ O + O ₂ , (b) C ₆ H ₁₁ OOH → C ₆ H ₁₁ O + OH, (c) CHOC ₄ H ₈ COOOH → OH + 3CH ₂ CHO and output y_1 (time at point of inflection). In (c) the mean f_0 is added to f_i for comparison with scatter plot.	219
6.12 First-order component function f_i for (a) 2C ₆ H ₁₁ OO → 2C ₆ H ₁₁ O + O ₂ , (b) C ₆ H ₁₁ OO + HO ₂ → C ₆ H ₁₁ OOH + O ₂ and scatter plot for (c) 2C ₆ H ₁₁ OO → C ₅ H ₁₀ CO + C ₆ H ₁₁ OH + O ₂ and output y_2 (O ₂ ratio). In (c) the mean f_0 is added to f_i for comparison with scatter plot.	220
6.13 Second-order component function f_{ij} for y_2 (O ₂ ratio) for 2C ₆ H ₁₁ OO → C ₅ H ₁₀ CO + C ₆ H ₁₁ OH + O ₂ and 2C ₆ H ₁₁ OO → 2C ₆ H ₁₁ + O ₂	222
8.1 Accuracy of the RS-HDMR metamodel for $N = 2\,048$ based on the probability density function (pdf) using 1 000 random points (a) using a first-order HDMR metamodel and (b) using a second-order HDMR metamodel.	238

8.2 Accuracy of the RS-HDMR metamodel (transformed output using $\log(y)$) for $N = 256$ based on the probability density function (pdf) using 1000 random points (a) using a first-order HDMR metamodel and (b) using a second-order HDMR metamodel.	239
B.1 GUI-HDMR main window.	263
B.2 GUI-HDMR setup window. Input and output files can be loaded and settings can be applied for the HDMR analysis.	265
B.3 GUI-HDMR main window. The summary of the settings is shown and the progress of the calculation.	266
B.4 GUI-HDMR results window. The optimal order is shown for the component functions which were approximated by polynomials. In order to assess the accuracy of the HDMR metamodel the relative error is also stated.	268
B.5 GUI-HDMR 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.	269
B.6 GUI-HDMR 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.	270

List of Tables

2.1	Generation of a Latin hypercube sequence for two parameters ($n = 2$) and four intervals ($N = 4$)	22
2.2	Pearson correlation coefficients calculated for the data sets presented by the scatter plots in figure 2.8 and 2.9.	34
2.3	Spearman rank correlation coefficients calculated for the data sets presented by the scatter plots in figure 2.8 and 2.9.	35
3.1	Relative error for the HDMR metamodel.	76
4.1	Number of terms required for the approximation of the RS-HDMR component functions considering a model with four input parameters and a second-order HDMR expansion.	93
4.2	First- and second-order sensitivity indices for the Ishigami function calculated analytically.	96
4.3	First- and second-order sensitivity indices and the sum of their absolute errors (AE) for the Ishigami function using the cut-HDMR approach.	97
4.4	Sensitivity indices first- and second-order and the sum of their absolute errors (AE) for the Ishigami function using the RS-HDMR approach (including optimisation of the polynomial order).	100
4.5	Input parameters and uncertainty ranges used in this study.	114

4.6	Percentage of points within a given relative error for the cut-HDMR metamodel.	116
4.7	Order of polynomials used for the approximation of first and second-order component functions of the MISKAM flow model using a sample size $N = 512$	118
4.8	Percentage of points within a given relative error for the RS-HDMR model replacement using a sample size of $N = 512$ and optimal-order polynomials.	118
4.9	First- and second-order sensitivity indices calculated on the basis of cut-HDMR using 641 model runs.	123
4.10	First- and second-order sensitivity indices calculated on the basis of RS-HDMR using 512 model runs.	124
5.1	First-order sensitivity indices for the Sobol' g-function calculated analytically.	135
5.2	Second-order sensitivity indices for the Sobol' g-function calculated analytically.	135
5.3	First- and second-order sensitivity indices for the Sobol' g-function using the RS-HDMR approach with various samples sizes N	136
5.4	Range of values for the structure function coefficient c_0 as found in the literature.	151
5.5	Physical and chemical parameter ranges for the Lagrangian stochastic model.	154
5.6	Accuracy of the RS-HDMR metamodel (first- and second-order) in terms of the coefficient of determination r^2 for all considered model outputs (peak NO_x and O_3 concentrations at different distances (x in m) away from the source) using different sample sizes N	155

5.7 First-order sensitivity indices for $[NO_x]$ at the plume centre at different distances x away from the point source using the RS-HDMR approach with a samples size of $N = 512$	157
5.8 First-order sensitivity indices for $[O_3]$ at the plume centre at different distances x away from the point source for very high initial NO concentrations using the RS-HDMR approach with a samples size of $N = 512$. The numbers in brackets indicate the ranking of the parameter.	160
5.9 First-order sensitivity indices for $[O_3]$ at the plume centre at different distances x away from the point source using the RS-HDMR approach with a samples size of $N = 128$ and $N = 512$	162
5.10 First-order sensitivity indices for $[O_3]$ at the plume centre at different distances x away from the point source for lower initial NO concentration using the RS-HDMR approach with a samples size of $N = 512$. The numbers in brackets indicate the ranking of the parameter.	162
5.11 Overview of the three scenarios using different sets of parameter ranges for the 23 enthalpies of formation.	171
5.12 Nominal values and uncertainty ranges for the enthalpies of formation for Scenario 2.	172
5.13 First- and second-order sensitivity indices for scenario 3. When a reaction is indicated, the parameter is the A -factor for the forward reaction rate.	177
5.14 Relative Error calculated for a set of 2000 random samples for the metamodel based on RS-HDMR and a combination of Screening and RS-HDMR.	181
5.15 First-order sensitivity indices for scenario 2. Where a reaction is indicated, the parameter is the A -factor for the forward reaction rate.	181

5.16 First-order sensitivity indices for all 3 scenarios. Where a reaction is indicated, the parameter is the <i>A</i> -factor for the forward reaction rate.	183
6.1 First-order sensitivity indices (first eight parameters) for the Sobol' g-function calculated analytically.	191
6.2 Selected second-order sensitivity indices for the Sobol' g-function calculated analytically.	191
6.3 First-order (first eight inputs) and second-order (selected input combinations) sensitivity indices for the Sobol' g-function using the RS-HDMR approach.	193
6.4 First-order (first eight inputs) and second-order (selected input combinations) sensitivity indices for the Sobol' g-function using the combined screening + RS-HDMR approach.	197
6.5 Potentially important reactions. The reaction number refers to the ranking from the Morris analysis for output y_2	210
6.6 First- and second-order sensitivity indices using $N = 131\,072$ and 33 inputs for output y_1 (time at point of inflexion). Parameters x_i and x_j are the <i>A</i> -factors for the forward reaction rate.	211
6.7 First- and second-order sensitivity indices using $N = 131\,072$ and 33 inputs for output y_2 (O_2 ratio). Parameters x_i and x_j are the <i>A</i> -factors for the forward reaction rate.	211
6.8 Accuracy of the RS-HDMR metamodel (first- and second-order) in terms of the coefficient of determination r^2 for both model outputs (y_1 time at the point of inflexion and y_2 O_2 ratio) using different sample sizes N	213
6.9 Set-up (four most important parameters) which led to the maximum value (41 %) of y_2 . Parameters are the <i>A</i> -factors for the forward reaction rate in $\text{cm}^{-3} \text{molecule}^{-1} \text{s}^{-1}$	223

6.10 First- and second-order sensitivity indices using $N = 131\,072$ and 43 inputs for output y_1 (time at point of inflexion). Parameters x_i and x_j are the A -factors for the forward reaction rate.	224
6.11 First- and second-order sensitivity indices using $N = 131\,072$ and 43 inputs for output y_2 (O_2 ratio). Parameters x_i and x_j are the A -factors for the forward reaction rate.	224
8.1 First- and second-order sensitivity indices calculated analytically and estimated based on RS-HDMR for the non-transformed and transformed model output.	237

1 Introduction

The use of computational modelling as a design tool is increasing within engineering applications. Models can often reduce the burden of expensive experimental programmes within the design process and can be used to explore situations and parameter regions that cannot be reached by experiment. The model, depending on its purpose and field of application, will combine many different processes. However, it is up to the modeller to decide which processes to include and how to parameterise them. The difficulty is in defining the values of parameters because not all of them will be precisely known. Some parameters will only be defined in terms of a range of values with several orders of magnitude. Consequently, these uncertainties have to be taken into account during the modelling process. Due to the uncertainties in the input parameters the model response is also assigned with uncertainty. A quantification of model accuracy and the confidence that can be placed in model predictions is therefore crucial in many areas of engineering applications. In this work the focus is on two areas: urban air pollution modelling and combustion modelling. These two fields are of course interconnected since most of the pollution in urban areas is caused by emissions from transport and industrial sources.

1.1 Motivation

Nowadays transport plays a fundamental role in the lives of societies and individuals. However, on the other hand motor vehicles are a major contributor to ground level concentrations for most of the pollutants in urban areas. According to Krzyzanowski et al. (2005), monitoring data are published by the World Health Organization (WHO), traffic-related air pollution affects a number of health outcomes such as mortality, morbidity, allergic illness, cancer, pregnancy, birth outcomes and male fertility. Furthermore, transport-related air pollution increases the risk of death and of non-allergic respiratory symptoms and disease. A review focused on recent studies of air pollution and health is given in Brunekreef & Holgate (2002).

The monitoring of air quality in urban areas is essential and decisions have to be made in order to improve the air quality, especially in so called hot spots (areas with high concentrations). Several guidelines and standards exist for ozone (O_3), nitrogen dioxide (NO_2) and particulate matter recommended by the World Health Organization (World Health Organization 2000) and the European Union (EU) (Commission of the European Communities 2005). Modelling air pollution would provide a possibility to predict concentrations for future years taking into account emission controls and new or changed source emissions. Furthermore, scenario testing could be undertaken to determine source contributions and control strategies.

To assess the air quality in urban areas a combination of measurement and modelling is required. A number of measurement sites exist in the UK, which are published by the Department for Environment, Food and Rural Affairs (Defra). The aim of modelling air quality is to estimate the concentration of a pollutant at a point or an area, for example near a busy road. For the prediction of the concentration of a certain pollutant one has to have information about emissions from road traffic, meteorological conditions and local street geometry. Due to the complexity of chemical

reactions in the atmosphere and the complexity of pollutant emission sources and emission rates the modelling is not a straight forward process and not all emissions from all sources can actually be measured. Therefore, a lot of parameters have to be estimated which increases the uncertainty in the modelling process.

Chemical mechanisms describing the combustion of fossil fuels can be very complex. For some well known systems such as methane combustion, elementary reaction rates have been extensively measured using experiments and evaluated rate data has been published for the majority of important reactions (Hughes et al. 2001a, Hughes et al. 2001c). However, for other fuels the elementary pathways are less well understood and significant uncertainties exist in the mechanisms describing complex fuels such as gasoline, kerosene or diesel (Battin-Leclerc 2008, Hughes et al. 2008, Pilling 2008). It is also often the case that at early design stages, many process parameters can only be estimated or are not known at all, so that any procedures to select alternatives for further development have to incorporate uncertainties. Much research is undertaken to develop cleaner and more efficient combustion technologies in order to reduce the CO₂, NO_x and particulate emissions. A shift from hydrocarbon fossil fuels to bio-fuels (derived from recently dead biological material) is highly desirable (van Thuijl et al. 2003). However, bio-fuels are mixtures of hundreds of different chemical compounds and many of them have never been previously investigated (Cathonnet 2003). Consequently, a large number of uncertainties have to be taken into account in order to model the combustion process.

Models used for the prediction of air pollutants vary in complexity from relatively simple statistical models to highly complex models such as computational fluid dynamic (CFD) models (Andria et al. 2000). Statistical models calculate a relationship between measured concentrations and pollutant emissions and extrapolate this to other locations and into the future based on estimated emissions. The more complex CFD models attempt to recreate mathematically the actual processes of

pollutant emission and transport in the atmosphere. CFD modelling approaches to represent flow and dispersion in urban areas are becoming more and more common, especially due to increasing computing power. Examples of micro-scale CFD models include the CHENSI (Sini et al. 1996), MIMO (Ehrhard et al. 2000) and MISKAM (Eichhorn 1996) models. Furthermore, CFD models have various advantages over empirically based models. For example, CFD models can predict the concentration of pollutants over a domain rather than at a specific point. However, they are computationally more expensive to run than statistical models and they have a large number of parameters which may be subject to uncertainties (Collett & Oduyemi 1997, Vardoulakis et al. 2003).

In order to make predictions of concentrations, chemical kinetics have to be incorporated in a computationally efficient way within reactive flow models. Since decisions are derived from the results of the models it is a big concern whether a model and its results are correct. This concern is addressed through model validation and model verification (Jakeman et al. 2006). Verification refers to a process to assure that the model is correct and validation of the model refers to the general comparison of modelling results against monitoring data. Discrepancies between measured concentrations and results from a CFD model are due to a number of reasons, for example uncertainties in the traffic flow (such as speed and proportion of the vehicles), uncertainties in the emission for vehicles, meteorological data uncertainties, uncertainties in the turbulence parameters, uncertainties in reaction rates etc (Vardoulakis et al. 2002). An important part of the model validation process is carried out by sensitivity analysis and uncertainty analysis.

1.2 Aim and Objectives

Sensitivity analysis (SA) and uncertainty analysis (UA) play a crucial role in improving the confidence in air pollution models and chemical sub-models particularly if they are used in a decision making context. Traditional methods for sensitivity and uncertainty analysis are no longer practical because of the complexity of most models in environmental and safety engineering and their large number of uncertain parameters.

The aim of this PhD study is therefore the development of efficient methods for global sensitivity and uncertainty analysis based on a set of tools called high dimensional model representation (HDMR) (Rabitz et al. 1999). In particular, the developed methods should be easy to use, applicable to a wide range of models and should be able to deal with complex models in a computationally efficient way. Their effectiveness will be demonstrated for a variety of analytical test functions and a range of applications within environmental and safety engineering. Applications include the study of:

- the flow field in a 3D street canyon using a micro-scale CFD model,
- a reactive atmospheric turbulent plume using a Lagrangian stochastic model,
- the sulphur-nitrogen chemistry interactions within a one dimensional premixed methane flame model (high temperature),
- isothermal cyclohexane oxidation under fuel rich conditions (low temperature).

One objective of this work is to achieve a better physical insight into these models by applying the HDMR method. Discrepancies between experimental results and simulations can be investigated. Furthermore, important parameters can be highlighted and ranked by their importance. This enables the user to focus on key parameters

and uncertainty ranges for these parameters can be revised and improved. If the experimental results can not be recovered by the simulations taking into account the uncertainties in the inputs, then this might be an indication of problems within the model physics (e.g. missing reactions in the chemical scheme) or that important assumptions have not been met when setting up the model (e.g. a too small model resolution). By applying the HDMR method to the above mentioned case studies it is hoped a better understanding of the model will be gained so that more trust can be placed in its predictions and to highlight potential weaknesses of the model so that improvements can be made.

Special attention is paid in this work to the comparison of two different HDMR approaches in order to identify the one which is most suitable for a large range of applications. Another objective is the development of a software package with a graphical user interface based on the HDMR approach so that global uncertainty and sensitivity analysis can be performed in an easy way by all interested users.

1.3 Outline of the Thesis

An overview of widely used methods for uncertainty and sensitivity analysis is given in chapter 2. The main focus here is on global methods which are able to explore the whole input parameter space. The most important sampling based and variance based methods are described and compared against each other. Advantages and disadvantages are highlighted and the need for a new approach is emphasised.

Chapter 3 introduces the high dimensional model representation (HDMR) method. The principles of the approach are explained and two different HDMR approaches are presented: cut-HDMR, which depends on the value at a specific reference point and RS-HDMR, which depends on the average value over the whole domain. Vari-

ous methods are suggested in order to assess the accuracy of the constructed HDMR expansion, which can also be used as a surrogate. This is crucial since the calculation of the sensitivity indices is entirely based on the HDMR component functions. The calculation of the sensitivity indices for first-order and higher order effects is stated for both HDMR approaches and the importance of the plots of the HDMR component functions is highlighted.

An optimisation method for the RS-HDMR approach as an extension to the existing set of HDMR tools is developed in chapter 4. This is followed by a comparison of the cut-HDMR and the RS-HDMR approach using the analytical Ishigami function. Both approaches are also applied in order to perform global sensitivity analysis on a computational fluid dynamics (CFD) model simulating the turbulent air flow in a 3D street canyon. Consequently, the performance of the cut-HDMR and RS-HDMR approach are studied and advantages and disadvantages of both methods are highlighted. Four uncertain parameters are considered, among them the background wind direction and the roughness lengths for inflow, wall and surface. Discrepancies between the experimental results and the model simulations are investigated and the effects of the input parameters on various model output such as the turbulent kinetic energy (TKE) and the wind components at measuring points in the street canyon are studied. It is hoped that for example the inflow roughness length has only very little effect on the TKE in the street canyon, because it is very difficult to determine. The background wind direction is usually determined quite far away from the measuring site and here it is also hoped that its influence on the assigned model outputs is only small. The prediction of the TKE and the wind components in the street canyon is very important, because the flow field can also be used in connection with dispersion models to calculate concentrations and concentration fluctuations of pollutants.

Another extension for the RS-HDMR approach is developed in chapter 5 in order

to further improve the method particularly for cases where a large number of uncertain input parameters (i.e. >20) is investigated. The improved method is tested on the analytical Sobol' g-function and then sensitivity analysis is performed on two case studies. The first focuses on a Lagrangian stochastic model which is used to investigate a reactive turbulent atmospheric plume. The HDMR method is applied in order to study the uncertainties in 22 physical and chemical parameters and their possible interactions. A comparison with experimental results (a wind tunnel study) is made and discrepancies are highlighted. It is also of concern whether or not the simple mixing scheme applied here is suitable for accurately predicting concentrations and how the number of particles used in the Lagrangian simulations influences the results. The effectiveness of the HDMR tools is also demonstrated for a case study of sulphur-nitrogen chemistry in a one dimensional premixed methane flame model under high temperature. Uncertainties in rate constants and thermodynamic data are considered leading to a study of 176 input parameters. There is a lack of consensus in the literature as to the nominal values and uncertainty ranges for the thermodynamic data. Therefore, it is investigated in three scenarios how different parameter ranges will affect their importance on the predicted final NO mole fraction.

In chapter 6 the focus is on systems with significant higher order effects, i.e. where a pair of parameters acts together upon the model output. A two step algorithm is proposed where a screening method is used in order to identify unimportant parameters first. The HDMR method is then applied with a reduced input space dimension. The performance of this two step approach is illustrated using the analytical Sobol' g-function, followed by a case study which focuses on the low temperature isothermal cyclohexane oxidation under fuel rich conditions. Uncertainties in 238 rate constants are considered. The analysis aims to investigate the important features of the oxidation process as well as possible factors underlying qualitative discrepan-

cies between simulations and experiments. The particular feature of interest is the characteristic of quadratic autocatalysis which is observed experimentally and leads to the maximum rate of reaction occurring at 50 % oxygen consumption.

Chapter 7 introduces the software GUI-HDMR which has been developed on the basis of the HDMR approach. The software combines RS-HDMR tools and extensions developed in this work in one Matlab package equipped with a graphical user interface. A short documentation of the software is given so that it can be applied by interested users.

Limitations of the HDMR method and future work is discussed in chapter 8 and a summary and conclusions are given in chapter 9.

2 Uncertainty and Sensitivity Analysis

According to Jakeman et al. (2006) the development and evaluation of environmental models should include discussions of model accuracy, limitations for a given application and scope for model improvement. An important part of the model evaluation process should therefore include some assessment of the impact of the lack of knowledge regarding model inputs on the predicted outputs of the model. This is generally achieved by sensitivity analysis (SA) and uncertainty analysis (UA). Uncertainty analysis investigates the uncertainty in the model output given the defined uncertainties in the model input. If enough is known about the model inputs it can be used to provide error bars on the model predictions. One useful purpose of this is to establish whether, within the error bars, the model is capable of reproducing measured values of target outputs. If the model physics is appropriate then uncertainty analysis methods provide a measure of confidence that can be placed in the predicted outputs. The main purpose of SA is to estimate the effects of each model input, either in isolation or through combined effects, on the model output and to identify the main contributors to the output uncertainty. This helps to identify where attempts to improve the model predictability should be focused.

This chapter aims to give an overview of what sensitivity analysis is and how to

perform it. A classification of different techniques is presented. Uncertainty analysis is not explicitly discussed in this work. However, uncertainty and sensitivity analysis are closely linked and are often performed together. A short review of the most commonly used global SA methods is given.

2.1 Overview of Sensitivity Analysis Methods

Sensitivity analysis does not simply deal with uncertainties in the input parameters, it is mainly used to increase the confidence in the model and its predictions (Saltelli 2000). According to Hamby (1994) SA may be conducted for several reasons such as the need to determine: which parameters require additional research for strengthening the knowledgebase, which parameters are insignificant and may be eliminated from the final model or which parameters contribute most to the output variability. Generally, SA tries to provide an understanding of how the model output responds to changes in the inputs. A good introduction to sensitivity analysis is given in the book of Saltelli et al. (2000).

There are a wide range of methods and tools available in order to perform sensitivity analysis. A classification of five different SA techniques is made in Campolongo et al. (2000b):

1. Factor screening,
 2. Differential analysis,
 3. Monte Carlo analysis,
 4. Response surface method (RSM),
 5. Fourier amplitude sensitivity test (FAST).
-

A similar categorisation is given in (Helton 1993), which states one informal approach (factor screening) and four formal approaches (differential analysis, Monte Carlo analysis, RSM and FAST).

Another classification of SA methods can be given in accordance with Campolongo et al. (2000b) depending on the setting (see also figure 2.1):

1. Local SA (e.g. differential analysis),
2. Factor screening (e.g. one at a time (OAT) design, Morris method),
3. Global SA (e.g. Monte Carlo analysis, RSM, method of Sobol', FAST).

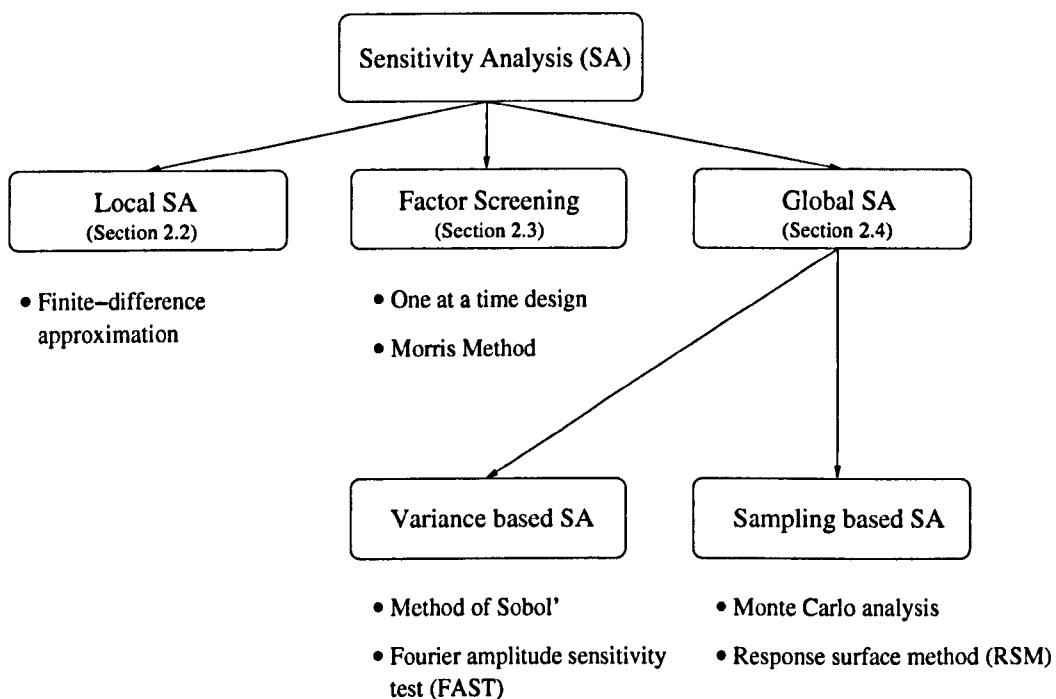


Figure 2.1: Possible classification of sensitivity analysis methods.

Local SA methods focus on the impact of a parameter variation in the vicinity of a single point. The sensitivity is often characterised through gradients. Local SA methods will not be discussed in detail in this work, however a short overview and summary is given in section 2.2.

Factor screening can be classified as a local or a global method depending on its set up and is discussed in section 2.3.

Global SA methods investigate the impact of a parameter over the whole range of its uncertainty. They can be further divided in two classes: (1) sampling based methods, which generate a mapping from uncertain inputs to uncertain outputs (e.g. Monte Carlo analysis, RSM) and (2) variance based methods, which use partial variances as an indicator of importance of an input factor (e.g. method of Sobol', FAST). Global SA methods are presented in section 2.4.

2.2 Local Methods

Commonly, only local sensitivity analysis is performed within model evaluation because of the complexity and computational expense of most environmental models. Local SA is usually carried out by calculating the partial derivatives of the output function with respect to the input parameters (Saltelli 2000). The input parameters can be varied within a small interval around the nominal value in order to compute the derivatives numerically or in certain cases partial derivatives can be defined in terms of common model outputs.

The simplest way in calculating local sensitivities is by using the finite-difference approximation (also known as brute force method or indirect method) (Turányi & Rabitz 2000). Here, one parameter is slightly changed at a time and a new model run is performed. The sensitivity measures are the partial derivatives $\partial y / \partial x_i$ (also called first-order local sensitivities) which can be approximated by the finite-difference:

$$\frac{\partial y}{\partial x_i} \approx \frac{y(x_i + \Delta x_i) - y(x_i)}{\Delta x_i} \quad i = 1, \dots, n. \quad (2.1)$$

The calculation of the local sensitivities requires $n + 1$ simulations of the model. In

Turányi & Rabitz (2000) it is stated that in most cases a 1% perturbation of the parameter x_i is a good choice.

Local methods provide only information related to a single point (e.g. at a nominal point or best estimate for a parameter) and do not cover the whole input parameter space. These methods are most appropriate where the uncertainties in the input space are small i.e. the parameters are well known, or where the model is not highly non-linear. If uncertainties in the input space are significant, then sensitivities calculated at the nominal point are often not representative of all regions of the input space (especially if the model is highly non-linear). In addition, knowledge of the influence of interactions between parameters is not explored using local SA methods.

More information and reviews of other local SA methods can be found for example in Rabitz et al. (1983), Turányi (1990) or Tomlin et al. (1997).

2.3 Screening Methods

Screening methods aim to identify unimportant parameters and to rank potentially important ones. They are usually applied in cases where the input space dimension is very high (Campolongo et al. 2000a). However, screening methods can not quantify the partial contributions to the overall variance or how much a given factor is more important than another one (Campolongo et al. 1999). Due to their easy set up and their relatively small computational requirements, they are usually applied first in a two step algorithm. When the input space dimension is narrowed down (i.e. by fixing the parameters, which were identified as unimportant), a second method (i.e. Sobol's method as described in section 2.4.5 or the HDMR method which is introduced in chapter 3) can be applied in order to calculate sensitivity indices.

Several screening techniques have been proposed in the literature. In this work the

focus is only on two of them: The one at a time (OAT) design (mostly a local SA method) and the Morris method (global SA method).

2.3.1 One at a Time (OAT) Design

Conceptually the simplest way of performing sensitivity analysis is to vary just one parameter at a time while the others remain fixed. The impact of changing the value of each parameter is evaluated in turn (Daniel 1973).

In the standard OAT design, the nominal value (best estimate) is used for each of the parameters and then usually two extreme values are proposed to represent the range of likely values for each parameter (the nominal value is usually in the middle between the two extremes). First, the model is simulated with all parameters at their nominal values. Then, successively the extreme input parameter values are applied and the magnitude of the difference between the outputs (all parameters at their nominal values - one parameter at its extreme value) are compared in order to find the parameters that significantly affect the model output (Campolongo et al. 2000a).

The computational effort is quite small with only $2n + 1$ required model runs (n is the number of the input parameters). However, one significant drawback of this method is that interactions between parameters cannot be estimated. In addition to this, many OAT experiments are local, because factors are only changed over a small interval around their nominal value. Consequently, the model behaviour is only investigated locally around the selected point (Campolongo et al. 2000a).

2.3.2 Morris Method

The one at a time design proposed by Morris (1991) counts as a global sensitivity experiment since it aims to cover the entire space over which the input factor may be varied. The method determines an importance ranking for parameters in terms of their mean effect on output variance as well as determining those parameters with linear additive effects and non-linear interactions. One factor at a time is varied over the space of the input parameters with p values in the set $\{0, 1/(p-1), 2/(p-1), \dots, 1\}$ which results in a p -level grid for the region of experimentation. In practical applications these values are re-scaled to values from within their uncertainty ranges. The elementary effect of the i th factor at a given point \mathbf{x} is defined as:

$$d_i(\mathbf{x}) = \frac{[y(x_1, \dots, x_{i-1}, x_i + \Delta, x_{i+1}, \dots, x_k) - y(\mathbf{x})]}{\Delta} \quad (2.2)$$

where Δ is a predetermined multiple of $1/(p-1)$. The total computational effort required is $r(n+1)$, where n is the number of input parameters and r the number of trajectories. One trajectory contains $n+1$ runs where each parameter has been changed according to the sampling strategy. An example for two parameters is shown in figure 2.2. The mean effect across r indicates the overall importance of a parameter and is given by:

$$\bar{d}_i = \frac{1}{r} \sum_{l=1}^r |d_i^l|. \quad (2.3)$$

In the literature (i.e. Campolongo et al. (1999)) it is stated, that r is typically in the range between 5 and 15. However, this is not a reliable statement. It would make much more sense to investigate whether or not the calculated value of the mean in

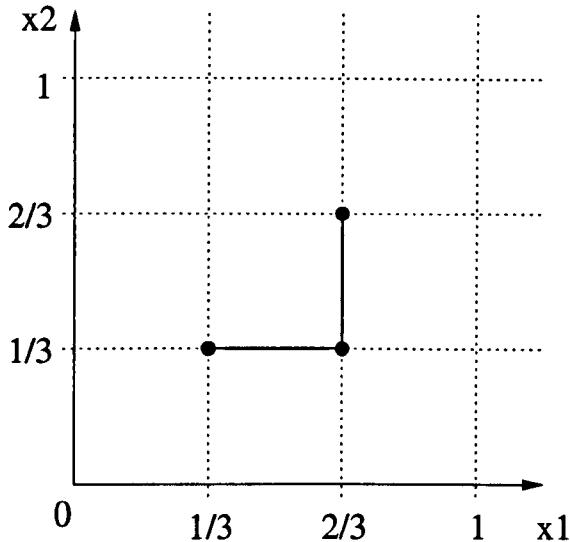


Figure 2.2: Example of a trajectory for two parameters ($n = 2$) and a four-level grid ($p = 4$).

equation (2.3) has reached convergence. This could be checked for the parameters, which were identified as potentially important. If the mean does not converge, then more trajectories have to be added, which consequently makes the Morris method more expensive. In chapter 6 the Morris method is applied to an analytical test function and to a real case study (low temperature isothermal cyclohexane oxidation under fuel rich conditions). Even in the test case scenario, ten trajectories have not proven to be enough to achieve convergence. In the case study, 100 trajectories were not sufficient in order to produce a reliable parameter ranking.

Additionally, the calculation of the standard deviation of the elementary effects gives an indication of non-linearities or parameter interactions (Morris 1991). However, the Morris method does not distinguish non-linear from interaction effects (Campolongo & Braddock 1999). An overall measure of the interactions of a particular parameter with the rest of the model is given, but it does not give specific information as to which part of the rest of the model it is interacting with. To overcome this problem an extended Morris method has been developed by Campolongo

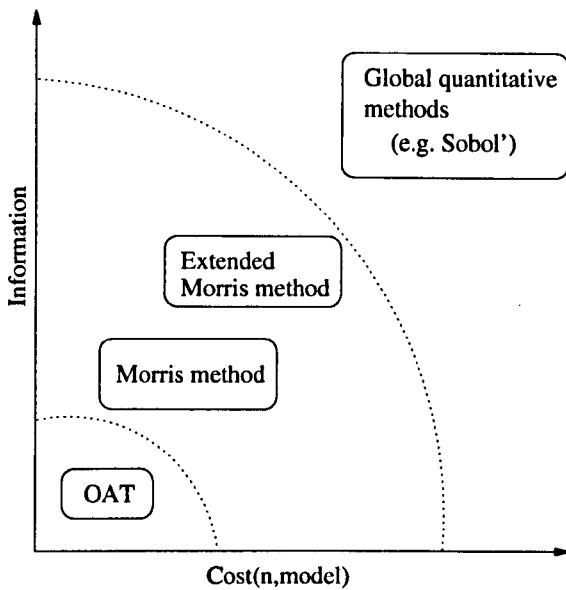


Figure 2.3: Overview of various classes of SA methods according to two properties: the amount of information produced in terms of the model sensitivity and the computational cost of the method. The computational cost is measured in terms of the number of required model evaluations and depends on the number of input parameters n and the complexity of the model (adapted from Campolongo et al. (1999)).

& Braddock (1999) in order to develop information on two-factor interactions, which on the other hand leads to an increase in model runs required.

An improved sampling strategy for the Morris method has also been developed by Campolongo et al. (2007) in order to cover the input space more efficiently. A large number of trajectories (up to 1000) is generated first and than a smaller number (e.g. $r = 10$) is chosen with the highest spread, which is based on a distance measure (e.g. Euclidean distance).

Note that the Morris method provides sensitivity measures that are only qualitative (Campolongo et al. 1999). To obtain a quantitative measure (e.g. in order to show how much more one parameter is important than another one), other methods have to be applied which usually require more model runs. The relationship between the computational cost of the method and the amount of information produced in

terms of the model sensitivity is shown in figure 2.3.

2.4 Global Methods

In order to explore the input parameters and their potential interactions over the whole input space, global SA methods are required. One global SA method, the Morris method, has been introduced already in the latter section in connection with screening methods. This section aims to provide a short review of the most commonly applied global SA methods. The focus is on two sampling based and two variance based approaches.

2.4.1 Overview

Global sensitivity analysis methods can be divided into sampling based methods and variance based methods. Sampling based methods for uncertainty and sensitivity analysis involve the generation and explorations of a mapping from uncertain inputs to uncertain predicted outputs (Helton & Davis 2002, Helton et al. 2006). In this work the focus is on two sampling based methods: the Monte Carlo (MC) method (or Monte Carlo analysis) and the response surface method (RSM) (Box & Wilson 1951, Box 1954). Both methods are based on Monte Carlo simulations, where a deterministic model (original model or surrogate) is iteratively evaluated using a set of random numbers as inputs. Variance based analysis uses partial variances as an indicator of the importance of the input factors x_i . Two such methods, the method of Sobol' (Sobol' 2001) and the Fourier amplitude sensitivity test (FAST) (Cukier et al. 1973), are introduced in section 2.4.5 and 2.4.6. Both methods offer an approach to determine variance based sensitivity indices, which are global and model independent.

Useful reviews of global SA methods can be found in Helton & Davis (2000*a*), Helton & Davis (2000*b*) and Saltelli et al. (2000). Although global methods may provide far more insight for non-linear models, they can require large numbers of model runs. In addition, the interpretation of such global sensitivity tests can be difficult because of the number of parameters involved and the potential complexity of the causal relationships (Jakeman et al. 2006). It is clear that more general methods for the analysis of non-linear responses and parameter interactions would be useful within the field of environmental and safety models. Such methods must be computationally efficient and capable of determining sensitivity indices that can be used for importance ranking of potentially large numbers of input parameters in an automatic way.

2.4.2 Sampling

Even variance based methods (e.g. the method of Sobol', introduced in section 2.4.5) involve Monte Carlo simulations. Due to the large variety of applications for MC simulations and because of their overall importance for uncertainty and sensitivity analysis this section focuses on different sampling methods first, before the focus is on individual global SA methods.

Monte Carlo simulations require sampling procedures. The generation of the random samples is a crucial task and the success of a Monte Carlo calculation depends on the quality of the random samples (Niederreiter 1992). Three different sampling procedures are discussed here: random sampling, Latin hypercube sampling and quasi-random sampling.

Random Sampling

A sample of the desired size N is generated from the joint probability distribution of the input parameter. Each sample point is selected independently of all other sample points. Random numbers generated by a computer code are pseudo-random, because a deterministic algorithm is used (Niederreiter 1992). The generated sequence is actually not random (hence they are called pseudo-random), however they share many properties with completely random numbers.

Latin Hypercube Sampling

The idea of Latin hypercube sampling (McKay et al. 1979) is to ensure the full coverage of the range of each input parameter. A square grid (Latin square) is used, which contains only one sample for each column and each row (also see example in figure 2.4). Hence, one has to first decide how many sample points to use. Once a Latin hypercube sequence has been generated it can not be extended by simply adding more sample points to it.

The range of each input parameter x_i is divided in N intervals and one random sample is generated for each of the intervals

$$(0, 1/N), (1/N, 2/N), \dots, ((N-1)/N, 1) \quad (2.4)$$

for each of the parameters. Afterwards the samples of each input parameter are randomly permuted.

An example on how to construct a Latin hypercube sequence for two input parameters with a sample size of $N = 4$ is given in figure 2.4. Table 2.1 presents the corresponding data set for the two input parameters x_1 and x_2 . The first two columns in table 2.1 show the random values for each of the two inputs for the four

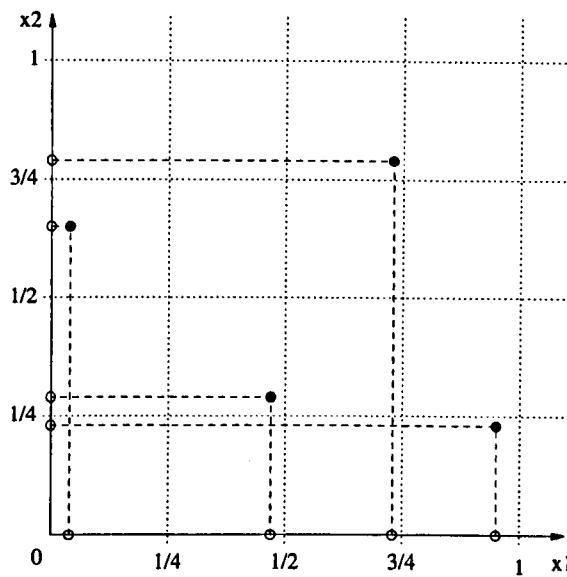


Figure 2.4: Example of a Latin hypercube sequence for two parameters ($n = 2$) and a samples size of $N = 4$.

Random values for		Random permutation for	
x_1	x_2	x_1	x_2
0.04	0.23	0.04	0.65
0.47	0.29	0.73	0.79
0.73	0.65	0.47	0.29
0.95	0.79	0.95	0.23

Table 2.1: Generation of a Latin hypercube sequence for two parameters ($n = 2$) and four intervals ($N = 4$).

intervals and the last two columns show the random permutation.

Quasi-Random Sampling

Quasi-random sequences are totally deterministic and are based on low discrepancy, hence they are also called low discrepancy sequences. The discrepancy is a measure for the uniformity (high uniformity equals low discrepancy) of a sequence and is computed by comparing the actual number of sample points in a given volume of multidimensional space with the number of sample points that should be there

assuming a uniform distribution (Morokoff & Caflisch 1995). Successive sample points are added in a position as far away as possible from the other sample points so that clustering (groups of points close to each other) can be avoided. The best known low discrepancy sequences are: Halton (Halton 1960), Faure (Faure 1992), Sobol' (Sobol' 1967) and Niederreiter (Niederreiter 1987, Niederreiter 1988). The principles of the Halton, Faure and Sobol' sequences are discussed in the following.

Halton sequence The Halton sequence (Halton 1960) is the simplest of the low discrepancy sequences. To generate a Halton sequence a consecutive set of non-negative integers is transformed into numbers in the interval $[0, 1)$ (zero is included in the sequence, but it is open at one because the sequence never reaches one). This can be described in a two step procedure (Galanti & Jung 1997):

1. The integers are expanded in an arbitrary base p , where p is a prime number greater than or equal to two. This means, that each integer is converted into the base p number system. For example, the integer four is represented by 100 ($4 = 1 \cdot 2^2 + 0 \cdot 2^1 + 0 \cdot 2^0$) in the base 2 system.
2. The base p number is transformed into a number in the interval $[0, 1)$ by reflecting about the decimal point. For example, the base two number 100 becomes 0.001. The binary fraction 0.001 represents $\frac{1}{8}$ ($0 \cdot \frac{1}{2^1} + 0 \cdot \frac{1}{2^2} + 1 \cdot \frac{1}{2^3} = \frac{1}{8}$). Thus the corresponding Halton number for integer four is $\frac{1}{8}$.

Each successive Halton number fills in the gaps of the existing sequence. Therefore, one does not need to define the sample size N in advance. In the multidimensional Halton sequence a different base p (prime number) is used for each of the dimensions. For example, the first dimension uses base two, the second dimension uses base three, the third dimension base five and so on. The Halton algorithm and a more detailed

example of how to generate a two dimensional Halton sequence is given in Galanti & Jung (1997).

Faure Sequence The Faure sequence (Faure 1992) is similar to the Halton sequence, but with two differences. Firstly, it uses the same base for each dimension. This base is the smallest prime number that is greater than or equal to the number of dimensions in the input parameter space. Secondly, the Faure numbers are reordered within each dimension (otherwise sequences would be identical across all dimensions).

Sobol' Sequence The Sobol' sequence (Sobol' 1967), like the Faure sequence, has the same base for all dimensions and it uses reordering of the sample points within each dimension. However, the Sobol' sequence uses only base two for all dimensions. This reduces the computational time especially when finer grid points are generated. The reordering task is more complex, because the same base is used across all dimensions.

Galanti & Jung (1997) state that the Sobol' sequence outperforms both Faure and Halton sequence. They demonstrated, that on average the Sobol' sequence exhibits better convergence properties than either the Faure or Halton sequence.

Comparison of Sampling Methods

Figure 2.5 shows the distribution of random, Latin hypercube and quasi-random (Halton sequence and Sobol' sequence) points for a sample size of $N = 1024$ in two dimensions. For both random sampling and Latin hypercube sampling (figure 2.5a and 2.5b), there are areas that are empty of points and areas where a clustering of points is clearly visible. The 1024 sample points of the two quasi-random sequences

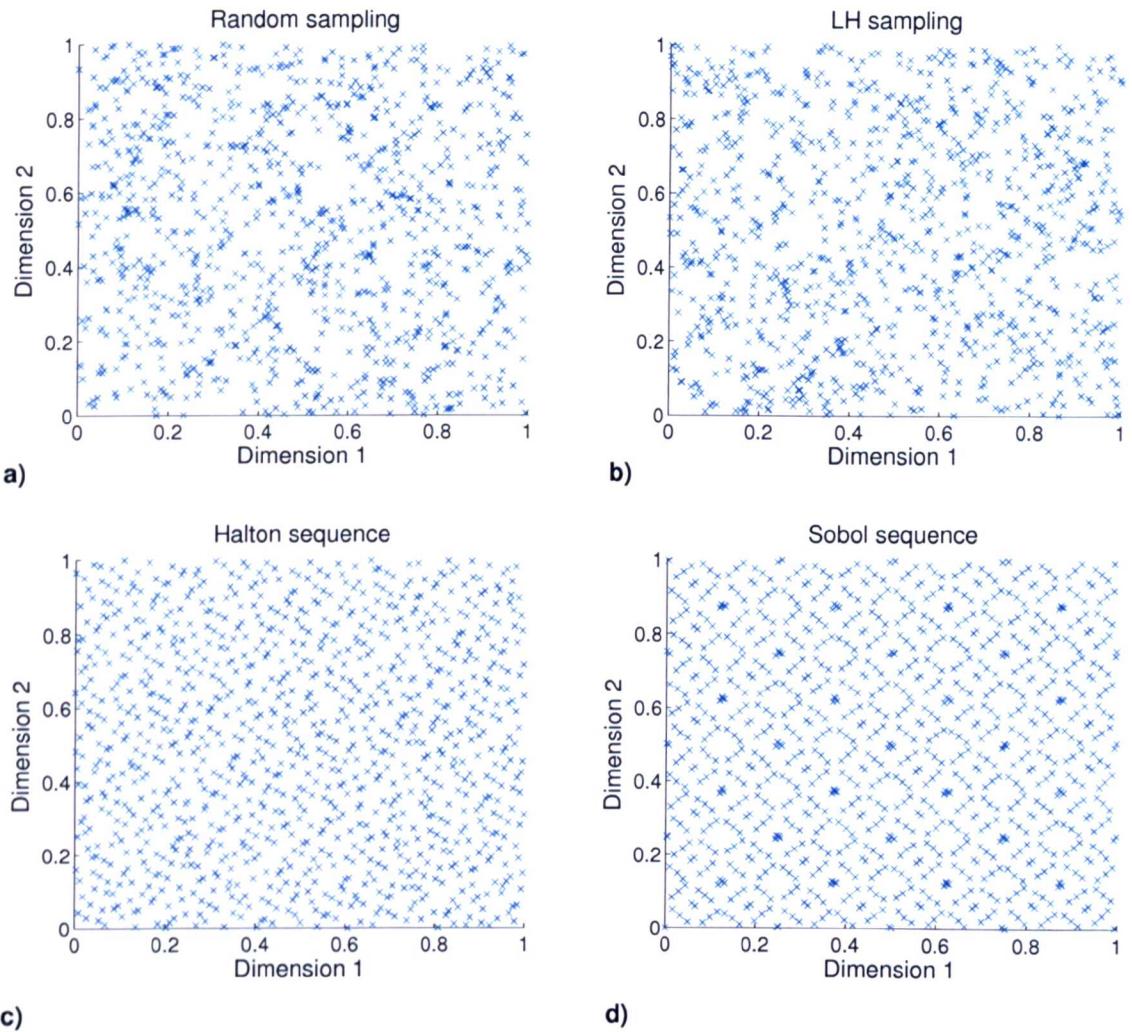


Figure 2.5: (a) 1024 random sampling points, (b) 1024 Latin hypercube sampling points, (c) 1024 points of the Halton sequence, (d) 1024 points of the Sobol' sequence. All sampling procedures are based on a uniform distribution in the interval $[0, 1]$.

(figure 2.5c and 2.5d) cover the unit square more uniformly than the points of the random sequences. In general it can be said that the more evenly the points are distributed throughout the domain, the more accurate the simulation. A good overview and comparative assessment of random and quasi-random sampling methods can be found in (Galanti & Jung 1997).

The efficiency of the quasi-random sequences can be demonstrated by using a simple

test model with two input parameters:

$$f(\mathbf{x}) = x_1 + x_2^4 \quad (2.5)$$

where x_1 and x_2 are uniformly distributed within $[0, 1]$. The four different sequences as shown in figure 2.5 are then used to simulate the model and the output values $f(\mathbf{x})$ are recorded. The arithmetic mean \bar{f} can be approximated by

$$\bar{f} \approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}) \quad (2.6)$$

for N samples $\mathbf{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \dots, x_n^{(s)})$ with $s = 1, 2, \dots, N$. The variance D of the model output can be approximated by

$$D \approx \frac{1}{N} \sum_{s=1}^N f^2(\mathbf{x}^{(s)}) - \bar{f}^2. \quad (2.7)$$

Figure 2.6 presents the mean and the variance of the model output for the simple test model. The “real” mean calculated by

$$\bar{f} = \int_0^1 f(\mathbf{x}) d\mathbf{x} = 0.7 \quad (2.8)$$

is quickly reached if using the Halton or the Sobol’ sequence. Approximately 200 samples are already enough to get very close to the mean and the value converges after that. The random sequence, as shown in figure 2.5a, delivers the worst performance. For a small sample size (e.g. $N < 200$) the approximated mean is far off the real value. The mean only converges after applying more than 800 samples. The performance of Latin hypercube sampling is better than using random sampling but is still worse than using any of the quasi-random sequences. As for the random sequence, more than 800 samples are required in order to reach convergence.

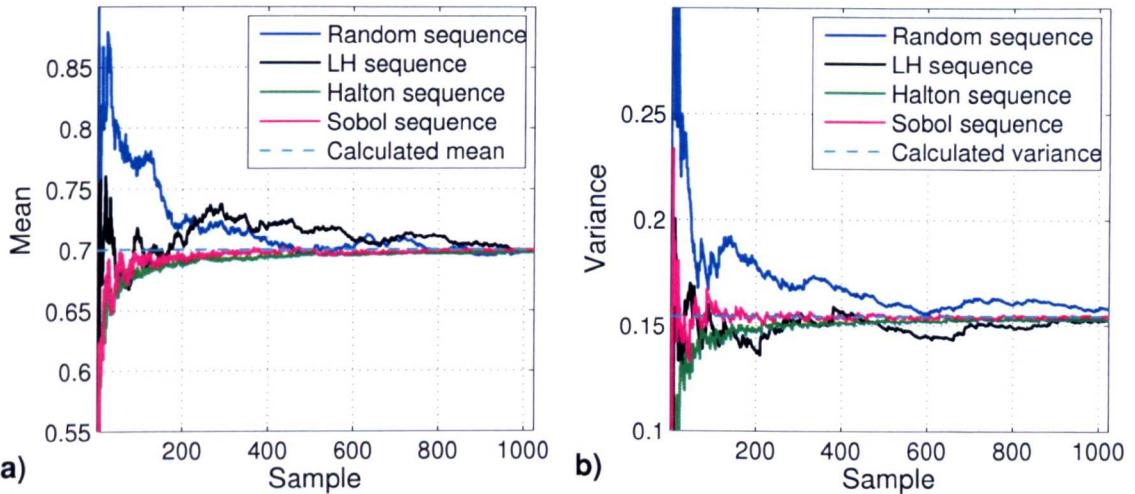


Figure 2.6: Plot of (a) the mean and (b) variance for the simple test model using different sampling procedures.

The estimated variance is presented in figure 2.6b. The "real" variance is calculated by

$$D = \int_0^1 [f(\mathbf{x}) - \bar{f}]^2 d\mathbf{x} = \frac{139}{900}. \quad (2.9)$$

The performance of the four different sequences in order to estimate the variance is similar to the performance in order to estimate the mean. The Sobol' sequence outperforms the random and the Latin hypercube sequence and using the simple test function it has also shown to be better than the Halton sequence. These results are in accordance with the study by Galanti & Jung (1997). Therefore, the Sobol' sequence will be used in this work as the quasi-random sequence of choice.

In Bratley et al. (1992) it is argued that in high dimensional problems ($n > 12$) quasi-random methods do not have an advantage over random methods anymore. However, Kucherenko (2007) argues that this is not true in general. The Sobol' method to estimate global sensitivity indices (see section 2.4.5) in connection with quasi-random sampling is used by Kucherenko (2007) in order to predict the efficiency of quasi-Monte Carlo algorithms (MC using quasi-random sampling). A

classification of some important types of integrable functions is developed:

- Type A: functions with not equally important parameters,
- Type B: functions with equally important parameters and with dominant low order terms,
- Type C: functions with equally important parameters and with dominant interaction terms.

Kucherenko (2007) shows that for functions of type A and B quasi-Monte Carlo is superior to Monte Carlo even in the high dimensional case. Only for functions of type C does quasi-Monte Carlo loses its advantage over Monte Carlo.

2.4.3 Monte Carlo Analysis

In general, Monte Carlo analysis involves five basic components (Helton & Davis 2000*a*, Helton & Davis 2000*b*):

1. Selection of the range and distribution for each of the input parameters,
2. Generation of a sample for the inputs in consistency with their distributions,
3. Evaluation of the model output for each sample point (mapping),
4. Presentation of uncertainty analysis results,
5. Determination of sensitivity analysis results.

Selection of the Range and Distribution

The selection of the distribution for the input parameters is described as the most important step, because these distributions determine both the uncertainty in the

model output $f(\mathbf{x})$ and the sensitivity of $f(\mathbf{x})$ to the elements of the model input \mathbf{x} (Helton & Davis 2000a, Helton & Davis 2000b). If enough data is available, then it might be possible to estimate distribution and distribution parameters. Some guiding principles are given for example in Firestone et al. (1997).

If not enough data is available, then usually a crude distribution such as uniform or normal is assumed for the input parameters. The uniform distribution has a constant probability. The probability density function for a continuous uniform distribution on the interval $[a, b]$ is given by:

$$P(x) = \begin{cases} 0 & \text{for } x < a \\ \frac{1}{b-a} & \text{for } a \leq x \leq b \\ 0 & \text{for } x > b \end{cases} \quad (2.10)$$

The general formula for the probability density function of the normal distribution is:

$$P(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \quad (2.11)$$

where μ is the mean (location parameter) and σ is the standard deviation (scale parameter). A plot of the probability density function for the uniform and normal distribution is shown in figure 2.7.

Sampling

Three different sampling strategies are considered here in connection with the Monte Carlo method and have been described in section 2.4.2. If quasi-random sampling is applied, then the method is also known as quasi-Monte Carlo analysis.

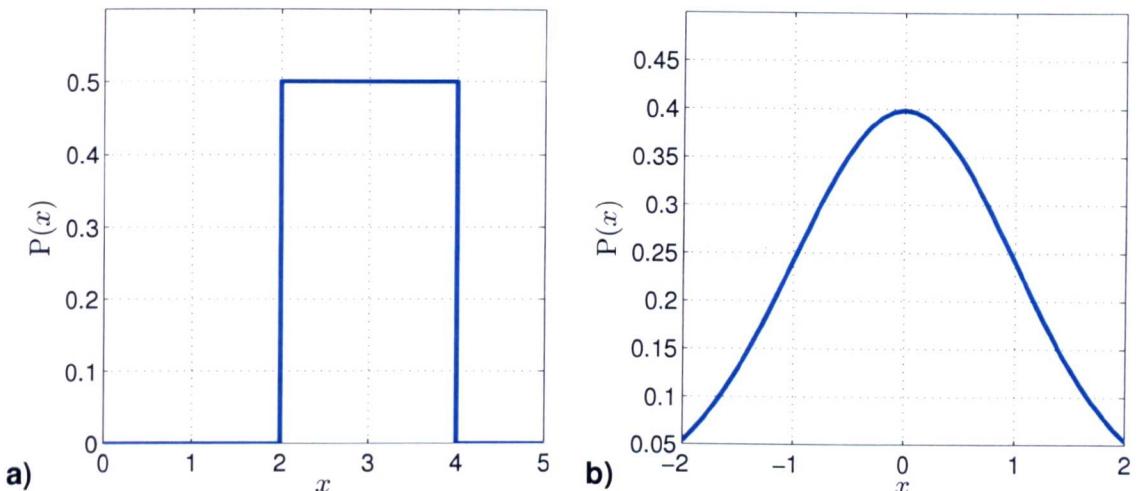


Figure 2.7: (a) Uniform distribution with $a = 2$ and $b = 4$, (b) Normal distribution with $\mu = 0$ and $\sigma = 1$.

Mapping

The third step of the Monte Carlo analysis is the evaluation of the model output for each of the samples. In spite of its simplicity this is the most time consuming part of the MC analysis, since the sample size N can easily exceed 10^3 .

Uncertainty Analysis

Presentation of the uncertainty results include characteristics such as mean, standard deviation, variance, probability density function (pdf), cumulative distribution function (cdf), complementary cumulative distribution function (ccdf) and box plot. Box plots and cdf show much more than means and standard deviation and are therefore preferable (Helton & Davis 2000a, Helton & Davis 2000b).

Probability density functions are used throughout this work to describe the characteristics of the model output and for example to describe the distribution of the input parameters. Both pdf and cdf are also used to assess the accuracy of the

HDMR metamodel (see for example section 3.3.1).

Sensitivity Analysis

Sensitivity analysis is usually more demanding than uncertainty analysis, since SA explores the effects of elements of the input \mathbf{x} on the elements of the output $f(\mathbf{x})$. A number of approaches exist and most of them will only be named in this context: scatter plots, regression analysis, correlation coefficient, rank transformation, variance decomposition, etc. Most of these approaches are explained for example in Helton & Davis (2000a) or in Mendenhall et al. (2003). Scatter plots and correlation coefficients are often used in this work and some more details are therefore given in the following paragraphs.

Scatter Plots Scatter plots are one of the most straightforward techniques for sensitivity analysis. They are used to provide a graphical display of the relationship between two variables (their correlation), the input parameter x_i and the output $f(\mathbf{x})$. The data points of these two variables are displayed in a two-dimensional graph with the response variable usually on the y -axis. Scatter plots may reveal relationships between the model input and the model output, such as non-linear relationships and thresholds (Helton 1993). The pattern shown in the scatter plot can provide the following information about the relationship between the two variables:

- Strength,
 - Shape (linear or non-linear),
 - Direction (positive or negative),
 - Presence of outliers.
-

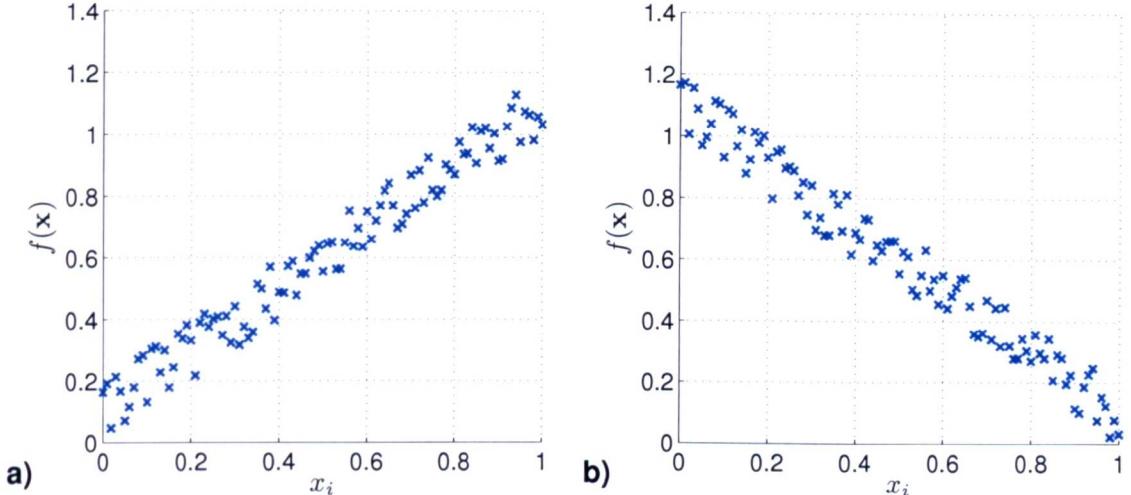


Figure 2.8: Scatter plots showing (a) a strong positive linear relationship, (b) a strong negative linear relationship.

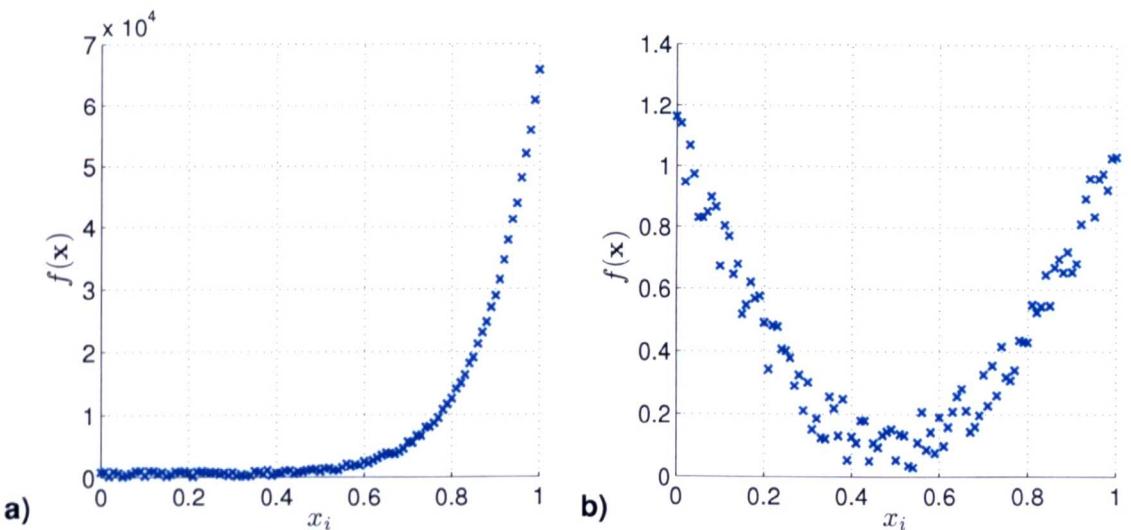


Figure 2.9: Scatter plots showing (a) a strong nonlinear (monotonic) relationship, (b) a strong nonlinear (non-monotonic) relationship.

Some examples of scatter plots are given in figure 2.8 and 2.9. The two scatter plots in figure 2.8 show a strong linear relationship between the input x_i and the model output $f(\mathbf{x})$. The scatter plot in figure 2.9a represents a non-linear relationship with a monotonic behaviour. The scatter plot in figure 2.9b shows also a non-

linear relationship between the input x_i and the output $f(\mathbf{x})$, however with a non-monotonic behaviour.

One significant drawback of using scatter plots is, that one needs to examine a large number of plots, at least one for each input parameter. This is especially problematic if the model possesses a large input parameter space and various outputs. The drawbacks of using scatter plots are discussed in more detail in connection with introducing the plots of the HDMR component functions in section 3.4.3 and in section 4.3.4 when the plots HDMR component functions are compared with scatter plots.

Pearson Correlation Coefficient The Pearson correlation coefficient (or Pearson product moment coefficient of correlation) is a measure of the strength of the linear relationship between two variables (e.g. input and output). It is defined as (Mendenhall et al. 2003):

$$r = \frac{\sum_{s=1}^N (x_i^{(s)} - \bar{x}_i)(f(\mathbf{x}^{(s)}) - \bar{f})}{\sqrt{\left[\sum_{s=1}^N (x_i^{(s)} - \bar{x}_i)^2 \right]} \sqrt{\left[\sum_{s=1}^N (f(\mathbf{x}^{(s)}) - \bar{f})^2 \right]}} \quad (2.12)$$

with

$$\bar{f} = \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}), \quad \bar{x}_i = \frac{1}{N} \sum_{s=1}^N x_i^{(s)} \quad (2.13)$$

for the range from +1 to -1. A correlation of +1 means that there is a perfect positive linear relationship between the two variables. A correlation of -1 means that there is a perfect negative linear relationship between the two variables. A correlation of 0 means there is no linear relationship between the two variables.

The square of the correlation coefficient r^2 (also known as the coefficient of deter-

mination) is the fraction of the variance in the output $f(\mathbf{x})$ that is accounted for by a linear fit of the input parameter x_i to $f(\mathbf{x})$. The value of r^2 is in the range from 0 to 1. For example, a r value of 0.9815, which has been calculated in this case on the basis of the data set shown in the scatter plot in figure 2.8a, results in a r^2 value of 0.9633. This means that 96.33 % of the total variation in the output $f(\mathbf{x})$ can be explained by the linear relationship between the input x_i and $f(\mathbf{x})$.

Scatter plot figure:	2.8a	2.8b	2.9a	2.9b
r	0.9815	-0.9812	0.7297	-0.0107
r^2	0.9633	0.9627	0.5325	0.0001

Table 2.2: Pearson correlation coefficients calculated for the data sets presented by the scatter plots in figure 2.8 and 2.9.

The remaining r and r^2 values for the data sets presented in the scatter plots from figure 2.8 and figure 2.9 are given in table 2.2. The strong linear relationship shown in the scatter plots in figure 2.8a and 2.8b is confirmed by the high absolute value of r and r^2 . The sign for the r value also indicates a positive or negative relationship. The strong relationship (non-linear, monotonic) presented by the scatter plot in figure 2.9a can not be confirmed by calculating the Pearson correlation coefficient. The r^2 suggests that only 53.25 % of the total variance is caused by the input x_i . This is because a linear relationship is assumed between the input and the output. The Pearson correlation coefficient calculated for the data set shown in figure 2.9b is nearly zero, even if there is a strong relationship (non-linear, non-monotonic) visible in the scatter plot.

The Pearson correlation coefficient is easy to calculate and simple to understand. However, it gives only the correct answer if the model is linear. This is a significant drawback of this method and if one can not assure that the model of investigation is linear the sole use of the Pearson correlation coefficient can provide misleading answers.

Spearman Rank Correlation Coefficient In principle, the Spearman rank correlation coefficient r_s is simply a special case of the Pearson product moment coefficient. However, the input data set for x_i and the output data set for $f(x)$ are converted into rankings \tilde{x}_i and $\tilde{f}(x)$ before calculating the coefficient. If there are no ties in either of the data sets of the input and the output, then the expression for r_s reduces to the simpler expression (Mendenhall et al. 2003):

$$r_s = 1 - \frac{6 \sum_{s=1}^N (\tilde{x}_i^{(s)} - \tilde{f}(x^{(s)}))^2}{N(N^2 - 1)} \quad (2.14)$$

for the range from +1 to -1. The Spearman rank correlation coefficient has basically the same meaning as the Pearson correlation coefficient, but can also cope with non-linear (albeit monotonic) relationships between the input and the output. However, Campolongo et al. (2000b) states that care must be employed when interpreting the results of analysis based on rank transformation, since any conclusion drawn using ranks does not translate easily to the original model.

Scatter plot figure:	2.8a	2.8b	2.9a	2.9b
r_s	0.9809	-0.9820	0.9027	-0.0032
r_s^2	0.9621	0.9643	0.8149	0.0000

Table 2.3: Spearman rank correlation coefficients calculated for the data sets presented by the scatter plots in figure 2.8 and 2.9.

Table 2.3 presents the r_s and r_s^2 values based on the data shown in the scatter plots in figure 2.8 and figure 2.9. For the two linear cases in figure 2.8a and 2.8b one gets nearly the same values as by using the Pearson correlation coefficient. The Spearman rank correlation coefficient can also handle monotonic non-linear relationships and therefore the values for r_s and r_s^2 based on the data set displayed in figure 2.9a are higher than using the Pearson correlation coefficient. However, as well as the Pearson correlation coefficient, the Spearman rank correlation coefficient fails to

identify the strong non-linear (non-monotonic) relationship shown by the scatter plot in figure 2.9b. If one can not guarantee that the assumptions made by applying the correlation coefficient are correct, then one can not rely on the calculated r_s and r_s^2 values.

2.4.4 Response Surface Method (RSM)

The basic idea of this methodology is to develop a response surface approximation to the original model. In this way a metamodel is constructed, which can then be used as a surrogate for the original model in order to perform uncertainty and sensitivity analysis. The term metamodel is used throughout this work when referring to the model approximation. According to Helton (1993) the response surface method involves six steps:

1. Selection of the range and distribution for each of the input parameters,
2. Development of an experimental design to define the combination of parameters for which the model will be evaluated,
3. Evaluation of the model output for each design point,
4. Construction of a response surface approximation to the original model,
5. Uncertainty analysis using the surrogate for the original model,
6. Sensitivity analysis using the surrogate for the original model.

Selection of Range and Distribution

The first step is similar to the first step in Monte Carlo analysis. It is necessary to determine at least the range of the input parameters x_i at this stage, since this will

have an impact on the selection of the design points in the second step. For the uncertainty analysis in step five the distribution of the input parameters x_i will also be required.

The complexity of the response surface model is controlled by the complexity of the original model. Therefore, the aim is not to add parameters to the response surface unnecessarily (Downing et al. 1985). A screening method, such as the one at a time design as discussed in section 2.3, can be used to determine the most influential parameters. The screening would take place after assigning the range for each of the input parameters.

Development of an Experimental Design

An experimental design is selected to define the combination of parameters for which the model will be evaluated. Different types of experimental designs are available such as factorial design or fractional factorial design (Box et al. 1978). Several other design methods especially for computer experiments are discussed for example in Sacks et al. (1989*b*) and Sacks et al. (1989*a*). The design chosen depends on many factors: the number of independent parameters under consideration, the possible presence of non-linear effects, the possible importance of parameter interactions and the computational expense of the model (Helton 1993).

Model Evaluation

The model is evaluated for each of the design points. In contrast to the Monte Carlo method described in section 2.4.3 the points are not random. They have been chosen using an experimental design which assures that a specific structure exists between the values of individual input parameters x_i . Again, this can be the most time consuming part depending on the number of simulations to be performed and

the individual run time of the model.

Construction of a Response Surface

The results from the model evaluations are used to construct a response surface approximation to the real model. This is usually based on least square methods using either a first-order model

$$f(\mathbf{x}) \approx b_0 + \sum_{i=1}^n b_i x_i \quad (2.15)$$

or a second-order model

$$f(\mathbf{x}) \approx b_0 + \sum_{i=1}^n b_i x_i + \sum_{i=1}^n b_{ii} x_i^2 + \sum_{1 \leq i < j \leq n} b_{ij} x_i x_j. \quad (2.16)$$

It is also possible to use polynomials or splines for the approximation.

Uncertainty Analysis

Two different approaches can be used to estimate the uncertainty in $y = f(\mathbf{x})$ (Downing et al. 1985, Helton 1993). Using the first approach and a first-order model, the expected value E of the output can be approximated by:

$$E(y) \approx b_0 + \sum_{i=1}^n b_i E(x_i). \quad (2.17)$$

If the input parameters are uncorrelated, than the variance D of the output can be calculated as follows

$$D(y) \approx b_0 + \sum_{i=1}^n b_i^2 D(x_i). \quad (2.18)$$

The second approach would be to use the same methods as stated in step four of the Monte Carlo analysis (section 2.4.3). However, in this case the response

surface approximation is used instead of the original model. The computational effort involving the surrogate is much smaller and therefore a large sample size N can be applied to present the results of the uncertainty analysis.

Sensitivity Analysis

If a first-order model is used to construct the response surface, then the following normalisation can be used to assess the sensitivity of the model output $y = f(\mathbf{x})$ to the individual parameters x_i (Helton 1993):

$$\frac{y - E(y)}{E(y)} \approx \sum_{i=1}^n \left[\frac{b_i E(x_i)}{E(y)} \right] \left[\frac{x_i - E(x_i)}{E(x_i)} \right] \quad (2.19)$$

$$\frac{y - E(y)}{SD(y)} \approx \sum_{i=1}^n \left[\frac{b_i SD(x_i)}{SD(y)} \right] \left[\frac{x_i - E(x_i)}{SD(x_i)} \right] \quad (2.20)$$

where SD stands for the standard deviation.

The coefficient in equation (2.19) $b_i E(x_i)/E(y)$ represents the importance of the input parameter x_i with respect to equal-sized perturbations from their expected values. The coefficient in equation (2.20) $b_i SD(x_i)/SD(y)$ indicates the importance of the input parameter x_i with respect to perturbations from their expected values that are equal to fixed fractions of their standard deviations (Helton 1993).

It is also possible to apply standard methods for sensitivity analysis as used in connection with Monte Carlo analysis (section 2.4.3, step five) such as scatter plots or correlation coefficients. However, instead of using the original model the response surface approximation is used as a surrogate.

More detailed information regarding the response surface method can be found in Box (1954), Box & Draper (1987), Myers et al. (1989) and Khuri & Cornell (1996).

2.4.5 Method of Sobol'

Sobol' developed the most general global SA method by introducing sensitivity indices (Sobol' 1993, Sobol' 2001, Sobol' & Kucherenko 2005). The method of Sobol' is a variance-based method based on the decomposition of $f(\mathbf{x})$ into summands of increasing dimensionality in the unit cube $K^n = [0, 1]^n$:

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j) + \dots + f_{12\dots n}(x_1, x_2, \dots, x_n). \quad (2.21)$$

Equation (2.21) is known as ANOVA decomposition or ANOVA representation (ANOVA - Analysis Of Variances) with the following properties (Sobol' 2001):

1. Integral null property: the integral of every summand $f_{i_1, \dots, i_s}(x_{i_1} \dots x_{i_s})$ over any of its independent parameters is zero

$$\int_0^1 f_{i_1, \dots, i_s}(x_{i_1} \dots x_{i_s}) dx_{i_n} = 0, \quad \text{for } 1 \leq n \leq s. \quad (2.22)$$

2. Orthogonality: because of equation (2.22) any two different summands f_{i_1, \dots, i_s} and f_{j_1, \dots, j_t} are orthogonal

$$\int_{K^n} f_{i_1, \dots, i_s}(x_{i_1} \dots x_{i_s}) f_{j_1, \dots, j_t}(x_{j_1} \dots x_{j_t}) d\mathbf{x} = 0 \quad (2.23)$$

for at least one index differing in $\{i_1, \dots, i_s\}$ and $\{j_1, \dots, j_t\}$.

All terms of the ANOVA decomposition (2.21) can be calculated via multidimen-

sional integrals:

$$f_0 = \int_{K^n} f(\mathbf{x}) d\mathbf{x} \quad (2.24)$$

$$f_i = \int_0^1 \cdots \int_0^1 f(\mathbf{x}) dx^i - f_0 \quad (2.25)$$

$$f_{ij} = \int_0^1 \cdots \int_0^1 f(\mathbf{x}) dx^{ij} - f_i - f_j - f_0 \quad (2.26)$$

...

The constant f_0 is derived by integrating over all indices, f_i is derived by integrating over all indices but x_i (indicated by dx^i) and so on.

The total variance D can be obtained from

$$D = \int_{K^n} f^2(\mathbf{x}) d\mathbf{x} - f_0^2 \quad (2.27)$$

and the partial variances D_{i_1, \dots, i_s} can be calculated from each of the terms in equation (2.21):

$$D_i = \int_0^1 f_i^2(x_i) dx_i \quad (2.28)$$

$$D_{ij} = \int_0^1 \int_0^1 f_{ij}^2(x_i, x_j) dx_i dx_j \quad (2.29)$$

...

Squaring and integrating equation (2.21) over the whole domain K^n results in:

$$D = \sum_{i=1}^n D_i + \sum_{1 \leq i < j \leq n} D_{ij} + \dots + D_{1,2,\dots,n}. \quad (2.30)$$

The sensitivity indices are defined as

$$S_{i_1, \dots, i_s} = \frac{D_{i_1, \dots, i_s}}{D}, \quad 1 \leq i_1 < \dots < i_s \leq n \quad (2.31)$$

so that all its terms sum to one:

$$\sum_{i=1}^n S_i + \sum_{1 \leq i < j \leq n} S_{ij} + \dots + S_{1,2,\dots,n} = 1. \quad (2.32)$$

The first-order sensitivity index S_i measures the main effect of the input parameter 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 (Chan et al. 2000).

Based on the sensitivity indices calculated by equation (2.31) it is possible to rank the input parameters according to their importance.

Computation of the Partial Variances

The integrals in equations (2.28) and (2.29) can be approximated via Monte Carlo integration (Homma & Saltelli 1996, Chan et al. 2000).

The approximation of the zeroth-order term f_0 of the ANOVA-decomposition (2.24) for a given sample size N is straight forward:

$$f_0 \approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}) \quad (2.33)$$

where $\mathbf{x}^{(s)}$ is a sample point in the multi-dimensional space K^n with $s = 1, 2, \dots, N$.

The total variance D can then be estimated by:

$$D \approx \frac{1}{N} \sum_{s=1}^N f^2(\mathbf{x}^{(s)}) - f_0^2. \quad (2.34)$$

The evaluation of the first-order partial variances is given in Homma & Saltelli (1996):

$$D_i = \int_0^1 f_i^2(x_i) dx_i = \int_0^1 \left[\int_0^1 \cdots \int_0^1 f(\mathbf{x}) d\mathbf{x}^i - f_0 \right]^2 dx_i \quad (2.35)$$

$$= f_0^2 - 2f_0 \int_{K^n} f(\mathbf{x}) d\mathbf{x} + \int_0^1 \left[\int_0^1 \cdots \int_0^1 f(\mathbf{x}) d\mathbf{x}^i \right]^2 dx_i \quad (2.36)$$

$$= \int_0^1 \cdots \int_0^1 f(\mathbf{u}, x_i) f(\mathbf{v}, x_i) d\mathbf{u} d\mathbf{v} - f_0^2 \quad (2.37)$$

where \mathbf{u}^i and \mathbf{v}^i are projections of \mathbf{x} on $K^n - 1 = K^n$ minus the parameter x_i . The integrals can be estimated by Monte Carlo integration which leads to:

$$D_i \approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{u}^i, x_i)^{(s)} f(\mathbf{v}^i, x_i)^{(s)} - f_0^2. \quad (2.38)$$

The Monte Carlo integration in order to estimate D_{ij} is given by:

$$D_{ij} \approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{q}^{ij}, x_i, x_j)^{(s)} \times f(\mathbf{r}^{ij}, x_i, x_j)^{(s)} - D_i - D_j - f_0^2. \quad (2.39)$$

The variables \mathbf{q}^{ij} and \mathbf{r}^{ij} are projections of \mathbf{x} on $K^n - 2 = K^n$ minus the two inputs x_i and x_j .

Two sampling matrices, containing two different sets of random values, have to be provided for the calculation of the partial variances, both with the dimension $N \times n$. For the computation of D_i the output evaluated using the inputs from the first sampling matrix (also known as the base matrix) is multiplied with the

output evaluated using the inputs from the second sampling matrix (known as the resampling matrix), where the i th column is kept constant. The same applies for the computation of D_{ij} , however in this case the i th and j th column are kept constant in the two matrices.

To obtain the total variance and all the partial variances (full characterisation of the system), altogether 2^n Monte Carlo integrals have to be evaluated. For each Monte Carlo integral different input matrices have to be provided. For example the computation of D_i calls for different input matrices than the calculation of D_j (the first input matrix $[N \times (\mathbf{u}^i, x_i)]$ is identical, however the second matrix $[N \times (\mathbf{v}^i, x_i)]$ differs for the calculation of D_i and D_j). This is a significant drawback of the described method (Homma & Saltelli 1996).

The required number of model runs to estimate the sensitivity indices is (Rabitz & Aliş 1999):

$$N \cdot \sum_{i=0}^L \frac{n!}{(n-i)!i!} \quad (2.40)$$

where L is the order of the ANOVA decomposition. The full characterisation of the model requires

$$N \cdot 2^n \quad (2.41)$$

model runs. For example, if one wants to investigate a model with only four input parameters using a second-order ANOVA decomposition (which allows the investigation of up to second-order effects), $N \cdot (1 + 4 + 6)$ model runs are required according to equation (2.40). The number of samples N can be quite high ($\geq 10^3$), however assuming $N = 1\,000$ means that a total of 11 000 model runs are required in order to estimate the sensitivity indices up to second-order.

Using this example it becomes obvious, that unless the dimension of the model input n and the dimension L is very small, far too many model runs are required.

Therefore, a more efficient approach to estimate the sensitivity indices is described in section 3.4 when the high dimensional model representation (HDMR) method is introduced.

2.4.6 Fourier Amplitude Sensitivity Test (FAST)

The Fourier amplitude sensitivity test (FAST) was created in the 1970s (Cukier et al. 1973, Schaibly & Shuler 1973, Cukier et al. 1975) and offers an alternative approach to calculate the same sensitivity indices as the Sobol' method described in the previous subsection. A summary of the FAST method is given in Chan et al. (2000) and Helton (1993).

If assuming a uniform distribution for all the inputs x_i , then the expected value E and the variance D of the model output $y = f(\mathbf{x})$ can be calculated over multidimensional integrals as

$$E(y) = \int_{R^n} f(\mathbf{x}) d\mathbf{x} \quad (2.42)$$

$$D(y) = \int_{R^n} [f(\mathbf{x} - E(y))]^2 d\mathbf{x} \quad (2.43)$$

where R^n is the entire domain of the input \mathbf{x} .

The basic idea of the FAST approach is to convert the multidimensional integrals over all the uncertain model inputs $\mathbf{x} = (x_1, \dots, x_n)$ (equation (2.42) and (2.43)) into a one dimensional integral in s by using the transformation function G_i

$$x_i = G_i(\sin \omega_i s) \quad (2.44)$$

for $i = 1, \dots, n$ with $s \in (-\pi, \pi)$ and a set of integer frequencies $\{\omega_i\}$.

The multidimensional integral in (2.42) can now be expressed as a one dimensional

integral and the expected value of the model output $y = f(\mathbf{x})$ is given by:

$$E(y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) ds \quad (2.45)$$

with

$$f(s) = f(G_1(\sin \omega_1 s), \dots, G_n(\sin \omega_n s)). \quad (2.46)$$

Further, the variance D is now given by:

$$D(y) = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s)^2 ds - E(y)^2. \quad (2.47)$$

By using the properties of the Fourier series as given in Cukier et al. (1973) the variance can be approximated by:

$$D(y) \approx 2 \sum_{k=1}^{\infty} (A_k^2 + B_k^2) \quad (2.48)$$

where A_k and B_k are the Fourier coefficients stated as:

$$A_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \cos(ks) ds \quad (2.49)$$

$$B_k = \frac{1}{2\pi} \int_{-\pi}^{\pi} f(s) \sin(ks) ds. \quad (2.50)$$

The first-order partial variances $D_i(y)$ can be approximated by

$$D_i(y) \approx 2 \sum_{k=1}^{\infty} (A_{k\omega_i}^2 + B_{k\omega_i}^2) \quad (2.51)$$

where ω_i is the integer associated with G_i in the conversion from a multidimensional integral to a one-dimensional integral (see also equations (2.45)-(2.46)).

Usually the maximum harmonic M (also known as interference factor (Saltelli et

al. 1999)) considered is taken to be four or six (Cukier et al. 1975). Thus, $D_i(y)$ can be further approximated by

$$D_i(y) = 2 \sum_{k=1}^M (A_{k\omega_i}^2 + B_{k\omega_i}^2). \quad (2.52)$$

The frequency ω_i for each input parameter can be determined according to Cukier et al. (1975):

$$\omega_1 = \Omega_n \quad (2.53)$$

$$\omega_i = \omega_{i-1} + d_{n+1-i}, \quad i = 2, \dots, n \quad (2.54)$$

where Ω_n and d_n are listed in table VI of their paper.

The calculation of the total variance D and the partial variances D_i allows an estimate for the first-order sensitivity indices:

$$S_i = \frac{D_i(y)}{D(y)} \quad (2.55)$$

The minimum number of sample points required to compute $D_i(y)$ is given in Cukier et al. (1975) as

$$N = 4\omega_{\max} - 14 \quad (2.56)$$

with ω_{\max} the maximum frequency amongst the set of ω_i . Saltelli et al. (1999) states the minimum sample size at

$$N = 2M\omega_{\max} + 1 \quad (2.57)$$

and Fang et al. (2003) uses

$$N = 5\omega_{\max} + 1 \quad (2.58)$$

in order to define the sample size. Cukier et al. (1975) have derived an empirical

relation between the sample size N and the number of parameters n . This relation is found to be (for $M = 4$):

$$N \simeq 2.6n^{2.5} \quad (2.59)$$

Due to symmetry properties the required number of model runs is only $N/2$ (Cukier et al. 1975).

In general classic FAST (as introduced by Cukier et al. (1973)) is computationally more efficient than the method of Sobol', because the evaluation of the first-order sensitivity indices can be carried out for all factors using only one set of runs (Saltelli et al. 1999). For example, the minimum number of model runs required for a model with four input parameters would only be 42, according to equation 2.59 and considering the symmetric properties. However, the number of required model runs directly depends on the number of input parameters n (see empirical relation in equation (2.59)). For instance, considering a model with 30 parameters the required number of model runs would already exceed 6 000.

By using classic FAST only the first-order terms S_i can be explored. This makes the approach insufficient if the sum of the first-order indices is much less than one (Sobol' 2001). An extended FAST was introduced by Saltelli et al. (1999) and can be used to calculate the total effect indices $TS(i)$ as well. However, the total effect indices do not provide a complete characterisation of the system. The total effect index $TS(i)$ is defined as the sum of all sensitivity indices involving the input parameter in question. It is therefore more reliable than the first-order index, if one is interested in the overall effect of each input parameter upon the output. However, the FAST method loses its attractive feature, that only one set of samples is required in order to calculate all indices. Using extended FAST a new set of samples will be needed to evaluate each of the total effect indices $TS(i)$ (Saltelli et al. 1999).

2.5 Summary

A variety of methods for global sensitivity analysis have been introduced in this chapter. Each method has its advantages and disadvantages. A detailed comparison of the Monte Carlo analysis, FAST and RSM can be found for example in Helton (1993) or Helton & Davis (2003).

Screening methods, which can be local or global depending on their set up, are very easy to implement and their low computational requirements make them often a first choice in order to produce a parameter ranking. However, they can not quantify how much a given factor is more important than another. In order to obtain a quantitative measure, other methods are required. Also, for certain types of model systems their convergence properties can be poor, increasing their computational requirements.

Monte Carlo analysis is probably the conceptually simplest approach among the methods discussed here for global sensitivity analysis and therefore widely used. The two main advantages of this method are the full coverage of the range of each input parameter and the large variety of uncertainty and sensitivity analysis techniques available. Different sampling procedures can be applied such as Latin hypercube sampling or quasi-random sampling. Particularly the quasi-Monte Carlo approach (MC analysis using quasi-random sampling) achieves a significantly higher accuracy than the Monte Carlo method using random sampling (Niederreiter 1992). A major drawback of the Monte Carlo analysis is in the large number of model runs required. Therefore, it has limited applicability for computationally expensive models. Another drawback is the difficulty in interpreting the results (e.g. scatter plots) for a large number of parameters.

The response surface method (RSM) uses a surrogate of the original model in order to

perform sensitivity and uncertainty analysis. Thus, this methodology is even applicable for models, which are expensive to run. However, there are several drawbacks of the response surface method. One point is that the development of an appropriate experimental design can cause difficulties, because the form of the model might not be known (Helton 1993). It is also possible, that many outputs have to be considered which could require different experimental designs. Another point is that only a few values are used for each input parameter as design points. A major drawback is that the response surface method works only for a limited number of input parameters (typically less than ten) and a few distinct output variables (Helton 1993). Depending on the model which is used to create the response surface, the relationship between the input and output variables is only linear or quadratic, which may be insufficient for some non-linear models. Even if the use of a surrogate promises a less expensive method for uncertainty and sensitivity analysis, the results are only as good as the response surface approximation of the real model.

Sobol's method is based on the variance decomposition and is used for the calculation of sensitivity indices. The whole range of the input parameters is covered and the full model is used for the estimation of the total and partial variances. A significant drawback of Sobol's method is in the high number of required model runs, because for each estimation (e.g. each first-order sensitivity index S_i) a different set of samples is required.

The Fourier amplitude sensitivity test (FAST) method offers an alternative approach to calculate sensitivity indices. The full range of the input parameters is considered and the original model is used for the estimation of expected value and variance. Only one set of samples is required in order to calculate all first-order sensitivity indices, which makes FAST computationally more efficient than the method of Sobol'. However, the FAST method is quite complicated and difficult to explain and only the first-order indices can be determined if using classic FAST. Extended FAST en-

ables the calculation of total sensitivity indices $TS(i)$, however it requires a new set of samples to evaluate each of the $TS(i)$.

After reviewing traditional methods for global SA it can be said that there is a need for more general methods for the analysis of non-linear responses and parameter interactions within the field of environmental and safety engineering. Methods are required, which are computationally efficient and can handle a large number of input parameters at the same time. Traditional methods can not fulfil all of these requirements. The more recently developed High Dimensional Model Representation (HDMR) method provides a promising approach in order to cope with a large input space dimension very efficiently. The method produces a detailed mapping of the input parameter space to selected outputs which is fundamental to global SA. Sensitivity indices can be calculated in an automatic way that can then be directly used in importance ranking and to explore parameter interactions. The HDMR method is introduced in the next chapter and its effectiveness is demonstrated using various analytical test functions and case studies from the field of environmental and safety engineering in the chapters 4 to 6.

3 High Dimensional Model Representation (HDMR)

Sensitivity analysis in general tries to reveal the relationship between the model inputs and the model outputs. This is usually done by Monte Carlo methods or variance-based approaches as described earlier in chapter 2. These methods aim to deduce the contribution of the input parameters to the output uncertainties. Another approach is to produce a detailed mapping of the input parameter space upon the output, which can be fundamental to overall model analysis (Rabitz & Aliş 2000). Without any simplification the identification of the input-output relationship of a system with a high dimensional input would be computationally too expensive. The high dimensional model representation (HDMR) method introduced by Rabitz et al. (1999) can dramatically reduce the computational effort needed for the mapping and was mainly developed to express the input-output relationship of a complex model with a large number of input parameters.

HDMR is an expansion with a hierarchical form in terms of the input parameters (Rabitz et al. 1999, Rabitz & Aliş 1999, Li et al. 2000). The mapping between the input parameters x_1, \dots, x_n and the output $f(\mathbf{x}) = f(x_1, \dots, x_n)$ in the domain R^n

can be written in the following form:

$$f(\mathbf{x}) = f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j) + \dots + f_{12\dots n}(x_1, x_2, \dots, x_n). \quad (3.1)$$

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 parameter 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 parameters x_i and x_j upon the output $f(\mathbf{x})$. The higher-order terms reflect the cooperative effects of increasing numbers of input parameters acting together to influence the output $f(\mathbf{x})$. If there is no interaction between the input parameters, then only the zeroth-order term f_0 and the first-order terms $f_i(x_i)$ will appear in the HDMR expansion.

The notation of zeroth-, first-, second-order, etc. in the HDMR expansion should not be confused with the terminology of a Taylor series: the HDMR expansion is exact and always of finite order (Rabitz & Aliş 2000). Li et al. (2001*b*) highlights that HDMR expansions possess faster convergence properties than Taylor series and the relationship between both approximation methods is shown. The Taylor series has an infinite number of terms whereas the HDMR expansion has only a finite number of terms, and it is stated that each HDMR component function can be composed of an infinite subclass of the full multidimensional Taylor series. Therefore, it is likely that a truncated HDMR expansion gives a better approximation of $f(\mathbf{x})$ than any truncated Taylor series. Furthermore, it is argued that high-order HDMR component functions are usually smaller than low-order ones. This will be investigated when the HDMR method is applied to various case studies in the chapters 4 to 6.

The HDMR expansion is computationally very efficient if higher order input parameter effects are weak and can therefore be neglected. For many systems a HDMR

expression up to second-order

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^n f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j) \quad (3.2)$$

already provides satisfactory results and a good approximation of $f(\mathbf{x})$ (Li et al. 2001b). Therefore, the main focus in this work will be on an up to second-order HDMR expansion.

There are two commonly used HDMR expansions: cut-HDMR, which depends on the value of $f(\mathbf{x})$ at a specific reference point $\bar{\mathbf{x}}$ as described in the next section, and RS-HDMR, which depends on the average value of $f(\mathbf{x})$ over the whole domain and is explained in section 3.2.

3.1 Cut-HDMR

The cut-HDMR expansion can be constructed, when ordered sampling of the output $f(\mathbf{x})$ is possible. A reference point $\bar{\mathbf{x}} = (\bar{x}_1, \dots, \bar{x}_n)$ has to be selected first. This could be in the simplest case the midpoint for each of the parameter ranges (see also comments in section 3.1.1). The component functions of the cut-HDMR with respect to the reference point $\bar{\mathbf{x}}$ possess the following forms (Wang et al. 2001, Wang et al. 2005):

$$f_0 = f(\bar{\mathbf{x}}) \quad (3.3)$$

$$f_i(x_i) = f(x_i, \bar{\mathbf{x}}^i) - f_0 \quad (3.4)$$

$$f_{ij}(x_i, x_j) = f(x_i, x_j, \bar{\mathbf{x}}^{ij}) - f_i(x_i) - f_j(x_j) - f_0 \quad (3.5)$$

...

where

$$f_i(x_i, \bar{\mathbf{x}}) = (\bar{x}_1, \dots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \dots, \bar{x}_n) \quad (3.6)$$

$$f_{ij}(x_i, x_j, \bar{\mathbf{x}}) = (\bar{x}_1, \dots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \dots, \bar{x}_{j-1}, x_j, \bar{x}_{j+1}, \dots, \bar{x}_n) \quad (3.7)$$

...

The notation $f_i(x_i, \bar{\mathbf{x}}) = (\bar{x}_1, \dots, \bar{x}_{i-1}, x_i, \bar{x}_{i+1}, \dots, \bar{x}_n)$ denotes that all input parameters are at their reference point values except x_i . The same applies for $f_{ij}(x_i, x_j, \bar{\mathbf{x}})$. In this case all input parameters are at their reference point values except x_i and x_j .

The f_0 term is a constant representing the output response of the system at the reference point $\bar{\mathbf{x}}$. The higher-order component functions are calculated along its parameter axis through the reference point $\bar{\mathbf{x}}$, only the last term $f_{12\dots n}(x_1, x_2, \dots, x_n)$ in equation (3.1) is evaluated from the difference between $f(\mathbf{x})$ and all the other component functions. The input-output response of the model is evaluated along lines, surfaces, subvolumes and so on in the input space dimension. Thus, the higher order terms of the HDMR expansion are determined as “cuts” through the reference point and therefore the method is named cut-HDMR. Figure 3.1 shows an example of cut-lines and cut-surfaces for two input parameters, if using the midpoint as a reference point.

The cut-HDMR component functions are usually numerically represented as low-dimensional look-up tables. The value of $f(\mathbf{x})$ for an arbitrary point \mathbf{x} can be calculated by the following two steps:

1. Perform a low-dimensional interpolation over the HDMR expansion terms $f_i(x_i)$, $f_{ij}(x_i, x_j)$, ... with respect to the input values of the given point \mathbf{x} ,
2. Sum all interpolated values of the higher-order HDMR terms (first-order up

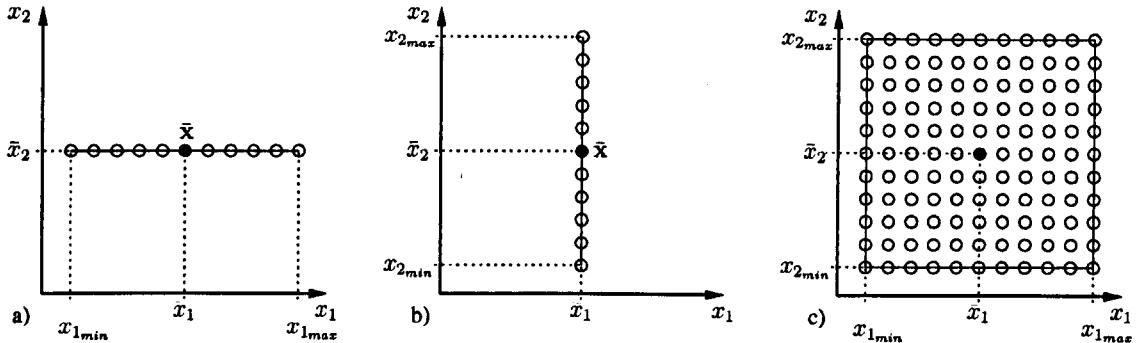


Figure 3.1: (a) Calculation of $f_1(x_1)$ along the x_1 -axis through the reference point, (b) calculation of $f_2(x_2)$ along the x_2 -axis through the reference point, (c) calculation of $f_{12}(x_1, x_2)$ along the surface of x_1 and x_2 through the reference point.

to highest-order) and then add the zeroth-order term as well.

A second-order cut-HDMR expansion requires the storage of the component functions in numerical tables for the first-order terms $f_i(x_i)$ and the second-order terms $f_{ij}(x_i, x_j)$. The interpolation in order to calculate the value for an arbitrary point \mathbf{x} has to be performed for a one-dimensional and two-dimensional function. A linear interpolation between the grid points is usually sufficient, however other methods such as spline interpolation can also be used. The general meaning and interpretation of the component functions is given in section 3.4.3.

The required number of model runs to construct the cut-HDMR component functions up to the L th order is given by:

$$\sum_{i=0}^L \frac{n!}{(n-i)!i!} (m)^i \quad (3.8)$$

where m is the number of sample points taken along each axis (Rabitz & Aliş 2000).

The lower order terms are subtracted in the process to determine the component functions and therefore the calculation of the model response at the reference point

is not required for the higher-order terms, since the value of the expansion is zero at this point (see equation (3.3)-(3.5)). Thus, the number of necessary model runs can even be reduced and is now given by:

$$\sum_{i=0}^L \frac{n!}{(n-i)!i!} (m-1)^i. \quad (3.9)$$

For example, a second-order cut-HDMR expansion ($L = 2$) with five input parameters ($n = 5$) and 11 grid points ($m = 11$) used for each of the parameter ranges would require $N = 1051$ model runs (1 zeroth order + 50 first order + 1 000 second order). It can be seen that the sampling effort directly depends on the number of input parameters and the number of grid points. This means that the cut-HDMR method can become computationally expensive if investigating models with a large input space dimension, especially if one is also interested in the higher-order effects.

3.1.1 Reference Point

In the convergence limit, the cut-HDMR is invariant to the choice of reference point $\bar{\mathbf{x}}$. In practice the reference point $\bar{\mathbf{x}}$ is usually chosen within the neighbourhood of the point of interest in the input space (Wang et al. 2005).

However, Sobol' (2003) investigated a class of models where a careless choice of the reference point $\bar{\mathbf{x}}$ can spoil the approximation. Therefore, he suggests a method on how to determine the reference point more wisely.

First of all the model output $f(\mathbf{x})$ is calculated for a moderate number N of quasi-random trial points uniformly distributed in R^n . Then a crude estimate of the mean can be calculated:

$$f_0^* \approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}). \quad (3.10)$$

The reference point $\bar{\mathbf{x}}$ is selected as the trial point $\mathbf{x}^{(s_0)}$ whose output is nearest to f_0^* :

$$|f(\mathbf{x}^{(s_0)}) - f_0^*| = \min_{1 \leq s \leq N} |f(\mathbf{x}^{(s)}) - f_0^*|. \quad (3.11)$$

The selection of the reference point requires N additional model runs which increases the computational effort of the cut-HDMR method. However, these additional model runs can also be used to assess the accuracy of the constructed HDMR metamodel as discussed in section 3.3.

3.1.2 Multicut-HDMR

If the input space dimension is too large, then the HDMR function expansion may not converge and cut-HDMR will not be able to accurately approximate the model output $f(\mathbf{x})$ by using only one single reference point $\bar{\mathbf{x}}$. To address this problem the Multicut-HDMR technique has been introduced in Li et al. (2004). Multicut-HDMR allows the use of multiple reference points which leads to an improvement in the accuracy of the metamodel. However, it also leads to an increase in the number of sample points required. Multicut-HDMR is not further discussed here and more detailed information can be obtained from Li et al. (2004).

3.2 RS-HDMR

In contrast to cut-HDMR, RS-HDMR requires a set of random sample points N over the entire domain R^n . The component functions are determined through an averaging process. First of all, the input parameters x_i are rescaled such that $0 \leq x_i \leq 1$ for all i . The output function is then defined in the unit hypercube $K^n = \{(x_1, \dots, x_n), i = 1, \dots, n\}$. The component functions of RS-HDMR possess the

following forms (Li et al. 2002a):

$$f_0 = \int_{K^n} f(\mathbf{x}) d\mathbf{x} \quad (3.12)$$

$$f_i(x_i) = \int_{K^{n-1}} f(\mathbf{x}) dx^i - f_0 \quad (3.13)$$

$$f_{ij}(x_i, x_j) = \int_{K^{n-2}} f(\mathbf{x}) dx^{ij} - f_i(x_i) - f_j(x_j) - f_0 \quad (3.14)$$

...

where $d\mathbf{x}^i$ stands for the product $dx_1 dx_2 \dots dx_n$ without dx_i and $d\mathbf{x}^{ij}$ stands for the same product, however without dx_i and dx_j . The last term $f_{12\dots n}(x_1, x_2, \dots, x_n)$ in equation (3.1) is evaluated from the difference between $f(\mathbf{x})$ and all the other component functions.

For the evaluation of f_0 , N samples of the n -dimensional input vector

$$\mathbf{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \dots, x_n^{(s)}), \quad s = 1, 2, \dots, N \quad (3.15)$$

are randomly generated uniformly distributed in K^n so that f_0 can be approximated by the average value of $f(\mathbf{x})$ for all $\mathbf{x}^{(s)}$:

$$f_0 = \int_{K^n} f(\mathbf{x}) d\mathbf{x} \quad (3.16)$$

$$\approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}). \quad (3.17)$$

The calculation of the higher order component functions can be obtained in various ways and is described in the following section.

3.2.1 Computation of the Higher Order Component Functions

The higher order component functions can be determined using two different approaches. The first uses direct determination, which requires sampling the output on a regular net in order to evaluate the integrals in equations (3.13) and (3.14) via Monte Carlo approximation. The second approach uses analytical basis functions or kernel smoothers in order to approximate the higher order component functions.

Direct Determination

The component functions are provided numerically at discrete values of the input parameters in low-dimensional look-up tables. The construction of the tables makes the evaluation of the integrals in the equations (3.12)-(3.14) necessary, which can be carried out by Monte Carlo approximation (Li et al. 2002*a*, Li et al. 2002*b*).

The direct determination of the first-order component function $f_i(x_i)$ requires different sets of random samples of $f(x_i, \mathbf{x}^i)$ at

$$(x_i, \mathbf{x}^i)^{(s)} = (x_1^{(s)}, \dots, x_{i-1}^{(s)}, x_i, x_{i+1}^{(s)}, \dots, x_n^{(s)}), \quad s = 1, 2, \dots, N. \quad (3.18)$$

The input parameter x_i is fixed at m distinct values so that mN random samples are required to construct the $f_i(x_i)$ table:

$$f_i(x_i) = \int_{K^{n-1}} f(\mathbf{x}) d\mathbf{x}^i - f_0 \quad (3.19)$$

$$\approx \frac{1}{N} \sum_{s=1}^N f((x_i, \mathbf{x}^i)^{(s)}) - \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}). \quad (3.20)$$

The second-order component functions $f_{ij}(x_i, x_j)$ are constructed similarly with dif-

ferent sets of random samples of $f(x_i, x_j, \mathbf{x}^{ij})$ at

$$(x_i, x_j, \mathbf{x}^{ij})^{(s)} = (x_1^{(s)}, \dots, x_{i-1}^{(s)}, x_i, x_{i+1}^{(s)}, \dots, x_{j-1}^{(s)}, x_j, x_{j+1}^{(s)}, \dots, x_n^{(s)}) \quad (3.21)$$

for $s = 1, 2, \dots, N$ with m distinct values for x_i and x_j it takes m^2N samples to construct the second-order term $f_{ij}(x_i, x_j)$:

$$f_{ij}(x_i, x_j) = \int_{K^{n-2}} f(\mathbf{x}) d\mathbf{x}^{ij} - f_i(x_i) - f_j(x_j) - f_0 \quad (3.22)$$

$$\begin{aligned} &\approx \frac{1}{N} \sum_{s=1}^N f((x_i, x_j, \mathbf{x}^{ij})^{(s)}) - \frac{1}{N} \sum_{s=1}^N f((x_i, \mathbf{x}^i)^{(s)}) - \\ &\quad \frac{1}{N} \sum_{s=1}^N f((x_j, \mathbf{x}^j)^{(s)}) - \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}). \end{aligned} \quad (3.23)$$

The required number of samples to construct the RS-HDMR expansion up to second-order is given by

$$N + nmN + \frac{n(n-1)m^2N}{2} \quad (3.24)$$

which makes the direct approach prohibitively expensive (Li et al. 2002a). For example, if considering a model with five input parameters ($n = 5$), 11 distinct values ($m = 11$) for each of the inputs and a set of 1000 random samples for $\mathbf{x}^{(s)}$, $(x_i, \mathbf{x}^i)^{(s)}$ and $(x_i, x_j, \mathbf{x}^{ij})^{(s)}$, then it would require a total of 1 266 000 model runs (1 000 zeroth-order + 55 000 first-order + 1 210 000 second-order).

Approximation of the Component Functions

The sampling effort can be dramatically reduced by approximating the RS-HDMR component functions by either a set of suitable analytical functions or numerically by using kernel smoothers (Li et al. 2001a).

In this work the focus is only on the analytical approximation of the component functions, where orthonormal polynomials, spline functions or simple polynomials can be used as basis functions (Li et al. 2002a). The RS-HDMR component functions now read as

$$f_i(x_i) \approx \sum_{r=1}^k \alpha_r^i \varphi_r(x_i) \quad (3.25)$$

$$f_{ij}(x_i, x_j) \approx \sum_{p=1}^l \sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_p(x_i) \varphi_q(x_j) \quad (3.26)$$

...

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 basis functions.

Using a second-order RS-HDMR expansion, the model output can now be expressed as

$$f(\mathbf{x}) \approx f_0 + \sum_{i=1}^n \sum_{r=1}^k \alpha_r^i \varphi_r(x_i) + \sum_{1 \leq i < j \leq n} \sum_{p=1}^l \sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_p(x_i) \varphi_q(x_j). \quad (3.27)$$

The coefficients α_r^i and β_{pq}^{ij} can be determined by minimisation of

$$\min_{\alpha_r^i} \int_0^1 \left[f_i(x_i) - \sum_{r=1}^k \alpha_r^i \varphi_r(x_i) \right]^2 dx_i \quad (3.28)$$

$$\min_{\beta_{pq}^{ij}} \int_0^1 \int_0^1 \left[f_{ij}(x_i, x_j) - \sum_{p=1}^l \sum_{q=1}^{l'} \beta_{pq}^{ij} \varphi_p(x_i) \varphi_q(x_j) \right]^2 dx_i dx_j \quad (3.29)$$

which leads to a linear equation

$$\mathbf{A}\mathbf{y} = \mathbf{b} \quad (3.30)$$

where \mathbf{A} is a constant nonsingular matrix, \mathbf{b} is a vector whose elements are integrals over a product of $f(\mathbf{x})$ and the basis functions and \mathbf{y} is the vector of coefficients for

the basis functions (Li et al. 2002a).

For example, the (r, r') -entry for the matrix \mathbf{A} to determine the coefficient α_r^i is given by

$$A_{r,r'} = \int_0^1 \varphi_r(x_i) \varphi_{r'}(x_i) dx_i \quad r, r' = 1, 2, \dots, k. \quad (3.31)$$

If applying a polynomial expansion order of $k = 2$, one would get the following linear equation in order to obtain the α_r^i coefficients:

$$\underbrace{\begin{bmatrix} \int_0^1 f_i(x_i) \varphi_1(x_i) dx_i \\ \int_0^1 f_i(x_i) \varphi_2(x_i) dx_i \end{bmatrix}}_b = \underbrace{\begin{bmatrix} \int_0^1 \varphi_1(x_i) \varphi_1(x_i) dx_i & \int_0^1 \varphi_2(x_i) \varphi_1(x_i) dx_i \\ \int_0^1 \varphi_1(x_i) \varphi_2(x_i) dx_i & \int_0^1 \varphi_2(x_i) \varphi_2(x_i) dx_i \end{bmatrix}}_A \cdot \underbrace{\begin{bmatrix} \alpha_1^i \\ \alpha_2^i \end{bmatrix}}_y \quad (3.32)$$

In Li et al. (2002a) it is also demonstrated that the approximation of the HDMR component functions by orthonormal polynomials provides the best accuracy if compared with splines and normal polynomials. Therefore, the focus of this work will be on orthonormal polynomials only.

Orthonormal Polynomials Polynomials are orthogonal in the domain $[a, b]$ if they satisfy the following conditions:

$$\int_a^b \varphi_k(x) dx = 0 \quad k = 1, 2, \dots \quad (3.33)$$

$$\int_a^b \varphi_k(x) \varphi_l(x) dx = 0 \quad k \neq l. \quad (3.34)$$

Additionally, if the polynomials have the unit norm

$$\int_a^b \varphi_k^2(x) dx = 1 \quad k = 1, 2, \dots \quad (3.35)$$

$$(3.36)$$

then they are called orthonormal.

Different orthogonal polynomials are known, such as Jacobi, Chebyshev or Legendre polynomials (Hochstrasser 1972) and could potentially be used as basis functions.

In this work Legendre polynomials are used to provide the set of orthonormal basis functions. Legendre polynomials are orthogonal in the domain $[-1, 1]$. Because orthonormal polynomials are required for the domain $[0, 1]$ (for the RS-HDMR approach the inputs x_i are rescaled between 0 and 1), one can use the shifted Legendre polynomials, which are also widely applied in optimal control theory (see for example Horng & Chou (1986) or Wang & Chang (1985)):

$$\varphi_1(x) = \sqrt{3}(2x - 1) \quad (3.37)$$

$$\varphi_2(x) = 6\sqrt{5} \left(x^2 - x + \frac{1}{6} \right) \quad (3.38)$$

$$\varphi_3(x) = 20\sqrt{7} \left(x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20} \right) \quad (3.39)$$

...

The calculation of the orthonormal basis using shifted Legendre polynomials is illustrated in the appendix (section A.2) and a list of up to tenth-order polynomials is provided.

Due to the orthonormal character of the basis functions the matrix \mathbf{A} in equation (3.30) is the identity matrix which allows the computation of the α_r^i coefficients as

follows:

$$\alpha_r^i = \int_0^1 f_i(x_i) \varphi_r(x_i) dx_i \quad (3.40)$$

$$\begin{aligned} \alpha_r^i &= \int_0^1 \left[\underbrace{\int_{K^{n-1}} f(\mathbf{x}) d\mathbf{x}^i - f_0}_{f_i(x_i)} \right] \varphi_r(x_i) dx_i \end{aligned} \quad (3.41)$$

$$= \int_0^1 \int_{K^{n-1}} f(\mathbf{x}) \varphi_r(x_i) d\mathbf{x}^i dx_i - f_0 \underbrace{\int_0^1 \varphi_r(x_i) dx_i}_0 \quad (3.42)$$

$$= \int_{K^n} f(\mathbf{x}) \varphi_r(x_i) d\mathbf{x}. \quad (3.43)$$

Using Monte Carlo integration for the approximation of the high dimensional integral one gets:

$$\alpha_r^i \approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}) \varphi_r(x_i^{(s)}). \quad (3.44)$$

The estimation of the β_{pq}^{ij} coefficients is similar and leads to:

$$\beta_{pq}^{ij} = \int_{K^n} f(\mathbf{x}) \varphi_p(x_i) \varphi_q(x_j) d\mathbf{x}. \quad (3.45)$$

$$\approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}) \varphi_p(x_i^{(s)}) \varphi_q(x_j^{(s)}). \quad (3.46)$$

By using analytical basis functions only one set of random samples N is required in order to determine all RS-HDMR component functions. This is a dramatic reduction of the sampling effort in comparison to the direct approach. Moreover, it is only necessary to store the expansion coefficients α_r^i and β_{pq}^{ij} . The value of $f(\mathbf{x})$ for an arbitrary point \mathbf{x} can then be calculated by the following two steps:

1. Calculate the HDMR expansion terms $f_i(x_i)$, $f_{ij}(x_i, x_j)$, ... with respect to the input values of the given point \mathbf{x} by using the analytical equations of the shifted Legendre polynomials and the corresponding expansion coefficients,

2. Sum all calculated values of the higher-order HDMR terms (first-order up to highest-order) and then add the mean (zeroth-order term).

The approximation of the HDMR component functions by orthonormal polynomials does not involve the storage of data in numerical tables and no interpolation is required.

3.2.2 Variance Reduction Methods

The determination of the expansion coefficients for the analytical basis functions (see equation (3.44) and (3.46)) is carried out by Monte Carlo integration. The error of the Monte Carlo integration controls the accuracy of the RS-HDMR metamodel and therefore it is necessary to assure that this error is kept small.

Before introducing methods which are able to control the accuracy of the RS-HDMR metamodel one needs to focus on some details of the Monte Carlo integration process. The theoretical foundations of the Monte Carlo integration can be found in Kalos & Whitlock (1986) or Robert & Casella (2004). The following summary is based on the one given in Li et al. (2003b).

Monte Carlo Integration

The random variable

$$F_N = \frac{1}{N} \sum_{s=1}^N F(\mathbf{x}^{(s)}) \quad (3.47)$$

with the function $F(\mathbf{x})$ and the independent random parameters $\mathbf{x} = (x_1, x_2, \dots, x_n)$ with a uniform probability density function has the following expectation value $E(F_N)$

$$E(F_N) = \int_{K^n} F(\mathbf{x}) d\mathbf{x}. \quad (3.48)$$

The variance D and the standard deviation (standard error) σ are given by:

$$D(F_N) = \frac{1}{N} D(F(\mathbf{x})) \quad (3.49)$$

$$\sigma(F_N) = \sqrt{D(F_N)} \quad (3.50)$$

$$= \frac{1}{\sqrt{N}} \sigma(F(\mathbf{x})). \quad (3.51)$$

The value F_N can be used as an estimate (empirical average) of the integral in equation (3.48). From equation (3.51) it is obvious that the Monte Carlo integration error (standard deviation σ) can be reduced by either increasing the sample size N or by decreasing the variance of $F(\mathbf{x})$.

According to Li et al. (2003b) the required sample size N to achieve a good accuracy of the expansion coefficients is mainly dependent on the order of the component function and not so much on the dimension of the input parameter vector \mathbf{x} . This means, that the determination of the higher order RS-HDMR component functions generally requires additional samples. A few hundred samples might be sufficient in order to calculate the first order component functions $f_i(x_i)$. However, to achieve a good accuracy for the second order component functions $f_{ij}(x_i, x_j)$ thousands of samples may be required (Li et al. 2003b). For practical applications the sample size N is often restricted by time and cost considerations and therefore an increase in the sample size is not always affordable. Another way to improve the accuracy of the Monte Carlo integration is to reduce the variance of the integrand. This is achieved by variance reduction methods and two methods are introduced in the following sections: the correlation method and the ratio control variate method.

Correlation Method

The correlation method for variance reduction of Monte Carlo integration is introduced in Kalos & Whitlock (1986) and was suggested for use in RS-HDMR by Li et al. (2003b). The determination of the expansion coefficients becomes an iterative process. For example, consider the integral for the coefficient α_r^i (equation 3.43):

$$\alpha_r^i = \int_{K_n} f(\mathbf{x})\varphi_r(x_i)d\mathbf{x}. \quad (3.52)$$

In order to reduce the variance of the integrand $f(\mathbf{x})\varphi_r(x_i)$ a reference function $h(\mathbf{x})$ is introduced which satisfies two conditions:

1. $f(\mathbf{x}) - h(\mathbf{x})$ is almost constant or vanishes in the whole domain,
2. The integral

$$\int_{K_n} h(\mathbf{x})\varphi_r(x_i)d\mathbf{x} \quad (3.53)$$

is analytically known.

The introduction of $h(\mathbf{x})$ leads to:

$$\alpha_r^i = \int_{K_n} (f(\mathbf{x}) - h(\mathbf{x}))\varphi_r(x_i)d\mathbf{x} + \int_{K_n} h(\mathbf{x})\varphi_r(x_i)d\mathbf{x} \quad (3.54)$$

The variance comes only from the first term in equation (3.54). As $f(\mathbf{x}) - h(\mathbf{x})$ is almost constant or zero it is expected that the variance of the integrand of the first term in equation (3.54) is smaller than the variance of the integrand in equation 3.52.

A truncated RS-HDMR expansion (for example up to second-order) can be used as $h(\mathbf{x})$:

$$h(\mathbf{x}) = f_0 + \sum_{i=1}^n \sum_{r=1}^k \bar{\alpha}_r^i \varphi_r(x_i) + \sum_{1 \leq i < j \leq n} \sum_{p=1}^l \sum_{q=1}^{l'} \bar{\beta}_{pq}^{ij} \varphi_p(x_i) \varphi_q(x_j). \quad (3.55)$$

The coefficients $\bar{\alpha}_r^i$ and $\bar{\beta}_{pq}^{ij}$ are the initial values and can be determined by direct Monte Carlo integration as given in equations (3.43) - (3.46). If the truncated RS-HDMR expansion is a good approximation of $f(\mathbf{x})$, then the difference $f(\mathbf{x}) - h(\mathbf{x})$ should be very small. In Li et al. (2003b) it is shown that

$$\int_{K_n} h(\mathbf{x}) \varphi_r(x_i) d\mathbf{x} = \bar{\alpha}_r^i. \quad (3.56)$$

This leads to the approximation of α_r^i by Monte Carlo integration

$$\alpha_r^i \approx \frac{1}{N} \sum_{s=1}^N (f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})) \varphi_r(x_i^{(s)}) + \bar{\alpha}_r^i \quad (3.57)$$

with better accuracy than given by equation 3.44.

The approximation of β_{pq}^{ij} is similarly and leads to

$$\beta_{pq}^{ij} \approx \frac{1}{N} \sum_{s=1}^N (f(\mathbf{x}^{(s)}) - h(\mathbf{x}^{(s)})) \varphi_p(x_i^{(s)}) \varphi_q(x_j^{(s)}) + \bar{\beta}_{pq}^{ij} \quad (3.58)$$

The first terms in equations 3.57-3.58 are corrections for the initial values $\bar{\alpha}_r^i$ and $\bar{\beta}_{pq}^{ij}$. The resulting values α_r^i and β_{pq}^{ij} can be reused again as new initial values for the construction of a new $h(\mathbf{x})$ which leads to an even smaller difference $f(\mathbf{x}) - h(\mathbf{x})$.

The procedure in how to determining the expansion coefficients can be summarised in the following steps:

1. Define $h(\mathbf{x})$ by using for example a second-order HDMR expansion (equation (3.55)) and determine its expansion coefficients by direct Monte Carlo

integration:

$$\bar{\alpha}_r^i \approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}) \varphi_r(x_i^{(s)}) \quad (3.59)$$

$$\bar{\beta}_{pq}^{ij} \approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}) \varphi_p(x_i^{(s)}) \varphi_q(x_j^{(s)}) \quad (3.60)$$

2. Calculate a correction for the initial values $\bar{\alpha}_r^i$ and $\bar{\beta}_{pq}^{ij}$ by using equation (3.57) and (3.58).
3. Set the calculated corrections as the new initial values:

$$\bar{\alpha}_r^i = \alpha_r^i \quad (3.61)$$

$$\bar{\beta}_{pq}^{ij} = \beta_{pq}^{ij} \quad (3.62)$$

4. Go back to step 2 and repeat a pre-defined number of times or until a defined convergence criterion is reached.

In this way the determination of the expansion coefficients becomes an iteration procedure which is expected to be convergent if the initial value $h(\mathbf{x})$ is close to $f(\mathbf{x})$ and if the sample size N is large enough.

Ratio Control Variate Method

The ratio control variate method applied by Li & Rabitz (2006) for RS-HDMR is even more stable under iteration than the correlation method. The approach is similar, however instead of using the difference of $f(\mathbf{x}) - h(\mathbf{x})$ as presented in equation 3.54 for the calculation of α_r^i , the ratio of $f(\mathbf{x})$ and $h(\mathbf{x})$ is used so that

$$\alpha_r^i = \frac{\int_{K_n} f(\mathbf{x}) \varphi_r(x_i) d\mathbf{x}}{\int_{K_n} h(\mathbf{x}) \varphi_r(x_i) d\mathbf{x}} \cdot \int_{K_n} h(\mathbf{x}) \varphi_r(x_i) d\mathbf{x}. \quad (3.63)$$

Again, a truncated RS-HDMR expansion can be used as $h(\mathbf{x})$. The approximation of α_r^i by Monte Carlo integration leads to

$$\alpha_r^i \approx \frac{\sum_{s=1}^N f(\mathbf{x}^s) \varphi_r(x_i^{(s)})}{\sum_{s=1}^N h(\mathbf{x}^s) \varphi_r(x_i^{(s)})} \cdot \bar{\alpha}_r^i. \quad (3.64)$$

The approximation of β_{pq}^{ij} is given by:

$$\beta_{pq}^{ij} \approx \frac{\sum_{s=1}^N f(\mathbf{x}^s) \varphi_p(x_i^{(s)}) \varphi_q(x_j^{(s)})}{\sum_{s=1}^N h(\mathbf{x}^s) \varphi_p(x_i^{(s)}) \varphi_q(x_j^{(s)})} \cdot \bar{\beta}_{pq}^{ij}. \quad (3.65)$$

Similar to the correlation method the expansion coefficients α_r^i and β_{pq}^{ij} are determined iteratively and the initial values $\bar{\alpha}_r^i$ and $\bar{\beta}_{pq}^{ij}$ are calculated by direct Monte Carlo integration.

3.3 Accuracy of the HDMR Metamodel

After the metamodel has been constructed using either the cut-HDMR or RS-HDMR approach, it is necessary to validate its accuracy. A good accuracy of the metamodel ensures that one can rely on the constructed HDMR expansion and shows that the model behaviour has been captured well. Another important reason for assessing the accuracy is that the calculation of the sensitivity indices (see section 3.4) relies entirely on the constructed metamodel.

If cut-HDMR has been used to construct the metamodel, then an additional set of samples (e.g. $N = 1000$ using random sampling) is required in order to show the accuracy. The additional set of samples becomes necessary because the cut-HDMR component functions are calculated along grid lines, grid planes and so on (see section 3.1 for details). Therefore, the solution at the grid points for the metamodel

is exact and the set of ordered samples which has been used for the construction of the cut-HDMR expansion cannot be used to assess its accuracy. The calculation of the metamodel response for an arbitrary point involves interpolation over the low dimensional numerical tables (see two step algorithm in section 3.1) which leads to an error. The error mainly depends on the grid size and the more grid points are used the smaller the error and the more accurate the metamodel will be.

If the metamodel has been constructed using the RS-HDMR approach in connection with analytical basis functions, one does not require an additional set of random samples in order to assess the accuracy. The same set can be used to show the accuracy of the metamodel, because the RS-HDMR approach uses a set of random (or quasi-random) samples for the construction of its component functions. One set of random samples is used for fitting the polynomials for all component functions (see section 3.2.1). However, if one wants to assure that the metamodel is tested against a set of samples which has not been used for the construction of the component functions, then an additional set of random samples is required (e.g. $N = 1000$). The RS-HDMR approach does not involve interpolation procedures like cut-HDMR, however the calculation of the expansion coefficients (see equation (3.44) and (3.46)) is based on Monte Carlo integration which controls the accuracy of the metamodel (see also section 3.2.2).

In the following section, five different methods are introduced to verify the accuracy of the constructed HDMR metamodel. All methods can be applied in connection with both approaches (cut-HDMR and RS-HDMR) and are based on one set of random (or quasi-random) samples.

3.3.1 Probability Density Function Estimate and Empirical Cumulative Distribution Function

A common way of presenting the accuracy of the constructed HDMR expansion is to compare the statistics of the response from the metamodel and the original model. This can be achieved by using the probability density function (pdf) estimate and/or the empirical cumulative distribution function (cdf) as applied for example in Balakrishnan et al. (2005). If the characteristics have been well captured, then the pdf/cdf of the metamodel and the original model should show good agreement. A large number of random samples (i.e. $N = 10\,000$) can be used in order to construct the pdf/cdf from the response of the metamodel because its computational running time is far less than the one of the original model.

An example for a pdf-plot is shown in figure 3.2. It is illustrated that in this case a first-order HDMR metamodel is not capable of capturing the model behaviour. The two pdfs show significant differences. The pdf produced on the basis of the model output has two modes and in contrast the pdf constructed on the basis of the response from the first-order metamodel has only one mode. However, when using a second-order HDMR metamodel (figure 3.2b) then both pdfs show a nearly perfect match. This affirms the high accuracy of the second-order HDMR metamodel.

The cdf-plot using the same data set is presented in figure 3.3 and confirms that a first-order HDMR metamodel is not sufficient to capture the behaviour of the original model. Again, when using the second-order metamodel both plots overlap (figure 3.3b) nearly perfectly.

Both, pdf-plot and cdf-plot, are only graphical methods to illustrate the accuracy of the HDMR metamodel. However, it is also possible to determine a quantitative measure based on the plots, for example by calculating the surrounded area of the

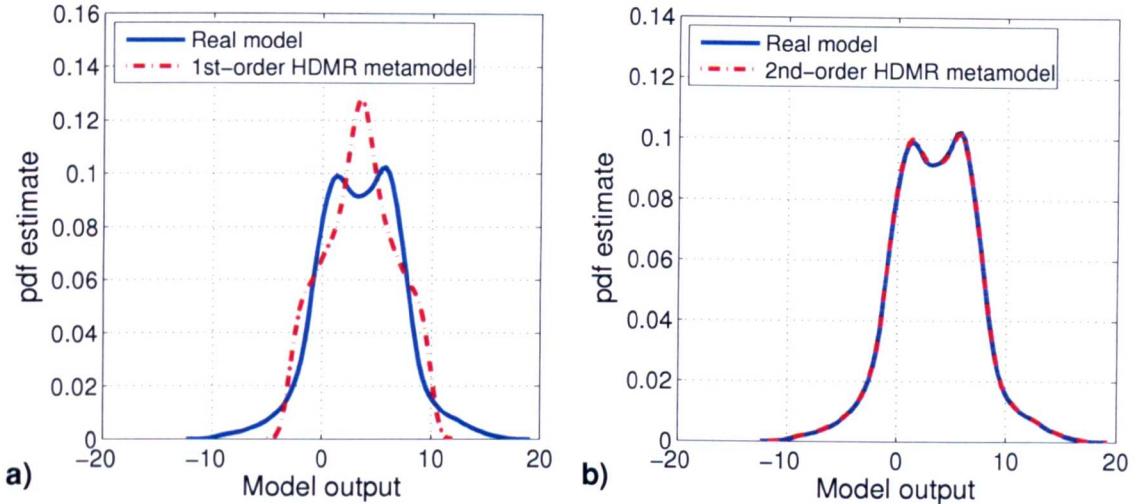


Figure 3.2: Example probability density function (pdf) estimate using a (a) first- and (b) second-order HDMR metamodel.

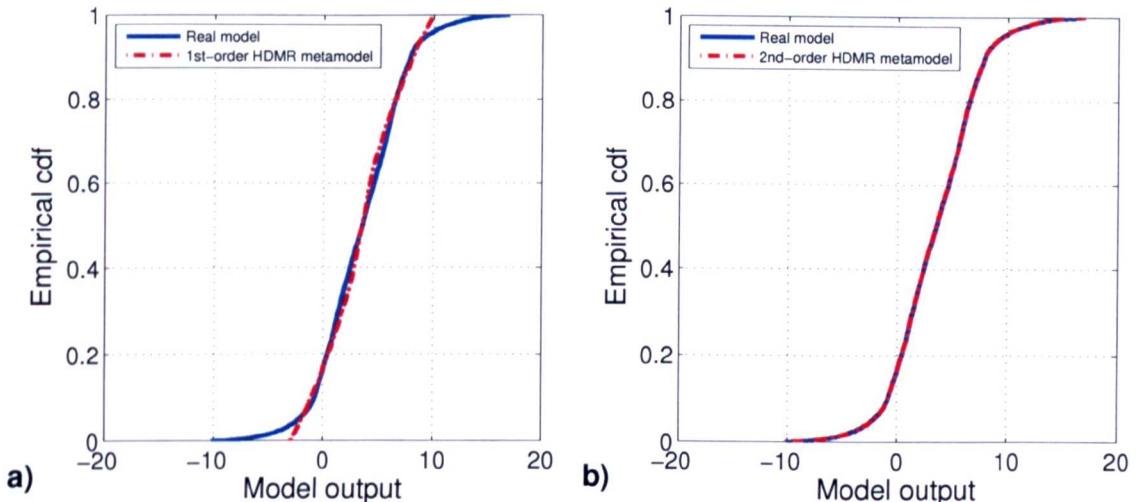


Figure 3.3: Example empirical cumulative distribution function (cdf) using a (a) first- and (b) second-order HDMR metamodel.

two curves.

The shift between two densities (surrounded area) has also been used in Borgonovo (2007) to introduce a new uncertainty importance measure. Here, the influence of the entire input distribution on the entire output distribution is evaluated by

measuring the difference between the unconditional and conditional pdfs.

3.3.2 Residual Variance

A very simple way of showing whether or not the model behaviour has been captured well by the HDMR metamodel is by calculating the residual variance. First, the variance of the model output is calculated and then compared with the variance of the response from the metamodel. If the variance produced by the metamodel is much smaller than the variance of the model output, than the accuracy of the metamodel is not satisfactory. Only if all the variance produced by the original model can be explained by the HDMR metamodel, can its accuracy be assumed to be sufficient. However, one has to ensure that a large enough sample size is used when calculating the variance so that convergence has been reached.

3.3.3 Relative Error

Another very common way of assessing the accuracy of the metamodel is by calculating the relative error (RE), which is defined as

$$RE = \left| \frac{f(\mathbf{x}^{(s)}) - \hat{f}(\mathbf{x}^{(s)})}{f(\mathbf{x}^{(s)})} \right| \quad (3.66)$$

with

$$\hat{f}(\mathbf{x}^{(s)}) = f_0 + \sum_{i=1}^n f_i(x_i^{(s)}) \quad (3.67)$$

for an up to first-order HDMR expansion, and

$$\hat{f}(\mathbf{x}^{(s)}) = f_0 + \sum_{i=1}^n f_i(x_i^{(s)}) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i^{(s)}, x_j^{(s)}) \quad (3.68)$$

for an up to second-order HDMR expansion.

The relative error is given in % and is usually calculated in the range of 1% RE up to 20% RE. Table 3.1 shows an example for the representation of the accuracy in terms of the relative error. The table illustrates that for example if using the first-order HDMR expansion, only 10.5% of the tested samples are within the 5% RE range. However, the accuracy significantly improves if using the second-order HDMR metamodel and now 92.8% of the tested samples are within the 5% RE range.

HDMR metamodel	1 % RE (%)	5 % RE (%)	10 % RE (%)
1st-order	2.3	10.5	19.4
2nd-order	71.3	92.8	95.8

Table 3.1: Relative error for the HDMR metamodel.

Despite its common usage, the relative error is not always a reliable measure. For example, if many of the output values $f(\mathbf{x})$ have very small absolute values, then the relative error may become very large even if the absolute error is quite small. Consequently, not many of the tested samples will be within the given relative error range, even if the metamodel provides a reasonably good approximation of the model output. Therefore, Li et al. (2006) suggests to define a threshold for the very small values. If the response of the metamodel is within the threshold, then this should be seen as a correct answer. Another possibility would be to use the absolute error instead of the relative error for small output values. Special care has also to be taken if the model output value is exactly zero. In this case one has to assign a value for the relative error or use the absolute error instead.

3.3.4 Scatter Plots and Coefficient of Determination

In order to prevent the drawbacks of using the relative error measure one can also use a scatter plot in order to illustrate the accuracy of the metamodel graphically.

Here, the model output is plotted against the response of the HDMR metamodel. If the plot produces a straight line ($f(\mathbf{x}^{(s)}) \approx \hat{f}(\mathbf{x}^{(s)})$ for $s = 1, \dots, N$), then the model behaviour is very well captured by the metamodel.

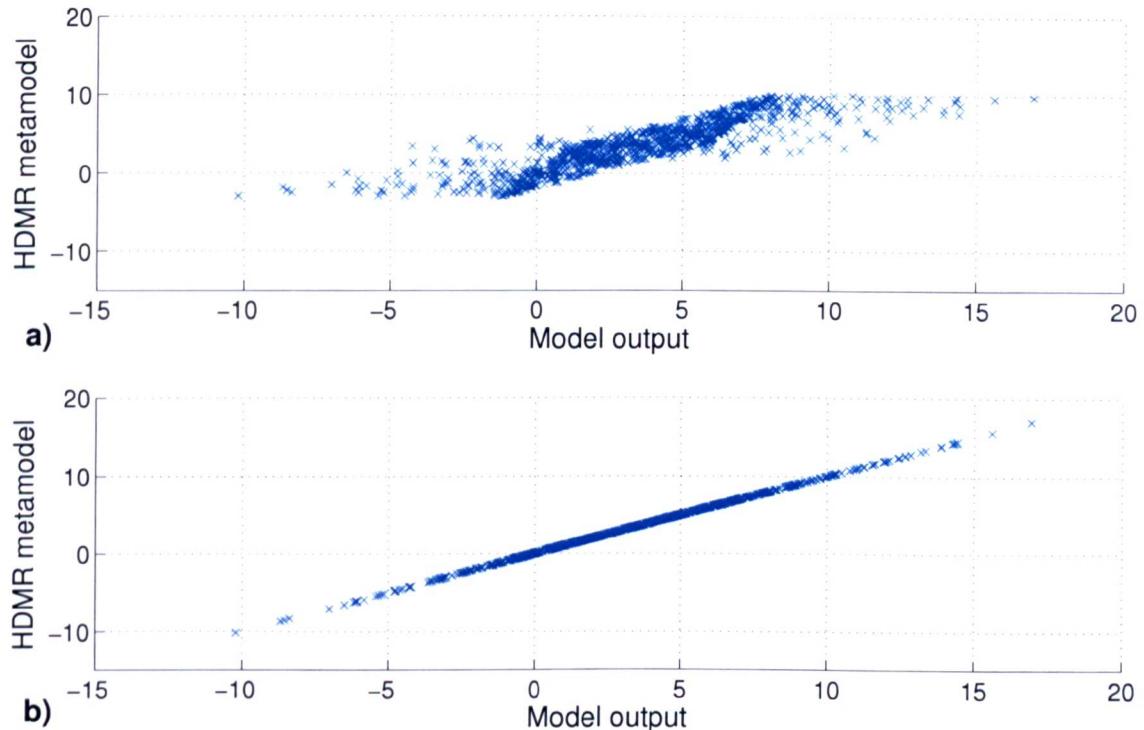


Figure 3.4: Example scatter plot for the model output and (a) the first-order HDMR metamodel and (b) the second-order HDMR metamodel.

Figure 3.4 shows an example of such a scatter plot and illustrates the relationship between the model output and a first- and second-order HDMR metamodel. The first-order metamodel cannot represent the model output sufficiently, shown by the fact that the sample points do not form a straight line. In contrast to this the second-order metamodel represents the model behaviour very well and all sample points are in line.

The quality of the metamodel can also be expressed by calculating the coefficient of determination r^2 (square of the Pearson correlation coefficient) between $f(\mathbf{x}^{(s)})$ and $\hat{f}(\mathbf{x}^{(s)})$ for $s = 1, \dots, N$. The r^2 value for the dataset shown in figure 3.4a is

0.7551 and for figure 3.4b it is 0.9999, indicating that a second-order HDMR model can be used as a surrogate. However, one of the advantages of using a scatter plot is, that it can illustrate in which output ranges the metamodel fails to capture the behaviour of the real model. In the example shown in figure 3.4 the first-order HMDR metamodel provides reasonably good results for output values between one and six. However, especially the extreme values can not be captured by the first-order HDMR expansion. This would lead to the assumption that the extreme values are mainly affected by parameter interactions, since the second-order metamodel is able to represent the model output over the entire range.

3.3.5 Sample Plots

Another simple way in presenting the accuracy of the HDMR metamodel for a small number of sample points is to plot the model output and the response of the metamodel for a few random input combinations in one figure. An example is given in figure 3.5 and it is shown that a first-order HDMR metamodel is not sufficient to represent the model behaviour. Especially for large absolute values is the difference between the model response and the metamodel quite big. The second-order metamodel produces a very good representation of the model output and the response of the model and the metamodel shows a good match for the 100 random input combinations.

The plot of individual sample points is only good for illustration purposes because of the small number of samples involved. In addition, other methods such as the relative error or a pdf plot should be used as well since here a large sample size can be considered.

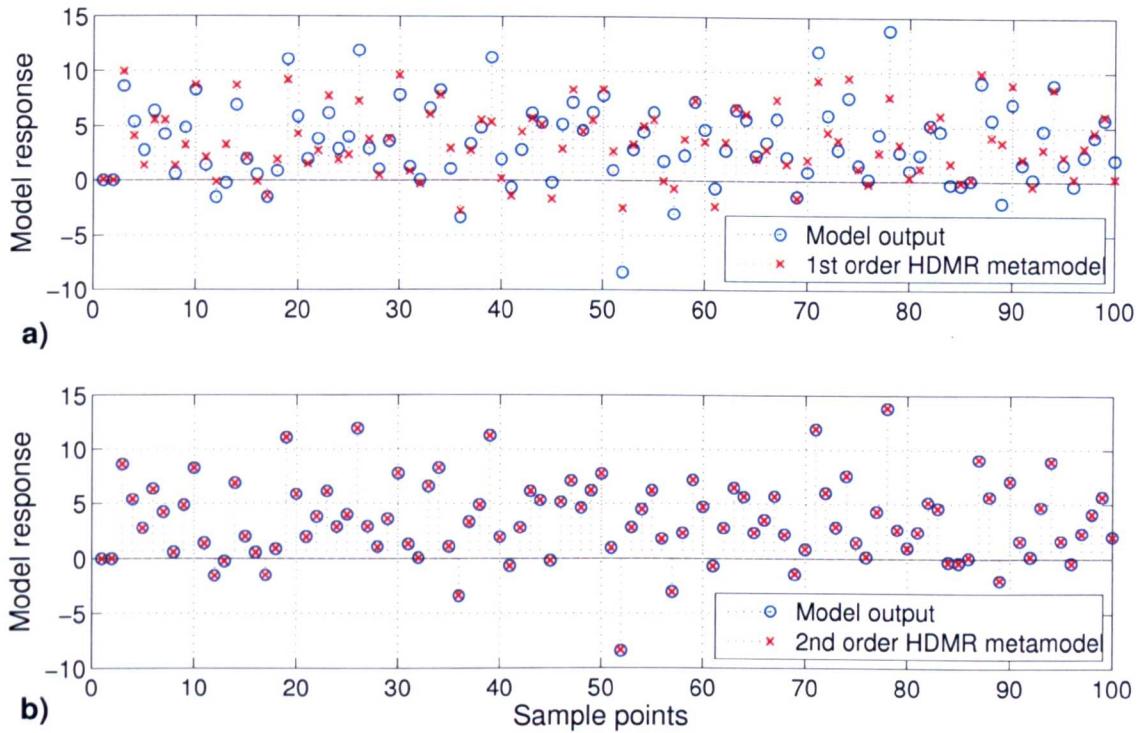


Figure 3.5: Comparison of the response from the original model and (a) the first-order HDMR metamodel and (b) the second-order HDMR metamodel.

3.4 Sensitivity Analysis Using HDMR

Once the HDMR metamodel has been constructed and its accuracy has been validated using one or more of the methods introduced in the latter section, global sensitivity analysis can be performed. This can be done in the traditional way as it is performed for example in connection with Monte Carlo analysis (see section 2.4.3) by using the cut-HDMR or RS-HDMR metamodel instead of the original model. Especially if the model is computationally expensive, the metamodel provides a fast fully functional equivalent and a large number of model runs can be performed in a small amount of time. Scatter plots, correlation coefficients etc. can then easily be applied involving only the HDMR metamodel.

Another possibility would be to use the constructed metamodel instead of the orig-

inal model within the method of Sobol' (see section 2.4.5) in order to calculate variance based sensitivity indices. Again, the computational effort is far less, because the metamodel is very quick to simulate. However, the HDMR expansion also allows to calculate variance based sensitivity indices, which are in principle the same as introduced by Sobol', easily from its structure. This can be achieved in a very efficient way without the need for further model runs (either original model or metamodel).

The next two sections describe how to determine sensitivity indices directly from the cut-HDMR and RS-HDMR expansion. The calculation is based on the ANOVA decomposition. Cut-HDMR can be converted into the ANOVA decomposition and RS-HDMR already gives the ANOVA decomposition.

3.4.1 Sensitivity Indices based on cut-HDMR

There are several relationships between cut-HDMR and ANOVA-HDMR¹. The relationship between the component functions of ANOVA-HDMR and cut-HDMR is given in Rabitz & Aliş (2000):

$$f_{i_1, \dots, i_l}^{ANOVA}(x_{i_1}, \dots, x_{i_l}) = \int_{K^n} f_{i_1, \dots, i_l}^{cut}(x_{i_1}, \dots, x_{i_l}) dx. \quad (3.69)$$

An efficient two step approach to calculate the ANOVA-HDMR component functions involving the cut-HDMR functions can be given as follows (Rabitz & Aliş 2000):

1. Evaluate the cut-HDMR component functions as stated in section 3.1 and

¹ANOVA-HDMR is a specific member of high dimensional model representation and is the same as the ANOVA decomposition used in statistics (Rabitz et al. 1999, Rabitz & Aliş 1999)

approximate the function $f(\mathbf{x})$ by

$$f(\mathbf{x}) \approx f^{cut}(\mathbf{x}) = f_0^{cut} + \sum_{i=1}^n f_i^{cut}(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}^{cut}(x_i, x_j) + \dots \quad (3.70)$$

2. Compute the ANOVA-HDMR expansion of $f^{cut}(\mathbf{x})$:

$$f_0^{ANOVA} = \int_{K^n} f^{cut}(\mathbf{x}) d\mathbf{x} \quad (3.71)$$

$$f_i^{ANOVA}(x_i) = \int_0^1 \dots \int_0^1 f^{cut}(\mathbf{x}) d\mathbf{x}^i - f_0^{ANOVA} \quad (3.72)$$

$$\begin{aligned} f_{ij}^{ANOVA}(x_i, x_j) &= \int_0^1 \dots \int_0^1 f^{cut}(\mathbf{x}) d\mathbf{x}^{ij} - f_i^{ANOVA}(x_i) \\ &\quad - f_j^{ANOVA}(x_j) - f_0^{ANOVA} \end{aligned} \quad (3.73)$$

...

where $d\mathbf{x}^i$ means integrating over all indices but x_i and so on.

The calculation of the cut-HDMR component functions does not require any integrations at all and the integration involved using the second step only calls for integrations of dimensions $1, \dots, L$ (where L is the order of the cut-HDMR expansion). These integrals can be calculated with only a few points very accurately using quadrature techniques (Rabitz et al. 1999, Rabitz & Aliş 2000). On the other hand the calculation of the ANOVA-HDMR as introduced in section 2.4.5 involves many Monte Carlo integrals using a large number of sample points, which is very time consuming.

The ratio of the computational costs of the cut-HDMR to that of the ANOVA-HDMR is given by (Rabitz & Aliş 2000):

$$\rho = \frac{\sum_{i=0}^L \frac{n!}{(n-i)!i!} (m-1)^i}{\sum_{i=0}^L \frac{n!}{(n-i)!i!} N} \quad (3.74)$$

Usually a cut-HDMR expansion up to second order ($L=2$) is sufficient. Assuming that $m \approx 10$ and $N \gg 10^3$ the ratio will satisfy $\rho \ll 1$. In this way, the introduced two step algorithm based on the cut-HDMR component functions provides an efficient alternative approach to the method of Sobol' in calculating the ANOVA decomposition.

Once the terms of the ANOVA decomposition are determined, one can continue to calculate the partial variances and sensitivity indices as described by equations (2.28)-(2.31) in section 2.4.5 for the method of Sobol'. An example for the conversion of the cut-HDMR component functions into ANOVA component functions and the calculation of the partial variances is given in the appendix in section A.1 for a model with four input parameters.

In contrast to the method of Sobol' and equation (2.27), the total variance D is calculated as the sum of the partial variances. If a second-order HDMR expansion is used, then the total variance is simply the sum of all first- and second-order partial variances:

$$D = \sum_{i=1}^n D_i + \sum_{1 \leq i < j \leq n} D_{ij}. \quad (3.75)$$

This also implies, that all first- and second-order sensitivity indices add up to exactly one (see equation (2.31)-(2.32)). However, care must be taken because this can lead to a wrong assumption, for example, in cases where higher-order effects (≥ 3) are present but not considered if using only a second-order HDMR expansion. Therefore, it is crucial to ensure that the accuracy of the cut-HDMR metamodel is sufficient, so that the behaviour of the real model can actually be represented by a HDMR expansion which neglects such higher order effects.

It is also possible to perform additional model runs using one set of random samples $\mathbf{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \dots, x_n^{(s)})$ with $s = 1, 2, \dots, N$ in order to estimate the total variance

as

$$D \approx \frac{1}{N} \sum_{s=1}^N f^2(\mathbf{x}^{(s)}) - f_0^2. \quad (3.76)$$

In this case potential higher order effects are not neglected. However, one has to consider the computational effort in performing those additional model runs.

3.4.2 Sensitivity Indices based on RS-HDMR

The RS-HDMR component functions are independent and give unique physical information on the relationships amongst the model input parameters for their actions on the output properties. This means, that the individual component functions of the RS-HDMR expansion have a direct statistical correlation interpretation, that permits the model output variance D to be decomposed into its input parameter contributions (Li et al. 2002*b*):

$$D = \int_{K_n} [f(\mathbf{x}) - \bar{f}]^2 d\mathbf{x} \quad (3.77)$$

$$= \int_{K_n} [f(\mathbf{x}) - f_0]^2 d\mathbf{x} \quad (3.78)$$

$$= \int_{K_n} \left[\sum_{i=1}^n f_i(x_i) + \sum_{1 \leq i < j \leq n} f_{ij}(x_i, x_j) + \dots \right]^2 d\mathbf{x} \quad (3.79)$$

$$= \sum_{i=1}^n \int_{K_n} f_i^2(x_i) dx_i + \sum_{1 \leq i < j \leq n} \int_{K_n} f_{ij}^2(x_i, x_j) dx_i dx_j + \dots \quad (3.80)$$

$$= \sum_{i=1}^n \int_0^1 f_i^2(x_i) dx_i + \sum_{1 \leq i < j \leq n} \int_0^1 \int_0^1 f_{ij}^2(x_i, x_j) dx_i dx_j + \dots \quad (3.81)$$

$$= \sum_{i=1}^n D_i + \sum_{1 \leq i < j \leq n} D_{ij} + \dots \quad (3.82)$$

The overall variance of the model output D can be determined using Monte Carlo integration and one set N of random samples $\mathbf{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \dots, x_n^{(s)})$ with $s =$

$1, 2, \dots, N$:

$$D = \int_{K^n} [f(\mathbf{x}) - f_0]^2 d\mathbf{x} \quad (3.83)$$

$$\approx \frac{1}{N} \sum_{s=1}^N f^2(\mathbf{x}^{(s)}) - f_0^2 \quad (3.84)$$

where f_0 is the zeroth-order term of the RS-HDMR expansion:

$$f_0 = \int_{K^n} f(\mathbf{x}) d\mathbf{x} \quad (3.85)$$

$$\approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}). \quad (3.86)$$

If the RS-HDMR component functions are approximated by orthonormal polynomials as described in subsection 3.2.1, then the partial variances D_i, D_{ij}, \dots can be calculated as follows:

$$\begin{aligned} D_i &= \int_0^1 f_i^2(x_i) dx_i \approx \int_0^1 \left[\sum_{r=1}^{k_i} \alpha_r^i \varphi_r(x_i) \right]^2 dx_i \\ &= \sum_{r=1}^{k_i} (\alpha_r^i)^2 \end{aligned} \quad (3.87)$$

$$\begin{aligned} D_{ij} &= \int_0^1 \int_0^1 f_{ij}^2(x_i, x_j) dx_i dx_j \approx \int_0^1 \int_0^1 \left[\sum_{p=1}^{l_i} \sum_{q=1}^{l'_j} \beta_{pq}^{ij} \varphi_p(x_i) \varphi_q(x_j) \right]^2 dx_i dx_j \\ &= \sum_{p=1}^{l_i} \sum_{q=1}^{l'_j} (\beta_{pq}^{ij})^2 \\ &\dots \end{aligned} \quad (3.88)$$

Note that only one set of random samples N is necessary to obtain the overall variance and all partial variances.

The RS-HDMR expansion already provides the ANOVA decomposition and addi-

tionally a very detailed input-output mapping which can be used for other purposes as well (e.g. black-box metamodel). The ANOVA decomposition as described in subsection 2.4.5 only provides the possibility to determine the partial variances and sensitivity indices, but not the component functions itself.

3.4.3 Plot of the HDMR Component Functions

Global sensitivity analysis on the basis of the HDMR component functions is not only computationally more efficient than the method of Sobol' (or other traditional methods), another advantage is in the possibility of exploring the component functions of the HDMR expansion. The component functions have a physical meaning and can be plotted to show the influence of one parameter (first-order component function) or the interactive effect of a pair of parameters (second-order component function) upon the model output over the whole parameter range.

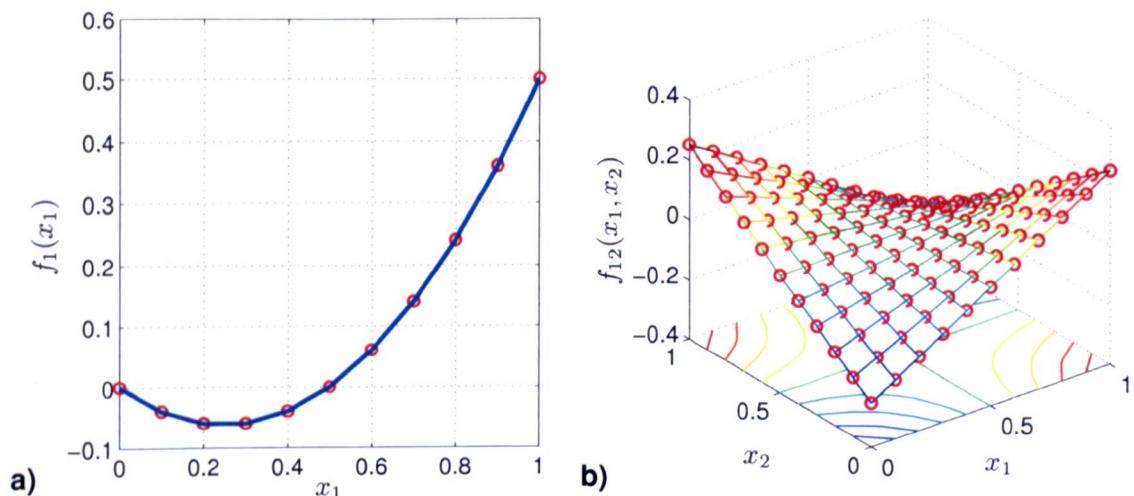


Figure 3.6: (a) Example of a first-order cut-HDMR component function and (b) a second-order cut-HDMR component function with linear interpolation between the exact solutions (indicated by the red circles) using $m = 11$ grid points.

An example plot of a first-order and second-order cut-HDMR component function is

presented in figure 3.6. The solution is exact at the grid points and a linear interpolation has been performed in between. The accuracy of the component functions can be improved by increasing the number of grid points m for each parameter range.

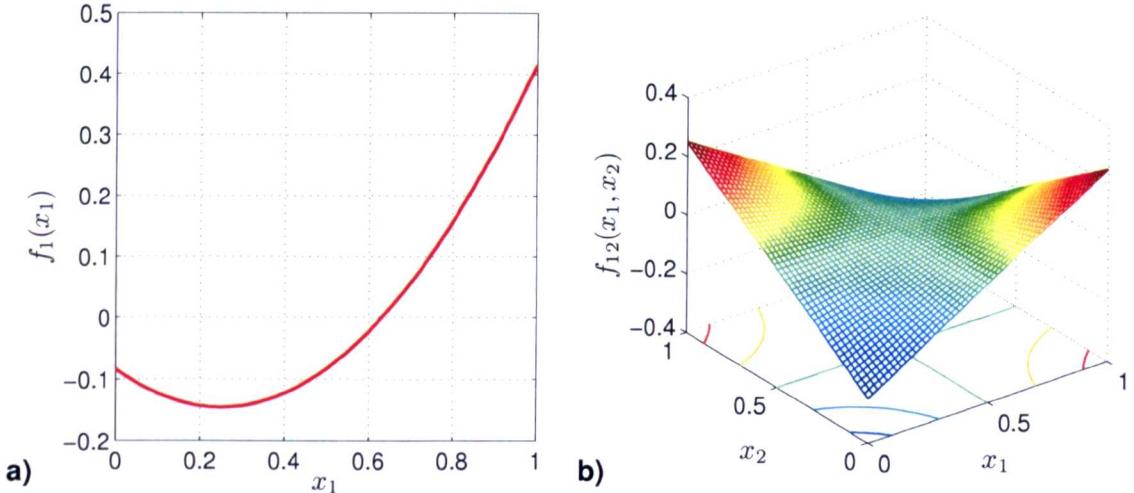


Figure 3.7: (a) Example of a first-order RS-HDMR component function and (b) a second-order RS-HDMR component function using second-order orthonormal polynomials.

Figure 3.7 shows an example of a first- and second-order RS-HDMR component function. Here the graph may not contain any exact points, because orthonormal polynomials are used to approximate the component functions via a fitting procedure. In this chapter it has been shown that the fitting error (Monte Carlo integration error) can be reduced by either increasing the sample size N and/or by applying a variance reduction method.

The sensitivity indices can provide a ranking of the most important parameters and they can show how much one parameter or parameter interaction is more important than another one. However, they cannot provide the amount of information that is available by investigating the plot of the first- and second-order component functions. The plot reveals the kind of relationship (linear, non-linear, monotonic, non-monotonic, etc.) which exists between the input (or combination of inputs) of

interest and the output. Unlike scatter plots, the first-order component functions are not obscured by the influence of the other parameters, because the HDMR method is able to isolate the effect of one parameter or parameter interactions.

The isolated effect of one parameter (plot of the first-order component function) is shown over the whole uncertainty range. This can be used to demonstrate the sensitivity of this parameter at certain smaller intervals within the whole range. In this way the plot of the HDMR first-order component functions reveals also information which are comparable with the ones obtained from local sensitivity analysis (e.g. gradient at a certain point).

The advantages of using the plot of the component functions instead of scatter plots is further discussed when applying the HDMR method to the case studies in the chapters 4-6.

3.5 Summary

The HDMR method is a novel and promising approach in order to perform global sensitivity analysis very efficiently. It was mainly developed to express the input-output relationship of a complex model with a large number of input parameters. Two different HDMR approaches have been introduced: cut-HDMR, which depends on the value of $f(\mathbf{x})$ at a specific reference point $\bar{\mathbf{x}}$ and RS-HDMR, which depends on the average value of $f(\mathbf{x})$ over the whole domain.

3.5.1 HDMR Approaches

Cut-HDMR can be applied when ordered sampling of the output $f(\mathbf{x})$ is possible. A reference point $\bar{\mathbf{x}}$ has to be selected first. Some guidelines on how to determine

the reference point were given. The component functions are calculated along lines, surfaces, subvolumes and so on using a defined grid for each of the parameters. The values are stored numerically in low dimensional look-up tables. A low dimensional interpolation has to be performed within the tables in order to calculate the metamodel response for an arbitrary point x . The number of required model runs strongly depends on the number of input parameters. This is a significant drawback, particularly, if one is interested in investigating higher order effects (≥ 2) which can require a large number of model runs.

RS-HDMR requires only one set of random (or quasi-random) samples. The component functions can be approximated by analytical basis functions such as orthonormal polynomials. The optimal coefficients (expansion coefficients) for these polynomials are calculated via Monte Carlo integration. Two variance reduction methods, correlation and ratio control variate, were introduced in order to reduce the Monte Carlo integration error. It is only required to store the expansion coefficients and analytical functions in order to calculate the metamodel response for an arbitrary point x .

Various approaches were described to assess the accuracy of the constructed cut-HDMR or RS-HDMR metamodel. All methods are based on a set of random samples. For cut-HDMR, additional model runs have to be performed in order to assess its accuracy. For RS-HDMR the same set of random samples can be applied, which has been used to construct the metamodel. Sometimes it is advisable to use more than one of the methods introduced in order to validate the constructed metamodel (e.g. one graphical method such as the probability function estimate + coefficient of determination).

Variance based sensitivity indices can be calculated using either of the two HDMR approaches without requiring any further model runs. The cut-HDMR component

functions can be converted into the ANOVA decomposition. The partial variances are calculated by squaring and integrating over the ANOVA component functions and the total variance is simply assumed to be the sum of all partial variances. This assumption is only reasonable if one can ensure that the model behaviour is fully captured by the cut-HDMR metamodel. RS-HDMR already gives the ANOVA decomposition and if orthonormal polynomials are used for the approximation of the component functions, then the partial variances can easily be obtained from the expansion coefficients. The total variance is calculated by Monte Carlo integration over the whole domain.

3.5.2 HDMR in Comparison to Other Global SA Methods

The HDMR method allows the calculation of variance based sensitivity indices, which are in principle the same as the Sobol' sensitivity indices (section 2.4.5), in a more efficient way. For example, if using a second-order RS-HDMR expansion it is possible to estimate all first- and second-order sensitivity indices with only one set of random samples. The method of Sobol requires one set of samples for the estimation of each of the first- and second-order sensitivity indices. For instance, if investigating a model with only four input parameters and assuming a sample size of $N = 1\,000$, the method of Sobol' would require a sampling effort of 11 times (1 zeroth-order + 4 first-order + 6 second-order) the sample size ($11N = 11\,000$ required model runs), whereas the RS-HDMR method could estimate all sensitivity indices up to second-order from the one set of samples ($1\,000$ required model runs). Unlike the method of Sobol' the HDMR approach allows the construction of a metamodel, which can be used as a surrogate. In this way the HDMR method has something in common with the response surface method (section 2.4.4). Both methods develop an approximation to the full model. However, unlike the RSM

the HDMR method does not require assumptions about the model behaviour and it also works with a large number of input parameters. Compared with RSM, which commonly uses only linear and quadratic terms, HMDR is able to provide a more accurate model approximation. The HDMR method is based on the variance decomposition approach and the partial variances and sensitivity indices can easily be calculated from the HDMR expansion without the need to perform further model runs. The RSM method does not provide this possibility. Further, the plot of the HDMR component functions up to second-order can reveal important information about the input-output relationship. The influence of one parameter (first-order) or pair of parameters (second-order) upon the output is shown over the whole parameter range. Unlike scatter plots, the plot of the first-order component functions is not obscured by the influence of other parameters, because the HDMR method is able to isolate the independent effect of a single parameter. None of the introduced global SA methods in section 2.4 provide this feature.

Monte Carlo analysis (section 2.4.3) is conceptually a very simple approach and therefore widely used. However, the HDMR method is also conceptually simple and straight forward. A large variety of uncertainty and sensitivity analysis methods are available for Monte Carlo analysis. All these tools can also be applied in connection with the HDMR method if using the HDMR metamodel as a surrogate. The computational expense to simulate the metamodel is very small and large numbers of model runs can be performed. Monte Carlo analysis does not directly allow the calculation of variance based sensitivity indices or the investigation of parameter interactions. The use of the Pearson correlation coefficient and the Spearman rank correlation coefficient which are often applied in Monte Carlo analysis require assumptions about the input-output relationship (linear or monotonic non-linear respectively). Sensitivity indices do not require these assumptions and their application is therefore universal. The Fourier amplitude sensitivity test (section 2.4.6) also provides the

possibility to calculate sensitivity indices. However, the concept is relatively complicated and classic FAST allows only the determination of the first-order indices. Extended FAST provides the possibility to calculate total effect indices, but they do not provide a complete characterisation of the system.

The HDMR method provides more information in terms of the model sensitivity than any other global method discussed in this work (section 2.4) and it usually requires only moderate computational costs. In the following chapters the set of HDMR tools will be extended and tested using analytical functions and case studies.

4 Comparison of cut-HDMR and RS-HDMR in Practice

Two HDMR approaches, cut-HDMR and RS-HDMR, have been introduced in the last chapter. This chapter aims to assess and compare the performance of the two HDMR approaches using an analytical test function and a case study. The Ishigami function has been chosen as the analytical test model because of its highly non-linear and non-monotonic behaviour. The case study focuses on the simulation of the turbulent flow field in a 3D street canyon using a micro-scale computational fluid dynamics (CFD) model. Before the HDMR methods are applied, the RS-HDMR approach is extended by an optimisation algorithm, which automatically chooses the best polynomial order for the approximation for each of the component functions.

4.1 Optimisation Method for RS-HDMR

The RS-HDMR approach as described in section 3.2 uses the same polynomial order for all first- 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. If the input-output relationship, represented by a component function, is highly non-linear, then

higher-order polynomials are theoretically the best choice for the approximation. On the other hand it has to be taken into account, that higher-order polynomials have more terms than lower-order polynomials, and each term has its own Monte Carlo integration error. This can lead to a poor approximation, especially if using only a small sample size N . Therefore, it might be better to use lower-order polynomials in connection with a small sample size N , even if the relationship is highly non-linear. If increasing the sample size N , then a higher-order polynomial might be more suitable for the same component function. However, if the input-output relationship for a parameter or parameter interaction is only linear, then it is sufficient to approximate the corresponding component function by a first-order polynomial independently of the sample size N . Sometimes the contribution of a certain RS-HDMR component function to the overall value of $f(\mathbf{x})$ is zero and could therefore even be excluded from the HDMR expansion function completely. Otherwise one would only add the Monte Carlo integration error to the HDMR expansion, which could decrease the overall accuracy of the metamodel especially if the sample size N is small.

Polynomial order	Number of terms in:				
	f_i	$\sum f_i$	f_{ij}	$\sum f_{ij}$	$\sum f_i + \sum f_{ij}$
First (linear)	1	4	1	6	10
Second	2	8	4	24	32
Third	3	12	9	54	66
:			:		
Tenth	10	40	100	600	640

Table 4.1: Number of terms required for the approximation of the RS-HDMR component functions considering a model with four input parameters and a second-order HDMR expansion.

If the component functions are approximated by higher-order polynomials, then the number of terms in the HDMR expansion grows significantly, especially for the second-order component functions. Table 4.1 presents the number of terms in

the HDMR expansion up to second-order assuming a model has only four input parameters and if using the same polynomial order for all component functions. For example, if using third-order polynomials exclusively for all component functions, then 66 expansion coefficients (see equations (3.44)-(3.46)) have to be estimated by Monte Carlo integration. This means there are 66 possible sources for a Monte Carlo integration error. This could be reduced if the “right” polynomial order is chosen for each of the component functions, and if component functions are excluded if they do not contribute to the overall output value.

In order to decide which component function is approximated best by which polynomial order, an optimisation algorithm based on the least square method was developed as an extension to the already existing set of RS-HDMR tools. The idea is to calculate the sum of the square errors using the results of the original model runs (set of samples of the size N) and the approximation of the component functions by either polynomials of up to a certain maximum order (here we assume a maximum order of ten) or by excluding the component function. The smallest sum of square error indicates the best approximation order for the corresponding component function. For a second-order RS-HDMR expansion the optimisation process can be carried out in two steps:

1. Compute the optimal polynomial expansion order for all first-order component functions:

$$\min_{k_i \in \{0, 1, \dots, 10\}} \sum_{s=1}^N \left[f(\mathbf{x}^{(s)}) - \left(f_0 + f_i(x_i^{(s)}) \right) \right]^2 \quad (4.1)$$

with

$$f_i(x_i) = \begin{cases} \sum_{r=1}^{k_i} \alpha_r^i \varphi_r(x_i) & : k_i \in \{1, \dots, 10\} \\ 0 & : k_i = 0 \end{cases} \quad (4.2)$$

2. Compute the optimal polynomial expansion order for all second-order compo-

nent functions using the results of the first step:

$$\min_{l_i = l'_j \in \{0, 1, \dots, 10\}} \sum_{s=1}^N \left[f(\mathbf{x}^{(s)}) - \left(\hat{f}(\mathbf{x}^{(s)}) + f_{ij}(x_i^{(s)}, x_j^{(s)}) \right) \right]^2 \quad (4.3)$$

with

$$\hat{f}(\mathbf{x}^{(s)}) = f_0 + \sum_{i=1}^n \sum_{r=1}^{k_i} \alpha_r^i \varphi_r(x_i^{(s)}) \quad (4.4)$$

and

$$f_{ij}(x_i, x_j) = \begin{cases} \sum_{p=1}^{l_i} \sum_{q=1}^{l'_j} \beta_{pq}^{ij} \varphi_p(x_i) \varphi_q(x_j) & : l_i, l'_j \in \{1, \dots, 10\} \\ 0 & : l_i, l'_j = 0 \end{cases} \quad (4.5)$$

The computational effort to calculate the optimal order for all polynomials is low and therefore its calculation is worthwhile in order to improve the accuracy of the final metamodel. The optimisation method and its application have also been published in Ziehn & Tomlin (2008*b*).

4.2 Test Model: Ishigami Function

Before applying the cut-HDMR method as described in section 3.1 and the RS-HDMR method as described in section 3.2 with the optimisation approach developed in the latter section to a case study, their performance is first tested using a widely used analytical model - the Ishigami function (Ishigami & Homma 1990). This has the advantage of providing the opportunity to calculate the variance based sensitivity indices analytically and compare them against the results estimated by the HDMR method.

4.2.1 Properties of the Ishigami Function

The Ishigami function (Ishigami & Homma 1990) is an artificial analytical model with three input parameters:

$$f(\mathbf{x}) = \sin(x_1) + a \sin^2(x_2) + bx_3^4 \sin(x_1) \quad (4.6)$$

where x_i is uniformly distributed within $(-\pi, \pi)$ and the constants $a = 7$, $b = 0.1$. The model is highly non-linear, non-monotonic and possesses significant parameter interactions. The sensitivity indices can be calculated analytically on the basis of the ANOVA decomposition (see method of Sobol' in section 2.4.5) and are given in table 4.2. One interesting fact 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 .

S_1	S_2	S_3	$\sum S_i$	S_{12}	S_{13}	S_{23}	$\sum S_i + S_{ij}$
0.4424	0.3139	0	0.7563	0	0.2437	0	1.0000

Table 4.2: First- and second-order sensitivity indices for the Ishigami function calculated analytically.

As shown in table 4.2 the total variance of the model output can be explained by first- and second-order effects and therefore the sum of all first- and second-order sensitivity indices equals exactly one. This is an ideal test case for the cut-HDMR and RS-HDMR methods, because a second-order HDMR expansion should be able to capture all partial variances and both methods should be able to estimate the sensitivity indices shown in table 4.2 very accurately.

4.2.2 Application of cut-HDMR

The cut-HDMR approach was applied as described in section 3.1. The midpoint of each of the parameter ranges has been chosen to be the reference point $\bar{\mathbf{x}}$. To show the performance and accuracy of the cut-HDMR method, different grid sizes m have been tested. The necessary sampling effort depends directly on the grid size and can be calculated using equation (3.9).

N	301	1 261	4 921	19 441	analytical
m	11	21	41	81	
S_1	0.3115	0.3130	0.3137	0.3138	0.3139
S_2	0.4028	0.4316	0.4396	0.4417	0.4424
S_3	0	0	0	0	0
$\sum S_i$	0.7143	0.7446	0.7533	0.7555	0.7563
S_{12}	0	0	0	0	0
S_{13}	0.2857	0.2554	0.2467	0.2445	0.2437
S_{23}	0	0	0	0	0
$\sum S_i + S_{13}$	1.0000	1.0000	1.0000	1.0000	1.0000
$\sum AE$	0.0840	0.0234	0.0060	0.0016	0.0000

Table 4.3: First- and second-order sensitivity indices and the sum of their absolute errors (AE) for the Ishigami function using the cut-HDMR approach.

In order to compare the accuracy of the sensitivity indices estimated by the HDMR approach and sensitivity indices calculated analytically, the sum of the absolute error (AE) has been calculated for first-order

$$AE = |S_i - \hat{S}_i| \quad (4.7)$$

and for second-order

$$AE = |S_{ij} - \hat{S}_{ij}| \quad (4.8)$$

where S_i and S_{ij} are the analytical values and \hat{S}_i and \hat{S}_{ij} are the estimated values.

As illustrated by table 4.3 a grid size of $m = 11$ ($N = 301$) is already enough

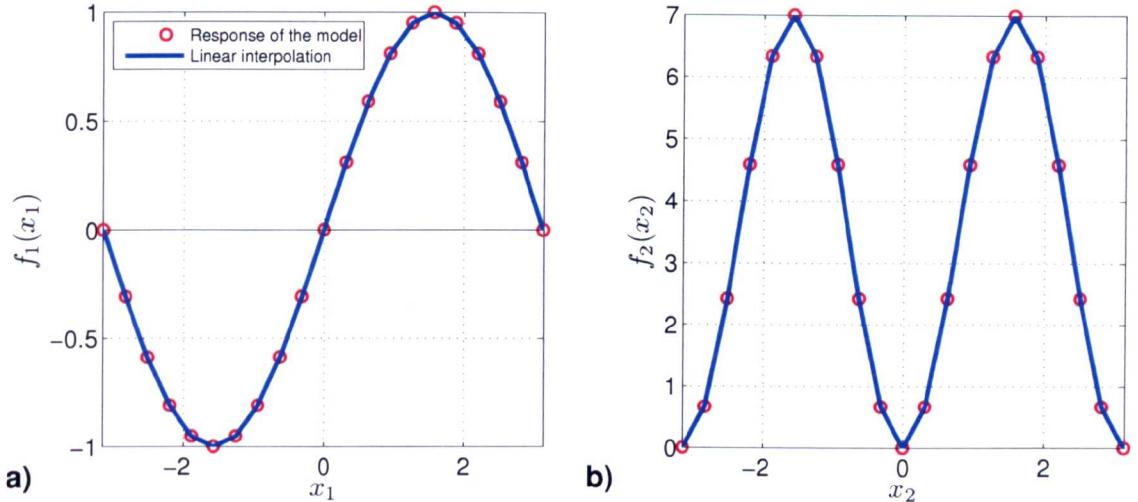


Figure 4.1: First-order component function (a) for input parameter x_1 and (b) for input parameter x_2 using a grid size of $m = 21$.

to produce the correct ranking for the importance of the parameters. However, the individual sensitivity indices estimated, especially for parameter x_2 and for the parameter interactions between x_1 and x_3 , are still different by a considerably large amount to the analytical ones ($\sum \text{AE} = 0.0840$). A larger grid size is necessary in order to get more accurate estimates for the sensitivity indices S_2 and S_{13} which results in a smaller absolute error ($\sum \text{AE} = 0.0234$). However, this leads to much larger sample sizes N and for a grid size of $m = 81$ already $N = 19\,441$ samples are required. This makes the cut-HDMR method computationally quite expensive.

The plot of the first-order cut-HDMR component functions for parameters x_1 and x_2 is shown in figure 4.1. Linear interpolation has been used between the cut-points, where the solution is exact. The plot of the second-order component function showing the interactive effect of parameter x_1 and x_3 is presented in figure 4.2. It can be seen that parameter x_3 has to be at the lower or upper end of its range in order to interact with parameter x_1 . Again, a linear interpolation has been performed between the cut-points.

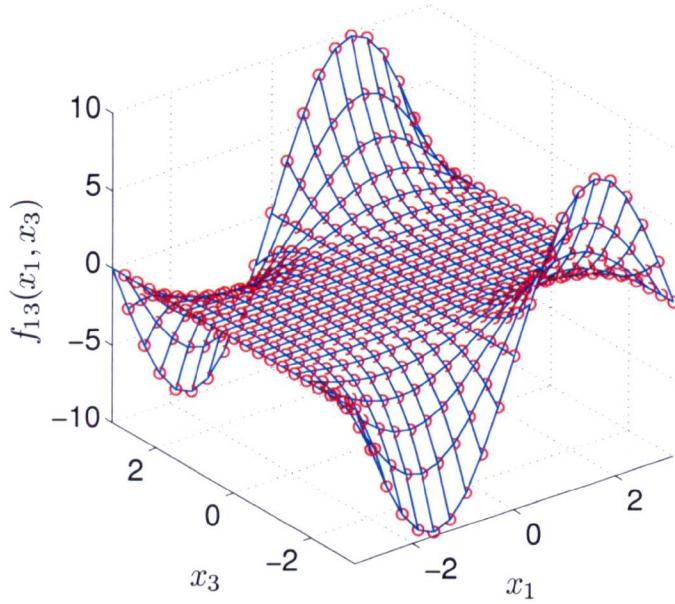


Figure 4.2: Second-order component function for input parameters x_1 and x_3 using a grid size of $m = 21$.

Another drawback of the cut-HDMR method is that the sum of the sensitivity indices is always one (see table 4.3). This is due to the fact, that the overall variance is simply calculated as the sum of all partial variances. This makes it difficult for the user to decide whether or not the "real" overall variance has completely been captured by the metamodel. To prevent this, it is necessary to use an additional set of samples (random or quasi-random) in order to calculate the overall variance independently (see equation 3.76). If the estimated partial variances are then normalised by the overall variance (calculated using the additional set of samples), then the user gets a more realistic estimate of the sensitivity indices.

4.2.3 Application of RS-HDMR

The RS-HDMR method has been applied including the optimisation approach which has been introduced in section 4.1. The maximum polynomial order for the approximation of the first-order component functions was set to be eight, and the maximum

order for the approximation of the second-order component functions was set to be five. Different sample sizes N have been investigated using the quasi-random Sobol' sequence.

N	256	512	1024	2048	analytical
S_1	0.3182	0.3171	0.3115	0.3142	0.3139
S_2	0.4401	0.4391	0.4426	0.4421	0.4424
S_3	0.0011	0	0	0	0
$\sum S_i$	0.7594	0.7562	0.7541	0.7563	0.7563
S_{12}	0	0	0	0	0
S_{13}	0.2180	0.2418	0.2431	0.2439	0.2437
S_{23}	0	0	0	0	0
$\sum S_i + S_{13}$	0.9774	0.9980	0.9973	1.0002	1.0000
$\sum \text{AE}$	0.0334	0.0094	0.0032	0.0008	0.0000

Table 4.4: Sensitivity indices first- and second-order and the sum of their absolute errors (AE) for the Ishigami function using the RS-HDMR approach (including optimisation of the polynomial order).

Table 4.4 presents the first- and second-order sensitivity indices estimated using various sample sizes. A sample size of $N = 256$ is already sufficient in order to identify the correct ranking of the parameters and parameter interactions. The first-order sensitivity indices for x_1 and x_2 are very accurate already, if compared with the analytical values and the sum of all absolute errors is only $\sum \text{AE} = 0.0334$. The sensitivity index for x_3 is supposed to be zero, since x_3 has no additive effects. However, for a small sample size the HDMR method overestimates the contribution for input parameter x_3 by a very small amount (0.11 % instead of 0 %). The contribution of the interaction of x_1 and x_3 is slightly underestimated. The component functions for $f_{12}(x_1, x_2)$ and $f_{23}(x_2, x_3)$ have been excluded from the HDMR expansion by the optimisation algorithm. In total 99.74 % of the overall variance can be explained by the second-order HDMR expansion. On increasing the sample size, the individual values for the sensitivity indices become even more accurate and the component function for x_3 is also correctly excluded.

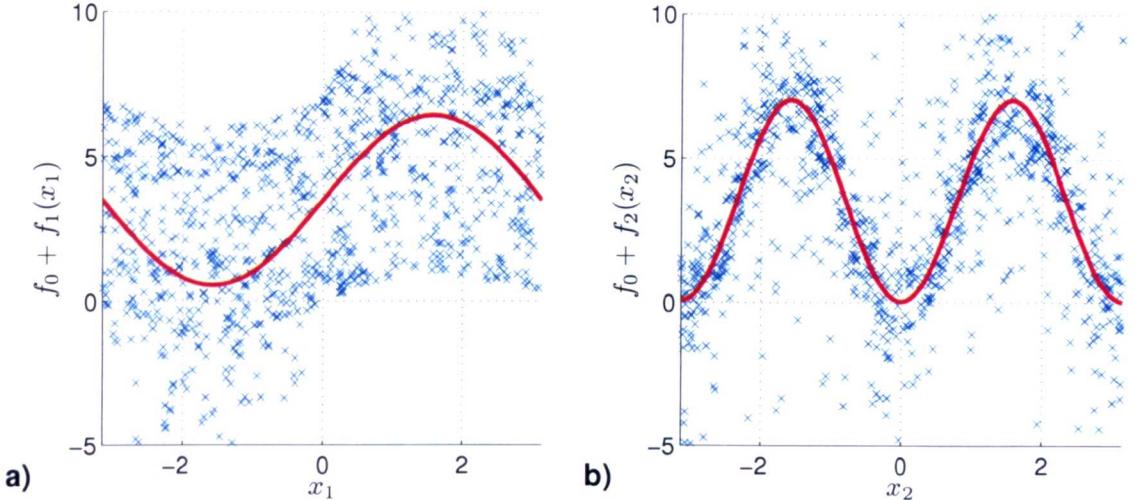


Figure 4.3: First-order component function (a) for input parameter x_1 (approximated by a 7th order polynomial) and (b) for input parameter x_2 (approximated by a 8th order polynomial) using $N = 512$.

The plot of the first-order component functions is shown in figure 4.3. It is demonstrated that in particular for $f_2(x_2)$ the plot is much smoother than that obtained using the cut-HDMR approach (see figure 4.1 for comparison). This is because RS-HDMR uses orthonormal polynomials to represent the component functions. The scatter plots are also shown in figure 4.3 for comparison and it can be seen that particularly for figure 4.3a the relationship between the input x_1 and the model output would not be as clearly visible if only investigating the scatter plot. The HDMR method is able to isolate the independent effect of one parameter (or parameter interactions) and therefore the plot of the component functions gives much more detailed information about the input-output relationship. The plot of the second-order component function $f_{13}(x_1, x_3)$ is illustrated in figure 4.4.

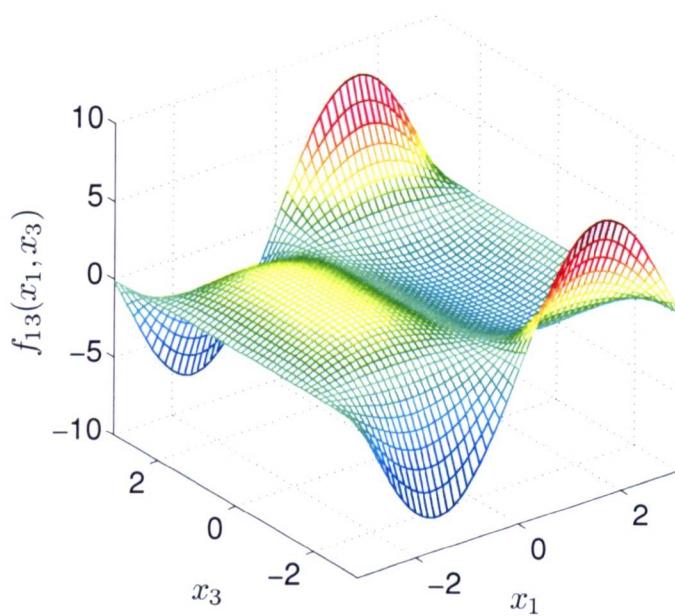


Figure 4.4: Second-order component function for input parameters x_1 and x_3 approximated by 4th order polynomials using $N = 512$.

4.2.4 Discussion and Conclusions

Both HDMR approaches, cut-HDMR and the extended RS-HDMR, are able to cope with the non-linear and non-monotonic behaviour of the Ishigami function. The first- and second-order sensitivity indices could be estimated with a very large accuracy if compared with the analytical values. The RS-HDMR method required only $N = 256$ model runs in order to provide a very good estimate for the first- and second-order sensitivity indices ($\sum \text{AE} = 0.0334$). The cut-HDMR method on the other hand could not provide the same accuracy at this small sample size ($m = 11, N = 301$) and the sum of the absolute error for the estimation of the first- and second-order sensitivity indices is much higher ($\sum \text{AE} = 0.0840$). The number of grid points m had to be increased in order to obtain a better estimate for the sensitivity indices. Using $m = 21$ and $N = 1\,261$ the absolute error reduces to $\sum \text{AE} = 0.0234$. However, the sampling effort of cut-HDMR is already five times higher than the one using RS-HDMR in order to achieve a comparable accuracy for the estimated

sensitivity indices.

Both HDMR approaches provide the possibility to plot the component functions up to second-order. In this way, the relationship between a selected parameter or parameter interaction and the model output becomes visible over the whole parameter range.

4.3 Case Study: Simulation of the Turbulent Flow Field in a 3D Street Canyon

Global sensitivity analysis using cut-HDMR and RS-HDMR is carried out in this section on the $k-\epsilon$ closure computational fluid dynamics model (CFD) MISKAM (Eichhorn 1996, Ries & Eichhorn 2001). The investigations allow a detailed representation of the effects of changing input parameters on the defined model outputs. The scenario studied is that of a complex street canyon in the city of York, UK. The results of this work have also been published in Ziehn & Tomlin (2008b) and Benson et al. (2008) and were presented at the 11th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO) (Ziehn & Tomlin 2007b, Benson et al. 2007).

4.3.1 Introduction

The improvement of air quality in urban areas is an important goal in order to reduce health related problems. Local authorities must monitor a range of pollutants and attempt to control their concentrations to levels below specified national standards (Department for Environment, Food and Rural Affairs (DEFRA) 2007). Due to the cost of real time monitoring equipment, on street measured data is often sparse. A

range of empirical and computational models are therefore usually employed during the screening of air quality, providing high resolution spatial coverage. If reliable, model predictions can help to site longer term monitoring equipment in potential hot spot areas, and to evaluate a range of proposed strategies for the future improvement of air quality. In order to reliably inform decision making, the model should be able to represent monitored pollutant concentrations during evaluation, as well as the impact of the main factors leading to elevated pollutant concentrations.

There has been significant progress in the development of suitable modelling strategies for predicting pollution dispersion at the local scale in recent years. A particular feature that has received attention is the influence of local building topologies on the flow and turbulence patterns that are established within networks of urban streets (Nakamura & Oke 1988, Louka et al. 2000, Dixon et al. 2006). Although empirical models may provide representation of some of these types of features, such as street canyon recirculation (Berkowicz 2000), there is a growing acknowledgement that in order to properly resolve 3-dimensional flow structures, computational fluid dynamics (CFD) models will be required. Such CFD models range from highly detailed representation of turbulent processes such as found in Direct Numerical Simulation (DNS) (Moin & Mahesh 1998) and Large Eddy Simulation (LES) (Liu & Barth 2002, Baker et al. 2004) to Reynolds Averaged Navier Stokes (RANS) models which include parameterisations of turbulent mixing e.g. MISKAM (Eichhorn 1996). It is fair to say that at the present time only the RANS approach provides the rapid simulation times required by operational models. However, since the models include parameterisations of important features, their evaluation benefits from the inclusion of sensitivity studies that highlight the impact of uncertain input parameters on predicted flow fields.

Most investigations on air pollution in urban areas have been focused on the street canyon. A street canyon is the simplest street configuration made up by a relatively

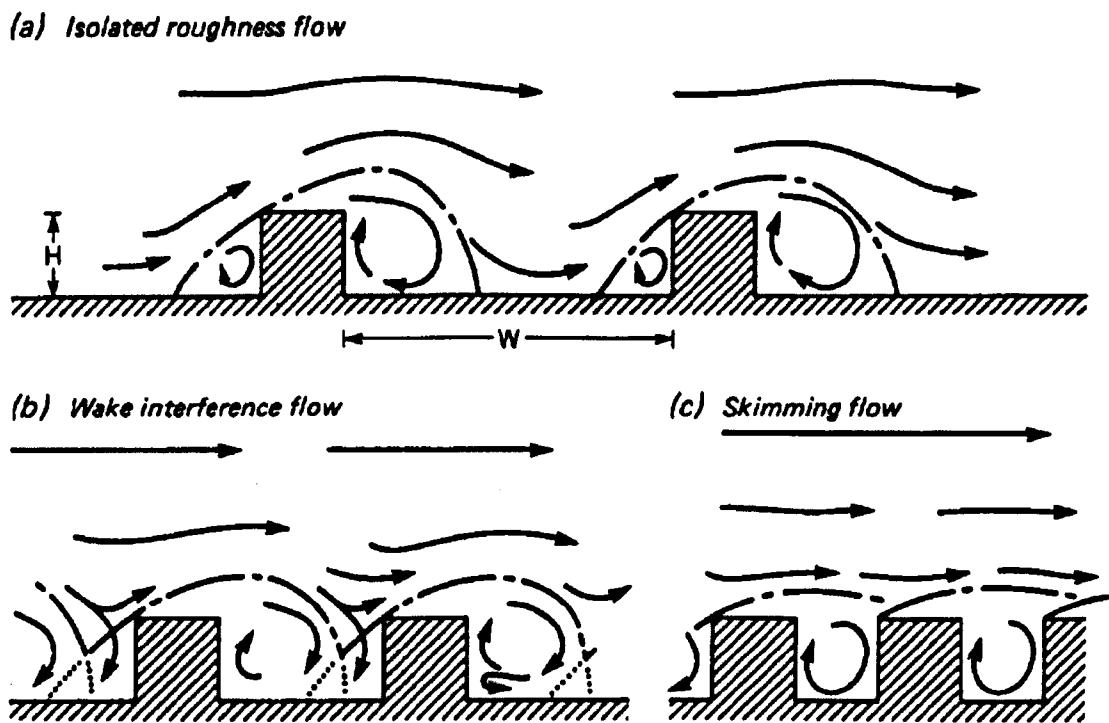


Figure 4.5: Flow regimes associated with perpendicular background winds in street canyons with different aspect ratios (Oke 1988).

narrow street between buildings that line up continuously along both sides (Oke 1988, Sini et al. 1996). The in-canyon flow patterns depend on the geometry of the street canyon, particularly the aspect ratio (building height to street width H/W). Three different flow regimes (figure 4.5) can be described for street canyons when the above-roof flow is perpendicular (90° to the street canyon alignment) (Oke 1988). Figure 4.5a shows "isolated roughness" flow which occurs for widely spaced buildings (i.e. $H/W < 0.3$), figure 4.5b presents "wake interference" flow for buildings that become more closely spaced (i.e. $H/W \approx 0.5$) and figure 4.5c shows "skimming" flow which is characterised by a single vortex circulation within the canyon (i.e. $H/W \approx 1.0$) (Hunter et al. 1992, Vardoulakis et al. 2003).

It is computationally feasible to develop a RANS model for real city geometries which contains street canyons and other forms such as street intersections (Dixon et

al. 2006). Important features of such models will include the use of an appropriate grid resolution to resolve building geometries, as well as proper definitions of model boundary conditions. These include the velocity and turbulence profiles coming into the model domain and profiles at the model surfaces such as near walls and roofs.

The amount of computational grid points required to resolve the near wall turbulent structures in street canyons using for example LES are too large. Additionally, the walls of atmospheric boundary layers flows are not smooth but consist of roughness elements. A computational mesh which resolves all the individual roughness elements is impossible (Bechmann et al. 2007). In order to solve the problems of the rough wall, approximate boundary conditions are necessary. Therefore, RANS models such as MISKAM use boundary conditions based on idealised empirical relationship known as the log-law. However, the parameters describing the flow close to the boundary are subject to uncertainties since they are difficult to estimate or measure over complex surfaces (Ro & Hunt 2007).

4.3.2 Model Description

A large field study was conducted in 2003 in the City of York, UK (Boddy et al. 2005). One aim of the study was to investigate the influence of background meteorology and building topologies on flow and turbulence patterns within urban street canyons. One of the canyons under investigation was Gillygate. Gillygate is a relatively narrow street with an aspect ratio (H/W) of approximately 0.8 leading to recirculating flow under a range of background wind directions. The "skimming" flow of the type as shown in figure 4.5c provides minimal ventilation of the canyon and is ineffective of removing pollutants (Berkowicz 2000). A schematic illustration of the flow and dispersion conditions in street canyons is shown in figure 4.6. It can be seen that the pollutants emitted from traffic in the street are primarily transported

towards the leeward side (upwind building). The traffic flow along Gillygate is quite high with significant periods of congestion and it therefore represents a potential pollution hot spot.

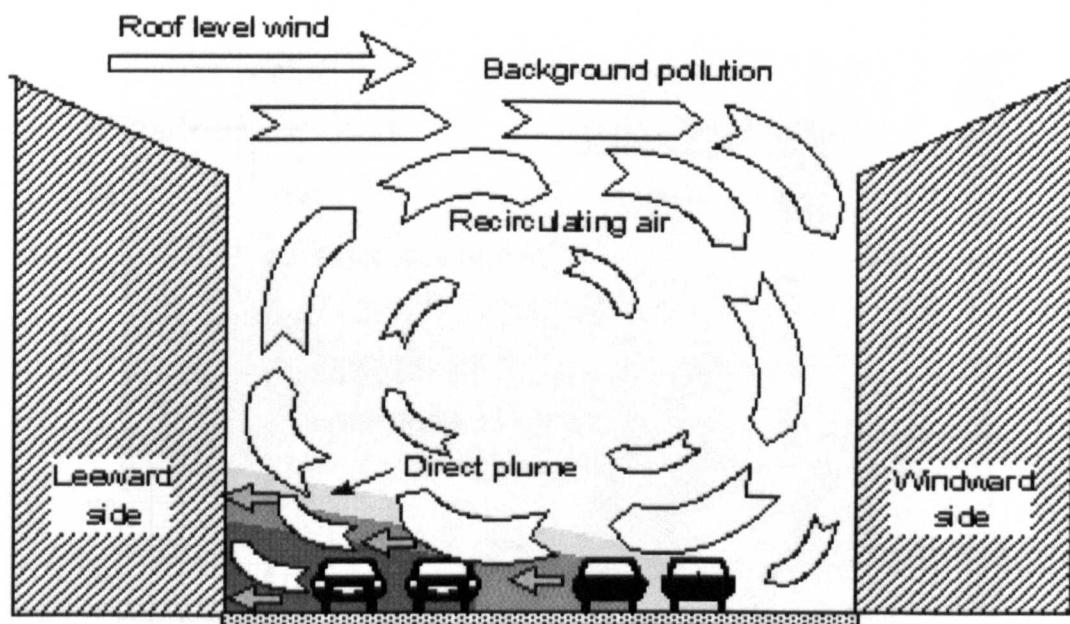


Figure 4.6: Schematic of the recirculating flow showing the transport of pollutants to the leeward side of the street canyon (Berkowicz 2000).

During the field experiment the in-street wind data were measured using Gill ultrasonic anemometers at different locations within the street. This study concentrates on data from two of the anemometers referred to as G3 and G4. Both anemometers were attached to a lamp post at heights of 5.5 m and 5.7 m respectively on opposite sides of the street canyon. The background wind speed and direction were measured using an anemometer attached to a mast in a nearby car park at a height of 19.5 m.

The flow field within the Gillygate street canyon was simulated using the micro-scale $k-\epsilon$ model MISKAM (Eichhorn 1996, Ries & Eichhorn 2001). MISKAM consists of a 3-dimensional non-hydrostatic flow model and an Eulerian dispersion model. This study takes only the first stage into account: the simulation of the

flow field. A comprehensive user guide (Eichhorn 2008) which also explains the MISKAM software in more detail can be found on the following website: <http://www.lohmeyer.de/eng/Software/default.htm>.

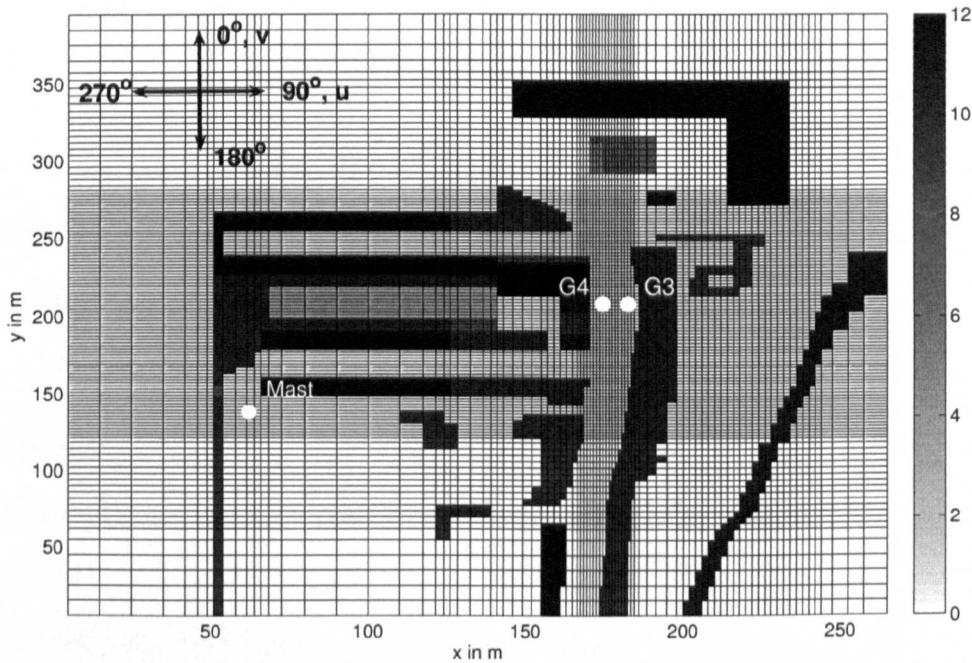


Figure 4.7: Grid and building configuration around Gillygate as used in MISKAM.

Figure 4.7 shows the grid and the building configuration of Gillygate and the surrounding area that were used for the simulation with MISKAM. The building heights in m are indicated in the legend. The measurement points are marked as G3, G4 (anemometer in-street data, one on each side of the road) and Mast (background wind speed and direction). A non-equidistant grid was used to enable a higher resolution within the area of interest (Gillygate). The grid has a resolution of 89 points (270 m) in the x-direction by 124 points (400 m) in the y-direction by 28 point (100 m) in the z-direction. The compass in figure 4.7 represents direction from and is oriented with respect to the street canyon.

The model assumes a logarithmic wind profile and neutral static stability at the

inflow boundaries. The logarithmic wind profile (law of the wall) is a mathematical relationship used to approximate the general logarithmic profile of wind speeds as they increase with increasing distance from the ground:

$$u = \frac{u_*}{\kappa} \ln \left(\frac{z + z_0}{z_0} \right) \quad (4.9)$$

where u is the mean wind speed at the height z above the surface, z_0 is the surface roughness length, u_* is the friction velocity and κ is the von Karman's constant. The assumptions made in this parameterisation are that the boundary grid point is in the region that has constant Reynolds stress and a logarithmic velocity profile.

The roughness length (z_0) is not a physically measurable length, rather a length scale describing the surface roughness. It can be considered as the height at which the horizontal wind speed equals zero. A high roughness length refers to a surface with for example buildings and trees and a low roughness length refers to a rather plain terrain. One difference between the standard law of the wall (Oke 1987) and equation (4.9) is that equation (4.9) uses $z + z_0$ instead of simply z . This is because if the height of the first grid point is the same as the roughness length then the log law cannot be applied (Benson et al. 2008).

If the model domain is very large, then the sensitivity to the inflow roughness length should be small, because the flow will adjust over the real roughness elements resolved within the model. However, this has computational implications and hence the model usually covers a restricted domain. It is therefore important to investigate the sensitivity of the model to the assumed inflow roughness length at the inflow boundary.

A comparison of the normalised flow components of the mean wind against the background wind direction U_m (mean wind speed measurement at the mast) for both the

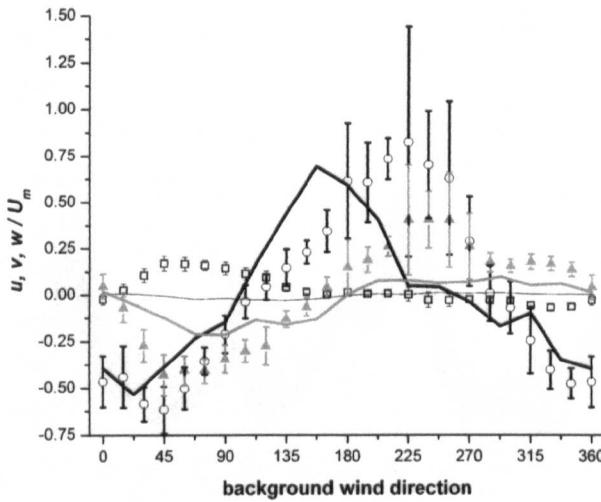


Figure 4.8: Normalised velocity components against background wind direction from experimental data and model data for G4 (model data for the u component are not shown). Experimental data: \square u component, \circ v component, \blacktriangle w component. The experimental data are 15 min averages from the field campaign and standard deviation of data is shown as error bars (Dixon et al. 2006).

field and model (MISKAM) data is shown in figure 4.8 for the measurement point G4. The along-street wind component v shows good agreement with the experimental data for background wind directions between 290° and 120° , however between 45° and 60° the modelled data are slightly outside the error bars of the experimental data. There is a disagreement between the model response and the experimental data for the along-street wind component v for background winds between 205° and 285° . Here, the model underestimates the normalised wind component. This can be explained by the effect of a tree near the mast, which is probably causing the background wind speed to be under-recorded. Therefore, the normalised values of the experimental values are higher (Dixon et al. 2006). For background winds between 120° and 165° the model overestimates the flow along the street.

Figures 4.9 and 4.10 present the comparison of experimental data and model results

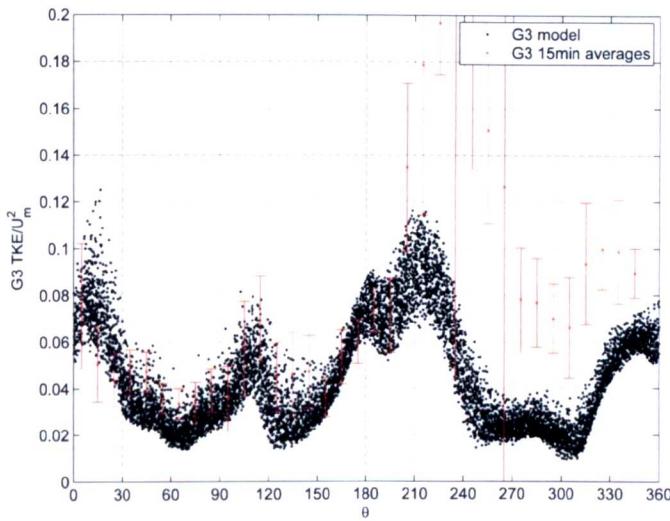


Figure 4.9: Normalised TKE against background wind direction θ from experimental data and model data for G3. The scatter for the model data is caused by the uncertainties in the roughness parameters (Benson et al. 2008).

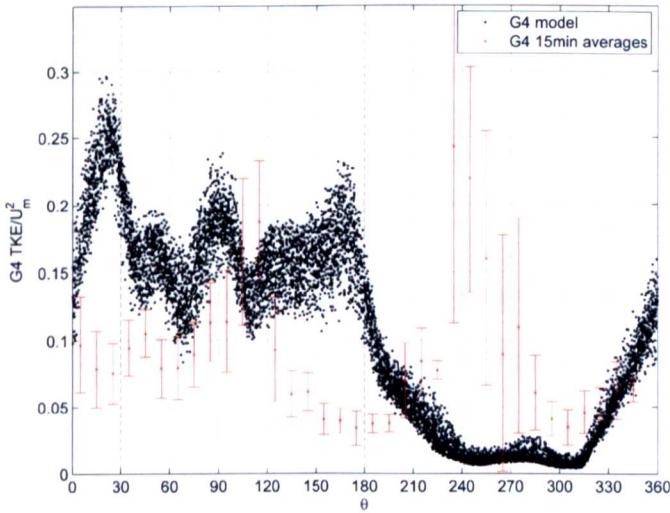


Figure 4.10: Normalised TKE against background wind direction θ from experimental data and model data for G4. The scatter for the model data is caused by the uncertainties in the roughness parameters (Benson et al. 2008).

for the turbulent kinetic energy TKE at the measuring points G3 and G4. For background winds between 200° and 270° the model underestimates the normalised TKE for both measuring points. This again can be explained by the effects of the tree near the mast, which slows down the flow compared to the true background flow. For all other background wind directions the agreement between experimental data and model results is quite good for G3 (figure 4.9). However, there is only poor agreement between the data at the measuring point G4 (figure 4.10), even taking into account the uncertainties in the model input parameters (table 4.5). For most of the background wind directions the model overestimates the TKE. This may be due to the fact that G4 is located quite close to the wall, suggesting the possibility of excessive TKE generation at the wall surfaces (Benson et al. 2008). This is also confirmed by the cross-section shown in figure 4.11 presenting the mean TKE and the u, w wind vectors for a background wind direction of 90°. The re-circulating flow pattern is clearly visible and high TKE can be found close to the windward wall showing how critical the location of the measurement point in the model is (Benson et al. 2008).

As it could be shown there is disagreement between the experimental data and the model for the normalised TKE and the wind components for various background wind directions at G4. Uncertainty and sensitivity analysis can help to explain the disagreement. However, if inconsistencies are very large, then the differences might come from the model physics or resolution rather than from parameter uncertainties.

A particular feature of interest within urban street canyons is the presence of a cross canyon vortex under background wind directions that have a significant cross canyon component (see also figure 4.6 and 4.11). In this study we therefore focus on background wind directions that are close to perpendicular (90°) to Gillygate and explore the sensitivity of selected model outputs to small changes in wind direction as well as the input roughness parameters.

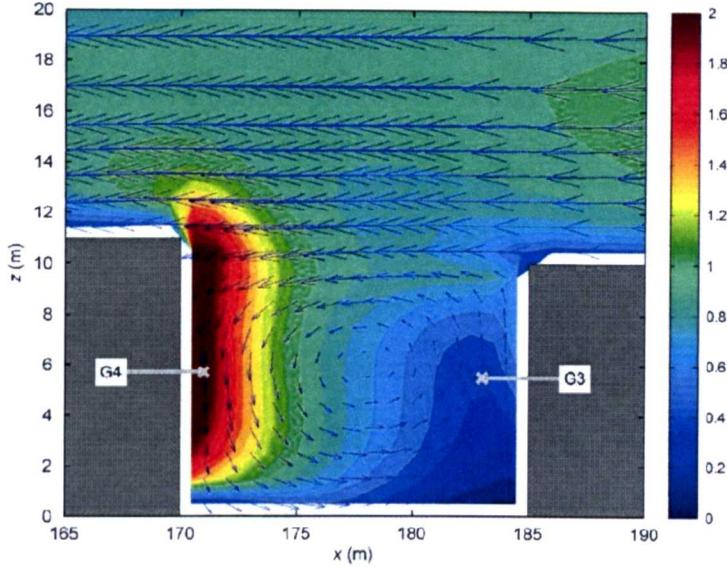


Figure 4.11: Street canyon un-normalised TKE and wind vector cross-section for a background wind direction of 90° (Benson et al. 2008).

The input parameters under consideration with ranges are shown in table 4.5. A uniform distribution is assumed for all parameters. The inflow roughness length is used by MISKAM to create a 1D wind profile which is then applied horizontally across the entire model domain. A roughness length can also be defined for buildings (wall roughness length) and the ground (surface roughness length) and is assumed to be uniform for the whole domain. The surface of the ground is assumed to be rougher than the surface for buildings because of the existence of potential objects such as cars, people etc. in a typical street. According to the MISKAM manual the maximum value for each of the roughness parameters is recommended to be half of the size of the smallest distance to the wall grid point which is 50cm in this study. The inflow roughness length is therefore restricted to the maximum value allowed although typical values for a small town would normally be considered larger (e.g. 60-100cm (Stull 1988)). Thus, the model's constraint on the inflow roughness length may limit the TKE at the inflow boundary which could lead to lower street canyon values (Benson et al. 2008).

Input parameter	Uncertainty range
x_1 Inflow roughness length	[5...50]cm
x_2 Surface roughness length	[0.5...50]cm
x_3 Wall roughness length	[0.5...10]cm
x_4 Background wind direction θ	[80...100] $^{\circ}$

Table 4.5: Input parameters and uncertainty ranges used in this study.

The selected outputs of interest are the wind components u, v, w normalised by the background wind speed U_m at G4 (y_1-y_3), the normalised turbulent kinetic energy (TKE/ U_m^2) at G4 (y_4) and the normalised turbulent kinetic energy (TKE/ U_m^2) at the mast (y_5). These outputs have been selected because the velocity and turbulence fields are also used for dispersion modelling (see example the Lagrangian stochastic plume model in section 5.4) in order to calculate concentrations and concentration fluctuations. Therefore, it is of interest to explore their uncertainties and to identify important parameters. An accurate prediction of pollutants and potential hot spot areas is only possible if the flow field is reliable.

A preliminary evaluation of the MISKAM model for the York study area was discussed in Dixon et al. (2006) where the sensitivities to inflow roughness length and model resolution were explored using a small ensemble of model simulations. The aim of this study is to establish and test techniques that will allow global sensitivity analysis of the model to be performed in a computationally efficient way. This will also allow to study the model physics in a more detailed way and will help to explain the discrepancies between experimental results and simulations.

4.3.3 Results

A fully functional metamodel is first constructed, using both HDMR approaches (cut-HDMR and RS-HDMR) for comparison. On the basis of these metamodels uncertainty and global sensitivity analysis can be performed in a computationally

efficient way. The sensitivity indices are calculated up to second-order to show the cooperative effects of two parameters acting together.

Construction of the Metamodel

The cut-HDMR method calls for a quantisation of the input variables and 11 equally spaced grid points were chosen for each of the inputs. The reference point $\bar{\mathbf{x}}$ was selected as the midpoint in each of the four input parameter ranges. According to equation (3.9) 41 model runs are necessary to construct a first-order metamodel and 641 model runs for a second-order metamodel. The results of the simulations are stored in low-dimensional look-up tables with respect to the chosen sampling points. The first-order metamodel requires only one-dimensional look-up tables (for the representation of the component functions $f_i(x_i)$) and the second-order metamodel one and two-dimensional look-up tables (for the representation of the component functions $f_i(x_i)$ and $f_{ij}(x_i, x_j)$). The value of $f(\mathbf{x})$ for an arbitrary point \mathbf{x} can be calculated by applying low-dimensional interpolation over the HDMR expansion terms $f_i(x_i)$ and $f_{ij}(x_i, x_j)$ with respect to the input values of the given point \mathbf{x} .

An example of the accuracy of the first- and second-order model replacement is shown in figure 4.12 for one output and 100 random points, which are different to the ones chosen to construct the model replacement. Another possibility to express the accuracy of the model replacement is the calculation of the relative error (RE) as shown in equation (3.66). The results for an ensemble of 2 000 random points are presented in table 4.6.

The first-order metamodel assumes that the output is only influenced by additive effects, which means that the input parameters are acting independently. As shown in figure 4.12 and table 4.6, the accuracy of a first-order cut-HDMR metamodel is not satisfactory for all of the outputs. Therefore, a second-order metamodel was

Output	1st-order metamodel			2nd-order metamodel		
	5 % RE	10 % RE	20 % RE	5 % RE	10 % RE	20 % RE
G4 u/U_m	84.9 %	94.3 %	98.8 %	98.1 %	100 %	100 %
G4 v/U_m	79.7 %	92.2 %	98.5 %	100 %	100 %	100 %
G4 w/U_m	93.2 %	99.6 %	100 %	100 %	100 %	100 %
G4 TKE/ U_m^2	92.1 %	99.2 %	100 %	100 %	100 %	100 %
Mast TKE/ U_m^2	94.4 %	99.4 %	100 %	100 %	100 %	100 %

Table 4.6: Percentage of points within a given relative error for the cut-HDMR metamodel.

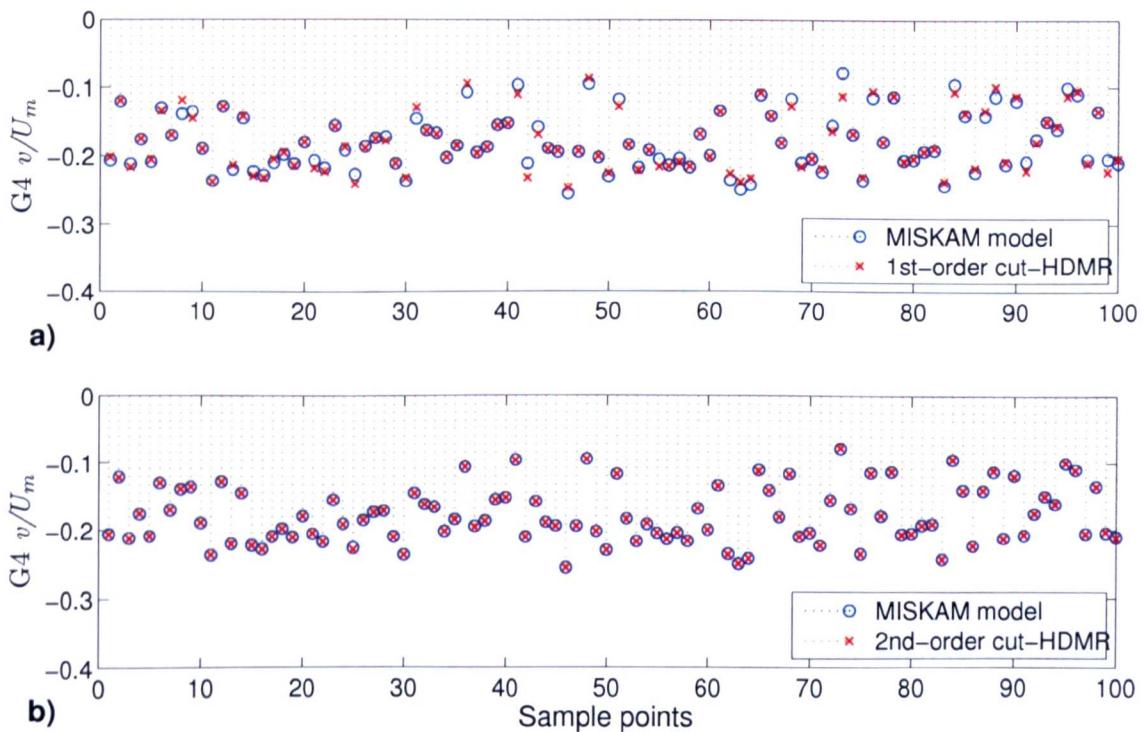


Figure 4.12: Comparison of the output response from the MISKAM model and (a) the first-order and (b) the second-order cut-HDMR metamodel for the normalised wind component v/U_m at the measurement point G4.

constructed, which includes the cooperative effects of two input parameters acting together upon the output. The second-order cut-HDMR expansion provides very accurate results in comparison with the ones obtained from the real model runs.

The RS-HDMR approach in connection with the approximation of the component

functions by orthonormal polynomials requires only one set of random samples. Different sample sizes N have been investigated. In this study a sample size of $N = 512$ using Sobol's quasi-random sampling method (Sobol' 1967) is applied to determine the coefficients for the orthonormal polynomials. A maximum order of three has been chosen for the approximation of the first- and second-order component functions in connection with the optimisation approach as introduced in section 4.1.

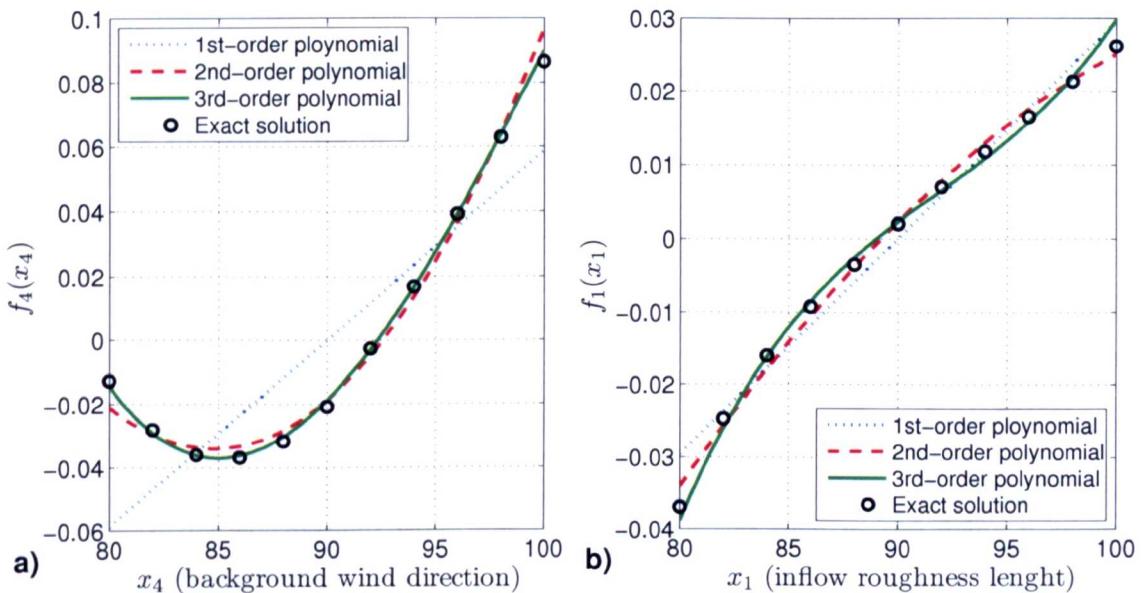


Figure 4.13: Comparison of different order approximations for first-order component functions for input parameter (a) x_4 (background wind direction) and (b) x_1 (inflow roughness length) and model output G4 w/U_m .

Figure 4.13 presents examples of the first-order component functions for two outputs. The figure shows how the approximation of the component functions would look if orthonormal polynomials different of orders were used. As shown in figure 4.13a, the first-order component function $f_4(x_4)$ for the output G4 w/U_m is best approximated by a third-order polynomial. In contrast, the first-order component function $f_1(x_1)$ for the same output (figure 4.13b) is best approximated by a second-order polynomial. In this case the approximation of the component function by a third-order polynomial results in a poorer accuracy, because a third-order polynomial has more

terms, each with its own Monte Carlo integration error. Therefore, if the third-order terms and the sample size N are quite small it is possible that the increased integration error outweighs any improvement in the overall fit. Experiments have confirmed, that if using a larger sample size (e.g. $N = 1024$) the approximation with a third-order polynomial gives slightly better results than the approximation with a second-order polynomial. The optimisation approach introduced in section 4.1 produces the same results for these examples in an automatic way, and the optimal order for all of the orthonormal polynomials used to approximate the RS-HDMR component functions can be found in table 4.7.

Output	f_1	f_2	f_3	f_4	f_{12}	f_{13}	f_{14}	f_{23}	f_{24}	f_{34}
1 G4 u/U_m	2	2	2	3	1	1	2	2	2	2
2 G4 v/U_m	2	2	2	2	1	1	2	2	1	2
3 G4 w/U_m	2	2	2	3	1	1	2	2	2	1
4 G4 TKE/ U_m^2	2	2	2	3	1	3	2	3	2	2
5 Mast TKE	2	2	3	3	1	3	1	3	2	2

Table 4.7: Order of polynomials used for the approximation of first and second-order component functions of the MISKAM flow model using a sample size $N = 512$.

Output	1st-order model replacement			2nd-order model replacement		
	5 % RE	10 % RE	20 % RE	5 % RE	10 % RE	20 % RE
1 G4 u/U_m	89.8 %	97.5 %	99.6 %	99.2 %	100 %	100 %
2 G4 v/U_m	79.1 %	93.5 %	98.8 %	99.9 %	100 %	100 %
3 G4 w/U_m	93.7 %	99.8 %	100 %	99.4 %	100 %	100 %
4 G4 TKE/ U_m^2	92.8 %	99.5 %	100 %	98.8 %	100 %	100 %
5 Mast TKE	93.8 %	99.4 %	100 %	96.7 %	99.8 %	100 %

Table 4.8: Percentage of points within a given relative error for the RS-HDMR model replacement using a sample size of $N = 512$ and optimal-order polynomials.

It can be seen that the component functions for a certain output are approximated using different polynomial orders. No component functions are found to be excluded.

All first-order component functions are represented by either second-order or third-order polynomials which indicates that the first-order response to all parameters is non-linear for all of the outputs. This has implications for the use of standard statistical tests such as Pearson correlations, because a certain relationship (e.g. linear) is assumed between the input and the output.

Table 4.8 shows the accuracy of the first- and second-order RS-HDMR metamodel if using optimal-order polynomials and a sample size of $N = 512$. The results obtained by applying cut-HDMR and RS-HDMR are very similar, and both approaches require nearly the same computational effort to construct a second-order metamodel for this case of four input parameters.

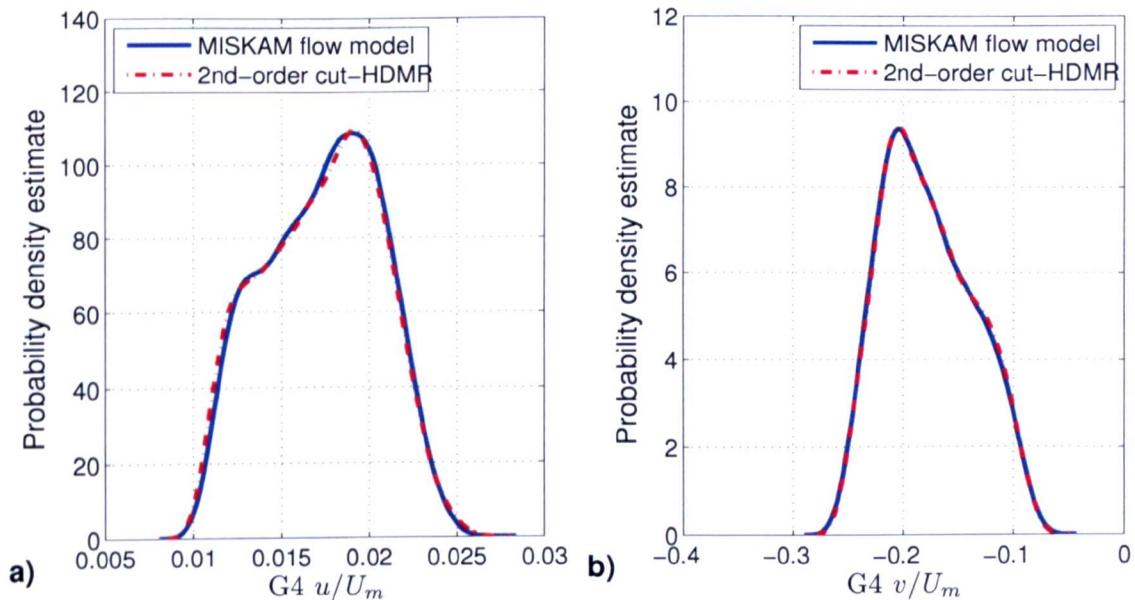


Figure 4.14: Comparison of the statistics between the MISKAM flow model and the second-order cut-HDMR metamodel.

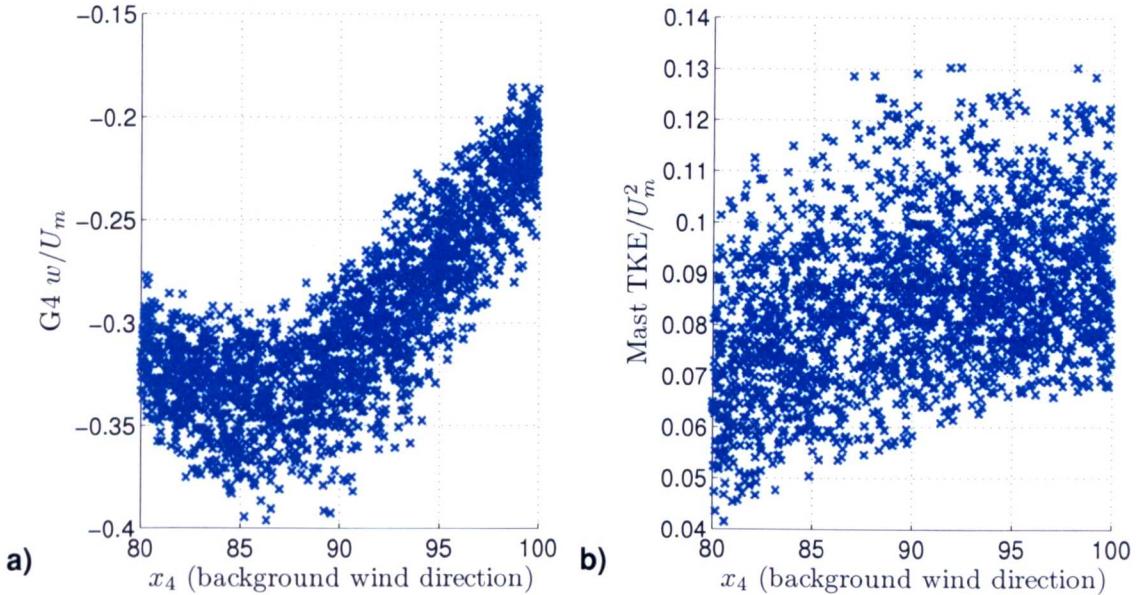


Figure 4.15: Scatter plot of x_4 (background wind direction) and (a) output $G4 w/U_m$ (normalised vertical wind velocity) and (b) Mast TKE/U_m^2 .

Uncertainty and Sensitivity Analysis

The fully functional metamodel of the MISKAM flow model based on the cut-HDMR and RS-HDMR approaches can now be used to perform uncertainty and global sensitivity analysis. Figure 4.14 shows the probability density function (pdf) estimate for outputs y_1 ($G4 u/U_m$) and y_2 ($G4 v/U_m$) of the MISKAM flow model. A total of 4 000 samples were used to perform the real model runs and another set of 4 000 random samples to calculate the output response of the second-order cut-HDMR metamodel for comparison. The statistics show very good agreement implying that the metamodel can be used instead of the MISKAM flow model in, for example, Monte Carlo type analysis. A large sample size can be applied because the metamodel is much less expensive to run than the real model. It is worth noting that the set of 4 000 real model runs performed here for evaluation purposes required parallel computing facilities.

For comparison, figure 4.15 shows two example Monte Carlo scatter plots using the metamodel. From figure 4.15a it is clear that a strong relationship exists between background wind angle and the output G4 w/U_m . This is expected since the orientation of the background wind strongly affects the strength of the recirculation vortex within the canyon. The response can be seen to be non-linear. This is an important finding, because the measurement of the background wind is usually obtained at quite a distance away from the street canyon.

The amount of scatter in figure 4.15b is quite large across the parameter range, suggesting that this output has a low first-order sensitivity to background wind direction. The exact shape of the first-order response is difficult to assess within the scatter.

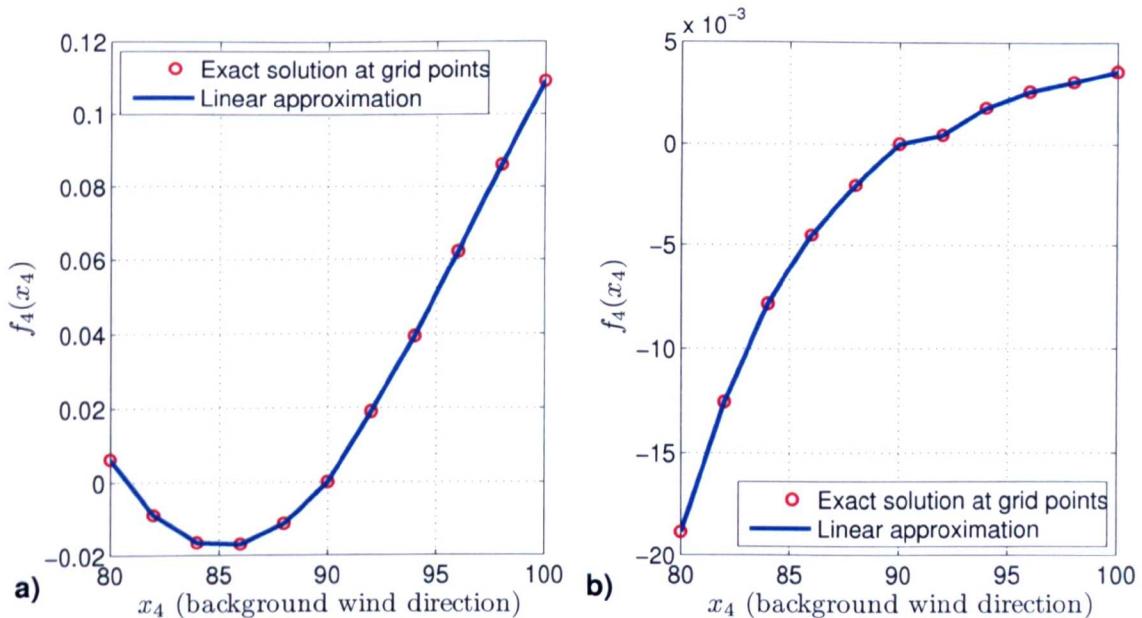


Figure 4.16: First-order component function (a) for input parameter x_4 (background wind direction) and output G4 w/U_m and (b) for input parameter x_4 (background wind direction) and output Mast TKE/ U_m^2 .

Figure 4.16 shows the corresponding first-order component functions from the cut-HDMR expansion. The component function in (a) describes the impact of the input

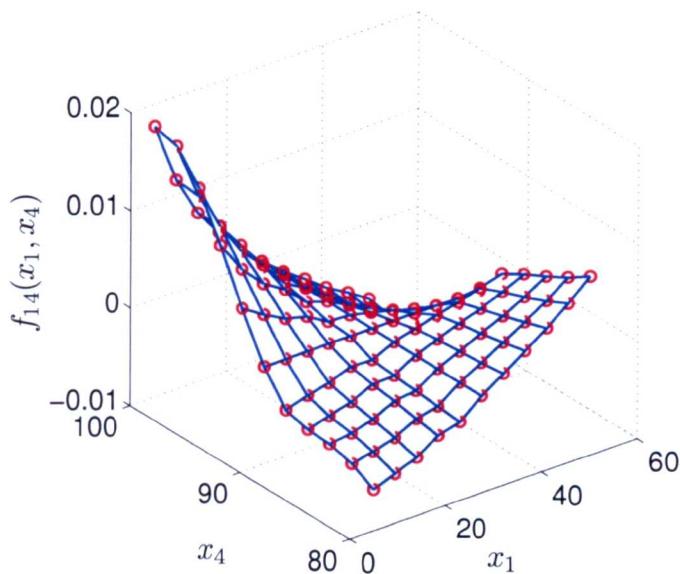


Figure 4.17: Second-order component function for input parameters x_1 (inflow roughness length) and x_4 (background wind direction) and output G4 TKE/ U_m^2 .

parameter x_4 (background wind direction) upon the output G4 w/U_m , acting independently and the first-order component function in (b) gives the effect of parameter x_4 (background wind direction) upon the output Mast TKE/ U_m^2 . The examples show that a strong non-linear relationship exists between the background wind angle and the normalised vertical velocity and the TKE respectively at the measurement location. The second-order component function in figure 4.17 presents the cooperative effects of the parameters x_1 (inflow roughness length) and x_4 (background wind direction) upon the output G4 TKE/ U_m^2 . The examples illustrate how much easier the component functions are to interpret than standard scatter plots.

In addition, the relationship between the HDMR component functions and the ANOVA decomposition can be used to calculate the partial variances and the sensitivity indices. A two step algorithm was given in section 3.4.1 if applying the cut-HDMR approach. The integrals in equation (3.71) and (3.72) were computed using the exact solutions at each of the grid points in connection with trapezoidal

numerical integration. Squaring and integrating over the ANOVA expansion functions provides the partial variances. The calculation of the partial variances on the basis of the RS-HDMR component functions is even simpler, since it only involves the expansion coefficients as described in equations (3.87) and (3.88).

Parameter	S_i, S_{ij}	Output:				
		1 G4 u/U_m	2 G4 v/U_m	3 G4 w/U_m	4 G4 TKE/U_m^2	5 Mast TKE
x_1 Inflow roughness length	S_1	0.0739	0.0075	0.1508	0.3197	0.4964
x_2 Surface roughness length	S_2	0.0414	0.0043	0.0757	0.1664	0.0116
x_3 Wall roughness length	S_3	0.1816	0.0435	0.0015	0.3474	0.3110
x_4 Background wind direction	S_4	0.6773	0.9139	0.7506	0.1194	0.1700
	$\sum S_i$	0.9743	0.9692	0.9787	0.9529	0.9889
x_1, x_2	S_{12}	0.0015	0.0001	0.0002	0.0007	0.0002
x_1, x_3	S_{13}	0.0009	0.0008	0.0015	0.0011	0.0048
x_1, x_4	S_{14}	0.0062	0.0134	0.0045	0.0353	0.0020
x_2, x_3	S_{23}	0.0010	0.0001	0.0003	0.0004	0.0003
x_2, x_4	S_{24}	0.0077	0.0014	0.0030	0.0081	0.0030
x_3, x_4	S_{34}	0.0083	0.0150	0.0118	0.0014	0.0009
	$\sum S_{ij}$	0.0257	0.0308	0.0213	0.0471	0.0111
	$\sum S_i + \sum S_{ij}$	1.0000	1.0000	1.0000	1.0000	1.0000

Table 4.9: First- and second-order sensitivity indices calculated on the basis of cut-HDMR using 641 model runs.

The first-order and second-order sensitivity indices determined on the basis of the cut-HDMR and the RS-HDMR component functions are given in table 4.9 and 4.10 respectively. Both approaches lead to very similar results with regard to the first and second-order sensitivity indices. However, there is one distinct difference. When using the cut-HDMR approach up to second-order, the overall variance D is assumed to be just the sum of the first- and second-order partial variances. Because of this all sensitivity indices for a certain output add up exactly to one. However, this means that higher-order effects, as far as they would exist, are not included in the overall variance. In fact it can be assumed that there are no crucial higher-order effects in the presented example of the MISKAM flow model, because the second-order cut-HDMR metamodel provides very accurate results and the statistics nearly perfectly

Parameter	S_i, S_{ij}	Output:				
		1 G4 u/U_m	2 G4 v/U_m	3 G4 w/U_m	4 G4 TKE/U_m^2	5 Mast TKE
x_1 Inflow roughness length	S_1	0.0878	0.0107	0.1677	0.2868	0.4773
x_2 Surface roughness length	S_2	0.0414	0.0062	0.0881	0.1846	0.0146
x_3 Wall roughness length	S_3	0.1557	0.0493	0.0006	0.3222	0.3285
x_4 Background wind direction	S_4	0.7036	0.9406	0.7813	0.1127	0.1525
	$\sum S_i$	0.9884	1.0069	1.0377	0.9063	0.9729
x_1, x_2	S_{12}	0.0006	0.0000	0.0002	0.0002	0.0002
x_1, x_3	S_{13}	0.0006	0.0003	0.0004	0.0061	0.0097
x_1, x_4	S_{14}	0.0074	0.0118	0.0043	0.0345	0.0020
x_2, x_3	S_{23}	0.0004	0.0001	0.0005	0.0045	0.0041
x_2, x_4	S_{24}	0.0067	0.0012	0.0024	0.0089	0.0032
x_3, x_4	S_{34}	0.0063	0.0154	0.0127	0.0010	0.0009
	$\sum S_{ij}$	0.0220	0.0288	0.0206	0.0553	0.0201
	$\sum S_i + \sum S_{ij}$	1.0104	1.0357	1.0583	0.9616	0.9931

Table 4.10: First- and second-order sensitivity indices calculated on the basis of RS-HDMR using 512 model runs.

match the statistics of the real model. When using the RS-HDMR approach the overall variance D can be estimated using the one set of samples $N = 512$ and equation (3.84). The sum of all the sensitivity indices does not add up to exactly one, however it is very close to one. The small discrepancy is due to the Monte Carlo integration error caused by the determination of the expansion factors α_r^i and β_{pq}^{ij} (see equations 3.44 and 3.46). The Monte Carlo integration error can be reduced by using a larger sample size N and/or by applying variance reduction methods as described in section 3.2.2. This would improve the accuracy of the metamodel and therefore result in even more precise sensitivity indices. Experiments regarding this have shown that the value of the sensitivity indices would not change significantly, however the sum of all the sensitivity indices adds up even closer to one.

The sensitivity indices in table 4.9 and 4.10 can be ranked to show which input parameter (or interaction of two input parameters) contributes most to the overall variance. For example, output y_1 (wind component u/U_m at G4) is mainly influenced by the background wind direction x_4 (rank: 1) and by the wall roughness length

x_3 (rank: 2). It is further influenced by the inflow roughness length x_1 (rank: 3) and the surface roughness length x_2 (rank: 4). All interactions (second-order effects) are negligible for output y_1 . In fact all outputs are mostly affected by input parameters acting independently. Only for output y_4 (normalised TKE at G4) does the interaction of two parameters (x_1 and x_4) have a significant influence on the overall variance.

Note that the considered outputs are normalised by the background wind speed U_m measured at the mast. This is necessary in order to compare the results of the simulations with experimental data, because the experimental data has large variations in mean wind speed U_m between different measuring periods (Benson et al. 2008). However, the sensitivities presented in table 4.9 and 4.10 are combinations of the effects of different outputs at separate model locations which makes the interpretation more difficult. For example, the normalised TKE at G4 is mainly influenced by the inflow roughness length x_1 (28.68 % using RS-HDMR) and the wall roughness length x_3 (32.22 % using RS-HDMR). However, if looking at the un-normalised TKE at G4 (as presented in Benson et al. (2008)) the influence of the wall roughness length is much stronger and contributes by over 60 % to the overall variance of TKE at G4. This is more realistic and confirms that for sensitivity analysis the un-normalised values for the wind components and TKE are actually better suited. In both cases (normalised and un-normalised) TKE at G4 is influenced by the inflow roughness length. However, the TKE in the street canyon should not very strongly be influenced by the inflow roughness and if it does it may be an indication that the model domain is not big enough (Dixon et al. 2006, Benson et al. 2008).

From table 4.9 and 4.10 it can be seen that the normalised wind components u, v, w are strongly influenced by the background wind direction (x_4) which contributes by around 70 %, 94 % and 78 % (obtained from RS-HDMR) respectively. The sensitivity of this parameter shows how important it is to specify the background wind

direction accurately. It also shows the limitations of determining the background wind direction far away from the measurement site. On the other hand, the inflow roughness length has only a small influence on the normalised wind components u, v, w which is a good feature of the model.

Due to the fact that the measuring point G4 is located so close to the wall, it is not surprising to see a strong influence of the wall roughness length on the normalised TKE. This would also explain the relatively poor agreement between experimental data and model data as shown in figure 4.10. The agreement between the normalised TKE from experimental data and model data is much better for the measuring point G3 (figure 4.9) which is located further from the wall. However, both figures 4.10 and 4.9 show only a fairly low degree of scatter, which suggests that if the background wind direction is well specified, then the output variability is quite low.

4.3.4 Discussion

It could be shown in this case study that the detailed input-output mapping produced by the HDMR method can be used for fundamental overall analysis of the model and for a computationally efficient determination of the sensitivity indices, including those of higher order. Special attention should be paid to the plots of the component functions, since they reveal useful information about the input-output relationship of the model and can be used for sensitivity analysis instead of more widely used scatter plots. In general scatter plots try to reveal the relationship between two variables (e.g. input and output). If the scatter plot shows a non-random structure a relationship is assumed to exist between the variables. Often, regression methods are used to establish the strength of the relationship. However, the examples shown here highlight the fact that often, the first-order sensitivity indices are highly non-linear. Measures such as the Pearson correlation cannot therefore be reli-

ably applied to standard outputs and either non-linear functions of the outputs must be developed, as in Manomaiphiboon & Russell (2004), or rank correlations must be used (MacDougall et al. 2005). In addition, it is sometimes very difficult to identify a relationship in the pattern of the scatter plot as demonstrated by the example in figure 4.15b. In some cases the scatter plot may appear to show a relationship that is caused by other variables than the one of interest. The interpretation of scatter plots is therefore problematic whereas the plots of the HDMR component functions up to second-order (e.g. see figures 4.16 and 4.17) show exactly how one input parameter or a pair of input parameters affect the output. By calculating the absolute magnitude of the component functions it becomes trivial to produce an importance ranking based on the sensitivity indices. In addition, the metamodel constructed on the basis of the HDMR method can also be used as a surrogate in Monte Carlo analysis. This provides the ability to use traditional uncertainty and sensitivity analysis methods and for example to compute output probability distributions as shown in figure 4.14. The computational effort involving the metamodel is quite low and therefore a large sample size can be applied to provide output statistics.

4.3.5 Comparison of cut-HDMR and RS-HDMR

Both HDMR approaches provide the possibility to construct a fully functional second-order metamodel with nearly the same computational effort for the number of input parameters tested here. The cut-HDMR method requires ordered sampling. The component functions are numerically presented in low-dimensional look-up tables over its variables. The RS-HDMR method uses one set of random samples in order to determine its component functions which are approximated by analytical basis functions. One advantage of the cut-HDMR method is that only 41 model runs were sufficient to construct a first-order metamodel of the MISKAM flow model.

The RS-HDMR method requires the whole set of 512 model runs to obtain comparable accuracy for the first-order model replacement. However, a drawback of the cut-HDMR method is that the number of model runs increases with the input space dimension (see equation (3.9)). Therefore, for larger numbers of input parameters it may become less efficient than RS-HDMR. Another drawback is in the storage of the component functions since they are provided in numerical look-up tables. However, for most applications this may not be prohibitive since only up to second-order tables have been required for applications to date. Exact component functions are provided at the points stored in the look-up table, however interpolation is required in order to determine the output response for an arbitrary point \mathbf{x} . The RS-HDMR method instead requires only the storage of the expansion coefficients. No interpolation is involved and only the analytical polynomial functions have to be calculated at a given point \mathbf{x} . There is however, no guarantee that the metamodel is exact anywhere since a fitting error may always be present. Nevertheless, the example has shown that using optimal-order polynomials, fitting errors and Monte Carlo integration errors can be kept small. In addition, the estimation of the sensitivity indices is easily achieved if the RS-HDMR approach is used in connection with orthonormal polynomials (see equations (3.87) and (3.88)).

4.3.6 Conclusions

The complexity of the street scale turbulent flow model restricts its application in connection with traditional global uncertainty and sensitivity methods such as Monte Carlo analysis since a large number of model runs are required. One run of the MISKAM flow model can take up to about 40 min on a 3 GHz PC. Local SA methods are not suitable, because of the high non-linearity of the input-output relationships and parameter interactions. HDMR methods have been shown to pro-

vide a straightforward approach to explore the input-output mapping. Both HDMR approaches provide an efficient way to calculate the sensitivity indices (in this study up to second-order) without the need for large numbers of model runs. In particular, the exploration of the interactive effects between input parameters makes the HDMR method extremely attractive for global sensitivity analysis. Because HDMR methods treat the model essentially as a black box, they could potentially be used for a wide range of applications in environmental modelling. The HDMR method is especially suitable for computationally expensive models with a large input space dimension. If ordered sampling of the inputs is possible, then the Cut-HDMR method can be applied. Otherwise the RS-HDMR approach provides a possibility to calculate the component functions by using only one set of random samples. As shown for this example an HMDR expansion up to second-order is usually sufficient to represent the real model.

4.4 Summary

The RS-HDMR tools have been extended by an optimisation algorithm to calculate the optimal polynomial order for the approximation of the component functions. This is a very useful approach and can increase the accuracy of the metamodel significantly. The optimal order is automatically determined under consideration of a maximum order provided by the user. Component functions can also be excluded from the HDMR expansion if they do not contribute to the overall output value.

Both HDMR approaches, cut-HDMR and extended RS-HDMR, have been first applied to an analytical model (Ishigami function) with three uncertain input parameters. In this case the sensitivity indices could be calculated analytically and compared with the estimates based on the HDMR component functions. Both HDMR approaches were able to provide very good estimates for the first- and second-order

sensitivity indices. The RS-HDMR approach has been shown to be more efficient than the cut-HDMR approach and only $N = 256$ model runs were required to construct the metamodel and to calculate the partial variances. In the next stage a micro-scale CFD model simulating the flow and turbulence in a 3D street canyon has been investigated. Four uncertain parameters were considered. Both HDMR approaches performed very well and a total of $N = 512$ (RS-HDMR) and $N = 641$ (cut-HDMR) model runs were sufficient in order to construct a very accurate second-order metamodel and to calculate the sensitivity indices. Based on these results important findings regarding the model behaviour and physics have been made. For example, it has been shown that the wind components u, v, w within the street canyon are very sensitive to changes in the background wind direction. This demonstrates the limitations of determining the background wind direction far away from a pollution hot spot location where model simulations are used as part of an air quality management strategy. However, if the wind direction is well specified the analysis showed the model simulations to be fairly robust for the measurement location G3 which was far enough away from the street canyon wall.

As a result it can be said that the RS-HDMR approach has advantages over the cut-HDMR approach. RS-HDMR requires only one set of random or quasi-random samples and the number of required samples N does not directly depend on the input space dimension n . Further, RS-HDMR does not require the storage of data in look-up tables or interpolation methods. Only the expansion coefficients have to be stored. The total variance is calculated on the basis of the random sample. Cut-HDMR would require an additional set of samples in order to calculate the total variance in this way. RS-HDMR is the more promising approach especially with the focus on models with a large input space dimension (e.g. $n \geq 10$) and therefore it will be solely used in connection with the case studies in the following two chapters.

5 Application of RS-HDMR for Systems with a Large Number of Input Parameters

The main focus of this chapter is on the investigation of systems with a large input space dimension (i.e. $n \geq 10$). Another extension to the existing set of RS-HDMR tools is presented in the next section. A threshold is introduced, which allows exclusion of unimportant component functions from the HDMR expansion. In this way large input space dimensions can be explored more efficiently. The extended RS-HDMR method is then tested using the analytical Sobol' g-function (Saltelli & Sobol' 1995). First- and second-order sensitivity indices are explored and compared with analytically calculated values. Further, the focus is on two practical applications. A reactive plume model is investigated in order to study the combined effects of uncertainties in the physical and chemical parameters. The second case study focuses on a one-dimensional low-pressure premixed methane flame model doped with sulphur and nitrogen species. Here, the effects of uncertainties in rate constants and thermodynamic data are investigated.

5.1 Threshold to Exclude Component Functions

The standard RS-HDMR approach has already been extended by an optimisation method introduced in section 4.1, which automatically chooses the best polynomial order for the approximation of each of the component functions. Here, this method is further extended by excluding 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 (see section 2.3) can be avoided. Component functions can be excluded, if they do not contribute to the overall output value and furthermore, a threshold can be defined to exclude component functions with only a very small contribution. This is particularly useful if a model has a large number of input parameters and only a few of them are important. The component functions of the HDMR expansion are approximated by orthonormal polynomials and the coefficients for these polynomials are computed by Monte Carlo integration. If a component function has only a very small influence on the output value, then it might be the case that the Monte Carlo integration error leads to an overestimation of the importance of the component function being considered. One might mainly include the Monte Carlo integration error in the HDMR expansion, rather than the real effect of this parameter or parameter interaction. This is especially true if the sample size N is only small and if higher order polynomials are used. In those cases it would be better to exclude these unimportant parameters and parameter interactions from the HDMR expansion by setting a threshold. For a second-order HDMR expansion a separate threshold can be defined for the exclusion of first- and second-order component functions.

To apply thresholds within the HDMR framework the sum of the square errors E_1 between the real model response and the mean f_0 is calculated

$$E_1 = \sum_{s=1}^N [f(\mathbf{x}^{(s)}) - f_0]^2 \quad (5.1)$$

as well as the sum of square errors E_2 between the real model response and the optimal polynomial approximation

$$E_{2i} = \sum_{s=1}^N [f(\mathbf{x}^{(s)}) - (f_0 + f_i(x_i^{(s)}))]^2 \quad (5.2)$$

for each of the first-order component functions $f_i(x_i)$. If the optimisation method described in section 4.1 is used to calculate the optimal order for the polynomial approximation, then E_1 and E_{2i} are already available and do not have to be calculated again. Following this, the relative error between E_1 and E_{2i} is calculated and compared with the threshold. If the relative error is smaller than the defined threshold, then the corresponding first-order component function will be excluded from the HDMR expansion. This process is similar for the exclusion of the second order component functions. However, here we use

$$f_0 + \sum_{i=1}^N f_i(x_i^{(s)}) \quad (5.3)$$

instead of f_0 for E_1 and

$$f_0 + \sum_{i=1}^N f_i(x_i^{(s)}) + f_{ij}(x_i^{(s)}, x_j^{(s)}) \quad (5.4)$$

instead of $f_0 + f_i(x_i^{(s)})$ for the calculation of E_{2i} .

5.2 Test Model: Sobol' g-function

The aim of this section is to demonstrate the usefulness of the threshold for excluding component functions from the HDMR expansion, which has been introduced in the last section. The Sobol' g-function (Saltelli & Sobol' 1995) is used as an analytical test model.

5.2.1 Properties of the Sobol' g-function

The Sobol' g-function (Saltelli & Sobol' 1995) is defined as

$$f(\mathbf{x}) = \prod_{i=1}^n g_i(x_i), \quad (5.5)$$

where

$$g_i(x_i) = \frac{|4x_i - 2| + a_i}{1 + a_i}. \quad (5.6)$$

The input parameter x_i is 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 . Here we consider 20 input parameters using $a_i = \{0, 1, 4.5, 9, 99, \dots, 99\}$. The spectrum of the first four values for a_i has been chosen according to Campolongo et al. (2000b). In addition to this, 16 parameters have been added with $a_i = 99$ in order to provide the high dimensional input space for this test case. The sensitivity indices can be calculated analytically on the basis of the ANOVA decomposition (see method of Sobol' in section 2.4.5) and the analytical values for the first-order sensitivity indices are given in table 5.1. It is clear that only the first four parameters (x_1 to x_4) can have an effect on the variance of the model output $f(\mathbf{x})$, since a_i has low values in those cases. Therefore, the first-order sensitivity indices for parameters five to 20 are zero.

S_1	S_2	S_3	S_4	$\sum S_i$
0.7165	0.1791	0.0237	0.0072	0.9265

Table 5.1: First-order sensitivity indices for the Sobol' g-function calculated analytically.

Table 5.2 presents the analytical values for the second-order sensitivity indices. Here, only the interactions of the first four parameters (x_1 to x_4) have an influence on the model output.

S_{12}	S_{13}	S_{14}	S_{23}	S_{24}	S_{34}	$\sum S_{ij}$
0.0597	0.0079	0.0024	0.0020	0.0006	0.0001	0.0726

Table 5.2: Second-order sensitivity indices for the Sobol' g-function calculated analytically.

As shown in table 5.1 the first-order sensitivity indices add up to nearly 93 %, which means that higher-order effects account for only around 7 % of the overall variance. Parameter x_1 is clearly the most important one ($a_1 = 0$) and parameter x_4 is the least important one ($a_4 = 9$) out of the first four input parameters. The interaction between parameter x_1 and x_2 is with nearly 6 % the most important second-order effect. All the other cooperative effects are negligible.

5.2.2 Application of RS-HDMR

The RS-HDMR approach is now applied in order to estimate the first- and second-order sensitivity indices. The maximum order for the approximation of the HDMR component functions is set to ten for first-order and three for second-order. The threshold for the exclusion of unimportant component functions was set to 0.5 %. Different sample sizes N have been investigated using the quasi-random Sobol' sequence.

N	256	512	1 024	2 048	analytical
S_1	0.6637	0.7026	0.7094	0.7118	0.7165
S_2	0.1784	0.1766	0.1774	0.1784	0.1791
S_3	0.0249	0.0259	0.0242	0.0234	0.0237
S_4	—	—	0.0073	0.0071	0.0072
$\sum S_i$	0.8670	0.9051	0.9182	0.9207	0.9265
S_{12}	—	0.0283	0.0531	0.0573	0.0597
S_{13}	—	0.0068	0.0075	0.0077	0.0079
S_{14}	—	—	0.0020	0.0020	0.0024
S_{23}	—	—	—	0.0021	0.0020
$\sum S_{ij}$	0	0.0351	0.0625	0.0691	0.0726
$\sum S_i + \sum S_{ij}$	0.8670	0.9402	0.9808	0.9898	0.9991

Table 5.3: First- and second-order sensitivity indices for the Sobol' g-function using the RS-HDMR approach with various samples sizes N .

Table 5.3 shows that even a small sample size of $N = 256$ is already enough to obtain the correct ranking for the three most important parameters. The threshold ensures that only these three component functions are approximated by orthonormal polynomials. The remaining 17 parameters are excluded from the HDMR expansion. Without applying a threshold another 11 parameters would have been approximated by polynomials where mainly the Monte Carlo integration error would have been added. Due to the small sample size, none of the higher-order effects can be accurately estimated. Thus, none of the second-order component functions are included in the HDMR expansion. The Monte Carlo integration error is too high in these cases so that the consideration of the approximated second-order component functions would not lead to an improvement in the accuracy of the HDMR metamodel. Again, the exclusion of the second-order component functions is controlled by the threshold.

An increase in the sample size N leads to a smaller Monte Carlo integration error and even second-order component functions can now be accurately estimated and included in the HDMR expansion. A sample size of $N = 2048$ gives almost the exact

first- and second-order sensitivity indices for the first four parameters if compared with the analytical values. By defining the value of the threshold the user can influence the complexity of the HDMR expansion. If one is only interested in the most important parameters and parameter interactions, then the value of the threshold should be greater than 1 %. For example, a threshold of 2 % in this test case would only include parameter x_1 and x_2 and the interaction of those two parameters into the HDMR expansion. All the other parameters and parameter interactions would have been excluded.

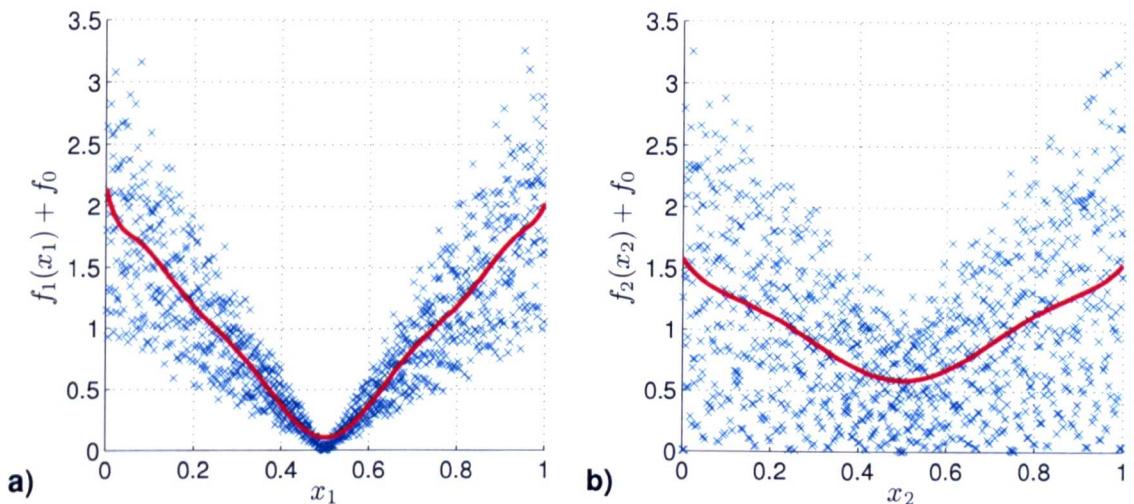


Figure 5.1: First order component function (a) for input parameter x_1 (approximated by a 10th order polynomial) and (b) for input parameter x_2 (approximated by a 6th order polynomial) using a sample size of $N = 1024$.

The plot of the component functions for the two most important parameters x_1 and x_2 is shown in figure 5.1. Here it can be seen why the Sobol' g-function is a tough test case for the RS-HDMR method. The relationship between the input parameter and the output is described by a non-monotonic V-shape curve with a non differentiable point. Higher order polynomials are required in order to approximate this shape. However, higher-order polynomials have more terms than lower-order polynomials and are therefore more likely to be effected by the Monte Carlo integration error. The

V-shape is most significant for parameter x_1 (figure 5.1a). Therefore, the accuracy of the first order sensitivity index S_1 for this parameter is not as accurate as for the other ones if using only a small sample size of $N = 256$. Only if using a larger sample size ($N \geq 512$) can the coefficients for the higher order polynomials be determined more accurately which results in a better estimate for the sensitivity index S_1 (table 5.3).

5.2.3 Discussion

By using the Sobol' g-function as a test case it can be demonstrated, that the RS-HDMR method can handle a large input space dimension. The method is able to identify the most important parameters and parameter interactions including their ranking with only small computational effort (e.g. $N = 512$). The highly non-linear and non-monotonic relationship between the input parameters and the model output could be approximated by orthonormal polynomials. However, higher-order polynomials (up to tenth-order) were required in order to deal with the complicated V-shape of the input-output relationship.

5.3 Introduction to Chemical Kinetics

The case studies investigated in this chapter partially involve complicated chemical mechanisms. The models provide interesting test cases for global sensitivity analysis using the HDMR method since a large number of uncertain parameters are considered and the mechanisms are usually highly non-linear. In the first case study the focus is on the effects of uncertainties in both, physical and chemical parameters and possible interactions. The second case study focuses on a more complex chemical scheme and investigates the effects of uncertainties in rate parameters and

thermodynamic parameters.

Studies are most often undertaken in order to develop reaction mechanisms which are accurate enough to be used within design applications or models informing environmental decision making. For example, atmospheric chemistry plays an important role in air pollution modelling. The formation of important secondary pollutants in the atmosphere such as nitrogen dioxide (NO_2) and ozone (O_3) can be described via chemical reactions and must therefore be part of any modelling system used for evaluating strategies to improve air quality. Combustion chemistry plays an important part in our daily life, e.g. burning fossil fuels for heating and for generating electricity, transportation systems (ground vehicles and aircraft) and industrial processes. However, the downside associated with using combustion for energy creation is in environmental pollution. Primary pollutants, such as nitrogen oxides, carbon monoxide and sulphur oxides are directly emitted in the atmosphere. This relates to health hazards, smog, acid rain, global warming and ozone depletion (Turns 2000). In order to predict the concentrations of primary and secondary pollutants it is important to be able to accurately model the complicated chemical processes in combustion and the atmosphere.

In order to get a better understanding of the chemical mechanisms and the parameters involved in these cases studies, this section aims to give a short overview of the principles in chemical kinetics. Chemical kinetics addresses the rate at which reactants are consumed and products and heat are formed. The time variable is introduced and rate of change of concentration of reactants or products with respect to time is followed. A good introduction into chemical kinetics can be found in Espenson (1995) or Upahyay (2006). Atmospheric Chemistry is discussed in detail for example in Seinfeld & Pandis (1998).

5.3.1 Rate of Reaction

The rate of reaction is the rate at which the reactants are transformed into the products of the reaction. It is the change in the concentration of the reactant and product that occurs during a given period of time:

$$\text{Rate} = -\frac{d[\text{Reactant}]}{dt} = \frac{d[\text{Product}]}{dt} \quad (5.7)$$

Here, [Reactant] and [Product] are the molar concentrations of reactant and product, respectively (Upahyay 2006).

5.3.2 Rate Law

A rate law is a mathematical equation that describes the progress of the reaction. There are two forms of a rate law for chemical kinetics: the differential rate law and the integrated rate law.

The differential rate law links the rate of reaction to the concentrations of the various species in the system. Differential rate laws can take on many different forms, especially for complicated chemical reactions. Each rate law contains a constant, k , called the rate constant which determines the speed of reaction given the reactant concentrations (Espenson 1995).

Rate Law for Zero-Order Reaction

A reaction is known as zero-order reaction when the rate of reaction is independent of the concentration of materials. The rate of reaction k is a constant. When the limiting reactant is completely consumed, the reaction stops abruptly.

The rate law for the general reaction



can be written as

$$\text{Rate} = -\frac{d[A]}{dt} = k. \quad (5.9)$$

Rate Law for First-Order Reaction

A first-order reaction depends on the concentration of only one reactant and can be represented in the general form



with the rate law

$$\text{Rate} = -\frac{d[A]}{dt} = k[A]. \quad (5.11)$$

Rate Law for Second-Order Reaction

A second-order reaction depends on the concentrations of one second-order reactant, or two first-order reactants.

The rate law for the general reaction

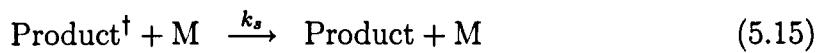


is

$$\text{Rate} = -\frac{d[A]}{dt} = k[A][B]. \quad (5.13)$$

5.3.3 Pressure Dependence of Reactions

Some reactions have an order which changes with pressure: they are third-order at low pressure and second-order at high pressure (Seinfeld & Pandis 1998). The rate of formation of a product in the general system



is

$$\frac{d[\text{Product}]}{dt} = \frac{k_a k_s [A][B][M]}{k_s[M] + k_r} \quad (5.16)$$

with M a third body and \dagger denotes vibrational excitation of the product (Seinfeld & Pandis 1998).

If $k_r \gg k_s[M]$ then the reaction is third-order

$$\frac{d[\text{Product}]}{dt} = \frac{k_a k_s}{k_r} [A][B][M] \quad (5.17)$$

and if $k_r \ll k_s[M]$ then the reaction is second-order

$$\frac{d[\text{Product}]}{dt} = k_a [A][B]. \quad (5.18)$$

5.3.4 Temperature Effect - Arrhenius Equation

The rate constant k introduced with the rate law in the previous section is often not constant but temperature dependent. It is well-known that raising the temperature of reactants can increase the reaction rate. This relationship between the rate a reaction proceeds and its temperature is determined by the Arrhenius equation, which was first proposed by van't Hoff and then extended by Arrhenius to:

$$k = AT^n e^{-\frac{E}{RT}} \quad (5.19)$$

where A is a constant (known as frequency factor, Arrhenius factor or pre-exponential factor), E is the activation energy, T the temperature, n a dimensionless exponent of temperature and R the gas constant (Upahyay 2006).

5.3.5 Photolysis

In atmospheric photochemistry the photon with energy $h\nu$ (where ν is the photons frequency) is a reactant in a chemical reaction (Seinfeld & Pandis 1998). For example the photolysis of NO_2 is given by:



The chemical compound is broken down by photons. The photon must have enough energy to break the bond. The photon's energy depends on its wave length and the shorter the wave length the higher its energy.

The rate constants in an atmospheric photolysis reaction are not temperature dependent, however they depend on the solar zenith angle (angle between the local zenith and the line of light to the sun). The variation of the photolysis rate with so-

lar zenith angle z can be described by an expression of the following form (Saunders et al. 2003):

$$J = l(\cos(z))^m \exp(-n \sec(z)). \quad (5.21)$$

The coefficients l, m, n can be calculated for each process by fitting the J values to the appropriate functions of the solar zenith angle. Values for l, m, n can be found for example in Saunders et al. (2003).

5.3.6 Chemical Thermodynamics

Chemical thermodynamics is concerned with the heat exchange accompanying transformations, such as mixing, phase transitions and chemical reactions. This includes calculations of such quantities as the heat capacity, enthalpy and entropy. Typically a thermodynamic database within modelling packages such as CHEMKIN (Kee et al. 1985) supply this type of information as NASA polynomial expressions comprising 14 coefficients (Gordon & McBride 1971, Burcat 1984, Burcat & McBride 1997, Burcat & Ruscic 2005) fitted for two different temperature ranges.

The NASA polynomials can be used to calculate the following functions for the heat capacity $C_p^0(T)$, the enthalpy $H^0(T)$ and the entropy $S^0(T)$:

$$\frac{C_p^0(T)}{R} = a_1 + a_2 T + a_3 T^2 + a_4 T^3 + a_5 T^4 \quad (5.22)$$

$$\frac{H^0(T)}{RT} = a_1 + a_2 \frac{T}{2} + a_3 \frac{T^2}{3} + a_4 \frac{T^3}{4} + a_5 \frac{T^4}{5} + \frac{a_6}{T} \quad (5.23)$$

$$\frac{S^0(T)}{R} = a_1 \ln T + a_2 T + a_3 \frac{T^2}{2} + a_4 \frac{T^3}{3} + a_5 \frac{T^4}{4} + a_7 \quad (5.24)$$

5.3.7 Summary

Chemical mechanisms as they are used in models for environmental and safety engineering are made up of complex couplings between species via reactions of the types introduced here which can lead to complicated non-linearities. The difficulty is usually in assigning values for the rate constants and photolysis rates because sometimes a wide range of values can be found in the literature. However, in other cases they have to be estimated because no studies are available. This may result in large uncertainty ranges for these parameters which have to be included when investigating the chemical mechanism. An atmospheric chemical mechanism comprising four species and eight reactions is introduced and investigated in the following section.

5.4 Case Study: Reactive Plume Model

An accurate prediction of concentrations and concentration fluctuations dispersing from a variety of sources such as traffic or fossil fuel burning power plants is fundamental to applications as for example the formation of secondary pollutants. In order to be able to predict concentrations of reactive pollutants in the atmosphere, models have to be developed which contain complex physical and chemical processes. The dispersion of plumes in urban areas is complicated because of the local topographical structures (e.g. densely packed buildings) which results in complex turbulent wind flows that can be difficult to model. Consequently, modelling the fluctuations in an urban area requires an accurate description of the in-street flow as well as a dispersion model that can account for the widespread inhomogeneities (Dixon & Tomlin 2007). The simulation of the turbulent flow field in an urban area (in particular a 3D street canyon) has already been investigated in section 4.3. A Lagrangian particle modelling framework is chosen here for the dispersion model. Due to the

lack of available data on secondary pollutants in simplified urban environments, it will be tested first for a reacting plume scenario where wind tunnel data has been obtained. The longer term plan will be to extend the framework to models of urban areas through linking with MISKAM.

In this case study a combined Lagrangian stochastic model with micro mixing and a chemical sub-model is used to investigate a reactive plume of nitrogen oxides (NO_x) released into a turbulent grid flow doped with ozone (O_3). Sensitivities to the model input parameters are explored for high NO_x model scenarios. A wind tunnel experiment is used to provide the simulation conditions for the first case study where photolysis reactions are not included and the main uncertainties occur in the parameters defining the turbulence scales, the source size σ and the reaction rate of NO (nitric oxide) with O_3 . At the next stage photolysis reactions are included in a chemical scheme consisting of eight reactions between species NO, O, O_3 and NO_2 . The extended RS-HMDR method is used to investigate the effects of uncertainties in the various model inputs resulting from the parameterisation of important physical and chemical processes in the reactive plume model, on the simulation of primary and secondary chemical species concentrations. Both independent and interactive effects of the parameters are studied. In total 22 parameters are assumed to be uncertain, among them the turbulence parameters, temperature dependant rate parameters, photolysis rates, temperature, fraction of NO in total NO_x at the source and background concentration of O_3 . The results of this case study were also presented at the 12th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO) (Ziehn et al. 2008a).

5.4.1 Introduction

Lagrangian stochastic models are now widely used to model mean concentrations and concentration fluctuations in atmospheric plumes (Franzese 2003, Sawford 2004, Cassiani et al. 2005). The theoretical basis and a review of Lagrangian stochastic models can be found in Wilson & Sawford (1996) In this study a combined Lagrangian stochastic model with a micro mixing sub-model (Dixon & Tomlin 2007) is used. This model was not only designed for open terrain simulations but also for predicting concentration fluctuations in urban areas. It is based on the marked particle model which uses the formulation of Thomson (1987) for inhomogeneous turbulent flows. The increments of velocity u_i and position x_i of each particle can be written as:

$$du_i = a_i(\mathbf{x}, \mathbf{u}, t)dt + b_{ij}(\mathbf{x}, \mathbf{u}, t)d\omega_j \quad (5.25)$$

$$dx_i = u_i dt \quad (5.26)$$

where a_i and b_{ij} are the drift and diffusion terms which are functions of position \mathbf{x} and turbulent velocity \mathbf{u} , dt is the time step and $d\omega_j$ are components of Gaussian white noise. The diffusion term b_{ij} is set to

$$b_{ij} = \sqrt{c_0 \epsilon} \delta_{ij} \quad (5.27)$$

where c_0 is a constant (structure function coefficient), δ_{ij} the Kronecker delta and ϵ the dissipation rate of turbulent kinetic energy k . The simplest and most common micro-mixing model is the Interaction by Exchange with the Mean (IEM) model, which is included here to calculate the evolution of the particle's concentration:

$$\frac{dC}{dt} = -\frac{1}{t_m}(C - \langle C \rangle) \quad (5.28)$$

where t_m is the mixing time scale and $\langle C \rangle$ is the local ensemble mean. In order to close the model the mixing time scale t_m has to be specified and is related to the turbulence time scale using

$$t_m = \alpha \frac{k}{\epsilon} \quad (5.29)$$

where α is a constant (mixing time scale coefficient). The turbulent flow field (k , ϵ and velocity) can be obtained for example from a MISKAM simulation. The model has been tested for a passive tracer against the open terrain wind tunnel data of Fackrell & Robins (1982) where the mean concentration and fluctuation intensity could be reproduced (Dixon & Tomlin 2007). The model was also coupled with the urban flow model MISKAM (Eichhorn 1996) and compared with the wind tunnel simulations of a two dimensional street canyon by Pavageau & Schatzmann (1999), also showing good agreement (Dixon & Tomlin 2007). A wider application of the model will depend on its reliability in different situations. One would hope that the turbulence parameters would not lead to large uncertainties in simulated concentrations. This has not been tested yet and moreover there is some argument in the literature about the values the turbulence parameters should adopt, which will be further discussed in section 5.4.2. Another concern is the application of the simple micro-mixing model. It is known that using the IEM model the mean concentration field is influenced by the micro-mixing (Cassiani et al. 2005), which is a significant drawback. However, the IEM model is computationally less demanding and useful for a large variety of applications. Global sensitivity analysis will help to investigate how strongly the simulated concentrations are effected by the mixing time scale when using the simple IEM model.

The Lagrangian stochastic model by Dixon & Tomlin (2007) can also be applied for reacting gas concentrations. Particles are initially placed randomly throughout the domain, with concentration assigned only to those particles that pass through the

source. The mixing model is then used to simulate the spread of concentration between neighbouring particles which is crucial to accurately predict the concentration fluctuations. Operator splitting is used so that after each flow and mixing time-step a chemical reaction step is performed, in this case using a stiff equation solver (a Rosenbrock method). The chemical rate equations and the Jacobian matrix have to be provided in the code.

It is important to investigate the combined effects of important physical and chemical parameters, because the simulated concentrations downwind from the source are influenced by both dispersion (atmospheric turbulence) and chemical reactions. Limited sensitivity studies exploring the impact of the turbulent parameters on passive plume dispersion were carried out in Dixon & Tomlin (2007) but only for a selected ensemble of simulations. The current work extends this previous analysis by applying the model to a reactive plume scenario and by extending sensitivity studies to the use of global sensitivity methods which can explore large regions of the input parameter space and the combined effects of physical and chemical parameters.

5.4.2 Model Validation and Uncertainty Analysis

The wind tunnel experiment by Brown & Bilger (1996) is used to provide the first case study for a reacting plume. The nominal mean axial velocity of the flow is $\bar{U} = 0.5 \text{ m s}^{-1}$. The flow field is doped with O₃ (background [O₃] = $2.45 \cdot 10^{13} \text{ molecule cm}^{-3}$) and well mixed. The velocity of the point source is matched to \bar{U} and consists of N₂ doped with NO (source [NO] = $1.26 \cdot 10^{16} \text{ molecule cm}^{-3}$). The reactants undergo the reaction NO + O₃ → NO₂ + O₂. Photolysis reactions are not included due to the absence of ultraviolet light. Using the nominal values for the source size ($\sigma = 0.008 \text{ m}$, assuming a Gaussian distribution), the mixing time scale coefficient ($\alpha = 0.75$) and the structure function coefficient ($c_0 = 5$) obtained from

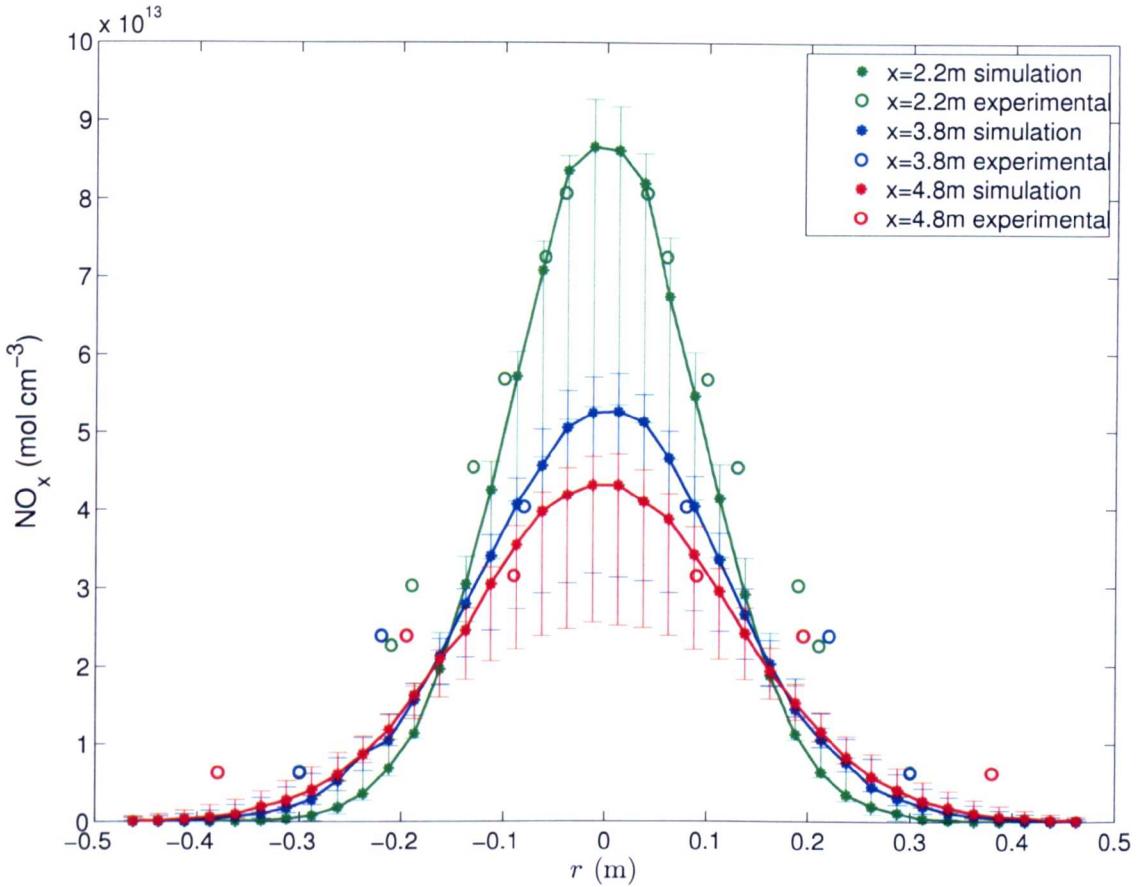


Figure 5.2: Radial profile of mean $[\text{NO}_x]$ at $x = 2.2 \text{ m}$, $x = 3.8 \text{ m}$ and $x = 4.8 \text{ m}$ from the point source obtained from simulations using the nominal values for σ , α and c_0 and from wind tunnel experiments. Error bars are based on 400 simulations varying σ , α and c_0 in their assigned ranges.

Dixon & Tomlin (2007), the model gives a good representation of the radial profile of the concentration of conserved scalar $\text{NO}_x = \text{NO} + \text{NO}_2$ compared to the wind tunnel experiment. However, the simulated profile is slightly narrower especially at longer distances from the point source as shown in figure 5.2. The results from the simulations show a very good match with the experimental data around the plume centre for both radial profiles of NO_x .

Error bars are also shown in figure 5.2 based on 400 simulations, where c_0 has been varied between $[3 \dots 6]$, α between $[0.6 \dots 3]$ and σ between $[0.008 \dots 0.016] \text{ m}$.

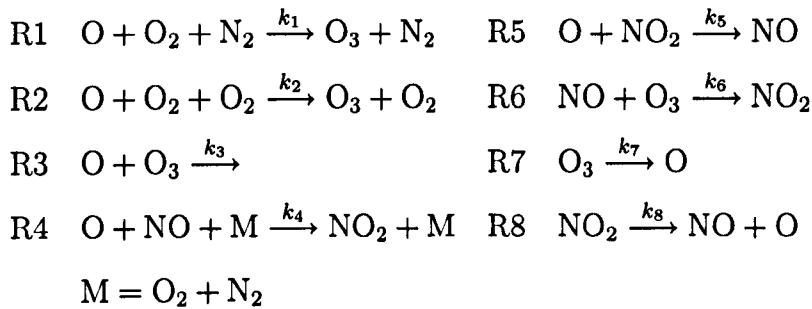
c_0	Reference
4.3 ± 0.3	Rizza et al. (2006)
3.0 ± 0.5	Du (1997)
5.5, 6.4	Lien & D'Asaro (2002)
5.0 ± 0.5	Reynolds (1998)

Table 5.4: Range of values for the structure function coefficient c_0 as found in the literature.

The structure function coefficient c_0 determines the effective turbulent diffusion in velocity space, and plays an important role in Lagrangian dispersion modelling. As pointed out in Rizza et al. (2006) a wide range of values are found in the literature, which have been determined in different ways and using different assumptions. The range for c_0 used here is estimated on the basis of a number of earlier studies such as Rizza et al. (2006), Lien & D'Asaro (2002) and Du (1997) (see also table 5.4). The lower value of three and the proposed universality of c_0 suggested in Du (1997) were questioned by Reynolds (1998) and therefore it is of interest to explore the impact of uncertainties in this parameter. The possible range for α describing the mixing time scale is based on Cassiani et al. (2005) who reviewed the range of values used in the literature in table 1 of their paper. The range for σ is estimated to represent the uncertainty in the effect of near field mixing on the initial source width. The error bars in figure 5.2 therefore represent the limits of uncertainty in the predictions of NO_x profiles based on the ranges of previous estimates for the turbulence constants. Figure 5.2 shows that although the peak values are represented we do not recover the experimental values at the plume edge for the radial NO_x profiles. The figure also shows that the nominal values selected give a good representation of the experimental data but that if broad input ranges are considered then the range of predictions is quite large. This is not a desirable feature and it indicates a high degree of sensitivity to the input parameters which will be explored further using sensitivity analysis in the following section.

5.4.3 Extended Chemical Mechanism

At the next stage the chemical reaction scheme has been extended in order to provide a more realistic atmospheric scenario. Eight reactions between the species NO, O, O₃ and NO₂ have been included and photolysis has also been considered for O₃ and NO₂:



A concentration of $[\text{N}_2] = 2.03 \cdot 10^{19}$ molecule cm⁻³ and $[\text{O}_2] = 5.45 \cdot 10^{18}$ molecule cm⁻³ has been used. The reaction scheme leads to the following four rate equations which are implemented in the Lagrangian stochastic model and solved by a Rosenbrock (Rosenbrock 1963) stiff solver :

$$\frac{d[\text{NO}]}{dt} = -k_4[\text{O}][\text{NO}] + k_5[\text{O}][\text{NO}_2] - k_6[\text{NO}][\text{O}_3] + k_8[\text{NO}_2] \quad (5.30)$$

$$\begin{aligned} \frac{d[\text{O}]}{dt} = & -k_1[\text{O}] - k_2[\text{O}] - k_3[\text{O}][\text{O}_3] - k_4[\text{O}][\text{NO}] - k_5[\text{O}][\text{NO}] \dots \\ & + k_7[\text{O}_3] + k_8[\text{NO}_2] \end{aligned} \quad (5.31)$$

$$\frac{d[\text{O}_3]}{dt} = k_1[\text{O}] + k_2[\text{O}] - k_3[\text{O}][\text{O}_3] - k_6[\text{NO}][\text{O}_3] - k_7[\text{O}_3] \quad (5.32)$$

$$\frac{d[\text{NO}_2]}{dt} = k_4[\text{O}][\text{NO}] - k_5[\text{O}][\text{NO}_2] + k_6[\text{NO}][\text{O}_3] - k_8[\text{NO}_2] \quad (5.33)$$

The physical and chemical processes in the reactive plume are then investigated in order to study the effects of possible uncertainties in the parameters. In total 22 parameters are assumed to be uncertain, among them c_0 , α and σ as above, as

well as the chemical reaction rate parameters for each reaction (*A*-factor, activation energy *E* and temperature coefficient for Arrhenius parameters or photolysis rate), temperature, background [O₃] and fraction of NO in total NO_x at the source. According to the Environment Agency¹ combustion processes emit a mixture of nitric oxide NO (90 %) and nitrogen dioxide NO₂ (10 %). This ratio is questioned for example by Carslaw et al. (2007). It is therefore of interest to investigate the effects of uncertainties in the NO fraction.

The ranges for the uncertain parameters were defined according to a minimum and maximum value, assuming equal probability throughout the ranges. The maximum and minimum values for the Arrhenius parameters are in accordance with Atkinson et al. (2004). The uncertainty ranges for the photolysis rates are obtained from Shetter et al. (1992) and Shetter et al. (1996) who used a chemical actinometer for the measurement. Uncertainties for the temperature were assumed to be within the range of [273 ... 293]K. The range for NO fraction in the source (initial concentration [NO] = $1.26 \cdot 10^{16}$ molecule cm⁻³) was varied from 100 % to 80 % in order to assess the impact of uncertainties of primary NO₂ in NO_x sources which is also of relevance to transport related emissions of NO_x (Carslaw et al. 2007). The concentration for ozone has been varied in between [$7.350 \cdot 10^{11} \dots 1.225 \cdot 10^{12}$] molecule cm⁻³. The ozone concentration has been chosen to be lower than the one in the experiment of Brown & Bilger (1996) in order to provide a more realistic atmospheric scenario. An overview of the ranges for all 22 parameters can be found in table 5.5. The very high initial NO concentration has also been lowered to [NO] = $1.26 \cdot 10^{15}$ molecule cm⁻³ in a second scenario.

¹<http://www.environment-agency.gov.uk/>

Parameter	Nominal Value	Minimum	Maximum	Unit	Reference
Mixing time scale coefficient α	0.75	0.6	3	-	Cassiani et al. (2005)
Structure function coefficient c_0	5	3	6	-	Table 5.4
Source size σ	0.008	0.008	0.016	m	-
A factor reaction R1	$5.60 \cdot 10^{-34}$	$4.9910 \cdot 10^{-34}$	$6.2833 \cdot 10^{-34}$	$\frac{\text{cm}^6}{\text{molecule}^2 \text{s}}$	Atkinson et al. (2004)
A factor reaction R2	$6.00 \cdot 10^{-34}$	$5.3475 \cdot 10^{-34}$	$6.7321 \cdot 10^{-34}$	ll	Atkinson et al. (2004)
A factor reaction R3	$8.00 \cdot 10^{-12}$	$6.6541 \cdot 10^{-12}$	$9.6181 \cdot 10^{-12}$	$\frac{\text{cm}^3}{\text{molecule s}}$	Atkinson et al. (2004)
A factor reaction R4 (k_0)	$1.00 \cdot 10^{-31}$	$7.9433 \cdot 10^{-32}$	$1.2589 \cdot 10^{-31}$	$\frac{\text{cm}^6}{\text{molecule}^2 \text{s}}$	Atkinson et al. (2004)
A factor reaction R4 (k_∞)	$3.00 \cdot 10^{-11}$	$1.5036 \cdot 10^{-11}$	$5.9858 \cdot 10^{-11}$	$\frac{\text{cm}^3}{\text{molecule s}}$	Atkinson et al. (2004)
A factor reaction R5	$5.50 \cdot 10^{-12}$	$4.7903 \cdot 10^{-12}$	$6.3148 \cdot 10^{-12}$	ll	Atkinson et al. (2004)
A factor reaction R6	$1.40 \cdot 10^{-12}$	$1.1645 \cdot 10^{-12}$	$1.6832 \cdot 10^{-12}$	ll	Atkinson et al. (2004)
Photolysis rate J O ₃ (R7)	0	$1.0000 \cdot 10^{-5}$	$4.5000 \cdot 10^{-5}$	s ⁻¹	Shetter et al. (1996)
Photolysis rate J NO ₂ (R8)	0	0.004	0.011	s ⁻¹	Shetter et al. (1992)
Temperature	298	273	293	K	-
n for reaction R1	2.6	2.1	3.1	-	Atkinson et al. (2004)
n for reaction R2	2.6	2.1	3.1	-	Atkinson et al. (2004)
E/R for reaction R3	2060	1860	2260	K ⁻¹	Atkinson et al. (2004)
n_0 for reaction R4	1.6	1.3	1.9	-	Atkinson et al. (2004)
n_∞ for reaction R4	-0.3	-0.6	0	-	Atkinson et al. (2004)
E/R for reaction R5	-188	-268	-108	K ⁻¹	Atkinson et al. (2004)
E/R for reaction R6	1310	1110	1510	$\frac{\text{K}^{-1}}{\text{molecule cm}^3}$	Atkinson et al. (2004)
Background concentration for O ₃				-	
Fraction of NO in total NO _x	1	0.8	1	-	Carslaw et al. (2007)

Table 5.5: Physical and chemical parameter ranges for the Lagrangian stochastic model.

5.4.4 Results

The RS-HDMR method has been applied in order to perform global sensitivity analysis and to calculate variance based sensitivity indices for the 22 input parameters based on the uncertainty ranges described in section 5.4.3. The outputs of interest for the sensitivity analysis are the concentrations of the conserved scalar NO_x and O_3 at the plume centre at different distances x away from the point source. A quasi-random sequence of up to $N = 512$ has been used in connection with the HDMR analysis and for the calculation of the sensitivity indices. The maximum order for the approximation of the component functions was chosen to be ten for first-order and three for second-order. Additionally, a threshold of 1% has been applied in order to exclude unimportant component functions from the HDMR expansion.

N	r^2 value in % for output:					
	NO_x ($x = 2.2$)	NO_x ($x = 4.8$)	NO_x ($x = 5.8$)	O_3 ($x = 2.2$)	O_3 ($x = 4.8$)	O_3 ($x = 5.8$)
128						
1st-order	94.26	94.94	94.06	95.39	93.26	93.86
2nd-order	93.41	94.89	93.93	95.19	90.84	90.81
256						
1st-order	98.49	98.44	98.65	96.81	96.18	95.56
2nd-order	98.50	98.32	98.63	96.81	96.18	95.12
512						
1st-order	98.55	98.93	98.98	97.33	96.73	96.54
2nd-order	98.56	98.91	98.98	97.33	96.73	96.46

Table 5.6: Accuracy of the RS-HDMR metamodel (first- and second-order) in terms of the coefficient of determination r^2 for all considered model outputs (peak NO_x and O_3 concentrations at different distances (x in m) away from the source) using different sample sizes N .

The accuracy of the results for the sensitivity analysis also depends on the set up of the Lagrangian stochastic model. Various numbers of particles (50 000 up to 1 000 000) and different simulation end times (200 – 1 200 s) have been tested and the mean concentrations after each time step were recorded. It was found that

a large number of particles (1 000 000) has to be used in connection with a long simulation end time of $t = 1\,200\text{ s}$ (12 000 time steps) to ensure that the mean concentrations reach convergence. This is definitely a drawback of the applied Lagrangian stochastic model, because the computational effort (run time of the model) increases significantly if using such a large number of particles and time steps. One simulation of the Lagrangian stochastic model using 1 000 000 particles takes about 20 hours on a 3 GHz PC. Therefore, the simulations had to be performed on the high performance computer facilities offered by the White Rose Grid ² at the University of Leeds where up to 30 simulations could be run in parallel.

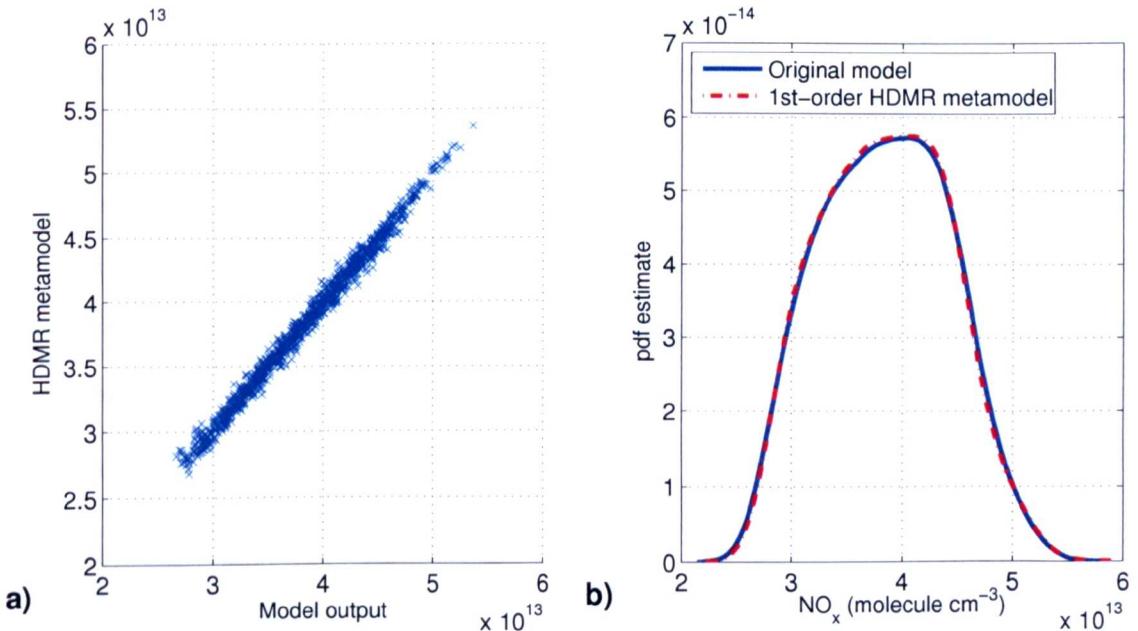


Figure 5.3: Accuracy of the RS-HDMR metamodel based on $N = 512$ model runs for $[\text{NO}_x]$ at the plume centre ($x = 4.8\text{ m}$) showing (a) the scatter plot between the original model response and the first-order RS-HDMR metamodel ($r^2 = 98.93\%$) and (b) the probability density function estimate for both the original model and the first-order RS-HDMR metamodel.

Due to the large computational requirements, a small sample size of $N = 128$ has been applied first in order to construct the HDMR metamodel. The accuracy of the

²<http://www.wrgrid.org.uk>

metamodel can be expressed by calculating the square of the Pearson correlation coefficient (r^2 value, see section 3.3.4) for the predicted output and the model output. As it can be seen from table 5.6 a small sample size of $N = 128$ is already enough to generate a reasonably accurate metamodel indicated by the r^2 value which is for all considered outputs greater than 93 %. An increase in the sample size N results in an even more accurate HDMR metamodel and a sample size of $N = 512$ seems to be sufficient for all considered outputs. Table 5.6 shows that there is hardly any difference in the accuracy if using a first- or second-order HDMR expansion. This indicates already that parameter interactions are negligible for all six investigated outputs. The accuracy of the constructed HDMR metamodel can also be shown with the help of a scatter plot or the probability density function (pdf) estimate as presented in figure 5.3. It is shown, that the data in the scatter plot in figure 5.3a form a straight line and the amount of scatter is quite small, indicating that a first-order HDMR metamodel approximates the original model very well. This is also confirmed by the plot of the pdf in figure 5.3b showing a nearly perfect match for both the original model and the metamodel.

Parameter	S_i ($x = 2.2 \text{ m}$)	S_i ($x = 4.8 \text{ m}$)	S_i ($x = 5.8 \text{ m}$)
Structure function coefficient c_0	0.7976	0.7941	0.7903
Mixing time scale coefficient α	0.1853	0.1918	0.1937
$\sum S_i$	0.9843	0.9874	0.9844

Table 5.7: First-order sensitivity indices for $[\text{NO}_x]$ at the plume centre at different distances x away from the point source using the RS-HDMR approach with a samples size of $N = 512$.

Table 5.7 presents the first-order sensitivity indices for $[\text{NO}_x]$. Since NO_x is a conserved tracer and the initial $[\text{NO}_x]$ is the same as in section 5.4.2, then these results effectively provide a sensitivity analysis of the variance expressed by the error bars in figure 5.2. No interactions between parameters (second-order effects) could be

found and only two parameters have an influence on the output variability. Hence, the HDMR expansion consists of only two first-order component functions.

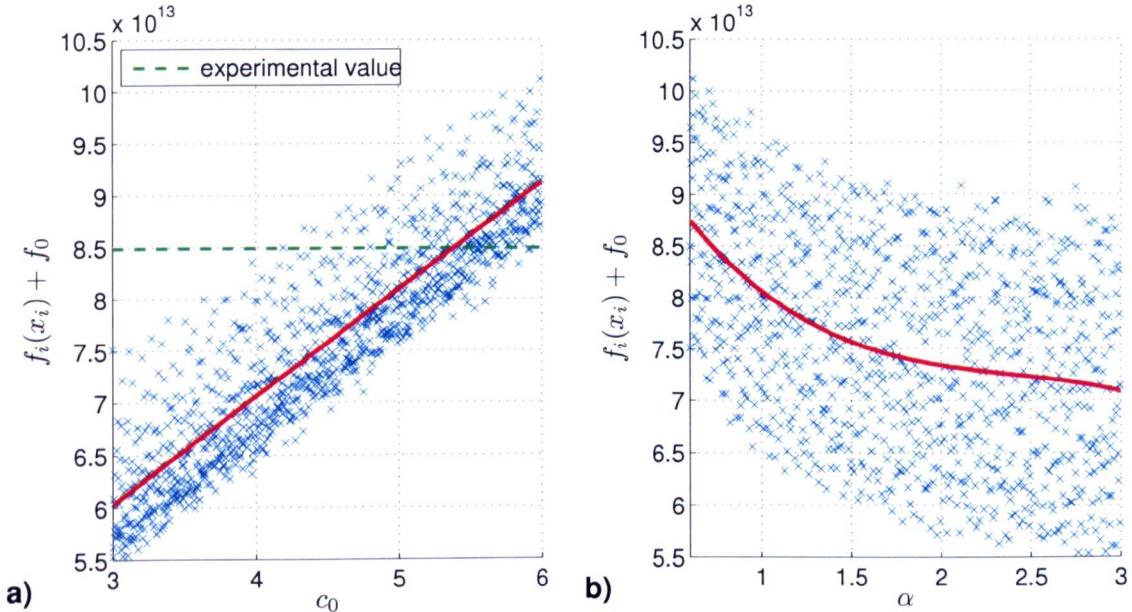


Figure 5.4: First-order component functions and scatter plots for (a) the structure function coefficient c_0 and (b) the mixing time scale coefficient α with respect to $[NO_x]$ at the plume centre for $x = 2.2\text{ m}$. The mean f_0 is added to f_i for comparison with the scatter plot.

The structure function coefficient c_0 is the most important parameter and is responsible for 80 % of the variance in peak $[NO_x]$. The mixing time scale coefficient α accounts for around 19 % of the overall variance. The sensitivity to the two parameters is similar for all three investigated distances x from the point source. There is a low sensitivity to the initial source size. HDMR component functions for each parameter allow the investigation of individual effects in more detail. The component functions for c_0 and α with respect to the NO_x concentration at the plume centre and for $x = 2.2\text{ m}$ are shown in figure 5.4. Scatter plots from a Monte Carlo simulation based on randomly selected input parameters are also shown for comparison. This highlights the advantage of the HDMR component functions, since they show directly the effect of each parameter on the selected output. It can be

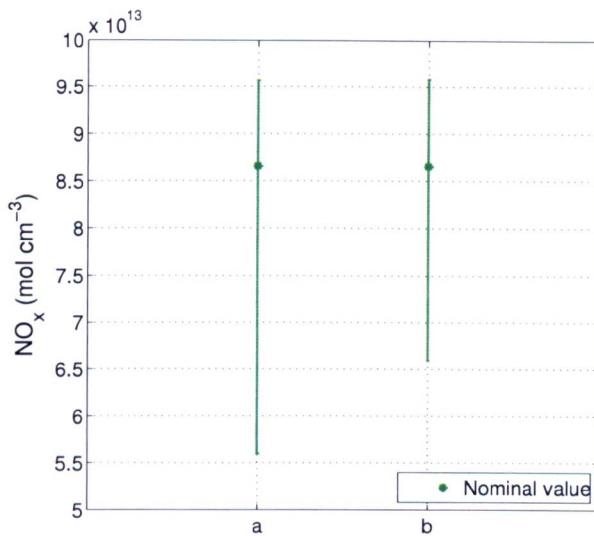


Figure 5.5: Error bars for peak $[NO_x]$ at 2.2 m from the point source using a: $c_0 = [3 \dots 6]$ and b: $c_0 = [4 \dots 6]$.

seen that a strong positive linear relationship exists between c_0 and $[NO_x]$, whereas the effect of α is negative and nonlinear. The experimental value of $[NO_x]$ at the plume centre is also highlighted in figure 5.4a and demonstrates that even taking into account uncertainties in other parameters, the experimental concentration is not recoverable with a value of c_0 smaller than four. This is in agreement with the studies by Reynolds (1998) who questioned a value as low as three for the structure function coefficient. With these findings it would make sense to narrow down the uncertainty range for c_0 to $[4 \dots 6]$ to see how this would effect the uncertainty in the model output. It can be assumed that reducing the uncertainty range for the most important parameter would also lead to a smaller uncertainty range for the $[NO_x]$ at the plume centre. This is confirmed by figure 5.5 showing the error bars for peak $[NO_x]$ at $x = 2.2$ m away from the source for both uncertainty ranges of c_0 . The model runs using the new uncertainty range can be performed using the HDMR metamodel. One advantage of the HDMR method is that a fully functional surrogate (here a first-order HDMR expansion is sufficient) is also constructed, which can

be used for further simulations and studies instead of the computationally expensive original model. This is very useful in this case, because the simulations with the original model are very time consuming (e.g. 20 hours for one run).

The mixing time scale coefficient α has a significant influence on peak $[NO_x]$ (see table 5.7). This indicates, that the IEM mixing model breaks one of the desirable properties of a mixing model, namely that the mean concentration should not be effected by the mixing model, although it contributes less than 20% to the overall variance.

Parameter	S_i ($x = 2.2$ m)	S_i ($x = 4.8$ m)	S_i ($x = 5.8$ m)
x_{20} E/R for $NO + O_3 \rightarrow NO_2$	0.4144 (1)	0.4381 (1)	0.4410 (1)
x_{12} Mixing time scale coefficient α	0.2971 (2)	0.2644 (2)	0.2538 (2)
x_{21} Background concentration for O_3	0.1058 (3)	0.0754 (4)	0.0721 (4)
x_{22} Fraction of NO in total $[NO_x]$	0.0590 (4)	0.0797 (3)	0.0859 (3)
x_{10} Temperature	0.0340 (5)	0.0363 (5)	0.0371 (5)
x_{11} Structure function coefficient c_0	0.0288 (6)	0.0357 (6)	0.0367 (6)
x_7 A -factor for $NO + O_3 \rightarrow NO_2$	0.0252 (7)	0.0263 (7)	0.0266 (7)
$\sum S_i$	0.9699	0.9651	0.9634

Table 5.8: First-order sensitivity indices for $[O_3]$ at the plume centre at different distances x away from the point source for very high initial NO concentrations using the RS-HDMR approach with a samples size of $N = 512$. The numbers in brackets indicate the ranking of the parameter.

Table 5.8 shows the first-order sensitivity indices for $[O_3]$ for the very high $[NO_x]$ scenario. In total seven parameters are responsible for the variance of the output. The most important parameter is the activation energy term E/R for the reaction $NO + O_3 \rightarrow NO_2$, which contributes more than 40% to the overall variance. Now that interactions between turbulent mixing and chemical reactions are included, the most important turbulence parameter becomes α which defines the mixing time scale. Parameter interactions were not found as shown by the total first-order sensitivity coefficients adding to almost one. The importance of the parameters, and

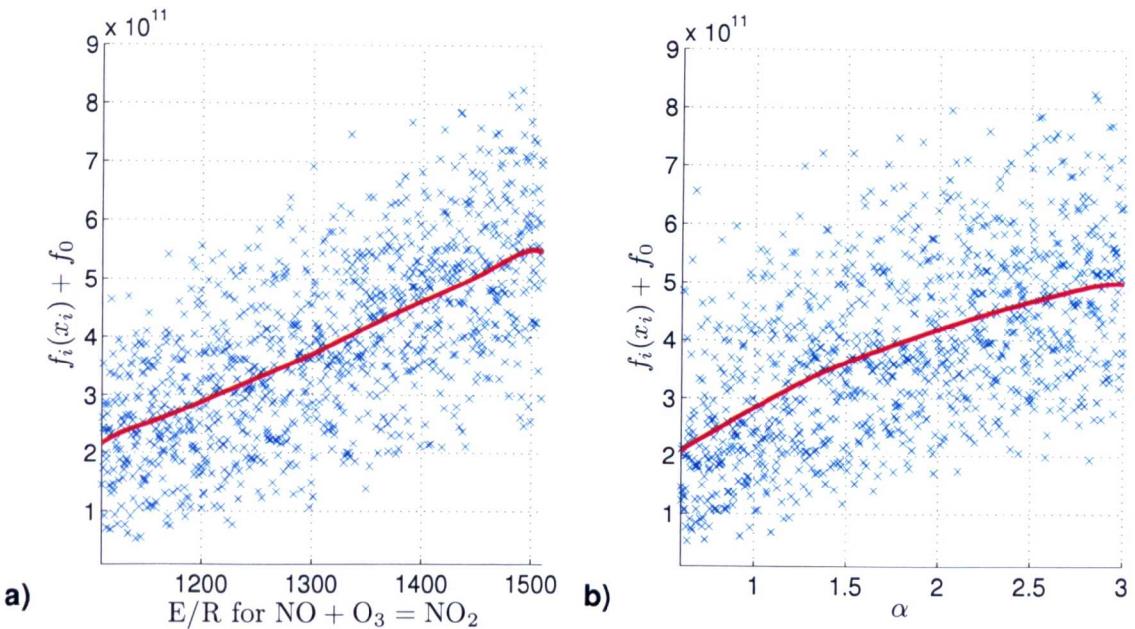


Figure 5.6: First-order component functions and scatter plots for (a) E/R for $\text{NO} + \text{O}_3 \rightarrow \text{NO}_2$ and (b) the mixing time scale coefficient α with respect to $[\text{O}_3]$ at the plume centre for $x = 2.2 \text{ m}$. The mean f_0 is added to f_i for comparison with the scatter plot.

even the ranking, changes with growing distance x from the source. For example, the fraction of NO in total $[\text{NO}_x]$ becomes more important at further distances x from the source and its rank changes from 4 to 3. The plot of the component functions for the two most important parameters is presented in figure 5.6. The relationship between E/R for the reaction $\text{NO} + \text{O}_3 \rightarrow \text{NO}_2$ and $[\text{O}_3]$ is nearly linear (positive) whereas the relationship between α and $[\text{O}_3]$ is fairly non-linear (positive).

Table 5.9 shows a comparison of the first-order sensitivity indices calculated based on a sample size of $N = 128$ and $N = 512$ for the O_3 concentration at the plume centre for the very high $[\text{NO}_x]$ scenario. It can be seen, that even if using a very small sample size the HDMR method is able to produce a reliable parameter ranking which is identical for both sample sizes. The individual values for the sensitivity indices differ quite a lot for some of the parameters for the two sample sizes and the sum of

x_i	First-order sensitivity indices S_i					
	$N = 128$		$N = 512$			
	($x = 2.2$)	($x = 4.8$)	($x = 5.8$)	($x = 2.2$)	($x = 4.8$)	($x = 5.8$)
x_{20}	0.3681	0.3506	0.4078	0.4144	0.4381	0.4410
x_{12}	0.2492	0.2152	0.2127	0.2971	0.2646	0.2538
x_{21}	0.0875	0.0579	0.0591	0.1058	0.0754	0.0721
x_{22}	0.0683	0.0914	0.1036	0.0590	0.0797	0.0859
x_{10}	0.0352	0.0332	0.0403	0.0340	0.0363	0.0371
x_{11}	0.0278	0.0317	0.0310	0.0288	0.0357	0.0367
x_7	0.0253	0.0272	0.0182	0.0252	0.0263	0.0266
$\sum S_i$	0.8677	0.8185	0.8787	0.9699	0.9651	0.9634

Table 5.9: First-order sensitivity indices for $[O_3]$ at the plume centre at different distances x away from the point source using the RS-HDMR approach with a samples size of $N = 128$ and $N = 512$.

Parameter	S_i ($x = 2.2$ m)	S_i ($x = 4.8$ m)	S_i ($x = 5.8$ m)
x_{21} Background concentration for O_3	0.6943 (1)	0.6097 (1)	0.5942 (1)
x_{20} E/R for $NO + O_3 \rightarrow NO_2$	0.1373 (2)	0.1838 (2)	0.1941 (2)
x_{12} Mixing time scale coefficient α	0.0905 (3)	0.1097 (3)	0.1110 (3)
x_{11} Structure function coefficient c_0	0.0132 (4)	0.0205 (4)	0.0224 (4)
x_{22} Fraction of NO in total $[NO_x]$	0.0111 (5)	0.0154 (5)	0.0165 (5)
x_{10} Temperature	0.0109 (6)	0.0145 (6)	0.0153 (6)
x_7 A factor for $NO + O_3 \rightarrow NO_2$	0.0089 (7)	0.0121 (7)	0.0127 (7)
$\sum S_i$	0.9742	0.9720	0.9725

Table 5.10: First-order sensitivity indices for $[O_3]$ at the plume centre at different distances x away from the point source for lower initial NO concentration using the RS-HDMR approach with a samples size of $N = 512$. The numbers in brackets indicate the ranking of the parameter.

the first-order sensitivity indices adds up only to less than 90 %. However, one can already identify the most important parameters with only reasonable computational effort.

In a second scenario the very high initial NO concentration has been lowered to $[NO] = 1.26 \cdot 10^{15}$ molecule cm $^{-3}$ and the results are presented in table 5.10 in form of the first-order sensitivity indices for $[O_3]$. Now the background concentration

for O_3 becomes the most important parameter, contributing by around 60 % to the overall variance. The second most important parameter is the activation energy term E/R for the reaction $NO + O_3 \longrightarrow NO_2$, which had been the most important parameter in the very high $[NO_x]$ scenario. The importance of the mixing time scale coefficient α is now much smaller and contributes only by around 10 %.

5.4.5 Conclusions

The variance in the mean concentration of the conserved scalar $[NO_x]$ at the plume centre is as expected only influenced by the turbulence parameters. The sensitivity to those parameters is constant throughout the investigated distances x from the point source with the structure function coefficient dominating the output variance. This is a significant result because there is considerable lack of agreement in the literature about the universality and value of this parameter. It could be shown that the mixing time scale coefficient α has a lower but still significant influence on the mean concentrations. This is clearly a disadvantage of using the simple IEM model as a mixing model. The application of the HDMR method has also suggested that the uncertainty range for the structure function coefficient c_0 has originally been chosen to be too wide. A comparison with experimental values (wind tunnel experiment by Brown & Bilger (1996)) has shown that the uncertainty range for c_0 can be reduced, because low values for c_0 do not allow the recovery of the experimental values. This has also highlighted the advantages of using the HDMR approach. On the one hand the effects of the uncertain parameters can be investigated over a large range (plot of the first-order HDMR component function). On the other hand new uncertainty ranges (as long as they are within the original range) can be assigned and the effects investigated without performing new model runs. This is because the HDMR metamodel can also be used as a surrogate and

a large number of simulations based on the HDMR metamodel can be performed almost instantly. The computational running time of the original model is quite high (e.g. 20 hours for one simulation) because a large number of particles (1 000 000) had to be used in order to reach convergence for the mean concentrations.

The variance in mean $[O_3]$ is influenced by both physical and chemical parameters. A total of seven parameters contribute independently to the overall variance with the mixing time scale coefficient now being the dominant turbulence parameter and the activation energy for the reaction $NO + O_3 \longrightarrow NO_2$ dominating the kinetic sensitivities for the very high $[NO_x]$ scenario. If the initial NO concentration is chosen to be lower, then the background concentration for O_3 becomes the most important parameter. The influence of mixing time scale coefficient is even smaller, which is a positive result and demonstrates that the simple IEM mixing model is sufficient in this case. Parameter interactions could not be identified and therefore a first-order HDMR expansion can be used to calculate the sensitivity indices. It has been shown that the HDMR approach is very efficient and in this case study 128 model runs were sufficient in order to identify the most important parameters and their ranking. A total of $N = 512$ model runs have been performed in order to construct an even more accurate HDMR metamodel which results in more accurate sensitivity indices. However, this did not change the importance ranking of the parameters identified already using only 128 model runs.

5.5 Case Study: Premixed Methane Flame Model

The effectiveness of the HDMR tools is demonstrated for a case study of a one dimensional low pressure premixed methane flame model doped with trace sulphur and nitrogen species. Uncertainties in rate constants and thermodynamic data for the SO_x extension are considered, leading to a study of 176 input parameters. Us-

ing the HDMR tools, the use of screening methods such as the Morris method (see section 2.3.2), which aim to identify unimportant parameters beforehand, can generally be avoided. However, in certain cases, a combination of a screening method and HDMR is computationally more efficient than using HDMR alone. Due to the non-linearity of the model, three different scenarios are investigated in order to highlight the importance of the uncertainty ranges chosen. The results have also been presented at the 5th International Conference on Sensitivity Analysis of Model Output (SAMO) (Ziehn & Tomlin 2007*a*) and been published in Ziehn & Tomlin (2008*a*)

5.5.1 Introduction

The use of complex chemical mechanisms is increasing within models describing a range of important chemical processes including combustion. Parameters describing the rates of chemical reactions and thermodynamics within such mechanisms can often be very uncertain. Highlighting the main parameters contributing to predictive uncertainty of the resulting models is therefore an important part of model development. It can help to focus kinetic studies on key parameters that affect the accuracy of the model with respect to validation data. Combustion models provide a challenge for uncertainty analysis since they usually contain a large number of uncertain parameters, many of which are derived from a limited number of experimental or theoretical studies, or are estimated based on structural additivity relationships by considering similar functional groups (Benson 1976). The uncertainty ranges can therefore be quite large for many parameters, and in addition, there is often a lack of consensus within the literature as to accepted nominal values as well as uncertainty ranges. This calls into question the use of local sensitivity information such as is usually available within commercial modelling packages (e.g. CHEMKIN (Kee et al. 1985)), since these give information in a small region of parameter space

close to the suggested nominal values. The exploration of wider input parameter ranges requires global sensitivity methods. However, the use of traditional methods for global uncertainty and sensitivity analysis such as Monte Carlo simulations becomes problematic due to the computational expense of the model and the difficulty in interpreting the results (e.g. scatter plots) for large parameter sets.

If the input space dimension is large then traditionally a screening method such as the Morris method (Morris 1991) is applied first, in order to identify unimportant parameters. In a second step a sampling or variance based method can be applied using only the potentially important parameters. The current study will investigate whether the extended HDMR method (HDMR tools including the optimisation approach and the threshold) is able to handle a large input space dimension without the need to apply a screening method beforehand.

The case study chosen to demonstrate the methods is a one dimensional low pressure premixed methane flame model doped with trace sulphur and nitrogen containing species and modelled using the CHEMKIN package PREMIX (Kee et al. 1991). Premixed flames are the simplest flames and a common method for mechanism evaluation since they can be modelled using a 1D reaction transport model. A range of experimental data sets exist for various problem types. If species profiles have been obtained in experiments then 1D flames provide a useful way of exploring detailed high temperature chemistry and thermodynamics and transport properties within the mechanisms.

It is known that trace amounts of sulphur in the fuel can have an impact on the extent of nitrogen oxide emissions (Hampartsoumian et al. 1998, Hughes et al. 2001a, Nimmo et al. 1998). Therefore suitable models describing the interaction of sulphur containing compounds with other species within flames are required. Sulphur chemistry is however, quite poorly understood at high temperatures, with

many reaction rates and thermochemical parameters being estimated in previous studies. An assessment of the resulting uncertainty in predictions of nitrogen oxide emissions is therefore of interest. The model also provides an interesting case study for the application of global sensitivity methods, since there is a lack of consensus in the literature as to the nominal values and uncertainty ranges for certain of its input parameters (Tomlin 2006). It therefore serves to highlight the importance of using global methods for the investigation of non-linear combustion models. Three model scenarios using different parameter uncertainty ranges are presented in order to explore this point.

5.5.2 Model Description

The mechanism under investigation is a sulphur chemistry scheme, which is linked to an updated version of the Leeds methane/NO_x mechanism (Hughes et al. 2001c, Hughes et al. 2001b). More detailed information can also be found on the University of Leeds website (<http://www.chem.leeds.ac.uk/Combustion/Combustion.html>). The focus is on the influence of sulphur containing compounds on the formation of nitrogen oxides (Hampartsoumian et al. 1998). The SO_x extension, is basically the same as used in Tomlin (2006), which was based on the scheme of Alzueta et al. (2001) and Glarborg et al. (1996). A total of 176 uncertain parameters are considered within the sulphur scheme, 153 of which are *A*-factors for the forward reaction rates of each elementary reaction, and 23 are enthalpies of formation (ΔH_f) which are calculated using NASA polynomials (see section 5.3.6). The reaction rates are expressed in standard Arrhenius form using units of cm³ molecule⁻¹ s⁻¹ for second-order rate constants. Reactions are treated as reversible, with reverse rates calculated from the appropriate equilibrium constants. In the current study all parameters are varied at the same time (or one at a time in case of the Mor-

ris method) in contrast to the earlier study of Tomlin (2006) which considered the rate parameters and thermodynamic parameters separately. The potential for interactions between the rate and thermodynamic parameters for the SO_x extension is therefore explored in this study.

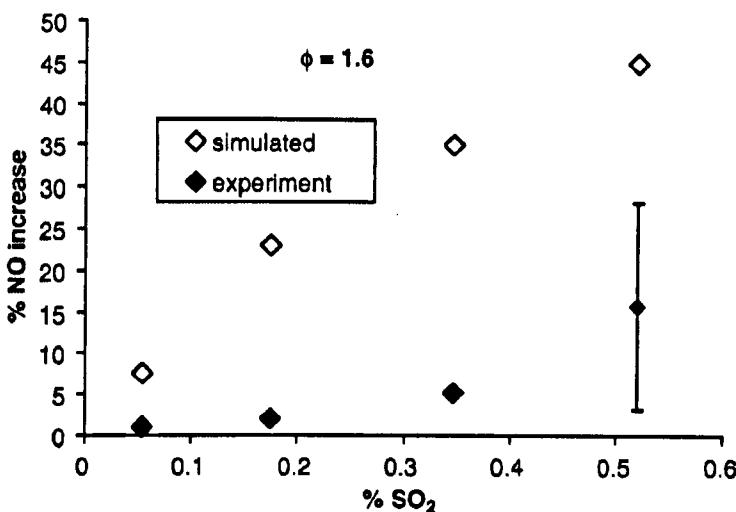


Figure 5.7: Comparison between model simulations and experimental measurements for the relative change in NO concentration on the addition of varying amounts of SO₂ to the fuel (Tomlin 2006).

For the modelling process the CHEMKIN package (Kee et al. 1985) is applied, which is commonly used for a range of applications including chemical mechanism validation in simplified flow environments. The simulation has been performed for a one dimensional low pressure flame (40 Torr) using PREMIX (Kee et al. 1991). Full details of the model scenario can be found in Tomlin (2006) but it should be pointed out that in this study we focus on the fuel-rich scenario with a flame stoichiometry of $\phi = 1.6$ and 0.5 % of SO₂ and 1.3 % of NH₃ added to the flame. In this case, the mechanism using the nominal parameter values tends to over predict the relative increase in NO mole fractions at the end of the flame on the addition of SO₂, compared to the experiments of Hughes et al. (2001a) as shown in figure 5.7. The final NO mole fraction within the burnt gas region is considered as the target

output variable within the following sensitivity studies. We are therefore exploring the main sources of uncertainty in predicting the final NO concentrations in the flame.

5.5.3 Investigated Scenarios and Experimental Setup

The output of interest for the methane/NO_x mechanism including the SO_x extension is the NO mole fraction in the burnt gas region of the flame. The uncertainty ranges for the 153 *A*-factors were taken from data evaluations where possible (Atkinson et al. 1997) as explained in Tomlin (2006). Unfortunately, this was the case for only 18% of the rate parameters. The majority of the parameters were derived from a low number of measurements (18%), measurements with no evaluation (7%), single theoretical studies (8%) or were estimates (49%) (Tomlin 2006). In many cases uncertainty ranges had to be assumed. For example, where data were derived from a single theoretical study or were estimated, a uncertainty factor of ten was used above and below the nominal value. Where parameters were derived from a low number of measurements a factor of two was applied. Due to the lack of information the uncertainty ranges for all input parameters were determined according to a minimum and maximum value assuming equal probability throughout the ranges. In contrast to the studies in Tomlin (2006), where rate parameters and thermodynamic parameters were considered separately, all parameters (153 *A*-factors + 23 enthalpies of formation) are varied in this study at the same time. This allows the investigation of possible interaction between rate and thermodynamic parameters for the sulphur scheme.

Scenarios

For the 23 enthalpies of formation three different sets of parameters are considered here (table 5.11). In the first case equal uncertainties of $\pm 10 \text{ kJ mol}^{-1}$ are applied to all 23 parameters. Here it is assumed that no information about the uncertainty of these parameters is available and hence a conservative estimate for the parameter ranges is adopted. In the second case the ranges were adapted according to those quoted in the Burcat table (Burcat & Ruscic 2005) if the uncertainty was listed, which was the case for half of the parameters. Table 5.12 presents the calculated ΔH_f at 298 K using the original NASA polynomial values along with the source of quoted polynomial data. The uncertainty ranges were applied across all temperatures. However, the Burcat table suggests to vary the enthalpy of formation for SN by $\pm 105 \text{ kJ mol}^{-1}$, which is a very large range. Therefore, a third case has been considered, where the range for the enthalpy of formation for SN was updated according to the study of Peebles & Marshall (2002). They suggest a higher nominal value with only a variation of $\pm 2 \text{ kJ mol}^{-1}$. In this case the upper value of the range was adapted to the upper value suggested by Peebles & Marshall (2002) which results in an uncertainty range for the enthalpy of formation for SN of $+13 \text{ kJ mol}^{-1}$. The range still includes the nominal value suggested in the Burcat table. All three investigated scenarios consider the same uncertainty ranges for the 153 A-factors and differ only in the ranges for the 23 enthalpies of formation. The intention behind including these three scenarios is to highlight issues that arise when conflicting information on input parameter ranges is available for non-linear models. This situation is not unique to sulphur chemistry within combustion mechanisms, but is commonly found in many comprehensive reaction schemes. In many cases, suggested nominal values can vary from different sources and therefore the modeller is forced to use judgement in deciding on upper and lower limits for the parameter. This can have consequences for parameter importance rankings as highlighted by the following

Scenario	Uncertainty ranges for ΔH_f
1	$\pm 10 \text{ kJ mol}^{-1}$ for all 23 parameters
2	According to Burcat table ($\Delta H_f \text{ SN} = \pm 105 \text{ kJ mol}^{-1}$)
3	According to Burcat table and study by Peebles & Marshall (2002) ($\Delta H_f \text{ SN} = +13 \text{ kJ mol}^{-1}$)

Table 5.11: Overview of the three scenarios using different sets of parameter ranges for the 23 enthalpies of formation.

results.

HDMR Setup

The RS-HDMR method and its extensions were applied to all three scenarios in order to estimate sensitivity indices and to show how a difference in the parameter ranges can affect the importance ranking of the parameters. An up to second-order HDMR expansion was found to be sufficient to produce an accurate metamodel in all cases. For the sampling process Sobol's quasi-random method (Sobol' 1967) was applied using a sample size of $N = 1024$, which means that 1024 model runs have to be performed in order to construct the HDMR expansion. This sounds quite a lot, however considering the large input space dimension of $n = 176$ this really is not. Other methods which are able to calculate sensitivity indices would require considerably more model runs. For example, the Fourier amplitude sensitivity test (FAST) would theoretically require more than half a million model runs according to equation (2.59) only for estimating the first-order effects. Extended FAST is able to calculate the total sensitivity indices $TS(i)$, however a new set of samples is required for each $TS(i)$. The method of Sobol' is probably out of the question, since it would require $N \cdot (1 + 176 + 15\,400)$ model runs in order to estimate the sensitivity indices up to second-order. Assuming $N = 1\,000$ this would result in more than 15 million model runs.

Species	ΔH_f (298) in kJ mol^{-1}	Source	Uncertainty range in $\pm \text{kJ mol}^{-1}$ for all temperatures
S	277.01	Burcat & Ruscic (2005)	0.25
SH	139.34	Glarborg et al. (1996)	0.52
H ₂ S	-20.50	Burcat & Ruscic (2005)	10.00
SO	5.01	Burcat & Ruscic (2005)	1.30
SO ₂	-296.87	Burcat & Ruscic (2005)	0.21
SO ₃	-395.80	Burcat & Ruscic (2005)	0.71
HSO ₂	-20.50	Glarborg et al. (1996)	10.00
HOSO	-236.33	Glarborg et al. (1996)	10.00
HOSO ₂	-391.23	Glarborg et al. (1996)	10.00
SN	263.61	Burcat & Ruscic (2005)	105.00
S ₂	128.50	Glarborg et al. (1996)	0.30
CS	280.36	Burcat & Ruscic (2005)	25.00
COS	-138.41	Burcat & Ruscic (2005)	1.00
HSO	-22.59	Glarborg et al. (1996)	10.00
HOS	0.01	Glarborg et al. (1996)	10.00
HSOH	-119.33	Glarborg et al. (1996)	10.00
H ₂ SO	-47.11	Glarborg et al. (1996)	10.00
HOSHO	-269.89	Glarborg et al. (1996)	10.00
HS ₂	104.61	Burcat & Ruscic (2005)	10.47
H ₂ S ₂	15.90	Burcat & Ruscic (2005)	10.00
CS ₂	116.95	Burcat & Ruscic (2005)	1.00
H ₂ SO ₄	-732.73	Burcat & Ruscic (2005)	10.00
HSOO	-255.88	Burcat & Ruscic (2005)	10.00

Table 5.12: Nominal values and uncertainty ranges for the enthalpies of formation for Scenario 2. A reference is given for the nominal values. The uncertainty ranges were taken from those quoted in the Burcat table (Burcat & Ruscic 2005) if the parameter was listed. For all other parameters an uncertainty range of $\pm 10 \text{ kJ mol}^{-1}$ has been assumed.

The maximum polynomial order for the approximation of the first-order component functions was set to be ten, and for the second-order terms it was set to be five. The optimal polynomial order for each of the component functions is then automatically calculated by the optimisation method as described in section 4.1. It is also possible to exclude component functions from the HDMR expansion, which can happen in two cases: First, if the component function does not contribute at all

to the overall output value, and second, if the component function has only a very small contribution. To control the exclusion of component functions with a small contribution a threshold can be defined and applied as described in section 5.1. The aim of excluding component functions from the HDMR expansion is to reduce the number of parameters so that one does not require a screening method beforehand.

5.5.4 Results and Discussions

First, an RS-HDMR metamodel was constructed for all three scenarios. Since the construction of the metamodel is similar in all three scenarios, only the details for the third scenario (uncertainty ranges of the enthalpy of formation according to the Burcat table and assuming $+13\text{ kJ mol}^{-1}$ for the enthalpy of formation for SN) are presented here.

Use of Threshold to Construct the HDMR Expansion for Scenario 3

The full second-order HDMR expansion would consist of a total number of 15 577 component functions (1 zeroth-order term + 176 first-order terms + 15 400 second-order terms). To reduce the number of terms in the HDMR expansion a threshold can be applied so that only component functions which have a significant contribution towards the overall output value are considered. For example, using a threshold of 1% for the first- and second-order component functions, only five of the 176 first-order component functions and none of the 15 400 second-order component functions are approximated by optimal-order polynomials. Therefore, the resulting HDMR expansion consists of only six terms (1 zeroth-order + 5 first-order terms + 0 second-order terms).

The accuracy of the constructed HDMR metamodel can be shown by calculating

the relative error (see section 3.3.3). The same set of samples which has been used to construct the HDMR component functions (Sobol' sequence with $N = 1024$) can also be used for the calculation of the relative error. However, in order to test the metamodel with a set of samples which have not been involved in the construction of the HDMR expansion, additional real model simulations are required and another 2 000 model runs have been performed for this purpose. Applying the threshold of 1%, the resulting first-order metamodel gives 99.05% of the tested samples within the 5% RE range. This supports the view that it is possible to build an accurate metamodel by including only five parameters. In this way we can already identify the five most important parameters, which are the A -factors for the forward reaction rates for $\text{SO} + \text{NH} = \text{NO} + \text{SH}$, $\text{SO} + \text{N} = \text{NO} + \text{S}$, $\text{SO} + \text{OH} = \text{SO}_2 + \text{H}$ and $\text{SH} + \text{NH} = \text{NS} + \text{H}_2$ and ΔH_f SO. It is actually quite surprising that only five of the 176 considered uncertain parameters drive the uncertainty in the model output and that parameter interactions are negligible.

The optimisation method as described in section 4.1 automatically calculates the optimal order for the polynomial approximation of the five first-order component functions. The polynomial order indicates the kind of relationship we can expect for the parameter-output relationship. Two of the component functions were approximated by first-order polynomials, indicating a linear relationship. The other three component functions were approximated by higher-order polynomials, indicating a non-linear relationship. None of the second-order component functions were approximated by polynomials and therefore one can assume that any parameter interactions are so small that they were excluded because of the 1% threshold.

If the threshold is decreased, then more component functions will be included in the HDMR expansion. For example, a threshold of 0.001% gives 51 first-order component functions and four second-order component functions. For the first-order RS-HDMR metamodel 99.50% of the tested samples are within the 5% RE range

and for the second-order metamodel 99.65 %. The accuracy can also be presented by using a scatter plot for the response of the metamodel and the original model or by comparing their empirical probability density functions as shown in figure 5.8. It is illustrated that the data points in figure 5.8a form a straight line with only a low amount of scatter. The coefficient of determination is $r^2 = 98.50$, which confirms the high accuracy of the first-order HDMR metamodel. The plot of the pdfs in figure 5.8b shows a very good match between the original model and the metamodel. The choice of the threshold depends on the aim of the investigation. A larger threshold will provide only the most important component functions so that one is able to identify the most important parameters. A smaller threshold ensures that less important component functions are also included.

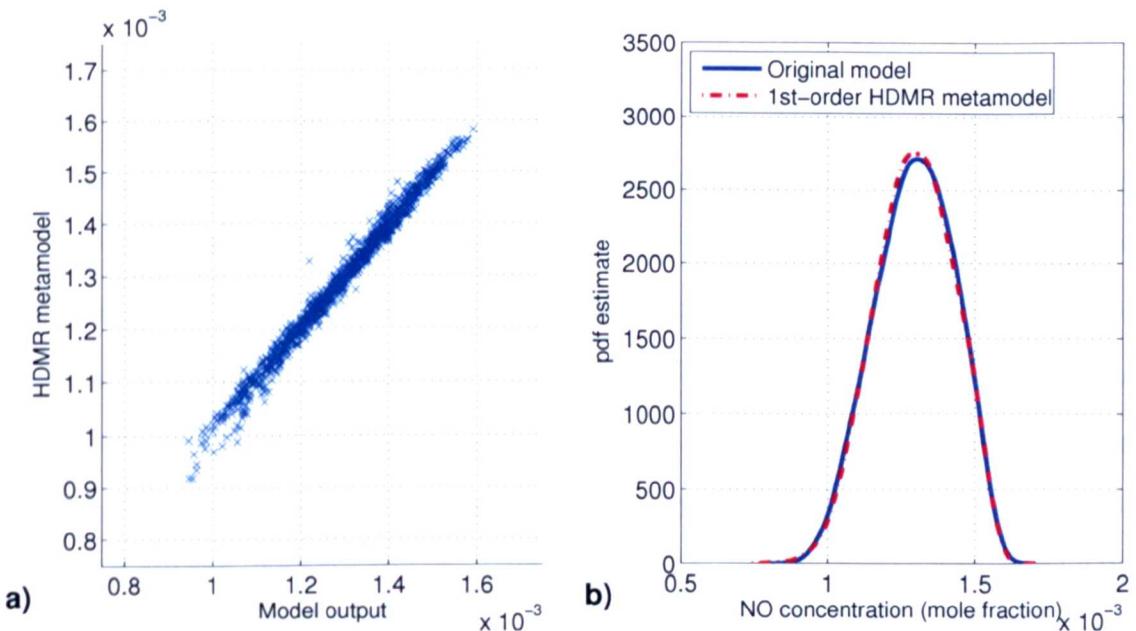


Figure 5.8: Accuracy of the RS-HDMR metamodel based on $N = 1\,024$ model runs for NO concentration (mole fraction) showing (a) the scatter plot between the original model response and the first-order RS-HDMR metamodel ($r^2 = 98.50\%$) and (b) the probability density function estimate for both the original model and the first-order RS-HDMR metamodel.

The results demonstrate that the application of a threshold in order to select HDMR

component functions enables one to reduce the number of HDMR terms (and therefore the number of parameters) dramatically. The accuracy of the metamodel for the 0.001 % threshold provides confidence in the constructed HDMR expansion and thus the calculated sensitivity indices. The use of a threshold provides a similar function to the use of a screening method, since one excludes unimportant parameters before actually calculating the sensitivity indices and the parameter ranking. However, unlike the two step algorithm where a screening method is used to identify unimportant parameters in the first step, and then a second method is used to calculate the sensitivity indices in a reduced input space dimension, here only a single set of model runs is used and only one method.

Sensitivity Indices and Component Functions for Scenario 3

Table 5.13 presents the five highest ranked first-order sensitivity indices and the four second-order sensitivity indices for the third scenario if using a sample size of $N = 1024$ and a threshold of 0.001 % as described previously. Since only four of the second-order component functions have been approximated by polynomials, only four second-order effects are seen in the table. Using these selected input ranges, the most important parameter is the A -factor for the forward reaction rate of $\text{SO} + \text{NH} = \text{NO} + \text{SH}$ which contributes almost 60 % to the overall variance of the model output. The top five first-order sensitivity indices already add up to 0.97 and therefore account for most of the output variance. They are also the same five parameters that were identified using the higher threshold of 1 %. This also suggests a low importance for parameter interactions. As seen in table 5.13, the second-order effects are quite small and almost negligible. No significant interactions between the thermodynamic parameters and rate parameters were found.

All first- and second-order sensitivity indices add up to 0.9826 indicating that the

Parameter x_i	Rank	S_i
$\text{SO} + \text{NH} = \text{NO} + \text{SH}$	1	0.5956
$\text{SO} + \text{N} = \text{NO} + \text{S}$	2	0.2758
$\text{SO} + \text{OH} = \text{SO}_2 + \text{H}$	3	0.0735
$\Delta H_f \text{ SO}$	4	0.0140
$\text{SH} + \text{NH} = \text{NS} + \text{H}_2$	5	0.0111
$\sum S_i$		0.9784
Parameter x_i	Parameter x_j	S_{ij}
$\text{SO} + \text{OH} = \text{SO}_2 + \text{H}$	$\text{SO} + \text{NH} = \text{NO} + \text{SH}$	0.0018
$\text{SH} + \text{H} = \text{H}_2 + \text{S}$	$\text{SH} + \text{NH} = \text{NS} + \text{H}_2$	0.0015
$\text{S} + \text{CS}_2 = \text{CS} + \text{S}_2$	$\text{HS}_2 + \text{H} + \text{M} = \text{H}_2\text{S}_2 + \text{M}$	0.0005
$\text{H}_2\text{S} + \text{M} = \text{H}_2 + \text{S} + \text{M}$	$\text{S} + \text{NO}_2 = \text{NO} + \text{SO}$	0.0004
$\sum S_{ij}$		0.0042
$\sum S_i + \sum S_{ij}$		0.9826

Table 5.13: First- and second-order sensitivity indices for scenario 3. When a reaction is indicated, the parameter is the A -factor for the forward reaction rate.

majority of the output variance has been accounted for. An increase in the sample size N in order to construct the HDMR expansion would lead to an even higher accuracy of the metamodel. However, experiments using various sample sizes up to $N = 8\,192$ have shown that this does not change the values of the sensitivity indices significantly and that the ranking of the most important parameters is the same.

Plots of the four most important first-order component functions for scenario 3 are presented in figure 5.9. For comparison, the equivalent scatter plot for each parameter is also shown. This highlights the potential difficulty in interpreting scatter plots, since the scatter caused by variability in all other parameters, can often obscure the response of the parameter of interest. All four scatter plots (figure 5.9a-d) show a non-random structure, which suggests that a relationship exists between the input parameter and the model output. However, sometimes it is rather difficult to see whether or not this relationship is linear or nonlinear. The plot of the corresponding first-order component functions (here the mean f_0 is added to make the

plots comparable with the scatter plots) gives a much more detailed picture of the relationship between the input parameter and model output.

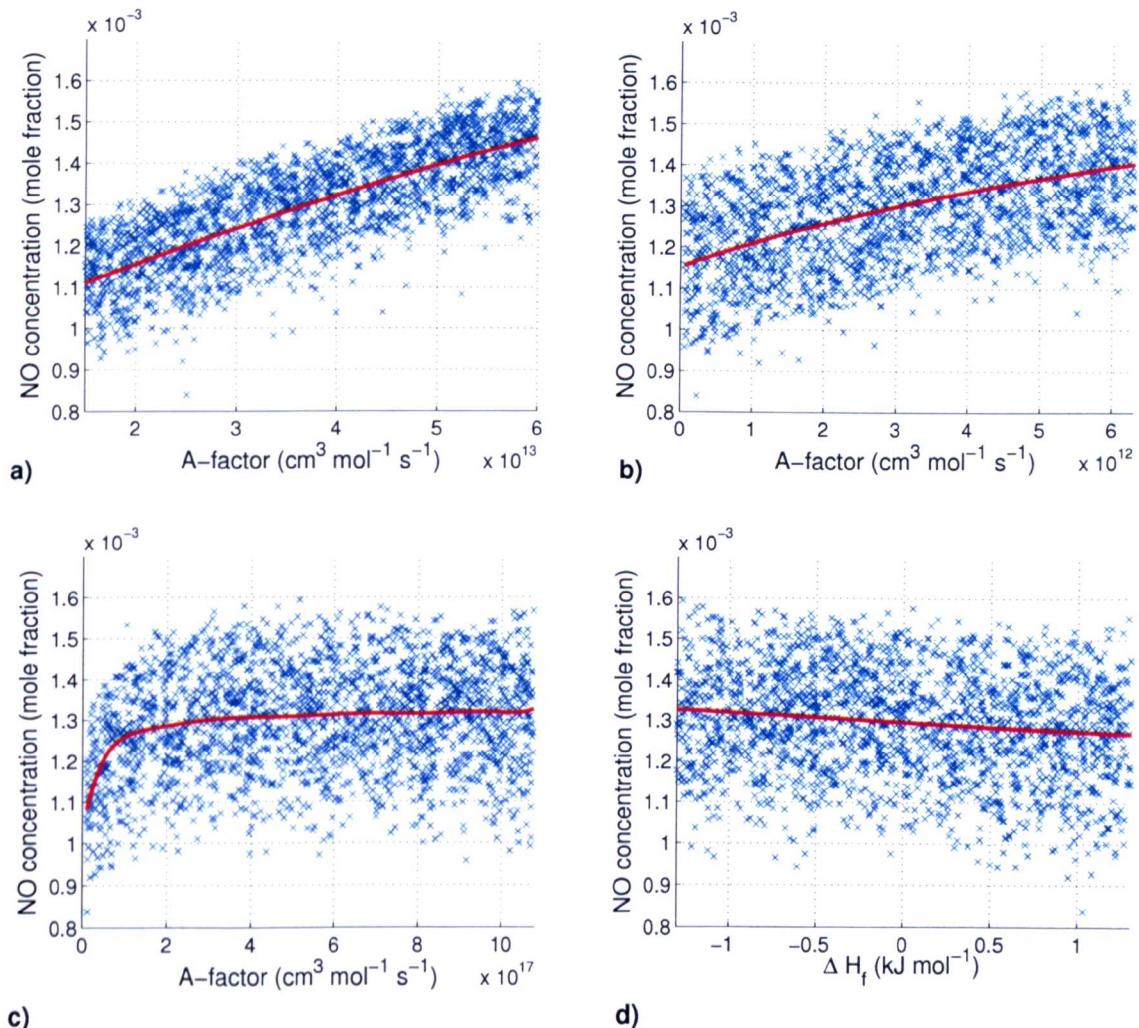


Figure 5.9: First-order component functions and scatter plots for (a) $\text{SO} + \text{NH} = \text{NO} + \text{SH}$, (b) $\text{SO} + \text{N} = \text{NO} + \text{S}$, (c) $\text{SO} + \text{OH} = \text{SO}_2 + \text{H}$, (d) SO (scenario 3).

In figures 5.9a and 5.9b a non-linear relationship is shown which is approximated by a second-order polynomial (figure 5.9a) and a third-order polynomial (figure 5.9b) respectively. The relationship between the *A*-factor for the forward rate for reaction $\text{SO} + \text{OH} = \text{SO}_2 + \text{H}$ and the NO mole fraction as shown in figure 5.9c is highly non-linear and therefore approximated by a higher order polynomial (ninth-order). The

shape of this relationship would not be so clearly visible if analysing only the scatter plot. The nominal value for the *A*-factor for this reaction is $1.077 \cdot 10^{17} \text{ cm}^3 \text{ mol}^{-1} \text{ s}^{-1}$ and the figure shows that the gradient of the first-order HDMR component function changes substantially around this value. It should be acknowledged that the uncertainty range used here is much wider than that recommended by the study of Blitz et al. (2000) but has been chosen as such since the nominal value is sourced from a single experimental study over a limited temperature range. For *A*-factor values lower than the nominal value the gradient is steep, indicating a strong sensitivity in this region. At values higher than the nominal value, the gradient is much lower, indicating a lower sensitivity. The gradient of the first-order component function indicates the first-order local sensitivity coefficient for the parameter at a particular nominal value assuming there are no higher order effects. This example demonstrates that the local sensitivity coefficient for this parameter changes substantially around the nominal value. This supports the view that global sensitivity information is required in order to fully evaluate the mechanism, since the local first-order sensitivity coefficient would be highly dependant on the nominal value chosen.

In contrast, the relationship between the change in the enthalpy of formation for SO and the model output is linear (in figure 5.9d) and is therefore approximated by a first-order polynomial. In this case the local first-order sensitivity coefficient does not change throughout the parameter range. It should be highlighted again that the polynomial order for the approximation of the component functions is automatically chosen by an optimisation algorithm (see section 4.1).

It is also possible to plot the second-order HDMR component functions (3D plot) to show the interactive effect of two parameters. However, since the second-order effects are very small, only one plot showing the interaction of the *A*-factors for reaction $\text{SO} + \text{OH} = \text{SO}_2 + \text{H}$ and $\text{SO} + \text{NH} = \text{NO} + \text{SH}$ upon the output (figure 5.10) is presented for demonstration purposes.

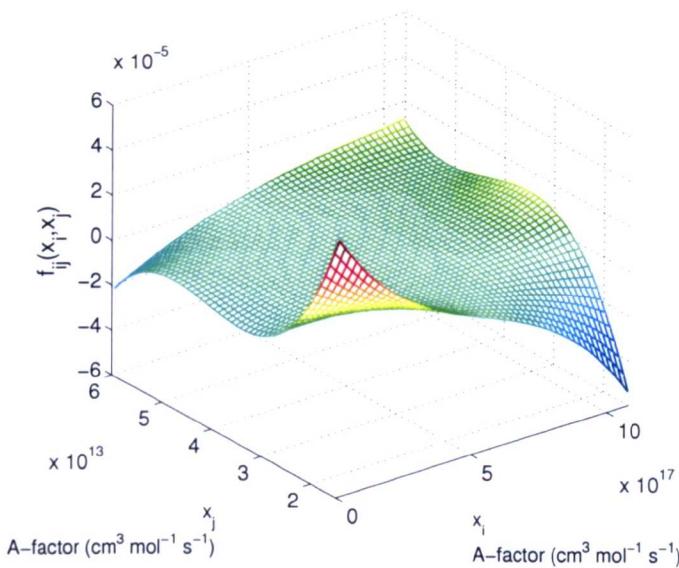


Figure 5.10: Second-order component function $f_{ij}(x_i, x_j)$ for x_i : $\text{SO} + \text{OH} = \text{SO}_2 + \text{H}$ and x_j : $\text{SO} + \text{NH} = \text{NO} + \text{SH}$ (scenario 3).

Prior Application of a Screening Method for Scenario 2

The construction of the HDMR metamodel and the calculation of the sensitivity indices for the first scenario (equal uncertainty ranges for all 23 enthalpies of formation parameters) is identical to the third scenario presented in the above sections. However, the construction of the HDMR expansion for the second scenario where the uncertainty ranges were adapted according to the Burcat table was more difficult. A sample size of $N = 1\,024$ was not sufficient to gain a satisfactory accuracy of the metamodel. It was necessary to increase the sample size up to $N = 4\,096$ in order to obtain reliable results. This of course requires more computational time since a higher number of model runs have to be performed. The necessary increase in the sample size N in comparison to scenarios 1 and 3 is probably due to the large uncertainty range of $\pm 105 \text{ kJ mol}^{-1}$ used for ΔH_f SN. This leads to a much larger region of parameter space which has to be used in the polynomial fitting.

Due to the large sample size required it is of interest to investigate an alternative

	RS-HDMR (N=1024)	RS-HDMR (N=4096)	Screening + RS-HDMR (N=1780+1024)
1st-order 5 % RE	85.85 %	90.40 %	89.90 %
2nd-order 5 % RE	86.25 %	94.80 %	95.30 %

Table 5.14: Relative Error calculated for a set of 2000 random samples for the meta-model based on RS-HDMR and a combination of Screening and RS-HDMR.

Parameter	RS-HDMR (N=4096)	Screening + RS-HDMR (N=1780+1024)
ΔH_f SN	0.3815	(1) 0.3855
SO + NH = NO + SH	0.3700	(2) 0.3719
SO + N = NO + S	0.1344	(3) 0.1365
SO + OH = SO ₂ + H	0.0410	(4) 0.0386
ΔH_f SO	0.0086	(5) 0.0087
$\sum S_i$	0.9524	0.9590
$\sum S_i + \sum S_{ij}$	0.9695	0.9755

Table 5.15: First-order sensitivity indices for scenario 2. Where a reaction is indicated, the parameter is the *A*-factor for the forward reaction rate.

approach where a screening method (the Morris method described in section 2.3.2) is applied beforehand in order to identify unimportant parameters. This can be used to reduce the number of parameters studied and effectively reduce the dimension of the fitting space. For the Morris method ten trajectories ($r = 10$) were used leading to $N = 1780$ model runs. Ten trajectories were enough in order to identify the most important parameters. However a reliable parameter ranking based on the Morris results could not be obtained, because ten trajectories were not sufficient to reach convergence for the mean of the elementary effects. A total of 47 potentially important parameters could be identified on the basis of the highest mean effect across the ten trajectories. The extended HDMR method was then applied using a sample size of $N = 1024$ using only the 47 pre-selected parameters. The results in

terms of model accuracy and the values for the sensitivity indices are nearly exactly the same as calculated using the HDMR involving all 176 uncertain parameters but using a sample size of $N = 4096$ (see table 5.14 and 5.15). This demonstrates that the development of the metamodel (i.e. the polynomial optimisation) was improved by reducing the dimension of the fitting space for a small sample size. In this case the combination of the screening and the HDMR method led to a decrease in the number of model runs required from $N = 4096$ to $N = 1780 + 1024$. Hence, whilst the extended HDMR method was capable of handling the large input space dimension effectively, in certain cases savings could be made by the prior application of a screening method. This may be valuable for situations where the basic model is computationally expensive and a limited number of simulations can be afforded.

The importance of a prior screening is discussed in detail in the following chapter, when the main focus is on investigations of systems which not only have a large input space dimension, but also significant higher order effects. It is shown, that in those cases a parameter screening is advisable.

Comparison of all Three Scenarios

Table 5.16 shows a comparison of the five highest ranked first-order sensitivity indices calculated for each of the three scenarios. It shows that a different ranking is calculated for each of the three scenarios. In scenario 1 the two most important parameters are the enthalpy of formation for SO and SO₂, which are both nearly equally important. A fairly high degree of uncertainty has been assigned to these parameters within scenario 1 which leads to their significant contribution to the overall output variance. Within the second and third scenarios a much lower degree of uncertainty as suggested by Burcat & Ruscic (2005) has been adopted for these enthalpies of formation and therefore they appear lower in the importance ranking.

In the second scenario ΔH_f SN becomes the most important parameter due to its extremely large range of uncertainty as suggested by Burcat & Ruscic (2005). The enthalpy of formation for SO is now only ranked fifth and the component function for ΔH_f SO₂ has been excluded from the HDMR expansion. The only change between scenarios 2 and 3 is ΔH_f SN. In the third scenario the ranking changes again and the *A*-factor for the reaction rate for SO + NH = NO + SH becomes the most important parameter with ΔH_f SN no longer highly ranked.

Parameter	Scenario 1 ($\Delta H_f = \pm 10\text{ kJ}$)		Scenario 2 (ΔH_f SN = $\pm 105\text{ kJ}$)		Scenario 3 (ΔH_f SN = $+13\text{ kJ}$)	
	S_i	(Rank)	S_i	(Rank)	S_i	(Rank)
SO + NH = NO + SH	0.2297	(3)	0.3631	(1)	0.5956	(1)
SO + N = NO + S	0.1007	(4)	0.1219	(3)	0.2758	(2)
SO + OH = SO ₂ + H	0.0255	(5)	0.0316	(4)	0.0735	(3)
ΔH_f SO	0.3082	(1)	0.0101	(5)	0.0140	(4)
SH + NH = NS + H ₂	0.0035		0.0019		0.0111	(5)
ΔH_f SO ₂	0.2874	(2)	0		0	
ΔH_f SN	0.0001		0.3479	(2)	0.0001	
$\sum S_i$	0.9771		0.8904		0.9784	
$\sum S_i + \sum S_{ij}$	0.9842		0.9055		0.9826	

Table 5.16: First-order sensitivity indices for all three scenarios. Where a reaction is indicated, the parameter is the *A*-factor for the forward reaction rate.

The comparison in table 5.16 shows that the importance of the parameters is critically dependent on the uncertainty ranges chosen. For example, ΔH_f SN only becomes important if using the large uncertainty range of $\pm 105\text{ kJ mol}^{-1}$ as suggested by the Burcat table (Burcat & Ruscic 2005). To further investigate the output response over the full parameter range the corresponding first-order component function for scenario 3 is plotted (figure 5.11a). The figure demonstrates why this parameter only becomes important when considering a large uncertainty range. The local sensitivities in the ranges as used for scenarios 1 and 3 are very small as indicated by the gradient of the component function between the vertical lines in

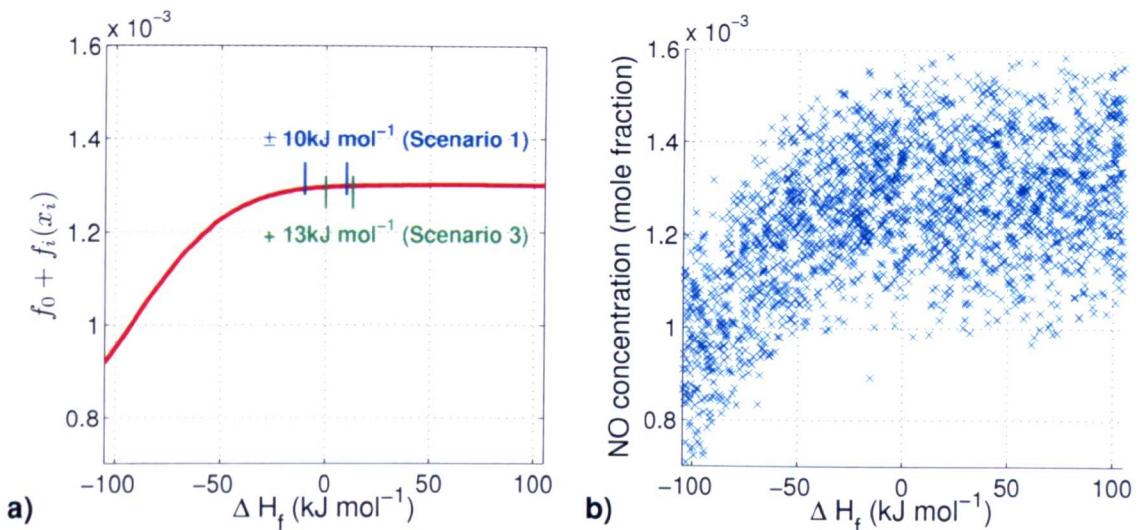


Figure 5.11: (a) First-order component function f_i and (b) scatter plots for SN (scenario 2). The mean f_0 is added to f_i to make the plot comparable with the scatter plot.

the figure. However, the region from -105 kJ mol^{-1} to approximately -25 kJ mol^{-1} , which is only covered by the large uncertainty range of $\pm 105 \text{ kJ mol}^{-1}$, shows a very steep gradient which explains the high sensitivity of this parameter in scenario 2. Figure 5.11b shows the scatter plot for the same input parameter for comparison. Again, the plot of the first-order component function gives us a much better idea of how the input parameter affects the output than the scatter plot. This example helps to demonstrate the power of the HDMR method and the value of being able to visualise the response of the output to individual parameters across the whole parameter range. If only sensitivity indices and the final importance ranking are used then the changes in local sensitivity coefficients across the parameter ranges are not apparent. The component functions allow the user to evaluate the potential response to the parameter if different uncertainty ranges were adopted. Given the lack of consensus between different sources for thermochemical parameters, this feature of HDMR is particularly valuable.

5.5.5 Conclusions

The RS-HDMR method provides a straightforward approach for the global sensitivity analysis of chemical kinetic models. However, an extension to the existing set of HDMR tools is necessary in order to explore the large input space dimensions that commonly result from the study of comprehensive kinetic schemes. The optimisation method (section 4.1) in cooperation with excluding component functions from the HDMR expansion via a threshold (section 5.1) has been proven to be very useful. The general outcome of this study is that the extended RS-HDMR method can handle a large input space dimension efficiently. In this case a sample size of $N = 1\,024$ was sufficient in order to construct an accurate metamodel and to calculate sensitivity indices for scenarios 1 and 3. Only in certain cases (here for scenario 2) was it computationally more efficient to apply a screening method beforehand. However, the use of the combined screening plus HDMR method, and the HDMR method used on its own, led to the same accuracy of the metamodel and to the same sensitivity indices. The use of the two step approach affected only the computational cost, but for this reason may be relevant in cases where the underlying model simulations are expensive.

The analysis of the plots of the first-order component functions has been found to be very helpful in order to explore the shape of the relationship between the input parameters and the target output. Whereas it is sometimes quite difficult to identify the relationship when using scatter plots, the HDMR component functions isolate the response to an individual parameter in the case of first-order component functions, or pairs of parameters in the case of second-order ones. This feature of HDMR becomes particularly useful where several uncertainty ranges are suggested for a particular parameter by different expert studies. The study shows that the final ranking of the input parameters is critically dependent on the input ranges chosen. This fea-

ture suggests that a unified framework where parameter ranges can be agreed and consistently updated as new experimental and theoretical studies become available would be a valuable contribution to model evaluation exercises. Such a framework is one of the aims of the PrIME data model web based co-laboratory (Djurisic et al. 2007). However, the HDMR component functions express the output response over the whole of the input range and can therefore be used to investigate the output response over particular regions of the input parameter space. A very important feature of the HDMR method is that a metamodel can be constructed with wide parameter ranges first. Then, if the ranges are narrowed down by subsequent studies, the metamodel can be used as a surrogate in order to update the output pdf or error bars without having to recompute the original model again. This is particularly useful if the original model is computationally expensive to run. This point could be illustrated in the previous case study (section 5.4) describing a reactive plume where the range of one parameter has been narrowed down after the model simulations had been performed already.

Several parameters in the present study exhibited highly non-linear responses within particular regions of their input space. This calls into question the use of local sensitivity indices for the evaluation of parameter importance in such non-linear models. On the other hand, the gradient of the HDMR first-order component functions can be used as an indication for the local sensitivity over wide parameter ranges. However, this is only correct if higher order effects are small, because for calculating the local sensitivity the partial derivative of the whole response is taken and not only of the first-order component function. Global sensitivity analysis can help to identify important parameters that contribute to the overall output variance and require better categorisation by further thermo-kinetic studies. The methods presented here therefore provide valuable tools for the evaluation and improvement of chemical kinetic models.

5.6 Summary

Another important extension to the existing set of HDMR tools has been developed in this chapter. Unimportant component functions, which have only a very small contribution towards the overall output value, can be excluded from the HDMR expansion via a pre-defined threshold. Thus, the number of the component functions approximated by orthonormal polynomials can be reduced. This is very useful if investigating models with a large number of input parameters in order to further increase the accuracy of the HDMR metamodel and successively the accuracy of the sensitivity indices.

The application of the threshold has been tested using the analytical Sobol' g-function with 20 input parameters. Unimportant first- and second-order component functions were correctly excluded from the HDMR expansion. This ensures, that those component functions which are affected by the Monte Carlo integration error are not considered for the construction of the metamodel. The extended HDMR method could cope very well with the non-linear and non-monotonic behaviour of the Sobol' g-function and a sample size of $N = 512$ was sufficient in order to identify the most important first- and second-order effects and their ranking. However, higher order polynomials (up to tenth-order) were required in order to accurately approximate the component functions. This is due to the complicated V-shape of the input-output relationship which also has a non-differential point.

A short introduction into chemical kinetics has been given, because the two case studies presented in this chapter contain complex chemical mechanisms. First, a reactive turbulent atmospheric plume model was investigated, highlighting the combined effects of uncertainties in physical and chemical parameters. In total 22 uncertain parameters were considered. No interactions between parameters could be found and a sample size of $N = 512$ was sufficient in order to produce a very

accurate metamodel. The HDMR analysis revealed that only two parameters are responsible for the large uncertainty in the peak NO_x concentration. Comparisons with experimental data revealed that the uncertainty range for the structure function coefficient c_0 could be narrowed down, which is also in agreement with various studies in the literature. It could also be shown that the mixing time scale coefficient α has a lower but still significant influence on the mean passive tracer concentration field, which is a shortcoming of the interaction by exchange with the mean (IEM) model. The second case study focused on a one-dimensional low-pressure premixed methane flame model with sulphur-nitrogen chemistry interactions. Uncertainties in rate constants and thermodynamic data for the sulphur scheme were considered, leading to a study of 176 input parameters. Three different scenarios have been investigated using different uncertainty ranges for the 23 thermodynamic parameters. It could be shown that the ranking of the parameters is critically dependent on the input ranges chosen. The HDMR analysis allows to explore the shape of the relationship between the input parameter and the output for different input ranges and it becomes clear in what ranges a certain parameter is most sensitive. The HDMR metamodel can also be used as a surrogate and when investigating parameter ranges which are narrower than the original ones, the HDMR metamodel can be used instead of performing a new set of simulations. This is particularly useful if the original model is computationally expensive to run. In comparison to other global sensitivity analysis methods such as the method of Sobol' or the Fourier amplitude sensitivity test (FAST), the HDMR method is very efficient. For two scenarios a sample size of $N = 1024$ was sufficient to construct an accurate metamodel. Only for one scenario was it computationally more efficient to apply a combined screening plus HDMR approach. The importance of parameter screening is further highlighted in the next chapter when the focus is on systems which have significant parameter interactions.

6 Application of RS-HDMR for Systems with Significant Parameter Interactions

The aim of this chapter is to investigate the application of the HDMR method for systems with a large number of input parameters and with significant parameter interactions. The exploration of parameter interactions usually requires much more samples than the investigation of the first-order effects. Especially in connection with a large input space dimension this can easily lead to a required sample size of $N > 10\,000$. Therefore, it might be computationally more efficient to follow a two step procedure. First a screening method such as the Morris method (see section 2.3.2) is applied in order to identify unimportant parameters. Then, the HDMR approach is used in connection with the reduced input space.

6.1 Need for Screening

It could be shown in the latter chapter of this work, that the HDMR method and its extensions can handle a large input space very efficiently. However, as shown in the case study in section 5.5 for the premixed methane flame model, it can sometimes

be computationally more efficient to apply a screening method beforehand in order to reduce the number of parameters. It is worth noticing that the case studies investigated in the latter chapter did not have significant higher order effects and a first-order HDMR expansion has been proven to be sufficient in order to cover the model behaviour and the variance of the model output. If one deals with models with a large input space dimension (i.e. greater than 20) where additional parameter interactions cause a significant amount of the variance in the output and one is interested in exploring them, then it is computationally more efficient to apply a screening method, such as the Morris method (see section 2.3.2) beforehand. Thus, unimportant parameters can be excluded and the HDMR method is then applied using only the reduced input space. This makes the RS-HDMR approach much more efficient for exploring parameter interactions, because the fitting of the polynomials takes place in a lower dimension.

The two step algorithm (screening + RS-HDMR) is tested in the following section using the analytical Sobol' g-function with 50 parameters, where parameter interactions account for more than 20 % of the overall variance in the model output. Section 6.3 focuses on a case study, where the HDMR method is applied in order to investigate the process of cyclohexane oxidation under low temperature fuel rich conditions. Here a total of 238 uncertain parameters are considered and significant parameter interactions are present.

6.2 Test Model: Sobol' g-function

The Sobol' g-function has already been introduced and applied in section 5.2 as an analytical test case in order to highlight the efficiency of the RS-HDMR method and its extensions. Here, the same function is used a test model ones again, but with a different configuration. The aim is to show, that a combination of the Morris screen-

ing method and the RS-HDMR method is a very efficient approach especially for systems with a large input space dimension and significant parameter interactions.

6.2.1 Properties of the Sobol' g-function

One very handy feature of the Sobol' g-function (see equation (5.5)) is, that the input parameter space can be arbitrarily extended and the importance of each parameter can be controlled by the value of a_i (see equation (5.6)). Here, we consider 50 input parameters using $a_i = \{4.5, 4.5, 1, 0, 1, 9, 0, 9, 99, \dots, 99\}$. The spectrum of a_i has been chosen according to Ratto et al. (2007). 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 .

S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	$\sum S_i$
0.0096	0.0096	0.0727	0.2906	0.0727	0.0029	0.2906	0.0029	0.7516

Table 6.1: First-order sensitivity indices (first eight parameters) for the Sobol' g-function calculated analytically.

S_{47}	S_{34}	S_{37}	S_{45}	S_{57}	$\sum S_{ij}$
0.0969	0.0242	0.0242	0.0242	0.0242	0.2209

Table 6.2: Selected second-order sensitivity indices for the Sobol' g-function calculated analytically.

The analytical values for the first-order sensitivity indices are presented in table 6.1. Only the first eight parameters ($i = 1..8$) and their interactions have an effect on the output and therefore table 6.1 focuses only on those parameters. The two most important parameters are x_4 and x_7 , which are equally important ($a_4 = a_7 = 0$) and each of them contributes by about 29 % to the overall output variance. The sum of the first-order effects covers 75.16 % of the output variance. The second-order

sensitivity indices based on the analytical calculation are shown in table 6.2. Only the five highest interactions are shown. The most important parameter interaction takes place between parameter x_4 and x_7 and contributes by nearly 10 % to the variance of the output. The sum of the second-order effects contributes by 22.09 %. This means, that 97.25 % of the entire model variance can theoretically be captured if using a second-order HDMR expansion.

6.2.2 Application of RS-HDMR

The RS-HDMR method and its extensions are now applied using the whole input space dimension ($n = 50$) in order to estimate the first- and second-order sensitivity indices. Only the first eight parameters ($x_1 - x_8$) of the 50 in total have an influence on the output. A threshold of 0.1 % has been applied to the first- and second-order component functions in order to exclude unimportant parameters. This results in the approximation of only eight first-order component functions (or six in the case of $N = 1024$). Consequently, the first-order sensitivity indices are only calculated for the first eight parameters as shown in table 6.3. Because the fitting of the polynomials takes place in such a high dimension, it is not possible to accurately estimate the second-order effects if using only a small sample size. As shown in table 6.3 the sample size N has to be at least 4 096 in order to get an estimate for most of the five highest second-order sensitivity indices. 8 192 runs are required to estimate all of the five highest second-order sensitivity indices. The sensitivity indices of the less important parameter interactions are not shown due to the lack of space. Most of the 1 225 second-order component functions have been excluded anyway, because their contribution is smaller than the applied threshold would allow. However, all estimated second-order sensitivity indices (approximately 20, depending on the sample size N) are included in $\sum S_{ij}$.

N	1 024	2 048	4 096	8 192	analytical
S_1	0.0000	0.0109	0.0100	0.0100	0.0096
S_2	0.0134	0.0153	0.0128	0.0098	0.0096
S_3	0.0606	0.0662	0.0684	0.0708	0.0727
S_4	0.2230	0.2580	0.2702	0.2804	0.2906
S_5	0.0636	0.0683	0.0701	0.0707	0.0727
S_6	0.0044	0.0033	0.0031	0.0030	0.0029
S_7	0.2320	0.2596	0.2697	0.2801	0.2906
S_8	0.0000	0.0041	0.0035	0.0033	0.0029
$\sum S_i$	0.6134	0.6873	0.7104	0.7309	0.7516
$S_{4,7}$	-	-	0.0818	0.0858	0.0969
$S_{3,4}$	-	-	0.0228	0.0224	0.0242
$S_{3,7}$	-	-	0.0199	0.0212	0.0242
$S_{4,5}$	-	-	-	0.0207	0.0242
$S_{5,7}$	-	-	0.0241	0.0212	0.0242
$\sum S_{ij}$	-	-	0.1945	0.2133	0.2209
$\sum S_i + \sum S_{ij}$	0.6134	0.6873	0.9048	0.9442	0.9725

Table 6.3: First-order (first eight inputs) and second-order (selected input combinations) sensitivity indices for the Sobol' g-function using the RS-HDMR approach.

From table 6.3 it can be seen that the RS-HDMR method is able to handle the large input space dimension. The most important first-order effects can be explored by using only a relatively small sample size of $N = 1\,024$. However, in order to investigate parameter interactions the sample size N has to be at least 4 096. This is a considerably large sample size and in order to achieve such a large number of model runs for practical applications it may be necessary to run the model in a parallel environment (e.g. computer cluster) particularly when the model is computationally expensive to simulate. However, other methods for global sensitivity analysis, such as the Fourier amplitude sensitivity test (FAST) and the method of Sobol' require much more model runs. Classic FAST is only able to calculate first-order sensitivity indices and according to equation (2.59) more than 20 000 model runs would be necessary. The method of Sobol' is able to calculate first- and second-order sensitivity indices, however this would lead $N \cdot (1 + 50 + 1\,225)$ model runs. Assuming $N = 1\,000$ the

method of Sobol would require more than one million model runs.

6.2.3 Combination of Screening and RS-HDMR

In order to investigate the efficiency of the combined screening plus HDMR approach, the Morris method is now applied first to identify potentially important parameters. The values of the unimportant parameters are then fixed and only the values of the potentially important parameters are varied in connection with the HDMR method.

Morris Method

The Morris method as introduced in section 2.3.2 has been applied to the Sobol' g-function with 50 input parameters. Ten trajectories ($r = 10$) have been used in order to calculate the mean effect and the standard deviation. Each trajectory contains $n + 1 = 51$ runs where each parameter has been changed according to the sampling strategy. The results of the Morris analysis can be seen in figure 6.1, where the standard deviation is plotted against the mean.

As shown in figure 6.1, the Morris method has identified eight potentially important parameters (x_1 to x_8) under the guideline that the higher the mean, the higher the effect of the parameter. A low standard deviation indicates that the effect of the concerning parameter is only linear. A high standard deviation indicates either non-linear effects or parameter interactions. Parameters x_4 and x_7 have been identified as the most important parameters, however according to the analytical values in table 6.1 they should be equally important and therefore their mean effect and standard deviation should be the same. This indicates, that ten trajectories are probably not enough in order to calculate a good average for the mean and standard deviation. However, here the interest is not in obtaining the exact ranking

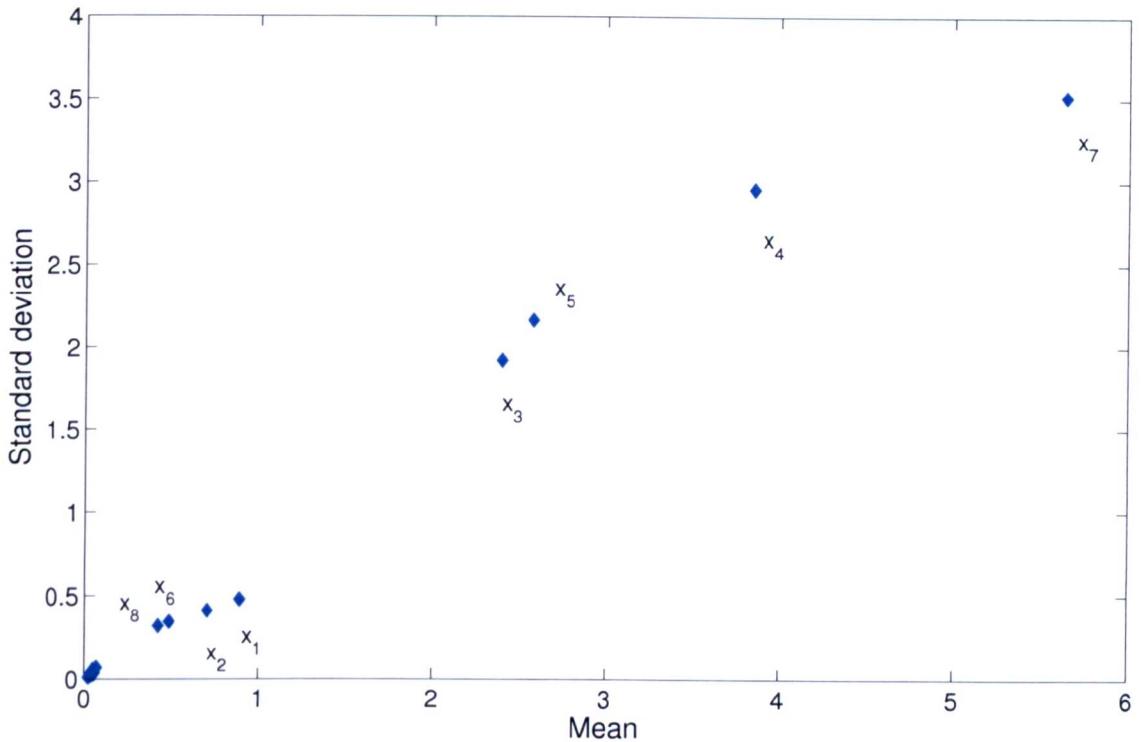


Figure 6.1: Results of the Morris analysis using $r = 10$ trajectories ($N = 510$). Only the first eight parameters $x_1 - x_8$ are potentially important. Parameters $x_9 - x_{50}$ are identified as unimportant.

of those parameters. The main focus is in reducing the input space dimension by identifying the unimportant parameters and $x_9 - x_{50}$ have been correctly identified as not important.

If one wants to produce a more reliable parameter ranking on the basis of the Morris analysis, more trajectories have to be used which also means an increase in the sample size N . Figure 6.2 shows the results of the Morris method using 10, 50 and 100 trajectories. It can be seen that the mean and standard deviation for the effects of the parameters which are equally important are actually quite different from each other, especially for the two most important ones. Only for the case where 100 trajectories are used, are the mean effect and standard deviation for the two most important parameters x_4 and x_7 nearly equal as one would expect. However,

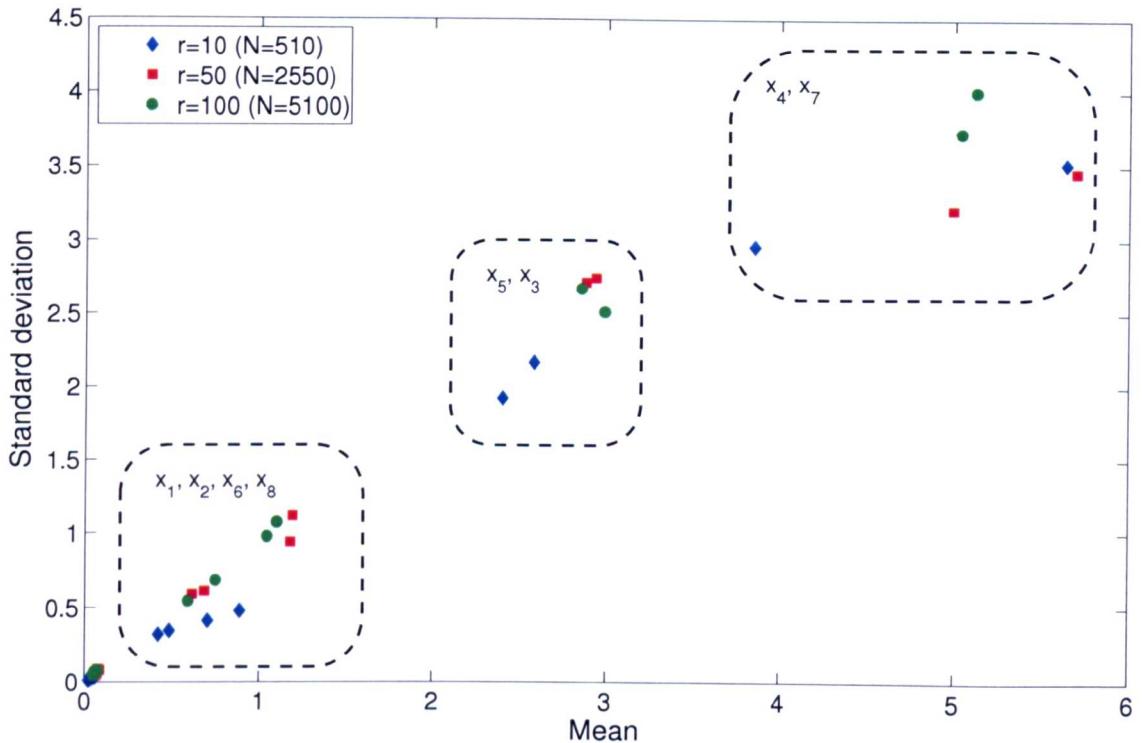


Figure 6.2: Results of the Morris analysis using 10, 50 and 100 trajectories.

the use of 100 trajectories is computationally quite expensive and requires a sample size of $N = 5\,100$. Considering this, the Morris method is not really efficient for this purpose, but has been shown to be efficient for screening out unimportant parameters.

RS-HDMR

Parameters x_1 to x_8 have been identified as potentially important by the Morris method, which leaves 42 out of the 50 parameters as unimportant. The HDMR method is now applied by varying only the first eight parameters (a_i with values between 0 and 9), the other parameters remain fixed at their nominal values. Reducing the dimensionality of the problem allows an easier fitting of the polynomials for the HDMR component functions even for smaller sample sizes N . Because of the smaller input space dimension the HDMR method can now calculate the first-

and second-order sensitivity indices very accurately even for smaller sample sizes. Table 6.4 shows, that 1 024 samples are already enough to produce the correct ranking for the first- and second-order sensitivity indices. Applying the HDMR software using 4 096 samples produces almost the exact sensitivity indices if compared with the analytical values.

N	512	1 024	2 048	4 096	analytical
S_1	0.0145	0.0129	0.0104	0.0098	0.0096
S_2	0.0000	0.0124	0.0141	0.0127	0.0096
S_3	0.0667	0.0667	0.0697	0.0707	0.0727
S_4	0.2467	0.2671	0.2791	0.2834	0.2906
S_5	0.0779	0.0746	0.0734	0.0731	0.0727
S_6	0.0053	0.0034	0.0028	0.0029	0.0029
S_7	0.2412	0.2643	0.2775	0.2828	0.2906
S_8	0.0051	0.0038	0.0036	0.0033	0.0029
$\sum S_i$	0.6575	0.7053	0.7303	0.7386	0.7516
$S_{4,7}$	0.0504	0.0844	0.0918	0.0918	0.0969
$S_{3,4}$	0.0143	0.0218	0.0200	0.0238	0.0242
$S_{3,7}$	0.0113	0.0181	0.0211	0.0241	0.0242
$S_{4,5}$	0.0167	0.0256	0.0184	0.0241	0.0242
$S_{5,7}$	0.0122	0.0249	0.0215	0.0241	0.0242
$\sum S_{ij}$	0.1127	0.1840	0.1955	0.2215	0.2209
$\sum S_i + \sum S_{ij}$	0.7702	0.8893	0.9258	0.9602	0.9725

Table 6.4: First-order (first eight inputs) and second-order (selected input combinations) sensitivity indices for the Sobol' g-function using the combined screening + RS-HDMR approach.

Figure 6.3a presents the plot of the first-order component function $f_4(x_4)$, which is approximated by a tenth-order polynomial. The component function is able to capture the extreme non-linear behaviour of the model. The scatter plot is also shown for comparison. The most important second-order effect contributes by over 9 % to the overall variance and the corresponding component function is shown in figure 6.3b.

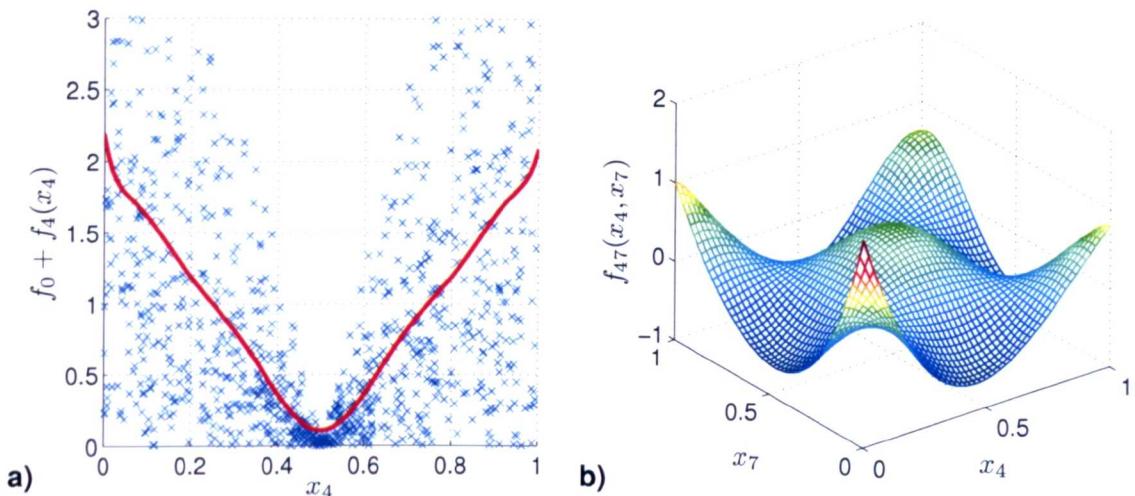


Figure 6.3: (a) First-order component function and scatter plot for parameter x_4 , (b) second-order component function for the parameter interaction between x_4 and x_7 for the Sobol' g-function using a sample size of $N = 4096$.

6.2.4 Discussion

The Sobol' g-function using 50 input parameters and the spectrum of a_i applied here is a difficult test case for global sensitivity analysis methods. However, the RS-HDMR method was able to correctly identify all of the important first-order effects and most of the important second-order effects by using a sample size of $N = 4096$. This is far less than other methods such as FAST (which can only provide first-order sensitivity indices) or the method of Sobol' would theoretically require. Due to the application of the threshold, a large number of unimportant first- and second-order component functions could be excluded from the HDMR expansion. This has further improved the accuracy of the metamodel and the estimated sensitivity indices. The prior application of a screening method for the Sobol' g-function with 50 parameters has been proven to be computationally more efficient. The Morris method required only $N = 510$ runs in order to identify eight potentially important parameters. The HDMR method could then accurately estimate the most important first- and second-order sensitivity indices based on the reduced input space dimension using only

$N = 2048$ model runs. In particular, the estimation of the parameter interactions (second-order effects) is more efficient in a smaller input space dimension. The application of the threshold still plays an important role, since many unimportant parameter interactions can be removed from the HDMR expansion even if the total number of parameters has already been narrowed down by the application of the screening method.

6.3 Case Study: Cyclohexane Oxidation

The focus in this study is on the investigation of the low temperature part of a chemical mechanism. This is different to the 1D premixed methane flame study in 5.5 where the focus was on the high temperature part. Here it was assumed that the flames are at steady state and the solutions are time independent. In general either burner stabilised flames or freely propagating flames are used to test the high temperature parts of a mechanism by evaluating their abilities to model species profiles, temperature profiles or flame speeds. In order to explore the low temperature behaviour of a mechanism either ignition behaviour can be explored for example in a closed vessel, shock tubes or rapid compression machines or flow reactors can be used to look at product analysis. Generally low temperature chemistry is more complex (i.e. more species and reactions have to be considered) but is very important in terms of engine applications (controlled ignition) for example for gasoline engines (induced ignition) or diesel engines (autoignition) and safety applications (uncontrolled ignition) (Pilling 1997). Because of the complexity and the wide range of applications for low temperature chemistry it is of interest to explore if global sensitivity analysis methods such as the HDMR method can be applied for all these different cases and if there are differences in terms of responses.

This case study presents a global sensitivity analysis of simulations of low temper-

ature isothermal cyclohexane oxidation under fuel rich conditions. A combination of the Morris method and the extended RS-HDMR method is applied. The analysis aims to investigate the important features of the oxidation process, as well as possible factors underlying qualitative discrepancies between simulations and experiments. The particular feature of interest is the characteristic of quadratic autocatalysis which is observed experimentally and leads to the maximum rate of reaction occurring at 50 % consumption of the deficient reactant (oxygen), with the fuel consumption exerting only a weak dependence. The kinetic mechanisms tested do not exhibit this characteristic when simulating the experimental conditions. The models also exhibit shorter induction times than those observed in the experiment. The results of this case study have also been presented at the 32nd International Symposium on Combustion (Ziehn et al. 2008*b*) and been submitted for publication (Ziehn et al. 2008*c*).

6.3.1 Introduction

Cyclohexane is an important surrogate compound for the use of cycloalkanes in transportation fuels with around 10 % in petrol, 15 – 70 % in kerosene and up to 35 % in diesel (McEnally & Pfefferle 2005, Silke et al. 2007). The oxidation of this class of hydrocarbons has been subject to several studies in the recent years. A summary and review of detailed low-temperature oxidation models is given in Battin-Leclerc (2008).

The oxidation of cyclohexane in fuel rich conditions presents also an autoignition hazard which is of concern in chemical process industries (Porter 2007). However, the kinetics of isothermal cyclohexane oxidation in fuel rich mixtures is poorly understood and not well captured in numerical simulations. The application of the HDMR method for global uncertainty and sensitivity analysis can help to investi-

gate the important features of the oxidation process and may be able to explain possible reasons for qualitative discrepancies between simulations and experiments.

Engine related experimental studies of the oxidation and autoignition of cyclohexane have been performed for example in a rapid compression machine over the compressed gas temperature range 670 – 870 K in Lemaire et al. (2001) and kinetic studies have been performed in a perfectly stirred, isothermal reactor at 807 K and 1050 K in El Bakali et al. (2000). Following these studies, numerical modelling of the cyclohexane oxidation process has been performed at both low and high temperatures for example by Silke et al. (2007) or Buda et al. (2006).

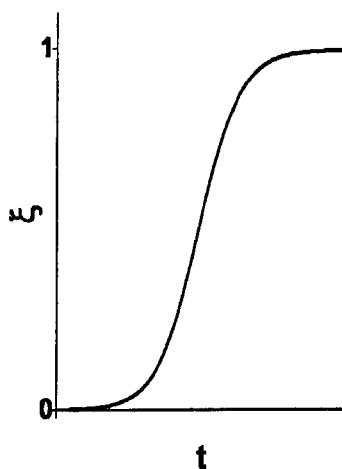


Figure 6.4: Theoretical curve of isothermal quadratic autocatalysis in a closed vessel (Snee & Griffiths 1989).

Due to the low minimum ignition temperature of MIT = 543 K (Jackson 1951) cyclohexane presents an industrial autoignition hazard. It is therefore necessary to assess the performance of the oxidation mechanism in these conditions. The oxidation studies performed in fuel rich conditions (Snee & Griffiths 1989) are of particular interest because they show very clearly the characteristics of quadratic autocatalysis (see figure 6.4) leading to the maximum rate of reaction occurring at 50 % consumption of the deficient reactant (oxygen), with the fuel consumption exerting only a very weak

dependence. Klai & Baronnet (1993) have reported the same autocatalytic profile for oxygen consumption during experimental studies of cyclohexane oxidation in a closed vessel at 635 K. It is of concern that zero-dimensional, numerical simulations using comprehensive models for cyclohexane (Silke et al. 2007, Buda et al. 2006) with existing kinetic data do not capture this fundamental kinetic property in a satisfactory way (see also figure 6.5).

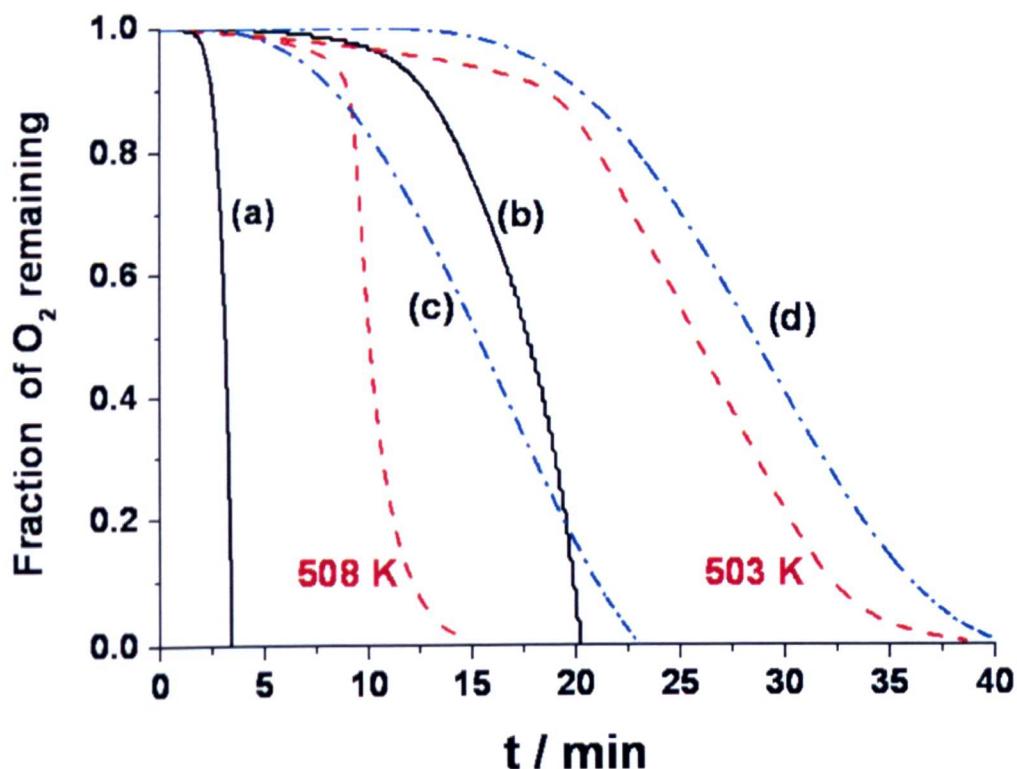


Figure 6.5: Experimental (Snee & Griffiths 1989) (dashed lines) and numerically simulated fractional O₂ consumption (a) EXGAS mechanism (Buda et al. 2006) (b) Silke mechanism (Silke et al. 2007) (c) Sample simulation using quasi-random inputs for A-factors (d) Sample simulation using quasi-random inputs for A-factors and wall losses. All simulations at T = 503 K.

Accurate modelling of cyclohexane oxidation is relevant to its commercial uses, and

comparison with available experimental data provides an important opportunity for evaluation of proposed mechanisms. The attention in this work is on the ability of the EXGAS generated cyclohexane scheme (Buda et al. 2006) to simulate the important kinetic features of alkane oxidation described above. Where discrepancies with experimental data are found, their possible sources are explored using global sensitivity methods. These allow the influence of uncertainties within the input parameters on output predictions to be investigated. The importance of parameters in terms of their contribution to the output variance can be ranked, helping to focus further kinetic studies on key parameters affecting model accuracy with respect to validation data, or to identify potential problems in the model structure. The latter may arise for example, from reaction steps that are missing from the mechanism.

HDMR methods have not yet been applied to models of low temperature oxidation kinetics. However, they have been successfully used to investigate the uncertainties for the high temperature part of a 1D premixed methane flame model (section 5.5). As already pointed out low temperature chemistry is usually more complex and the current case study therefore extends the scope of the methods and demonstrates their applicability to evaluating the main kinetic drivers of alkane oxidation, potentially highlighting sources of qualitative discrepancies with experiments (Snee & Griffiths 1989). Providing focus for future kinetic studies of key reaction steps is crucial, since as noted in Silke et al. (2007), cyclohexane is one of the simplest cycloalkanes, and its mechanistic data could be used in establishing rules for rate data estimation of relevance to larger fuels.

6.3.2 Evaluation of the Mechanism and Initial Comparison with Experiment

The present study focuses on the modelling of the isothermal kinetic response of fuel-rich mixtures as investigated experimentally in Snee & Griffiths (1989). The foundation of the low temperature scheme applied here originates from Buda et al. (2006) and comprises 499 species in 1 025 reversible reactions and 1 298 irreversible reactions. The mechanism was generated by the software EXGAS, which can be used to perform the automatic generation of detailed kinetic models for the gas-phase oxidation and combustion of linear and branched alkanes. The main features of the EXGAS software are described in Buda et al. (2005). The mechanism has been evaluated previously in Porter (2007) by constructing the equimolar cyclohexane+air $p - T_a$ ignition diagram and comparing it to the experimental cool flame boundaries in a closed vessel (Bonner & Tipper 1965). The qualitative structure of the cool flame boundaries as measured in the experiment is well captured by the $p - T_a$ diagram generated using the EXGAS mechanism. However, as has been noted in other cases (Hughes et al. 2006), the critical pressures tended to be under predicted, leading to the conclusion that the mechanism is too reactive in this region of operating conditions (Porter 2007).

For the global sensitivity studies, simulations of a reduced EXGAS scheme are used, comprising 60 species and 238 reactions (Fairweather et al. 2005). In the reduced scheme, reversible reactions from the original scheme are expressed as irreversible pairs since this is required for the purposes of redundant reaction removal. The 238 reaction scheme therefore represents the mechanism in irreversible form. The scheme was reduced to its skeletal form because using the comprehensive scheme would prohibit the subsequent HDMR analysis due to the numbers of simulations required to investigate higher numbers of input parameters, as well as the increased

cost of each simulation. The reduction was achieved using the automated application of local Jacobian and rate sensitivity based methods based on modifications of the well established KINALC software (Turányi 1997). Its validity has been assessed against the full mechanism over a wide range of temperature and pressure conditions (Porter 2007, Hughes et al. 2008) and the species profiles and carbon fluxes produced by the skeleton scheme have been shown to match well those of the full scheme for the conditions studied here.

Taking the initial stoichiometry,



the initial mixture approximated well to cyclohexane + “air” in 1 : 2 molar proportions, and was chosen to make direct comparisons with experimental results obtained at 1 atm (Snee & Griffiths 1989).

As outlined above, one would expect that for quadratic autocatalysis, the maximum rate of reaction $\frac{d\xi}{dt}$, where ξ is the fractional extent of reaction of the primary reactant, would occur at 50% reactant consumption. The characteristic sigmoidal curve of ξ versus t (illustrated schematically in figure 6.4) is observed widely in experiments, including for cyclohexane oxidation (Snee & Griffiths 1989) as shown in figure 6.5. However, the curves produced by the EXGAS model fail to show this important qualitative characteristic, and O_2 consumption accelerates until its virtual depletion. Isothermal conditions have been chosen specifically to remove the possible influence of uncertainties in the description of heat transfer processes. Therefore thermal effects cannot underlie this discrepancy and understanding the kinetic origin of this phenomenon requires further investigation. It is also clear that the time to $-(\frac{d\xi}{dt})_{\max}$ for O_2 are not well represented by the model. Again, the model appears to be over reactive in comparison to experimental data. Equivalent simulations using

the scheme of Silke et al. (2007) show a qualitatively similar acceleration (figure 6.5), and although this scheme appears to be less reactive than the EXGAS scheme under these conditions, it is still faster than the experiment as shown in the figure. The qualitative discrepancy is therefore not limited to a single reaction mechanism.

6.3.3 Selected Inputs and Outputs for Sensitivity Analysis

In order to explore these qualitative and quantitative discrepancies within the model, the selected target outputs are the time to $-(\frac{df}{dt})_{\max}$ for O₂ (y_1), and the percentage of O₂ remaining at the time of the maximum rate (y_2). Using the nominal values within the reaction scheme, y_2 is almost zero whereas the experiments indicate it should be close to 50 %. Global uncertainty techniques can establish whether the scheme is capable of reproducing the qualitative nature of the experimental profiles within certain bounds of uncertainty of the input parameters. If it cannot, then there may be key kinetic steps or physical processes missing from the model.

The inputs explored are the *A*-factors for the 238 irreversible reaction rates. The importance of wall losses for key radicals is also explored subsequently. A critical step in the evaluation of parameter importance is to define appropriate input distributions. Since a large number of parameters within the cyclohexane scheme were derived from estimates, input ranges were defined according to a minimum and maximum value, assuming equal probability throughout the ranges. In this study we are exploring qualitative discrepancies and therefore we have assumed conservatively large uncertainties of a factor five above and below the nominal value for each *A*-factor. It is accepted that in certain cases, where data evaluations are available, that this range may be an overestimation. However, in order to evaluate whether the qualitative behaviour can be recovered, for the large number of estimated reaction rates such uncertainties are justifiable.

6.3.4 Results

Since the number of uncertain parameters is considerably large, the Morris method is applied first in a two step algorithm. Screening methods are usually computationally less expensive, because only a low number of model simulations is required. The downside of applying the Morris method is, that it can not quantify how much one parameter is more important than another one. Additionally, in cases where interactions between parameters exist, the method fails to identify to which part of the model a parameter is interacting with. However, unimportant parameters can be identified and the RS-HDMR method is then used in the second step within a reduced input space dimension.

Parameter Screening Using the Morris Method

The Morris screening method (see section 2.3.2 and Hughes et al. (2006)) determines an importance ranking for parameters in terms of their mean effect on output variance as well as determining parameters with linear additive effects and non-linear interactions. It does not however provide the functional dependency of the output on individual parameters or parameter pairs as provided by the HDMR expansion. The Morris plot for output y_1 (time at point of inflexion for O₂) is presented in figure 6.6 and for output y_2 (O₂ ratio) in figure 6.7 for isothermal simulations at 503 K and 1 atm.

The Morris analysis for output y_2 identifies the same parameters as potentially important as the Morris analysis for output y_1 , but with a different ranking. A lower threshold of 1% of the highest mean effect across all parameters was used to exclude unimportant parameters in both cases. A subset of 33 A-factors were found to be important for the combined outputs which were used for further analysis. 100 trajectories ($r = 100$) have been used which leads to $N = 23\,900$ model simulations.

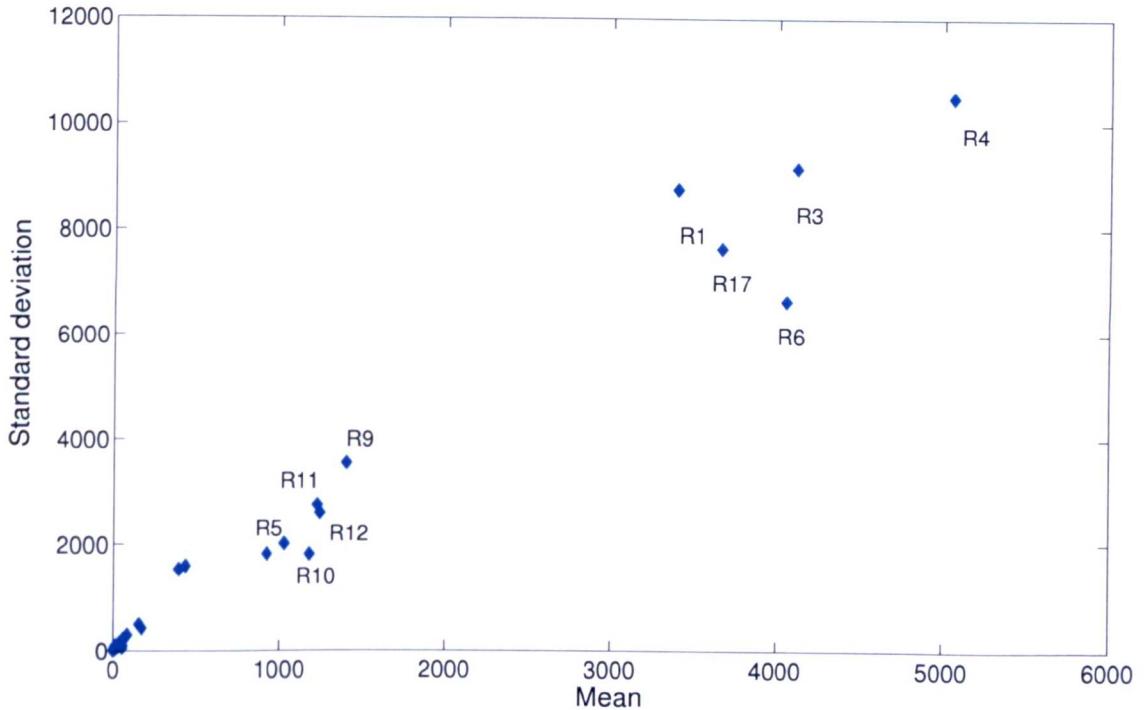


Figure 6.6: Morris plot for output y_1 (time at point of inflexion) using 100 trajectories. The ten most important parameters ranked by their mean are labelled.

Although this is not particularly efficient, it ensures that the correct number of potentially important parameters is identified. This cannot be achieved if using only a small number of trajectories (e.g. $r=10$) as suggested by Campolongo et al. (1999). It has to be highlighted, that even if using 100 trajectories the mean effect d_i (see equation 2.3) does not converge. However, because the HDMR method is applied in a second step it is not of importance here to obtain a reliable ranking for the most important parameters. The main focus is on identifying the unimportant parameters and potentially important ones and this can be achieved with 100 trajectories even if the mean effect d_i does not converge. Hughes et al. (2006) experienced a similar problem when applying the Morris method to investigate the uncertainties in rate parameters on the ignition delay time for the low temperature combustion of propane. A set of 1 000 trajectories was required in this case in order to ensure

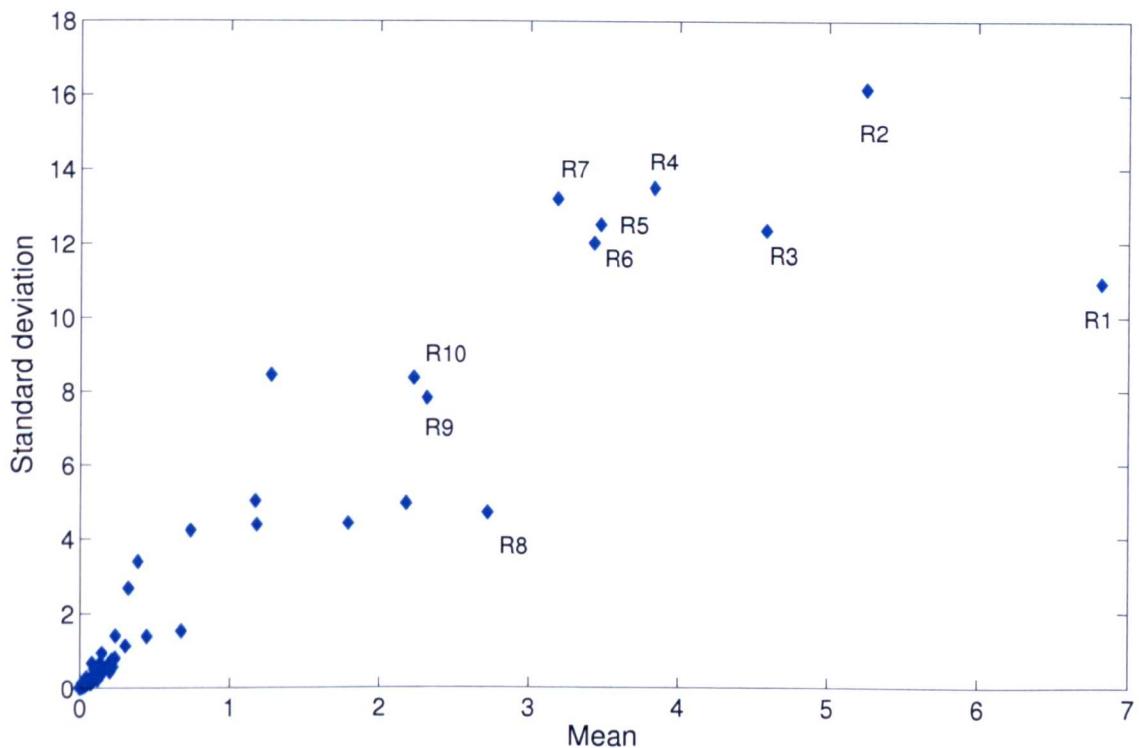


Figure 6.7: Morris plot for output y_2 (O_2 ratio) using 100 trajectories. The ten most important parameters ranked by their mean are labelled.

convergence. Consequently, it seems to be a common problem that a large number of Morris runs is required if investigating uncertainties in rate parameters for the low temperature part of a chemical mechanism. This also confirms the comments made in section 2.3.2 when the Morris method was introduced, that it is important to assess the convergence of the mean effect and that a general guide line for the number of trajectories of $5 \leq r \leq 15$ is not sensible.

It is worth pointing out that the Morris method explores the wider input space and is therefore more appropriate for initial parameter screening than local sensitivity methods. For example, a potential case may arise where the scheme shows low sensitivity to a parameter around its nominal value, but increasing sensitivity away from this value towards the edges of its range. Using a local method, this parameter would be screened out of the subsequent global HDMR analysis. Using the Morris

method, the whole of the input space of the parameter would be explored and hence the parameter would be retained if its average effect was above the chosen threshold. Table 6.5 presents the 20 most important reactions according to the Morris analysis for the combined outputs using the ranking for output y_2 .

No.	Reaction
R1	$C_6H_{11}OO + HO_2 \rightarrow C_6H_{11}OOH + O_2$
R2	$C_6H_{12} + OH \rightarrow C_6H_{11} + H_2O$
R3	$2C_6H_{11}OO \rightarrow C_5H_{10}CO + C_6H_{11}OH + O_2$
R4	$2C_6H_{11}OO \rightarrow 2C_6H_{11}O + O_2$
R5	$CH_2CHO + O_2 \rightarrow HCHO + OH + CO$
R6	$C_6H_{11}OOH \rightarrow C_6H_{11}O + OH$
R7	$C_6H_{11} + O_2 \rightarrow C_6H_{10} + OOH$
R8	$CH_2CHO + OOH \rightarrow O_2 + CH_3CHO$
R9	$CHOC_4H_8COOOH \rightarrow OH + 3CH_2CHO$
R10	$C_6H_{12} + OOH \rightarrow C_6H_{11} + H_2O_2$
R11	$CH_2CHO + O_2 \rightarrow CH_2CO + OOH$
R12	$C_6H_{11} + O_2 \rightarrow C_6H_{11}OO$
R13	$HCHO + OH \rightarrow CHO + H_2O$
R14	$2OOH \rightarrow H_2O_2 + O_2$
R15	$CH_2CO + OH \rightarrow CH_2OH + CO$
R16	$CH_2CO + OH \rightarrow CH_3 + CO_2$
R17	$C_6H_{12} + C_6H_{11}OO \rightarrow C_6H_{11} + C_6H_{11}OOH$
R18	$CH_3OO \rightarrow CH_3O + O_2$
R19	$OOH + HCHO \rightarrow CHO + H_2O_2$
R20	$C_6H_{10}OH \rightarrow C_2H_3 + C_2H_4 + CH_3CHO$

Table 6.5: Potentially important reactions. The reaction number refers to the ranking from the Morris analysis for output y_2 .

HDMR Analysis and Sensitivity Indices

The HDMR analysis is focused on isothermal simulations at 503 K and 1 atm using the 33 parameters identified by the Morris method, although the results are similar for other temperatures relating to the experimental profiles in Snee & Griffiths (1989). Table 6.6 and table 6.7 present the five most significant first and second-order sensitivity indices for output y_1 and output y_2 respectively. The results are

Parameter x_i		
No.	Reaction	S_i
R4	$2\text{C}_6\text{H}_{11}\text{OO} \rightarrow 2\text{C}_6\text{H}_{11}\text{O} + \text{O}_2$	0.1955
R6	$\text{C}_6\text{H}_{11}\text{OOH} \rightarrow \text{C}_6\text{H}_{11}\text{O} + \text{OH}$	0.1588
R9	$\text{CHOC}_4\text{H}_8\text{COOOH} \rightarrow \text{OH} + 3\text{CH}_2\text{CHO}$	0.1282
R3	$2\text{C}_6\text{H}_{11}\text{OO} \rightarrow \text{C}_5\text{H}_{10}\text{CO} + \text{C}_6\text{H}_{11}\text{OH} + \text{O}_2$	0.0932
R17	$\text{C}_6\text{H}_{12} + \text{C}_6\text{H}_{11}\text{OO} \rightarrow \text{C}_6\text{H}_{11} + \text{C}_6\text{H}_{11}\text{OOH}$	0.0759
$\sum S_i$		0.7447
Parameter x_i	Parameter x_j	S_{ij}
R4	R6	0.0377
R3	R4	0.0268
R4	R17	0.0221
R6	R9	0.0142
R9	R17	0.0141
$\sum S_{ij}$		0.1944
$\sum S_i + \sum S_{ij}$		0.9391

Table 6.6: First- and second-order sensitivity indices using $N = 131\,072$ and 33 inputs for output y_1 (time at point of inflexion). Parameters x_i and x_j are the A -factors for the forward reaction rate.

Parameter x_i		
No.	Reaction	S_i
R4	$2\text{C}_6\text{H}_{11}\text{OO} \rightarrow 2\text{C}_6\text{H}_{11}\text{O} + \text{O}_2$	0.1544
R1	$\text{C}_6\text{H}_{11}\text{OO} + \text{HO}_2 \rightarrow \text{C}_6\text{H}_{11}\text{OOH} + \text{O}_2$	0.0875
R3	$2\text{C}_6\text{H}_{11}\text{OO} \rightarrow \text{C}_5\text{H}_{10}\text{CO} + \text{C}_6\text{H}_{11}\text{OH} + \text{O}_2$	0.0283
R2	$\text{C}_6\text{H}_{12} + \text{OH} \rightarrow \text{C}_6\text{H}_{11} + \text{H}_2\text{O}$	0.0131
R12	$\text{C}_6\text{H}_{11} + \text{O}_2 \rightarrow \text{C}_6\text{H}_{11}\text{OO}$	0.0088
$\sum S_i$		0.3104
Parameter x_i	Parameter x_j	S_{ij}
R3	R4	0.1190
R1	R4	0.0616
R1	R3	0.0299
R1	R2	0.0194
R1	R12	0.0141
$\sum S_{ij}$		0.3607
$\sum S_i + \sum S_{ij}$		0.6711

Table 6.7: First- and second-order sensitivity indices using $N = 131\,072$ and 33 inputs for output y_2 (O_2 ratio). Parameters x_i and x_j are the A -factors for the forward reaction rate.

based on a quasi-random sample of the size $N = 131\,072$. This is a very large sample size, however it proved to be necessary in order to have enough confidence in the estimated sensitivity indices. Various samples of increasing sizes (512, 1 024, 2 048, ...) were used first and the accuracy of the constructed metamodel was assessed. Due to the fact that a satisfactory accuracy could not be reached for both outputs y_1 and y_2 the sample size was successively increased. However, even if using a large sample size of $N = 131\,072$ it was found that not the whole output variance could be explained by considering only up to second-order effects (second-order RS-HDMR expansion). It can be assumed that the estimated sensitivity indices up to second-order are trustworthy, because the accuracy of the metamodel and the values of the sensitivity indices for the most important parameters converged. This also means that higher order effects (≥ 3) are present which can not be accounted for if using a second-order HDMR expansion. This is in contrast to the applications of HDMR to sulphur chemistry in the premixed methane flame model under high temperature in section 5.5, where mostly first-order effects were observed and a smaller sample size could be used to efficiently calculate these effects.

Table 6.8 presents the accuracy of the first- and second-order HDMR metamodel for increasing sample sizes N for both model outputs in terms of the coefficient of determination r^2 . A small sample size of $N = 512$ is sufficient in order to estimate the most important first-order effects (S_i) for both outputs. The accuracy of the first-order HDMR metamodel does not significantly change even when a large sample size $N = 131\,072$ is used. However, a much larger sample size is necessary in order to estimate the second-order effects (S_{ij}) and $N = 4\,096$ samples are required in order to identify the most important parameter interactions. The values of highest ranked second-order sensitivity indices does not change significantly when further increasing the sample size up to $N = 131\,072$.

The accuracy of the constructed HDMR metamodel based on $N = 131\,072$ is also

N	r^2 value in % for output:		
	y_1 time at point of inflexion	y_2	O_2 ratio
512			
1st-order	75.07		26.49
2nd-order	84.64		31.11
1024			
1st-order	76.87		27.59
2nd-order	91.37		38.90
2048			
1st-order	76.91		26.91
2nd-order	90.64		39.47
4096			
1st-order	76.29		25.86
2nd-order	89.03		54.34
:			
131 072			
1st-order	75.94		27.01
2nd-order	94.28		60.31

Table 6.8: Accuracy of the RS-HDMR metamodel (first- and second-order) in terms of the coefficient of determination r^2 for both model outputs (y_1 time at the point of inflexion and y_2 O_2 ratio) using different sample sizes N .

shown in figure 6.8 in terms of a scatter plot and in figure 6.9 by comparing the pdf for both model outputs. In general, the accuracy of the metamodel is better for output y_1 (time at point of inflexion) and as shown in the scatter plot in figure 6.8c a second-order HDMR expansion provides a reasonable approximation of the model output indicated by the low amount of scatter and by the coefficient of determination $r^2 = 94.28\%$. This is also confirmed by the comparison of the pdf's (figure 6.9c) which show a good match. In contrast to the other case studies, in this work the pdf for output y_1 is not Gaussian shaped, it is skewed with a long tail. It can also be seen, that the majority of the performed model simulations have a very small time to $-(\frac{dx}{dt})_{\max}$ of around 75s. Only a couple of input parameter combinations allow a time to $-(\frac{dx}{dt})_{\max}$ of over 600s which is still far too small if compared with the experiment. The pdf for the first-order HDMR metamodel presented in figure 6.9a

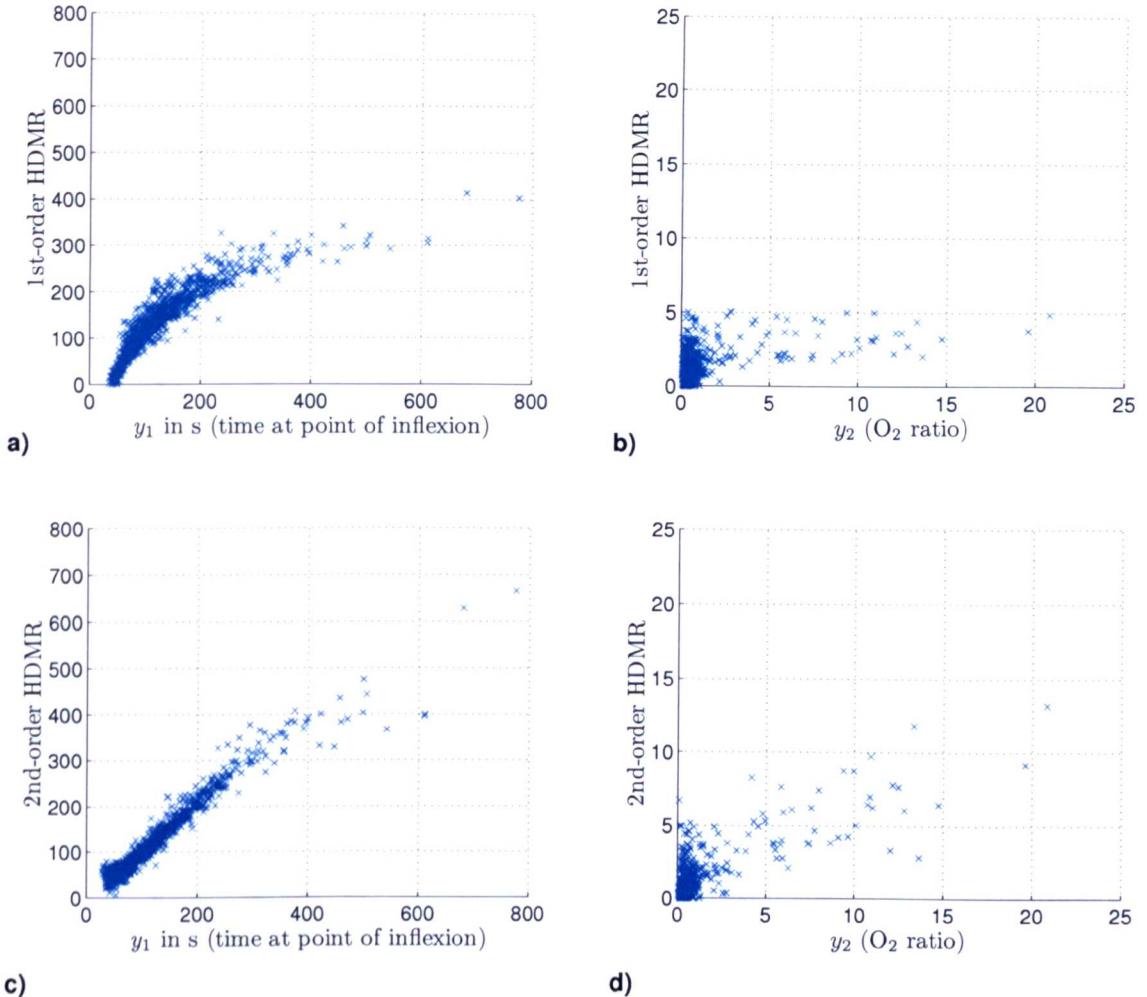


Figure 6.8: Accuracy of the RS-HDMR metamodel for $N = 131\,072$ based on the scatter plot (1 000 random points) for (a) the time at point of inflection using a first-order HDMR metamodel (b) the O_2 ratio using a first-order HDMR metamodel, (c) the time at point of inflection using a second-order HDMR metamodel and (d) the O_2 ratio using a second-order HDMR metamodel.

has Gaussian like shape and the tail is missing. This indicates, that higher-order effects are “within” the tail and it would suggest that model outputs with a tailed distribution are effected by parameter interactions. This will be further investigated in section 8.1 when the focus is on the exploration of higher order ≥ 3 effects. A second-order HDMR expansion is definitely not sufficient in order to construct a metamodel for output y_2 (O_2 ratio) as shown by the scatter plot in figure 6.8d. The

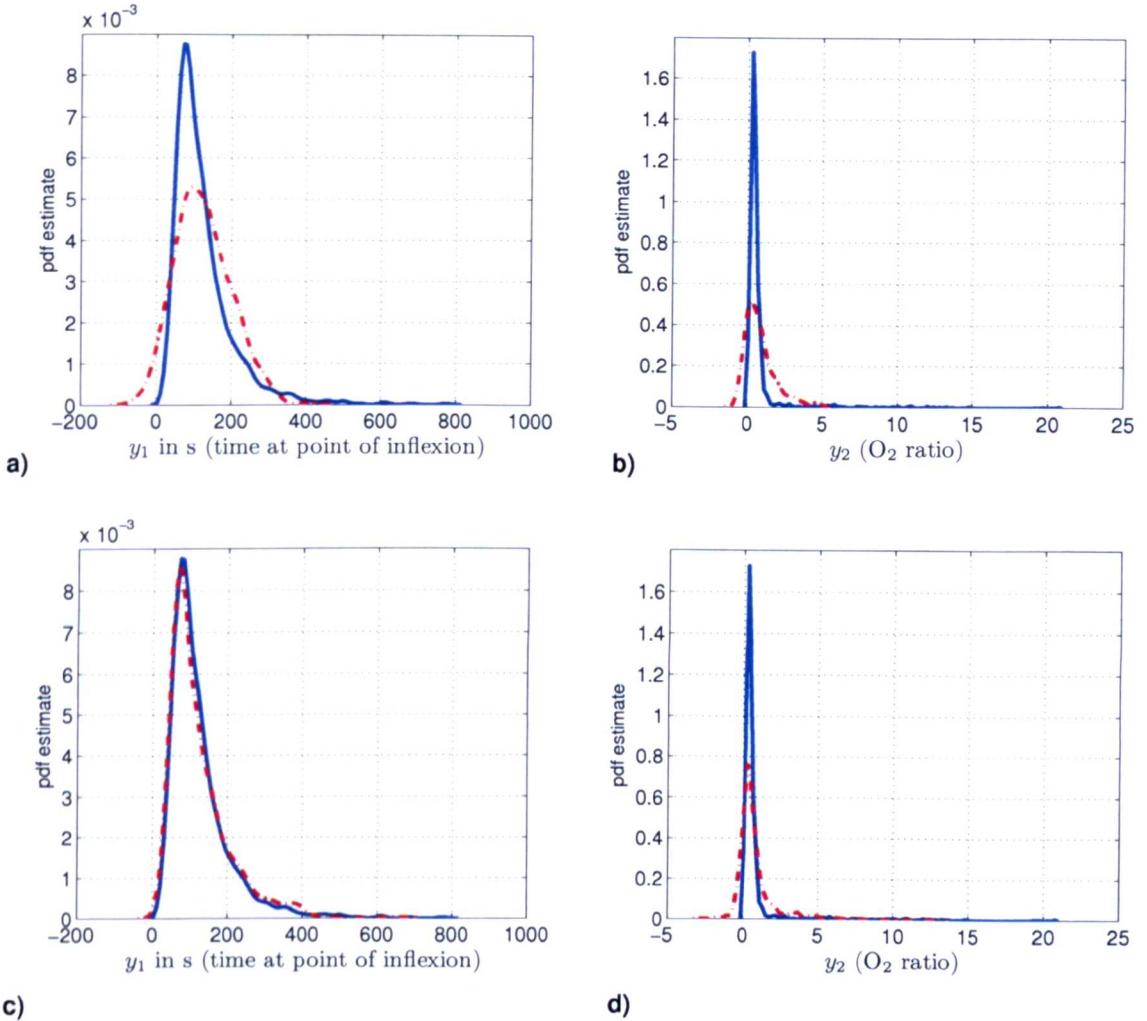


Figure 6.9: Accuracy of the RS-HDMR metamodel for $N = 131\,072$ based on the probability density function (pdf) using 1 000 random points for (a) the time at point of inflection using a first-order HDMR metamodel (b) the O_2 ratio using a first-order HDMR metamodel, (c) the time at point of inflection using a second-order HDMR metamodel and (d) the O_2 ratio using a second-order HDMR metamodel.

cloud of scatter is still quite big and the coefficient of determination $r^2 = 60.31\%$ confirms the poor accuracy. The majority of the points are close to zero, which means that oxygen is fully consumed at $-(\frac{d\xi}{dt})_{\max}$. Only a couple of points are in the region of 20 % O_2 ratio, which is still far from the experimental results (50 %). The pdf plot in figure 6.9b,d shows a very pointy distribution with a very long tail.

The HDMR metamodel is not able to match this distribution and it is assumed that significant higher order effects (≥ 3) exist, which of course can not be captured by a second-order HDMR expansion.

As shown in table 6.7, first-order effects account for almost 75 % of the variability of y_1 but only 31 % of y_2 . For y_1 , the sum of first- and second-order effects describes 94 % of the output variance and therefore higher order effects are deemed to be insignificant. For y_2 , the sum of first- and second-order effects describes only 67 % of the output variance and in this case effects of higher order are present. This is also in contrast to the statement by Li et al. (2001b) that usually higher order HDMR component functions are smaller than low order ones. It has been true for all the other investigated case studies in chapters 4 and 5, but it does not seem to apply for this low temperature chemistry example.

The calculation of the higher order effects (≥ 3) requires prohibitively large numbers of model runs and is therefore not discussed here. Any additional parameters that contribute only to these higher order effects would have been highlighted by the Morris analysis. However, the highly ranked parameters from the Morris analysis are present within the main first- and second-order effects. The presence of second and higher order effects indicates significant parameter interactions.

The largest first-order sensitivity for both outputs is assigned to the A -factor for the reaction of cyclohexylperoxy radicals ($C_6H_{11}OO$) to form the cyclohexylalkoxy radical ($C_6H_{11}O$). Within the EXGAS mechanism this reaction forms a major flux of carbon atoms under low temperature conditions as indicated in figure 6.10. The second ranked parameter in terms of y_1 is the A -factor for the secondary route to $C_6H_{11}O$ via cyclohexyl peroxide ($C_6H_{11}OOH$). Following the formation of $C_6H_{11}O$, a series of fast reactions take place leading to O_2 consumption and the formation of the lumped aldohydroperoxide. The A -factor for this reaction of aldohydroperox-

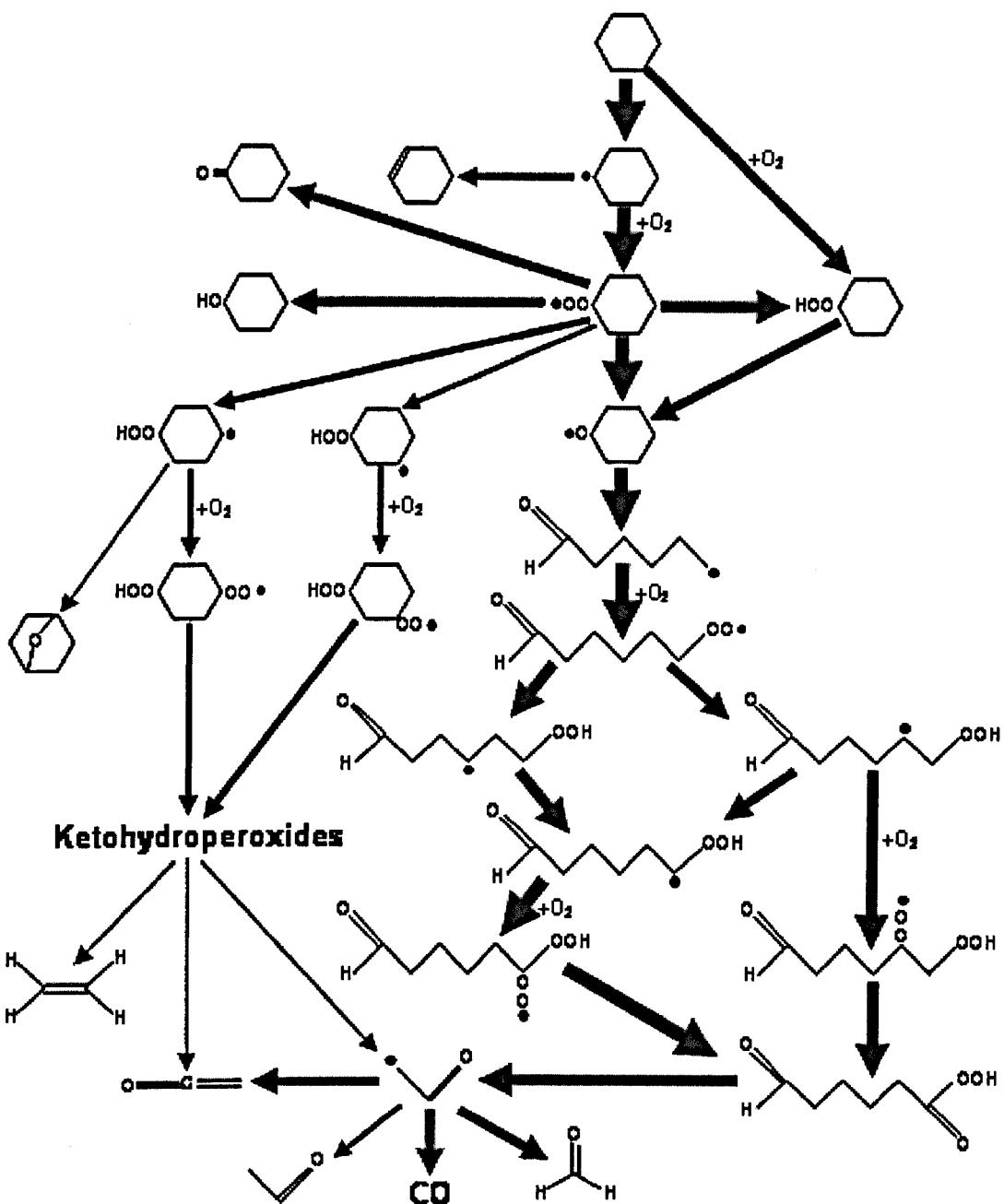


Figure 6.10: Carbon element flux analysis for isothermal oxidation using the 60 species 238 reactions mechanism: 1:2 cyclohexane in air, 470 K, 1 atm. Arrow thickness scaled to magnitude of element flux (Porter 2007).

ide also has a high first-order sensitivity with respect to y_1 . The time to $-(\frac{dy}{dt})_{\max}$ is therefore controlled by the route on the right of the flux diagram under these low temperature conditions (figure 6.10). At low temperatures and in agreement

with Snee & Griffiths (1989), within our simulations the isomerisation reactions of $C_6H_{11}OO$ have less significant carbon fluxes (figure 6.10), than at the higher temperatures studied in Silke et al. (2007), since they have strong positive temperature dependences. At $T = 503\text{ K}$, there are no significant sensitivities of either output to their A -factors (tables 6.6 and 6.7). One additional reaction of importance for y_2 is $C_6H_{11}OO + HO_2 \longrightarrow C_6H_{11}OOH + O_2$ (R1). This has the second highest first-order sensitivity coefficient for y_2 but also acts cooperatively with several other parameters to give significant second-order effects. Overall, peroxy-peroxy reactions seem to exhibit high sensitivities for both outputs and therefore are critical in determining the overall reactivity of the scheme at the temperatures studied.

HDMR Component Functions and Scatter Plots

Whilst the sensitivity indices give a useful overall picture of the parameters that contribute most to the output variance, they do not allow the exploration of the effect of each parameter across its input range. As shown in other case studies (e.g. high temperature premixed flame model in section 5.5), the indices themselves are also highly dependant on the uncertainty ranges chosen for the parameters, which in most cases have to be estimated. One advantage of the HDMR component functions is that for each parameter they allow individual effects or paired effects to be investigated automatically across the chosen input range. The shapes of the component functions demonstrate how sensitivities change across the range. Significant first-order component functions with respect to y_1 are shown in figure 6.11. The A -factor for $2C_6H_{11}OO \longrightarrow 2C_6H_{11}O + O_2$ (R4) had the highest overall first-order sensitivity index, but it can be seen that its sensitivity varies significantly across the input range. At the upper end of the range the parameter exhibits very low sensitivity, but at lower values, the sensitivity becomes strongly negative. The rate constant used in the Nancy EXGAS scheme is $1.046 \cdot 10^{-13} e^{(365/T)} \text{ cm}^{-3} \text{ molecule}^{-1} \text{ s}^{-1}$ and

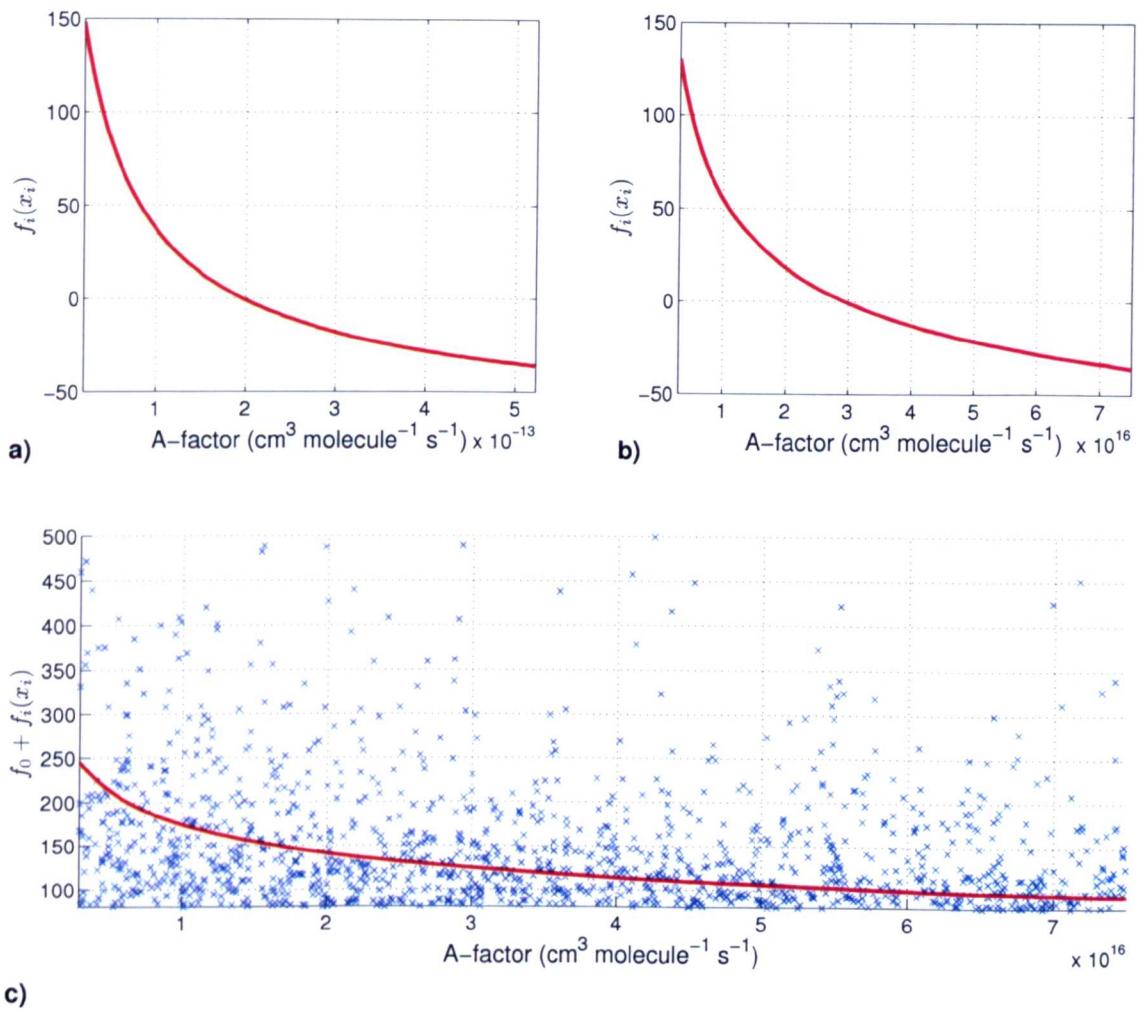


Figure 6.11: First-order component function f_i for (a) $2\text{C}_6\text{H}_{11}\text{OO} \longrightarrow 2\text{C}_6\text{H}_{11}\text{O} + \text{O}_2$, (b) $\text{C}_6\text{H}_{11}\text{OOH} \longrightarrow \text{C}_6\text{H}_{11}\text{O} + \text{OH}$ and scatter plot for (c) $\text{CHOC}_4\text{H}_8\text{COOOH} \longrightarrow \text{OH} + 3\text{CH}_2\text{CHO}$ and output y_1 (time at point of inflection). In (c) the mean f_0 is added to f_i for comparison with scatter plot.

Figure 6.11a implies that as the A -factor drops below its nominal value it has a strong negative sensitivity. The low temperature (atmospheric chemistry) measurements of Lightfoot et al. (1992) suggest a rate of $0.769 \cdot 10^{-13} e^{(-184/T)} \text{ cm}^{-3} \text{ molecule}^{-1} \text{ s}^{-1}$ which is about a factor of four lower than the Nancy EXGAS parameter value at 503 K. Given the high sensitivity to this reaction, further kinetic studies across wide temperature ranges would clearly be useful.

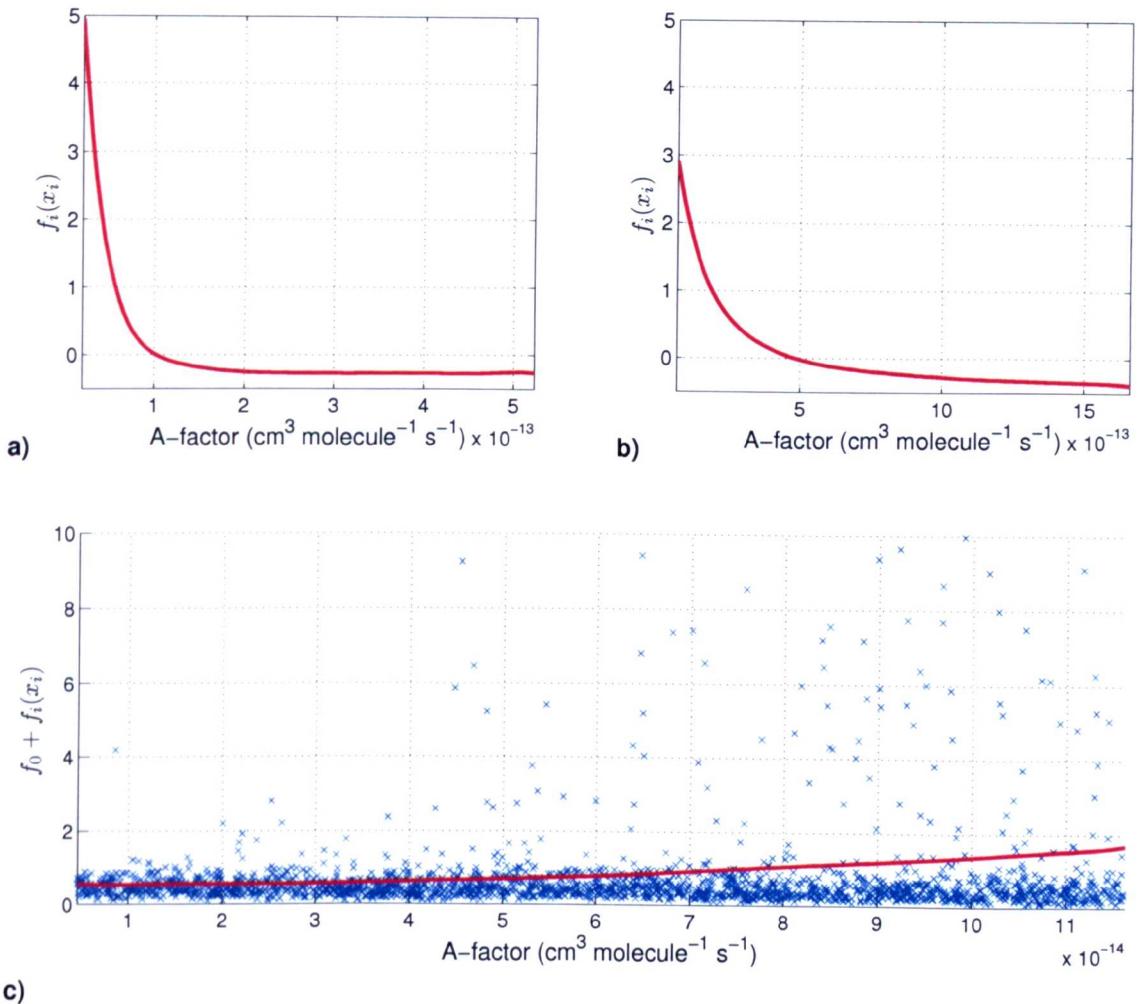


Figure 6.12: First-order component function f_i for (a) $2\text{C}_6\text{H}_{11}\text{OO} \rightarrow 2\text{C}_6\text{H}_{11}\text{O} + \text{O}_2$, (b) $\text{C}_6\text{H}_{11}\text{OO} + \text{HO}_2 \rightarrow \text{C}_6\text{H}_{11}\text{OOH} + \text{O}_2$ and scatter plot for (c) $2\text{C}_6\text{H}_{11}\text{OO} \rightarrow \text{C}_5\text{H}_{10}\text{CO} + \text{C}_6\text{H}_{11}\text{OH} + \text{O}_2$ and output y_2 (O₂ ratio). In (c) the mean f_0 is added to f_i for comparison with scatter plot.

Figures 6.11b,c show similar non-linear negative sensitivities for A -factors for reactions $\text{C}_6\text{H}_{11}\text{OOH} \rightarrow \text{C}_6\text{H}_{11}\text{O} + \text{OH}$ (R6) and $\text{CHOC}_4\text{H}_8\text{COOOH} \rightarrow \text{OH} + 3\text{CH}_2\text{CHO}$ (R9), highlighting the overall sensitivity to the rates through the alkoxy radical and aldehydroperoxide to final products. The scatter plot from a standard Monte Carlo simulation based on randomly selected input parameters is also shown in figure 6.11c. This highlights the advantage of the HDMR component functions, since they show directly the effect of each individual parameter on the selected

output, which is obscured in the scatter plot by that of other parameters.

Figure 6.12 shows the component functions for the highest ranked parameters with respect to y_2 . Again the A -factor for reaction $2\text{C}_6\text{H}_{11}\text{OO} \rightarrow 2\text{C}_6\text{H}_{11}\text{O} + \text{O}_2$ (R4) shows a highly non-linear negative first-order effect, although the component function shows that this saturates with respect to y_2 at high values of the A -factor. The A -factor for $\text{C}_6\text{H}_{11}\text{OO} + \text{HO}_2 \rightarrow \text{C}_6\text{H}_{11}\text{OOH} + \text{O}_2$ (R1) also shows a non-linear negative sensitivity with respect to the O_2 ratio. This reaction has also been studied previously because of its importance in the troposphere, with the rate suggested by Rowley et al. (1992) being faster than the rate used in the EXGAS scheme by over a factor of two. Studies at higher temperatures of relevance to combustion would be very useful. Figure 6.12c indicates a positive sensitivity to the A -factor for $2\text{C}_6\text{H}_{11}\text{OO} \rightarrow \text{C}_5\text{H}_{10}\text{CO} + \text{C}_6\text{H}_{11}\text{OH} + \text{O}_2$ (R3). This route is competitive to that through the alkoxy radical and an increase in its rate would have a positive effect on the O_2 ratio at the point of maximum reaction rate. The reason may be that a displacement to this channel reduces the build up of O_2 -consuming intermediates that are formed through the formation of the alkoxy radical. $\text{C}_6\text{H}_{11}\text{OH}$ produces the vinoxy radical, but via reactions involving the radical pool rather than O_2 consuming reactions. The relative product channels for the reaction of $\text{C}_6\text{H}_{11}\text{OO}$ seem therefore to be critical in terms of the rate of O_2 consumption.

This point can be further explored through the consideration of second-order sensitivities. The combined effect of reactions R3,R4 is seen (table 6.7) to have a major contribution to the output variance for y_2 . Figure 6.13 further explores this through the second-order component function for reactions R3,R4. It shows that a strong effect on y_2 could occur by shifting the relative rates of these two reactions in favour of reaction R3. The maximum value of y_2 simulated across all 131 072 runs was however 41 % (figure 6.5(c)). Table 6.9 shows the parameter configuration (sample of the Sobol' sequence) for the four most important parameters which led to the maxi-

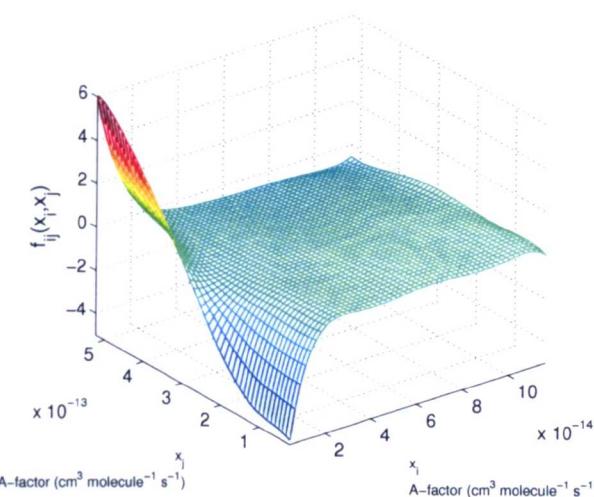


Figure 6.13: Second-order component function $f_{ij}(x_i, x_j)$ for y_2 (O_2 ratio) for x_i : $2C_6H_{11}OO \longrightarrow C_5H_{10}CO + C_6H_{11}OH + O_2$ and x_j : $2C_6H_{11}OO \longrightarrow 2C_6H_{11}O + O_2$.

mum value of y_2 . The values of parameter R4 (A -factor for reaction $2C_6H_{11}OO \longrightarrow 2C_6H_{11}O + O_2$) and R1 (A -factor for $C_6H_{11}OO + HO_2 \longrightarrow C_6H_{11}OOH + O_2$) are close to the lower limits of their uncertainty ranges. This is in accordance with their input-output relationships presented in figure 6.12a,b which suggests that a higher O_2 ratio would be possible if these parameters values are chosen closer to their lower limits. The value for parameter R3 (A -factor for $2C_6H_{11}OO \longrightarrow C_5H_{10}CO + C_6H_{11}OH + O_2$) is close the upper limit of its uncertainty range. Again, this is in accordance with the input-output relationship shown in figure 6.12c. In this case a positive relationship exists and a higher O_2 ratio would be possible if this parameter is chosen closer to its upper limit.

So it is possible to recover an S-shaped curve at 503 K within the uncertainty limits for the A -factors, but the 50 % O_2 consumption at $-(\frac{dc}{dt})_{max}$ is still not quite recovered and the induction time is still shorter than the experiment. This indicates, that there may be reactions missing in the chemical mechanism, which will be further investigated in the following section.

Parameter	Minimum	Maximum	Nominal Value	Set-up
R4	$2.092 \cdot 10^{-14}$	$5.230 \cdot 10^{-13}$	$1.046 \cdot 10^{-13}$	$2.583 \cdot 10^{-14}$
R1	$6.642 \cdot 10^{-14}$	$1.661 \cdot 10^{-12}$	$3.321 \cdot 10^{-13}$	$1.072 \cdot 10^{-13}$
R3	$4.650 \cdot 10^{-15}$	$1.163 \cdot 10^{-13}$	$2.325 \cdot 10^{-14}$	$1.067 \cdot 10^{-13}$
R2	$2.558 \cdot 10^{-17}$	$6.395 \cdot 10^{-14}$	$2.492 \cdot 10^{-16}$	$2.558 \cdot 10^{-17}$

Table 6.9: Set-up (four most important parameters) which led to the maximum value (41 %) of y_2 . Parameters are the A -factors for the forward reaction rate in $\text{cm}^{-3} \text{molecule}^{-1} \text{s}^{-1}$.

Possible Effect of Wall Losses

The influence of wall losses has been highlighted in previous work as a possible reason for the over reactivity of alkane oxidation mechanisms (Hughes et al. 2006). A second set of simulations was therefore performed where ten additional reactions were added to the skeletal mechanism, adding ten additional inputs (wall loss rates) to the 33 A -factors discussed above. Wall losses were added for peroxy radicals and peroxide species, in line with previous work (Hughes et al. 2006). Tests indicated that other radicals were too reactive for wall losses to be significant. The lower limit for each wall loss rate was zero and the upper limit was determined according to $\frac{3\gamma\bar{c}}{2d}$ (Baldwin & Howarth 1982), where d is the vessel diameter and \bar{c} is the radical average molecular velocity, with wall loss efficiencies (γ) of 10^{-3} for the peroxy radicals and 10^{-5} for the peroxide species. This is a conservatively large upper limit for an unstirred reactor under isothermal conditions but allows the possible effect of highly uncertain wall loss rates to be fully explored.

An example simulation from the quasi-random samples is shown in figure 6.5(d), indicating that the inclusion of wall losses can have the effect of increasing the induction time. However, table 6.10 and table 6.11 indicate that few of the wall losses exhibit strong first-order sensitivities for either output. The loss rate for $\text{C}_6\text{H}_{11}\text{OO}$ has the strongest first-order effect and acts cooperatively within second-

Parameter x_i No.	Reaction	S_i
R6	$C_6H_{11}OOH \rightarrow C_6H_{11}O + OH$	0.0613
R17	$C_6H_{12} + C_6H_{11}OO \rightarrow C_6H_{11} + C_6H_{11}OOH$	0.0562
R9	$CHOC_4H_8COOOH \rightarrow OH + 3CH_2CHO$	0.0145
R34	$C_6H_{11}OO \rightarrow wall$	0.0144
R4	$2C_6H_{11}OO \rightarrow 2C_6H_{11}O + O_2$	0.0139
$\sum S_i$		0.1799
Parameter x_i	Parameter x_j	S_{ij}
R6	R17	0.0346
R17	R34	0.0205
R6	R34	0.0144
$\sum S_{ij}$		0.1489
$\sum S_i + \sum S_{ij}$		0.3288

Table 6.10: First- and second-order sensitivity indices using $N = 131\,072$ and 43 inputs for output y_1 (time at point of inflection). Parameters x_i and x_j are the A -factors for the forward reaction rate.

Parameter x_i No.	Reaction	S_i
R4	$2C_6H_{11}OO \rightarrow 2C_6H_{11}O + O_2$	0.1486
R1	$C_6H_{11}OO + HO_2 \rightarrow C_6H_{11}OOH + O_2$	0.0273
R3	$2C_6H_{11}OO \rightarrow C_5H_{10}CO + C_6H_{11}OH + O_2$	0.0223
R6	$C_6H_{11}OOH \rightarrow C_6H_{11}O + OH$	0.0158
R43	$C_6H_{11}OOH \rightarrow wall$	—
$\sum S_i$		0.2441
Parameter x_i	Parameter x_j	S_{ij}
R3	R4	0.0935
R4	R6	0.0514
R1	R4	0.0224
R4	R43	0.0155
R1	R3	0.0104
$\sum S_{ij}$		0.3274
$\sum S_i + \sum S_{ij}$		0.5715

Table 6.11: First- and second-order sensitivity indices using $N = 131\,072$ and 43 inputs for output y_2 (O_2 ratio). Parameters x_i and x_j are the A -factors for the forward reaction rate.

order sensitivities as does that for C₆H₁₁OOH. The inclusion of wall losses clearly leads to significant higher order interactions between parameters, which is potentially due to their higher significance at lower radical concentrations, which in turn depends on the values of the *A*-factors chosen.

The inclusion of wall losses does lead to a higher number of simulations with values of *y*₂ close to 50 % and realistic induction times. Wall losses are clearly of potential importance and their interaction with other parameters should be more fully explored in further work.

6.3.5 Conclusions

It is clear from the component functions for both outputs that there are strong negative sensitivities for reactions of C₆H₁₁OO that lead to significant carbon fluxes through the alkoxy radical. This suggests that if this route was slower, the time to maximum reaction rate would be longer and the maximum reaction rate would not occur at complete oxygen consumption. In contrast, positive sensitivities were found for the rate of formation of C₅H₁₀CO and C₆H₁₁OH from C₆H₁₁OO. The low temperature product channels for reactions of the cyclohexylperoxy radical are therefore an important area for future kinetic studies. The effects of wall reactions of peroxy and peroxide species were not found to outweigh the impact of the main *A*-factors, but including wall losses led to significant higher order interactions between input parameters which forms an interesting area for further research.

The two step approach (Morris method + RS-HDMR) was able to highlight important parameters and parameter interactions (for two inputs). However, even if taking into account the uncertainties in the rate parameters the long time to $-(\frac{df}{dt})_{\max}$ and an O₂ ratio of 50 % as suggested by the experimental results could not be recovered. This suggested, that there were reactions missing in the scheme and therefore wall

losses were included by adding ten additional reactions. This inclusion of the wall losses led to a higher number of simulations with realistic induction times and a O_2 ratio close to 50 %.

The low temperature application of the alkane oxidation mechanism investigated in this study has been a challenge for the application of the HDMR method. In previous studies (e.g. sulphur-nitrogen chemistry interaction in the premixed methane flame model, section 5.5) first-order effects dominated the model output. Second-order effects (interaction between two parameters) were only small or could be neglected. Therefore, a second-order HDMR expansion could be used to accurately approximate the model output. In this study significant parameter interactions were discovered. The HDMR method has proven again to be very efficient in order to explore the first-order effects and $N = 512$ model runs were sufficient to obtain a reliable estimate for the most important first-order sensitivity indices. The exploration of the second-order effects was more demanding and $N = 4\,096$ model runs seemed to be sufficient in order to identify the most important parameter interactions between two inputs. However, it was not possible to construct an accurate HDMR metamodel for both outputs y_1 and y_2 on the basis of a second-order HDMR expansion even by increasing the sample size to $N = 131\,072$. The values of the first- and second-order sensitivity indices did not change significantly on further increasing the sample size. This indicated that higher order effects ≥ 3 were present, which cannot be incorporated if using only a second-order HDMR expansion. Therefore, it might be necessary to expand the HDMR approach so that potential third order effects can also be investigated. This will be discussed in section 8.1.

It has been shown in this work that it is likely that higher order effects may cause the tail in the probability density function (pdf) for the model output. All other case studies investigated in chapters 4 and 5 have shown that the model output had a more Gaussian like distribution and in all these cases first-order effects dominated

the model behaviour. In order to apply the HDMR method more efficiently for a system with a tailed distribution it might be useful to transform the model output by using for example $\log(y)$ to obtain a more Gaussian like shape. The transformation of the model output will also be further discussed in section 8.1.

6.4 Summary

The importance of parameter screening has been highlighted in this chapter. A two step approach has been introduced, where the Morris method is used first in order to identify unimportant parameters. However, the Morris method can only provide sensitivity measures that are qualitative. Therefore, the RS-HDMR method is used in a second step considering only the pre-selected potentially important input parameters. Thus, the input space dimension is reduced and the fitting of the polynomials is more accurate for a given sample size, especially for the approximation of the second-order component functions.

The proposed two-step approach has been tested using the analytical Sobol' g-function with 50 parameters, whereas only eight of them have an influence on the output variance. Around 75 % of the model output variance is caused by first-order effects, 22 % are caused by second-order effects. The RS-HDMR method was applied first considering all 50 input parameters. It could be shown that it is possible to accurately estimate the first- and second-order sensitivity indices, however a large sample size of $N = 8\,192$ was required. In order to investigate whether or not the computational effort could be reduced, the two-step approach has been applied. The Morris method was able to correctly identify the eight important parameters using only $N = 510$ model runs. However, using only ten trajectories ($r = 10$) the Morris method could not produce a reliable ranking for these parameters. Considering the identified eight parameters, the HDMR method was able to estimate the first- and

second-order sensitivity indices using only $N = 2048$ model runs. It could be shown that the estimation of the parameter interactions in particular is much more efficient in a smaller input space dimension. In total $N = 2558$ ($510 + 2048$) model runs were required using the two-step approach (Morris method + HDMR), in contrast to the $N = 8192$ required model runs using HDMR alone, to reach approximately the same accuracy for the sensitivity indices estimates.

The case study in this chapter focused on low temperature isothermal cyclohexane oxidation under fuel rich conditions. Uncertainties in 238 rate constants were considered. Since the number of uncertain parameters is so large, the Morris method was applied first in order to identify unimportant parameters. As a result a subset of 33 potentially important parameters could be found. However, 100 trajectories resulting in $N = 23900$ model runs were still required for this purpose. As discussed in section 6.3.4 even this large number of trajectories was not sufficient to obtain convergence for the mean effects. Hence, a reliable parameter ranking could not be obtained. The slow convergence of the Morris method was also observed by Hughes et al. (2006) when investigating the uncertainties in rate parameters on the ignition delay time for the low temperature oxidation of propane. Here, 1 000 trajectories had to be used in order to reach convergence. The slow convergence is probably due to the existence of significant higher order effects.

The extended RS-HDMR method was then applied using only the 33 identified parameters. Due to the existence of significant higher order effects up to $N = 13107$ model runs were performed. Since the model output variance could not completely be explained by a second-order HDMR expansion it is likely that higher order (≥ 3) effects exist. Therefore, the accuracy of the HDMR metamodel could not be used to justify the accuracy of the estimated sensitivity indices. Instead, various sample sizes (up to $N = 13107$) have been used for the estimation of the sensitivity indices to show that the values and the ranking do not significantly change on further increases

in sample size. In this way it can be assumed, that the calculated first- and second-order sensitivity indices are indeed a good estimate. Another ten parameters have been added to the ones identified by the Morris method in order to investigate the influence of potential wall losses. This led to further increases in the contribution of the higher order effects, but it also showed that the experimental results could be recovered for certain input combinations.

The existence of higher order effects is challenging for global sensitivity analysis methods. Theoretically, the HDMR method is able to consider higher order effects ≥ 3 , however it has to be investigated how efficiently this can be achieved. In this chapter (and in the other case studies) it has been shown that the HDMR method is very efficient when investigating first-order effects and even second-order effects can be estimated with reasonable efficiency. However, because a second-order HDMR expansion has been used, higher order effects ≥ 3 could not have been investigated. In cases where such higher order effects are present it might be worth to calculate the total effect indices $TS(i)$ as well, which are defined as the sum of all sensitivity indices involving the input parameter in question. One method to estimate total sensitivity indices would be the extended FAST method (Saltelli et al. 1999), however extended FAST requires a new set of samples to evaluate each of the total effect indices and it is questionable how efficient this will be.

7 GUI-HDMR Software

The usefulness of the RS-HDMR approach has been shown in this work by using analytical test functions and various case studies. In order to make the RS-HDMR tools and the extensions developed here easily available for all interested users, a software package has been developed. GUI-HDMR is a freely available Matlab toolbox with a graphical user interface. This chapter aims to give a short overview of its scope. A more detailed set of user documentation can be found in appendix B.

7.1 Introduction

The HDMR method provides a straightforward and efficient approach to explore the input-output mapping of a model. 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. Two extensions to the existing set of RS-HDMR tools have been developed in this work in order to further improve the method and to make it more applicable for complex models with a large number of input parameters. An optimisation algorithm has been introduced in section 4.1 to automatically calculate the best polynomial for the approximation of the component functions. Further, a threshold has been introduced in section 5.1 which allows the exclusion of unimportant component functions from the HDMR

expansion. The HDMR method can also be used as part of a two step approach, where the Morris method (section 2.3.2) is applied first to identify unimportant parameters.

Although the RS-HDMR method is conceptually simple, the application is not always that easy since many HDMR tools and extensions exist. Therefore, the GUI-HMDR software (Ziehn & Tomlin 2008c) has been developed to combine 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 users in all application fields without requiring comprehensive knowledge of global SA methods. The user has to supply only one set of input values (random or quasi-random) 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.

7.1.1 Graphical User Interface (GUI)

The GUI-HDMR software provides a main window from where the calculations can be started and progress shown. Three other windows can be accessed from the main window: a set up window, a window showing the optimal polynomial orders and the accuracy of the metamodel and a window for sensitivity analysis.

The set up window is used to load the sample input and output files into the workspace. If one wants to verify the constructed metamodel with a different sample set than the one used for constructing the metamodel, then a different set of sample input and output files can be provided for the accuracy test. Optionally the ranges for the input parameters can be loaded in order to produce the plots with the correct ranges. In the settings section one has to state the number of samples to be used in order to construct the HDMR metamodel and the number of samples to be

used for the accuracy test. The maximum polynomial order for the approximation of the component functions can be defined separately for first- and second-order. Orthonormal polynomials up to tenth order are supported. (Note that the computational time increases if using higher order polynomials. This is especially true for the second order component functions, because a large number of expansion coefficients have to be calculated). A variance reduction method can be chosen and the number of iterations can be entered for the calculation of the α and β expansion coefficients (see equations (3.44) and (3.46)). It is also possible to apply a threshold which ensures that only the component functions having a contribution larger than the threshold are considered. The threshold is given in % and is explained in section 5.1.

After the calculations have been finished and the metamodel has been successfully constructed, the results window gives an overview of how many component functions have been approximated and by what polynomial order. Furthermore the relative error can be calculated to show the accuracy of the first- and second-order metamodel.

The sensitivity analysis window presents the ranking of the five most important parameters (based on the first-order sensitivity indices, see equation (3.87) and (2.31)). The ranking of the five most import parameter interactions (based on the second-order sensitivity indices, see equation (3.88)-(2.31)) can also be displayed. Additionally all component functions can be plotted and compared with scatter plots. It is also possible to save all produced plots as Portable Network Graphics (.png), Portable Document Format (.pdf) and Encapsulated PostScript (.eps).

Finally, the whole Matlab workspace can be saved and loaded into the GUI-HDMR software again for further analysis and plots.

7.1.2 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. A simple example without using variance reduction methods and without applying a threshold is given in the user documentation appendix B.

7.2 Summary

The HDMR method is shown to be a very powerful tool in order to estimate variance based sensitivity indices. Only one set of Monte Carlo samples is required. For a more uniform coverage of the input space, and optimal convergence with sample size, a quasi-random sequence such as the Sobol' sequence is recommended.

The GUI-HDMR software has been tested in this work using two different analytical functions: the Ishigami function and the Sobol' g-function. Both models are highly non-linear, non-monotonic and have significant parameter interactions. The software has also been used for the case studies in chapters 4-6. It has been illustrated that even high input space dimensions can be investigated due to the HDMR extensions such as the optimisation method and application of a threshold for component function exclusion. However, if the model has significant parameter interactions it is computationally more efficient to apply a screening method beforehand in order to reduce the dimensionality and to improve the accuracy of the polynomial fitting.

The proposed GUI-HDMR software provides a very handy tool for global sensitivity analysis of complex models. The graphical user interface allows an easy set up and analysis of the results. The user has to provide only two external files containing the sample input values and the sample output values of the model runs. Therefore, the use of the GUI-HDMR software is not restricted to a certain type of model or a

specific platform but can be used for a large range of applications in all fields.

8 Limitations and Extensions of the RS-HDMR Approach and Future Work

The usefulness of the HDMR approach for global sensitivity analysis has been illustrated in this work using various analytical test functions and case studies. Two commonly used HDMR approaches, cut-HDMR and RS-HDMR, have been investigated. The main focus has been on the application of the RS-HDMR approach because here the computational effort (number of simulations) does not directly depend on the input space dimension n . Furthermore, RS-HDMR does not require the storage of data in look-up tables or interpolation methods. However, despite its effectiveness and universality in terms of fields of application, some limitations of the RS-HDMR approach, as introduced in chapter 3, exist and are therefore discussed in this chapter.

8.1 Investigation of Higher Order Effects

In this work the HDMR expansion has been used only up to second-order. Li et al. (2001b) argues, that for many systems a HDMR expression up to second-order

already provides satisfactory results and a good approximation of $f(\mathbf{x})$. However, it could be shown in the case study in section 6.3 (low temperature cyclohexane oxidation) that higher order effects (≥ 3) were present. For most scenarios the sum of the first- and second-order effects could only explain less than 70 % of the overall variance of one of the model outputs. In this case it is clearly highly desirable to investigate the higher order effects (≥ 3) as well in order to identify the missing important parameter interactions. The investigation of such higher order effects usually requires a large number of additional model runs which would make the HDMR approach computationally more expensive (Li et al. 2003b). For RS-HDMR this is due to the fact, that higher order component functions have more terms than lower order ones and each term has its own Monte Carlo integration error (see also section 4.1). For example, if assuming third-order orthonormal polynomials, then three terms have to be estimated for each first-order component function, nine terms for each second-order component function and 27 terms for each third-order component function.

Two approaches are introduced here in order to deal with higher order effects. The first uses a transformation of the model output so that a second-order HDMR expansion can efficiently be applied for the transformed data set. The second approach focuses on the inclusion of third-order effects into the HDMR expansion based on the products of the first- and second-order component functions.

8.1.1 Transformation of the Model Output

The investigations in the case study in section 6.3 have shown that higher order effects might be responsible for a tailed distribution of the model output. Other investigated case studies in this work did not have significant higher order effects and the distribution of considered outputs had a Gaussian like shape. In order to

apply the HDMR method more efficiently to systems with significant higher order effects, it can be useful to transform the model output by using for example the $\log(y)$ transformation. This can reduce the tail and the procedure will be demonstrated here using the simple test model:

$$f(\mathbf{x}) = \exp(\exp(x_1 + x_2^4)). \quad (8.1)$$

The output distribution for $f(\mathbf{x})$ is presented in figure 8.1 and its long tail is clearly visible. The RS-HDMR method has been applied to this model using a sample size of $N = 2048$ in order to estimate the sensitivity indices which are presented in table 8.1. The second-order effect (parameter interaction between x_1 and x_2) accounts for over 50 % of the output variance. Therefore, a first-order HDMR metamodel can not accurately approximate the model behaviour and as illustrated in figure 8.1a it fails to recover the long tail in the output distribution. The second-order HDMR metamodel incorporates the important parameter interactions and as shown in figure 8.1b it allows the recovery of the tail of the distribution. This confirms the assumption, that parameter interactions may cause a tailed output distribution.

Method	S_1	S_2	S_{12}	S
Analytical (non-transformed output)	0.1001	0.3249	0.5750	1.0000
RS-HDMR ($N = 2048$)	0.1020	0.3302	0.5764	1.0086
Analytical (transformed output)	0.4257	0.5308	0.0435	1.0000
RS-HDMR ($N = 256$)	0.4306	0.5354	0.0429	1.0089

Table 8.1: First- and second-order sensitivity indices calculated analytically and estimated based on RS-HDMR for the non-transformed and transformed model output.

The model output (equation (8.1)) is now transformed using for example $\log(y)$. This kind of transformation is often used in data analysis and has also been applied in connection with the HDMR method, for example in Ratto (2008), where the

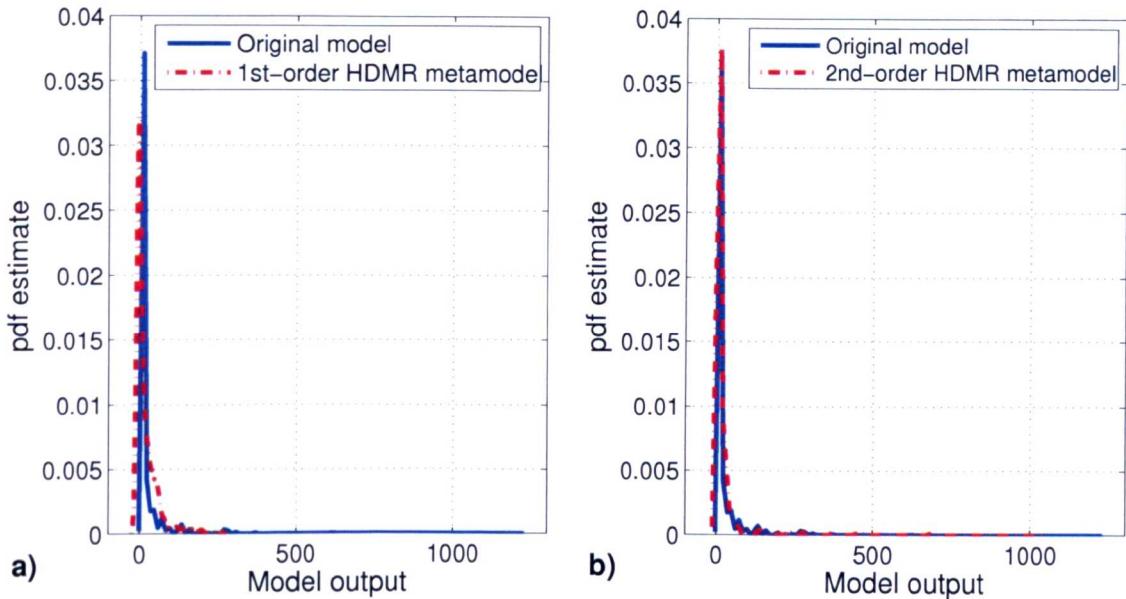


Figure 8.1: Accuracy of the RS-HDMR metamodel for $N = 2048$ based on the probability density function (pdf) using 1000 random points (a) using a first-order HDMR metamodel and (b) using a second-order HDMR metamodel.

HDMR expansion was constructed using the $\log(-y)$ transformation. The probability density function for the transformed output is presented in figure 8.2 and it can be seen that the tail of the pdf is very much reduced. A small sample size of $N = 256$ is now sufficient in order to construct a very accurate HDMR metamodel and to estimate the sensitivity indices (table 8.1). Here it can be seen, that the transformation of the model output also reduces the sampling effort.

On the one hand the transformation of the model output results in a distribution with a smaller tail, but on the other hand it also “removes” most of the parameter interaction between x_1 and x_2 . This interaction accounts now for only around 4 % instead of over 50 % for the non-transformed output (table 8.1). This has implications when interpreting the results of a transformed model output and one has to be aware that the results are only correct for the transformed output and do not necessarily relate to the original model output anymore. However, if the purpose

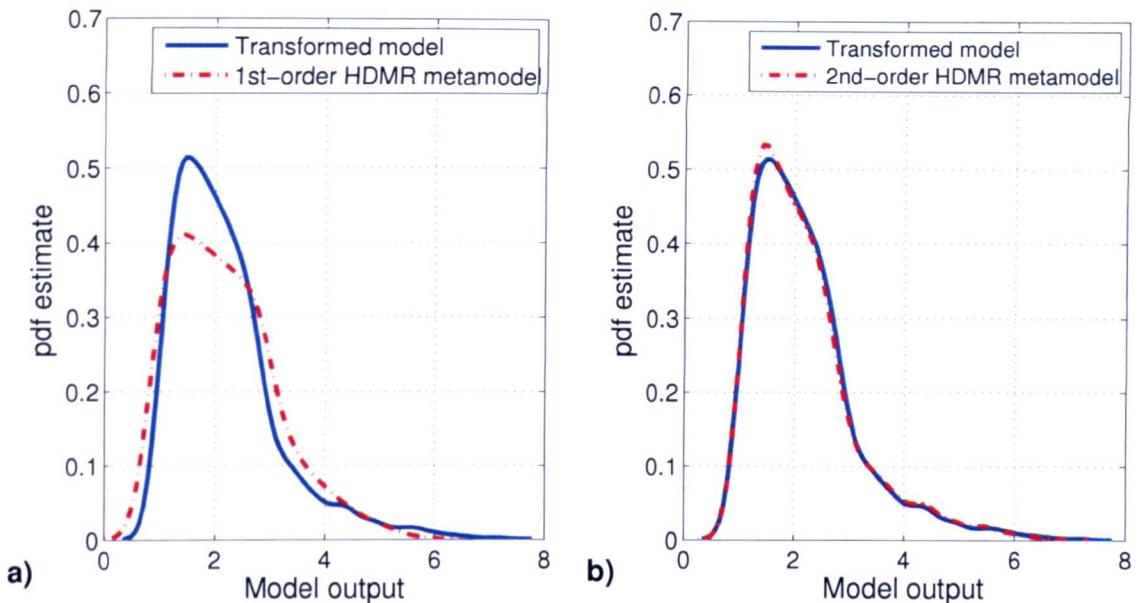


Figure 8.2: Accuracy of the RS-HDMR metamodel (transformed output using $\log(y)$) for $N = 256$ based on the probability density function (pdf) using 1 000 random points (a) using a first-order HDMR metamodel and (b) using a second-order HDMR metamodel.

is to identify key parameters, the use of output transformations could be extremely useful.

8.1.2 Inclusion of Third-Order Effects

The investigation of higher order effects (≥ 3) usually requires a large number of model runs as highlighted earlier. For example, the third-order terms can be calculated as follows (Li et al. 2003*b*):

$$\begin{aligned} f_{ijk}(x_i, x_j, x_k) &= \int_{K^{n-3}} f(\mathbf{x}) d\mathbf{x}^{ijk} - f_{ij}(x_i, x_j) - f_{ik}(x_i, x_k) - \dots \\ &\quad f_{jk}(x_j, x_k) - f_i(x_i) - f_j(x_j) - f_k(x_k) - f_0 \end{aligned} \quad (8.2)$$

$$f_{ijk}(x_i, x_j, x_k) \approx \sum_{p=1}^m \sum_{q=1}^{m'} \sum_{r=1}^{m''} \gamma_{pqr}^{ijk} \varphi_p(x_i) \varphi_q(x_j) \varphi_r(x_k) \quad (8.3)$$

where $d\mathbf{x}^{ijk}$ is the product dx_1, dx_2, \dots, dx_n without dx_i, dx_j, dx_k and m, m', m'' are the orders of the orthonormal polynomials.

The expansion coefficient γ_{pqr}^{ijk} can be obtained by Monte Carlo integration:

$$\gamma_{pqr}^{ijk} = \int_{K^n} f(\mathbf{x}) \varphi_p(x_i) \varphi_q(x_j) \varphi_r(x_k) d\mathbf{x} \quad (8.4)$$

$$\approx \frac{1}{N} \sum_{s=1}^N f(\mathbf{x}^{(s)}) \varphi_p(x_i^{(s)}) \varphi_q(x_j^{(s)}) \varphi_r(x_k^{(s)}) \quad (8.5)$$

with $\mathbf{x}^{(s)} = (x_1^{(s)}, x_2^{(s)}, \dots, x_n^{(s)})$ and $s = 1, 2, \dots, N$. Due to the high number of terms in each third-order component function the computational error might be large particularly for small sample sizes N .

In Li et al. (2003a) the so called low-order term product (lp)-RS-HDMR method is introduced to approximately represent the higher order RS-HDMR component functions as products of the lower order functions. It is possible to express the higher order (≥ 2) component functions solely as products of the first-order component functions. Hence, it is only required to compute the first-order RS-HDMR component functions which may require only a few hundred samples (Li et al. 2003b). The higher order component functions are no longer explicitly computed. It is also possible to express the higher order (≥ 3) component functions as products of the first-

and second-order component functions. More details and the mathematical foundations of (lp)-RS-HDMR along with an illustration of its utility in an atmospheric chemical kinetics model can be found in Li et al. (2003a).

8.1.3 Conclusions

The simple example in section 8.1.1 has demonstrated, that a transformation such as $\log(y)$ can be used to remove or reduce the tail of a model output distribution. The HDMR method can then be applied very efficiently requiring only a small number of model runs. For more complex models with higher order effects (≥ 3) this would possibly prevent the need to use a HDMR expansion greater than second-order. However, it has also been pointed out, that the interpretation of results based on the transformed data may cause difficulties. An efficient approach to include third-order effects into the HDMR expansion has been introduced in section 8.1.2. Here, only the theoretical idea was presented and it has to be tested in future work how efficient this approach really is and how well it works for analytical test functions and case studies.

8.2 Non-Uniform Input Distributions

In all investigated case studies in this work it has been assumed that the input parameters possess a uniform distribution. This is a valid assumption, if no other information is available. However, in some cases the user might have prior knowledge about the input parameter distribution and this should be included in the process of calculating the sensitivity indices. For example, for rate constants the log-normal distribution is often used and the heats of formation can be assigned with a normal distribution, so that the reverse rate constants are also log-normal

distributed (Phenix et al. 1998).

The formulation of RS-HDMR for uniform distributions cannot directly be applied to data with a non-uniform distribution. Hence, the RS-HDMR approach has been extended in Wang & Georgopoulos (2003) in order to work with non-uniform input distributions. Weighted orthonormal polynomials related to the pdf are introduced to approximate the RS-HDMR component functions. Different pdfs result in special formulas for the weighted orthonormal polynomials. The general formulas of RS-HDMR can be constructed, treating uniform RS-HDMR as a special case.

According to Wang & Georgopoulos (2003) the component functions of RS-HDMR with the input parameters having the pdf $w_i(x_i)$ are defined as follows:

$$f_0 = \int_{K^n} \prod_{i=1}^n w_i(x_i) f(\mathbf{x}) d\mathbf{x} \quad (8.6)$$

$$f_i(x_i) = \int_{K^{n-1}} \prod_{k=1, k \neq i}^n w_k(x_k) f(\mathbf{x}) d\mathbf{x}^i - f_0 \quad (8.7)$$

$$f_{ij}(x_i, x_j) = \int_{K^{n-2}} \prod_{k=1, k \neq i, j}^n w_k(x_k) f(\mathbf{x}) d\mathbf{x}^{ij} - f_i(x_i) - f_j(x_j) - f_0 \quad (8.8)$$

...

Weighted orthonormal polynomials with the following properties can be used to approximate the RS-HDMR component functions:

$$\int_0^1 w_i(x_i) \varphi_r^i(x_i) dx = 0 \quad \text{for all } r, i \quad (8.9)$$

$$\int_0^1 w_i(x_i) [\varphi_r^i(x_i)]^2 dx = 1 \quad \text{for all } r, i \quad (8.10)$$

$$\int_0^1 w_i(x_i) \varphi_p^i(x_i) \varphi_q^i(x_i) dx = 0 \quad p \neq q \quad (8.11)$$

The formulas for calculating the expansion coefficients α, β, \dots (see equations (3.44)-(3.46)) are the same as those for uniform RS-HDMR, because the sampling is drawn from the guidance by the weights $w_i(x_i)$. More detailed information regarding the RS-HDMR method with non-uniformly distributed input parameters can be found in Wang & Georgopoulos (2003).

8.3 Non-Independent Input Parameters

In the formulations of the HDMR component functions, as introduced in chapter 3, all input parameters were considered to be independent. Clearly, this restricts the application of the HDMR method and systems whose input parameters may not be independent can not be treated in this way. According to Zsély et al. (2008), correlations are common in kinetic and thermodynamic parameters. Rate parameters are correlated when obtained by fitting using a small mechanism and enthalpies of formation are correlated because they are obtained from measured reaction enthalpies. The work of Zsély et al. (2008) shows that it is important to consider potential correlations in input parameters when performing uncertainty and sensitivity analysis since the output uncertainty and the input-output relationships depends on these assumptions.

The HDMR method has been extended by Li et al. (2006) to treat systems with non-independent input parameters. The resulting formulae for the HDMR component functions are the same as the original ones except that the probability density functions are replaced by conditional probability density functions. In this way the HDMR component functions with independent input parameters are just a special case of the general treatment (Li et al. 2006).

8.4 Extensions and Improvements of the GUI-HDMR Software

The GUI-HDMR software (Ziehn & Tomlin 2008c) has been developed to combine RS-HDMR tools and their extensions (see section 4.1 and 5.1) in one Matlab package equipped with a graphical user interface. The software has been tested in this work using non-linear and non-monotonic analytical functions and it has been applied in the case studies in chapters 4-6.

The current version of the GUI-HDMR software uses only the relative error as a measure for the accuracy of the constructed metamodel. As already discussed in section 3.3 the relative error is not always the best choice in order to assess the accuracy of the HDMR metamodel. Therefore, it is planned to implement the r^2 -value (coefficient of determination) into the GUI-HDMR software to provide an additional measure for the accuracy of the metamodel. A scatter plot (model output versus response of the metamodel) for the first- and second-order HDMR expansion can also highlight regions in the output ranges where the HDMR metamodel fails to capture the behaviour of the real model. Additionally, the statistics of the real model and the metamodel can be compared by plotting its probability density function (pdf) estimate and/or the empirical cumulative distribution function (cdf). Therefore, pdf plot and scatter plot will also be included in the GUI-HDMR software.

The calculation of the optimal order (see section 4.1) and the threshold (section 5.1) are based on least square methods. In the next version of the GUI-HDMR software it is planned to use a statistical measure, such as the F-test, instead of the least square method in order to further improve those algorithms.

Variance reduction methods have been introduced in section 3.2.2 in order to reduce the Monte Carlo integration error when estimating the expansion coefficients (see

equations 3.57 - 3.58) for the orthonormal polynomials. In this way the calculation of the expansion coefficients becomes an iteration procedure. The GUI-HDMR software provides the possibility to enter a maximum number of iterations for the estimation of the α and β coefficients. However, in a next version of the software this is planned to be extended by a convergence criterion, so that the maximum number of iterations is only performed in case the convergence criterion is not fulfilled.

It is also planned to extend the HDMR expansion used in the GUI-HDMR software up to third-order (as described in section 8.1.2) so that third-order effects (interaction between three parameters) can be investigated. Further, it is planned to consider non-uniform distributions for the input parameters such as normal or log-normal by implementing the approach described in section 8.2 into the GUI-HDMR software. The treatment of non-independent input parameters (see section 8.3) will also be included.

Other possible ideas for extensions and improvements of the GUI-HDMR software have come from users of the software, which include an option to rotate the surface plots of the second-order HDMR component functions and to provide the possibility to save the plots of all the component functions in the Matlab format (.fig) as well.

8.5 Summary

The HDMR method as introduced in chapter 3 and the extensions developed for the RS-HDMR approach in section 4.1 and 5.1 provide already a universal approach for global sensitivity analysis for a large number of models in all fields. However, the application of the HDMR method was restricted to systems whose input parameters possess a uniform distribution and are independent. These limitations have been discussed and HDMR extensions addressing and solving these issues have already

been published and successfully applied. The investigation of higher order effects ($L \geq 3$) has also been discussed and (lp)-RS-HDMR (Li et al. 2003a) has been introduced, which offers a more efficient approach for calculating higher order effects. One aim for future work would be to include these extensions in the GUI-HDMR software. Other potential extensions and improvements of the GUI-HDMR software have also been discussed.

9 Summary and Conclusions

Uncertainty analysis and sensitivity analysis form an important part of the model evaluation process and should therefore be applied in order to increase the accuracy and reliability of computer models in environmental and safety engineering. However, after reviewing traditional methods for global sensitivity analysis, such as Monte Carlo analysis, it can be said that these methods do not fulfil all the requirements needed. The computational running time for a model can be large and the input space dimension for the uncertain parameters can be very high (e.g. 238 uncertain parameters for the cyclohexane oxidation model). Additionally, the model behaviour can be highly non-linear and parameter interactions might exist. Therefore, methods are required, which are able to handle all these difficulties and provide a quantitative measure in order to produce a reliable ranking of important parameters and parameter interactions.

A very common measure of the importance of parameters and parameter interactions is provided by variance based sensitivity indices, which were first introduced by Sobol' (1993). The determination of the sensitivity indices is based on the Analysis of Variance (ANOVA) decomposition, where the partial variances, caused by individual uncertain parameters or parameter interactions, are normalised by the overall variance. However, following the method of Sobol' the computation of the partial variances is computationally very expensive, because for the determination of

each partial variance a different set of random (or quasi-random) samples is required.

A more efficient approach in calculating the variance based sensitivity indices was introduced by Rabitz et al. (1999). The cut-HDMR was proposed, where the component functions are calculated along its variable axis through a reference point. The component functions are usually represented as low-dimensional look-up tables. In comparison to the method of Sobol' the sampling effort could be reduced, however one significant drawback of cut-HDMR is, that the number of model runs still directly depends on the number of uncertain parameters. Therefore, the cut-HDMR approach is only applicable for systems with a small number of input parameters. Consequently, the random sampling (RS)-HDMR approach has been introduced by Li et al. (2002a), where the component functions are approximated by orthonormal polynomials. Here, the number of required model runs no longer directly depends on the input space dimension and all partial variances can be estimated using only one set of random or quasi-random samples. In addition to the calculation of the sensitivity indices, cut-HDMR and RS-HDMR methods allow the plotting of individual component functions showing the effects of parameters and parameter interactions over the whole input range.

The performance of cut-HDMR and RS-HDMR has been compared in this work using the Ishigami function, a non-linear and non-monotonic analytical test model with three input parameters. Both HDMR approaches were able to correctly estimate the first- and second-order sensitivity indices. Furthermore, both methods have been applied to investigate the turbulent flow field in a 3D street canyon in the city of York using the micro-scale $k - \epsilon$ closure computational fluid dynamics (CFD) model MISKAM. Four main input parameters have been addressed which are the roughness lengths (inflow, surface and wall) and the background wind direction. Various outputs have been selected, among them the wind components and the turbulent kinetic energy (TKE) at certain measuring points. A small sample

size of $N = 512$ (RS-HDMR) and $N = 641$ (cut-HDMR) was sufficient in order to construct a very accurate metamodel based on a second-order HDMR expansion. It was shown that most of the first-order effects are non-linear. For some of the output parameters, interactions could be identified, which were responsible for more than 5 % of the overall output variance. Based on the uncertainty and sensitivity analysis results important insights into the model physics could be gained. For example, it has been shown that the inflow roughness length has a strong influence on the TKE in the street canyon, although not on the mean velocities. This might indicate, that the model domain needs to be large enough to resolve the effects of upwind roughness elements, even if dispersion within the street canyon is the main aim of the study. The HDMR analysis has also revealed that the wind components in the street canyon are strongly influenced by the background wind direction. The high sensitivity of this parameter shows how important it is to specify the background wind direction very accurately at the study site. This also illustrates the impact of practical limitations in determining a local reference background wind direction for pollution hot spot sites. Although both HMDR approaches performed very well when applied to the test function and the case study, RS-HDMR has proven to be the more promising approach. Therefore, the rest of the work has focused solely on the RS-HDMR method.

Two extensions to the existing set of RS-HDMR tools have been developed: The first is an optimisation approach, which automatically calculates the optimal order of the polynomials used for the approximation of the component functions. The second introduces a threshold which allows unimportant component functions to be excluded from the HDMR expansion. The usefulness of these extensions has been shown in this work in connection with analytical test functions and various case studies.

One important analytical test function applied in this work is the Sobol' g-function.

The Sobol' g-function is highly non-linear and non-monotonic and has often been used in the literature as a test model for calculating variance based sensitivity indices. One handy feature of the Sobol' g-function is, that the number of input parameters can be arbitrarily extended and the importance of each of the input parameters can be controlled by a constant. In this work, the Sobol' g-function has been used with a total of 20 and 50 input parameters. The RS-HDMR method with its developed extensions performed very well and first- and second-order sensitivity indices could be estimated very accurately. However, in the case where the Sobol' g-function has been used with 50 input parameters it has been found computationally more efficient to apply the Morris method (a screening method) first in order to identify unimportant parameters. The RS-HDMR method has then been applied with the reduced input space dimension (eight potentially important parameters) and sensitivity indices could be calculated using only a small number of model runs. Without prior screening an accurate estimation of the most important first- and second order sensitivity indices required at least $N = 4\,096$ model runs whereas the combination of the screening method + RS-HDMR required only $N = 2\,558$ model runs ($N = 510$ for the Morris method + $N = 2\,048$ RS-HDMR) in order to achieve about the same accuracy. The RS-HDMR method with the extensions developed in this work is able to handle the highly non-linear Sobol' g-function function, however if the input space dimension is very large (i.e. ≥ 50) then a combination of a screening method and RS-HDMR is computationally more efficient.

The RS-HDMR method and its extensions have also been used in this work to investigate a reactive plume of nitric oxide (NO) in a turbulent grid flow doped with ozone (O_3) by applying a combined Lagrangian stochastic model with a micro mixing sub-model. A chemical reaction scheme (including photolysis) consisting of eight reactions between the species NO, O, O_3 and NO_2 has been implemented and the effects of uncertainties in temperature dependent rate parameters, photolysis rates,

temperature and NO:NO₂ ratio at the source on the predicted pollutant profiles were studied. In total 22 parameters were assumed to be uncertain. The outputs of interest were the NO_x and O₃ concentration at the plume centre. A very accurate metamodel has been constructed using only $N = 512$ samples and a second-order HDMR expansion. No significant parameter interactions could be identified and the overall variance of the model output could therefore be explained by parameters acting independently.

The application of the HDMR method has revealed that the mixing time scale coefficient α has a significant influence on the mean concentrations. This is clearly a disadvantage of using the simple interaction by exchange (IEM) model as a mixing model, because the mean concentration should not be affected by the micro mixing. It has also been shown that the uncertainty range for the structure function coefficient c_0 has originally been chosen too wide based on existing literature studies. A comparison with experimental values has shown that the uncertainty range for c_0 can be reduced, because low values for c_0 do not allow the recovery of the experimental values. This has highlighted the advantages of using the HDMR approach. On the one hand the effects of the uncertain parameters can be investigated over a large range (plot of the first-order HDMR component function). On the other hand new uncertainty ranges (as long as they are within the original range) can be assigned and the effects investigated without performing new model runs. This is because the HDMR metamodel can also be used as a surrogate and a large number of simulations based on the HDMR metamodel can be performed almost instantly. The computational running time of the Lagrangian stochastic model is quite high (e.g. 20 hours for one simulation) because a large number of particles (1 000 000) had to be used in order to reach convergence for the mean concentrations.

Furthermore, the RS-HDMR method was applied to a one dimensional low pressure premixed methane flame model used to investigate the influence of fuel sulphur and

nitrogen on the emission of nitrogen oxides within the flame. A one dimensional steady state reaction advection diffusion model has been used where uncertainties in rate constants and thermodynamic data were considered. This led to a study of 176 input parameters (153 reaction rates and 23 enthalpies of formation). The final NO mole fraction within the burnt gas region was considered as the target model output. It was shown that despite the large number of input parameters, screening methods could be generally avoided by applying a threshold to exclude unimportant component functions. A sample size of $N = 1024$ was sufficient in order to construct an accurate metamodel and to calculate variance based sensitivity indices up to second-order. However, in one case, a combination of a screening approach and HDMR was shown to be computationally more efficient. Three different scenarios have been investigated where the uncertainty ranges for the 23 enthalpies of formation were changed according to different sources in the literature. The parameter importance ranking based on the first-order sensitivity indices was shown to be critically dependent on the uncertainty ranges chosen. In this high temperature study only a few of the 176 considered parameters had a significant influence on the model output and no significant parameter interactions were identified for this case.

The fourth case study investigated in this work focused on global sensitivity analysis of simulations of low temperature isothermal cyclohexane oxidation under fuel rich conditions. A well mixed non steady state box model with highly complex chemistry has been applied. The analysis was used to investigate the important features of the oxidation process, as well as possible factors underlying qualitative discrepancies between simulations and experiments. The chemical scheme compromised 60 species and 238 reactions, where the input parameters representing the temperature independent part of the rate constants (A -factors) were considered to be uncertain, i.e. 238 parameters. The selected target outputs were the time for the maximum rate of reaction for O_2 consumption and the percentage of O_2 remaining

at the time for the maximum rate. It was revealed that the model had significant higher-order effects and the calculation of the second-order sensitivity indices would have required a large number of samples N , especially in connection with the large input space dimension. Therefore, the Morris method was used in order to exclude unimportant parameters and as a result the HDMR method was applied using 33 potentially important parameters. It could be shown that even if taking into account the uncertainties in the rate parameters the long amount of time required for the maximum rate of reaction for O_2 consumption and the high percentage of O_2 remaining as suggest by the experimental results could not be recovered. This suggested that there were reactions missing in the scheme and consequently wall losses were included by adding ten additional reactions. The inclusion of wall losses led to a higher number of simulations with realistic induction times and higher O_2 percentage remaining.

When investigating a high temperature combustion case study (premixed methane flame model) it could be shown that only a few first-order effects dominated the model output. Second-order effects (interaction between two parameters) were only small or could be neglected. Therefore, a second-order HDMR expansion could be applied to accurately approximate the model output using only a small number of model runs. The low temperature combustion application (cyclohexane oxidation) has been a challenge for the application of the HDMR method, because of significant parameter interactions. The HDMR method has been proven again to be very efficient in order to explore the first-order effects and $N = 512$ model runs were sufficient to obtain a reliable estimate for the most important first-order sensitivity indices. The exploration of the second-order effects was more demanding and $N = 4096$ model runs seemed to be sufficient in order to identify the most important parameter interactions between two inputs. However, it was not possible to construct an accurate HDMR metamodel for both outputs on the basis of a second-order

HDMR expansion even by increasing the sample size to $N = 131\,072$. The values of the first- and second-order sensitivity indices did not change significantly on further increasing the sample size. This indicated that higher order effects ≥ 3 were present, which cannot be incorporated if using only a second-order HDMR expansion. For such cases it might be necessary to expand the HDMR approach so that potential third-order effects can also be investigated.

It has been shown in this work that it is likely that higher order effects may cause a tailed distribution of the model output. The three case studies investigated in chapters 4 and 5 (street canyon flow model, Lagrangian stochastic model, high temperature flame model) have shown that the model output had a more Gaussian like distribution without a tail and in all these cases first-order effects dominated the model behaviour. Only the case study in chapter 6 (low temperature cyclohexane oxidation) has shown a tailed output distribution. Due to the fact that only in the low temperature combustion case study significant higher order effects could be identified, it has been assumed that higher order effects may be responsible for a tailed distribution of the model output. This assumption was supported by the investigation of a simple test model, which also had mainly higher order effects and a tailed output distribution. In order to apply the HDMR method more efficiently for a system with a tailed output distribution it might be useful to transform the model output by using for example $\log(y)$ to obtain a more Gaussian like shape. This has been demonstrated for the simple test model, and it was shown that the transformation of the model output results in a distribution with a smaller tail and that only a small sample size is required in order to estimate the sensitivity indices. However, on the other hand the transformation also “removes” most of the parameter interactions. This has implications when interpreting the results of a transformed model output and one has to be aware that the results are only correct for the transformed output and complex inverse transformation would be required

for interpretation with respect to the original model output.

The application of the RS-HDMR approach to test functions and case studies has led to the development of the GUI-HDMR software. The software combines RS-HDMR tools and its extensions in one Matlab package equipped with a graphical user interface. This makes the RS-HDMR method easily available for all interested users in various fields. User documentation of the software is given in appendix B. Possible extensions of the software, such as the investigation of higher order effects (≥ 3), consideration of non-uniform input distributions and non-independent input parameters have been discussed because of their practical relevance. If enough data is available, then it might be possible to estimate the distribution of input parameters. In this work a uniform distribution has been assumed for all uncertain parameters, because parameter ranges were only defined by a minimum and maximum value. However, if it is possible to obtain the distribution for the model inputs, then this information should be included into the uncertainty and sensitivity analysis. Furthermore, it has been assumed in this work that model inputs are independent. A recent study by Zsély et al. (2008) has shown that correlations are common in kinetic and thermodynamic parameters. Consequently, this kind of information should also be included into uncertainty and sensitivity analysis studies.

It has been demonstrated in this work that the RS-HDMR method is a very powerful tool in order to estimate variance based sensitivity indices. Only one set of random or quasi-random samples is required. The method can cope with non-linearities and a high input space dimension. If the input space dimension is very large (i.e. ≥ 50) then a screening method can be used in combination with RS-HDMR. In most of the case studies and test functions investigated in this work a sample of $N \leq 1024$ was sufficient in order to accurately estimate sensitivity indices. Only when investigating higher order effects a larger sample size was required. The performed model runs are basically Monte Carlo simulations. They are independent from each

other and can therefore easily be performed within a parallel environment (e.g. high performance computing facilities) in order to speed up the overall simulation time. The constructed HDMR metamodel can also be used as a surrogate instead of the original model for example when input ranges have been altered (as long as they are within the original range). In this way no additional model runs have to be performed which can result in a enormous amount of time being saved if the original model is computationally expensive to evaluate (e.g. MISKAM model: up to 40 min, Lagrangian stochastic model: up to 20 hours for one simulation on a 3 GHz PC). Special attention should be paid to the plots of the HDMR component functions, since they reveal useful information about the input output relationship of the model. They can also be used for sensitivity analysis instead of more widely used scatter plots. The plots of the HDMR component functions up to second-order show exactly how one input parameter or pair of input parameters affect the output over the whole range of uncertainty. In particular they demonstrate the sensitivity relationship of parameters away from their nominal values and therefore provide useful information to link with further detailed studies of individual parameters (e.g. ab initio studies).

The RS-HDMR method treats the model as a black box and can therefore be applied in all sorts of fields. Potential fields of application for future work could be in the development of bio fuel models and the improvement of surrogate models. Here, the chemical mechanisms describing multi-step chemistry can often be very large. For example, the modelling of ethanol, which is used in combination with gasoline, includes the description of over 380 reactions involving 56 species (Marinov 1998). The chemical mechanism for n-heptane, which is considered a surrogate for liquid hydrocarbon fuels and can be used in combination with a methyl ester to represent a bio fuel type, features over 650 reactions and over 100 species (Lindstedt & Maurice 1995). The parameters describing the rates of the chemical reactions can often

be very uncertain, with the parameters sometimes being rough estimates of their true value. Due to the fact that there are so many parameters and very accurate calculations of each one can be time consuming, it is important to highlight the main parameters contributing to predictive uncertainty for the purpose of improving the model. It has been shown in this work, that the application of the RS-HDMR method can help to highlight key reactions and species so that these parameters can be calculated more accurately. This could also work in the opposite direction so that reactions are highlighted that are not included in current kinetic mechanisms and the sensitivity to their addition could be investigated.

The case studies in this work have demonstrated that it is possible to investigate important characteristics such as ignition behaviour (low temperature) and product formation (high temperature) of complex chemical mechanisms by applying HDMR methods. In the field of atmospheric modelling it is planned to combine the CFD model MISKAM with the Lagrangian stochastic model in order to predict pollutant concentrations and concentration fluctuations in urban areas. Both models have been investigated separately in this work and important physical insights have been gained. However, in order to accurately predict pollutants and potential hot spots in urban areas both flow and dispersion have to be modelled together. As demonstrated in this work, the HDMR method can help to highlight discrepancies between experimental data and model simulations and is able to show where model improvements are required.

A Mathematical Background and Additional Information

A.1 Example for the Conversion of cut-HDMR Component Functions

Considering a model with four input parameters, the first-order ANOVA expansion function $f_1^{ANOVA}(x_1)$ would be calculated as follows:

$$\begin{aligned} f_1^{ANOVA}(x_1) &= f_1^{cut}(x_1) + \int_0^1 f_{12}^{cut} dx_2 + \int_0^1 f_{13}^{cut} dx_3 + \int_0^1 f_{14}^{cut} dx_4 \\ &\quad - \int_0^1 f_1^{cut} dx_1 - \int_0^1 \int_0^1 f_{12}^{cut} dx_1 dx_2 - \int_0^1 \int_0^1 f_{13}^{cut} dx_1 dx_3 \\ &\quad - \int_0^1 \int_0^1 f_{14}^{cut} dx_1 dx_4 \end{aligned} \quad (\text{A.1})$$

and for the second-order ANOVA expansion function $f_{12}^{ANOVA}(x_1, x_2)$ one would get

$$\begin{aligned} f_{12}^{ANOVA}(x_1, x_2) &= f_{12}^{cut}(x_1, x_2) + \int_0^1 \int_0^1 f_{12}^{cut} dx_1 dx_2 - \int_0^1 f_{12}^{cut} dx_2 \\ &\quad - \int_0^1 f_{12}^{cut} dx_1 \end{aligned} \quad (\text{A.2})$$

The integrals in equation (A.1) and (A.2) can be computed using the exact solutions

at each of the grid points in connection with trapezoidal numerical integration.

Squaring and integrating over the ANOVA component functions provides the partial variances, for example:

$$D_1 = \int_0^1 f_1^2(x_1) dx_1 \quad (\text{A.3})$$

$$D_{12} = \int_0^1 \int_0^1 f_{12}^2(x_1, x_2) dx_1 dx_2 \quad (\text{A.4})$$

with $f_1(x_1) \hat{=} f_1^{ANOVA}(x_1)$ and $f_{12}(x_1, x_2) \hat{=} f_{12}^{ANOVA}(x_1, x_2)$.

The overall variance D is just the sum of all partial variances. Because a second-order cut-HDMR expansion is used in this work, the sum includes only first- and second-order partial variances. This means, that higher order effects (\geq third-order), as far as they exist, are not considered.

A.2 Shifted Legendre Polynomials

The shifted Legendre Polynomials are orthogonal in the domain $[0, 1]$ and can be generated from the Rodrigues formula (Hochstrasser 1972):

$$P_k(x) = (k!)^{-1} \frac{d^k}{dx^k} ((x^2 - x)^k) \quad (\text{A.5})$$

or more conveniently from the sum formula

$$P_k(x) = \sum_{j=0}^k \left[(-1)^j \frac{(2k-j)!}{j! ((k-j)!)^2} \right] x^{k-j} \quad (\text{A.6})$$

where k is the order of the polynomial.

In order to obtain orthonormality the shifted Legendre polynomials have to be mul-

tiplied by the factor $\sqrt{2k+1}$. This leads to the following list of orthonormal polynomials, which provides the basis for the approximation of the RS-HDMR component functions:

$$\varphi_1(x) = \sqrt{3}(2x - 1) \quad (\text{A.7})$$

$$\varphi_2(x) = 6\sqrt{5} \left(x^2 - x + \frac{1}{6} \right) \quad (\text{A.8})$$

$$\varphi_3(x) = 20\sqrt{7} \left(x^3 - \frac{3}{2}x^2 + \frac{3}{5}x - \frac{1}{20} \right) \quad (\text{A.9})$$

$$\varphi_4(x) = 140\sqrt{9} \left(\frac{1}{2}x^4 - x^3 + \frac{9}{14}x^2 - \frac{1}{7}x + \frac{1}{140} \right) \quad (\text{A.10})$$

$$\varphi_5(x) = 630\sqrt{11} \left(\frac{2}{5}x^5 - x^4 + \frac{8}{9}x^3 - \frac{1}{3}x^2 + \frac{1}{21}x - \frac{1}{630} \right) \quad (\text{A.11})$$

$$\begin{aligned} \varphi_6(x) = & 3150\sqrt{13} \left(\frac{22}{75}x^6 - \frac{22}{25}x^5 + x^4 - \frac{8}{15}x^3 + \frac{2}{15}x^2 - \frac{1}{75}x \dots \right. \\ & \left. + \frac{1}{3150} \right) \end{aligned} \quad (\text{A.12})$$

$$\begin{aligned} \varphi_7(x) = & 16632\sqrt{15} \left(\frac{13}{63}x^7 - \frac{13}{18}x^6 + x^5 - \frac{25}{36}x^4 + \frac{25}{99}x^3 - \frac{1}{22}x^2 \dots \right. \\ & \left. + \frac{1}{297}x - \frac{1}{16632} \right) \end{aligned} \quad (\text{A.13})$$

$$\begin{aligned} \varphi_8(x) = & 84084\sqrt{17} \left(\frac{15}{98}x^8 - \frac{30}{49}x^7 + x^6 - \frac{6}{7}x^5 + \frac{75}{182}x^4 - \frac{10}{91}x^3 \dots \right. \\ & \left. + \frac{15}{1001}x^2 - \frac{6}{7007}x + \frac{1}{84084} \right) \end{aligned} \quad (\text{A.14})$$

$$\begin{aligned} \varphi_9(x) = & 420420\sqrt{19} \left(\frac{17}{147}x^9 - \frac{51}{98}x^8 + \frac{48}{49}x^7 - x^6 + \frac{3}{5}x^5 - \frac{3}{14}x^4 + \dots \right. \\ & \left. - \frac{4}{91}x^3 - \frac{3}{637}x^2 + \frac{3}{14014}x - \frac{1}{420420} \right) \end{aligned} \quad (\text{A.15})$$

$$\begin{aligned} \varphi_{10}(x) = & \sqrt{21} (184756x^{10} - 923780x^9 + 1969110x^8 - 2333760x^7 \dots \\ & + 1681680x^6 - 756756x^5 + 210210x^4 - 34320x^3 + 2970x^2 \dots \\ & - 110x + 1) \end{aligned} \quad (\text{A.16})$$

B GUI-HDMR Software Documentation

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 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 ($1 \dots N$) and the columns stand for the different inputs ($i = 1 \dots n$) 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' 1967) 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 assess the accuracy of the constructed metamodel the relative error (RE) is 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 sensitivity indices of first- and second-order can be calculated. Component functions of first- and second-order can also be plotted and compared with scatter plots.

B.1 Graphical User Interface (GUI)

After uncompressed the file `gui_hdmdr_software.zip` the graphical user interface of the GUI-HDMR software can be started by typing `gui_hdmdr` into the Matlab command window¹. The main window as shown in figure B.1 will open. From the main window calculations can be started and progress is shown. Three other windows can be accessed 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.

In case a previous HDMR analysis has been performed using the GUI-HDMR soft-

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

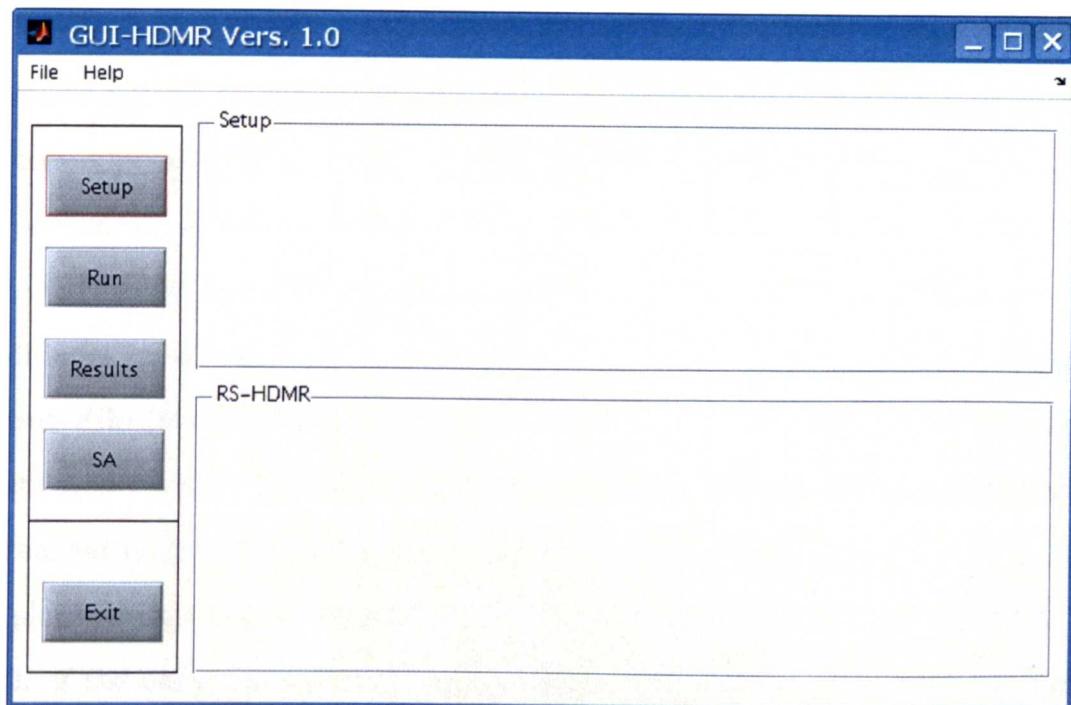


Figure B.1: GUI-HDMR main window.

ware 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.

B.1.1 Setup

From the main window the *Setup* window (figure B.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. In this case the metamodel is tested against values which were not used to construct 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**.

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, see also section B.1.3) and the scatter plots (see also section B.1.4).

The optimisation method (see section 4.1) which automatically calculates the optimal order for each of the polynomials is included in the GUI-HDMR software and one can set 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.

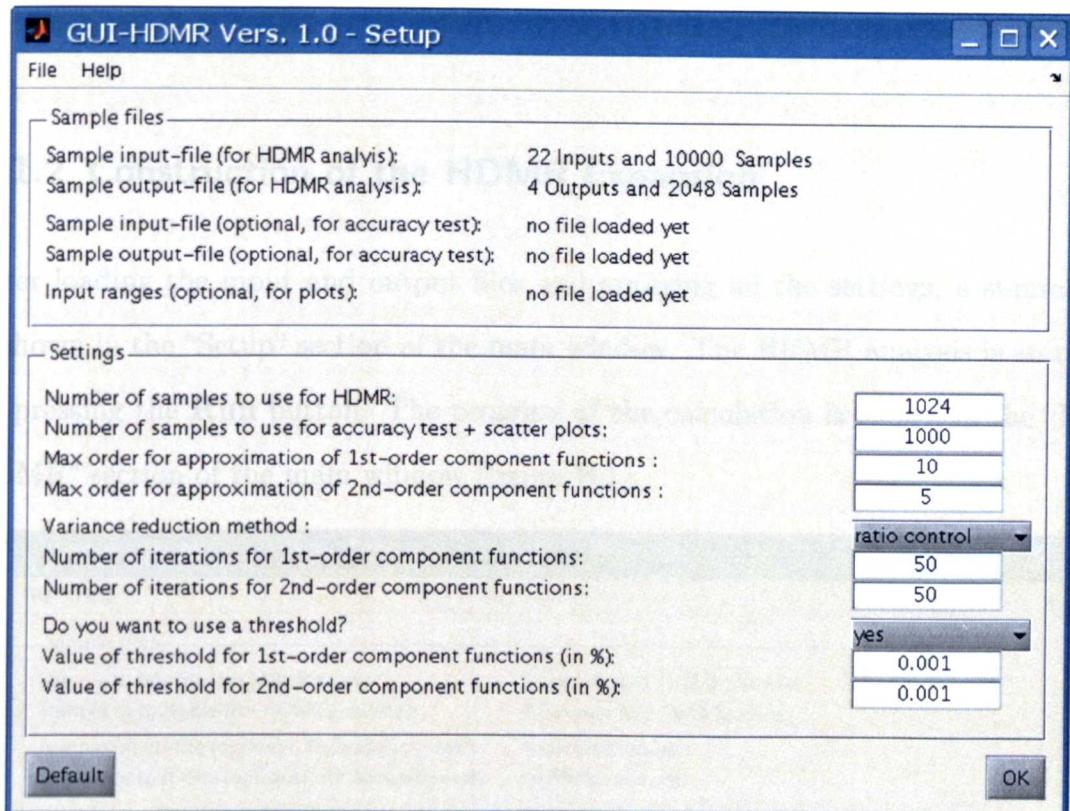


Figure B.2: GUI-HDMR setup window. Input and output files can be loaded and settings can be applied for the HDMR analysis.

A variance reduction method (see section 3.2.2) in order 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 5.1) 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.

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.

B.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 B.1).

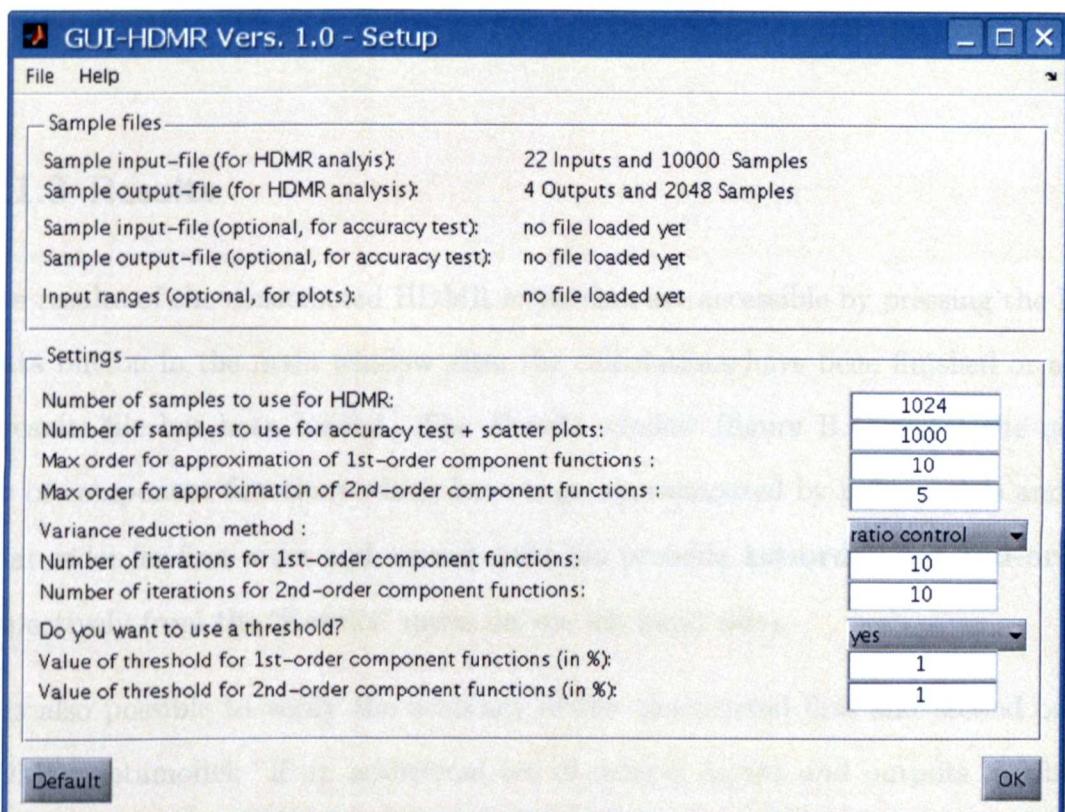


Figure B.3: GUI-HDMR main window. The summary of the settings is shown and the progress of the calculation.

If a variance reduction method has been applied, then the current number of the iteration process and the total number of iterations (figure B.3) is 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**. This will create a MAT-file called `workspace.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.

B.1.3 Results

The results of the constructed HDMR expansion are accessible 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 B.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. If an additional set of sample inputs and outputs has been provided (see section B.1.1), then the relative error is 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. 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) is calculated for the first order HDMR metamodel using the number of samples as defined in the setup (see section B.1.1). Clicking on **2nd-order** in the

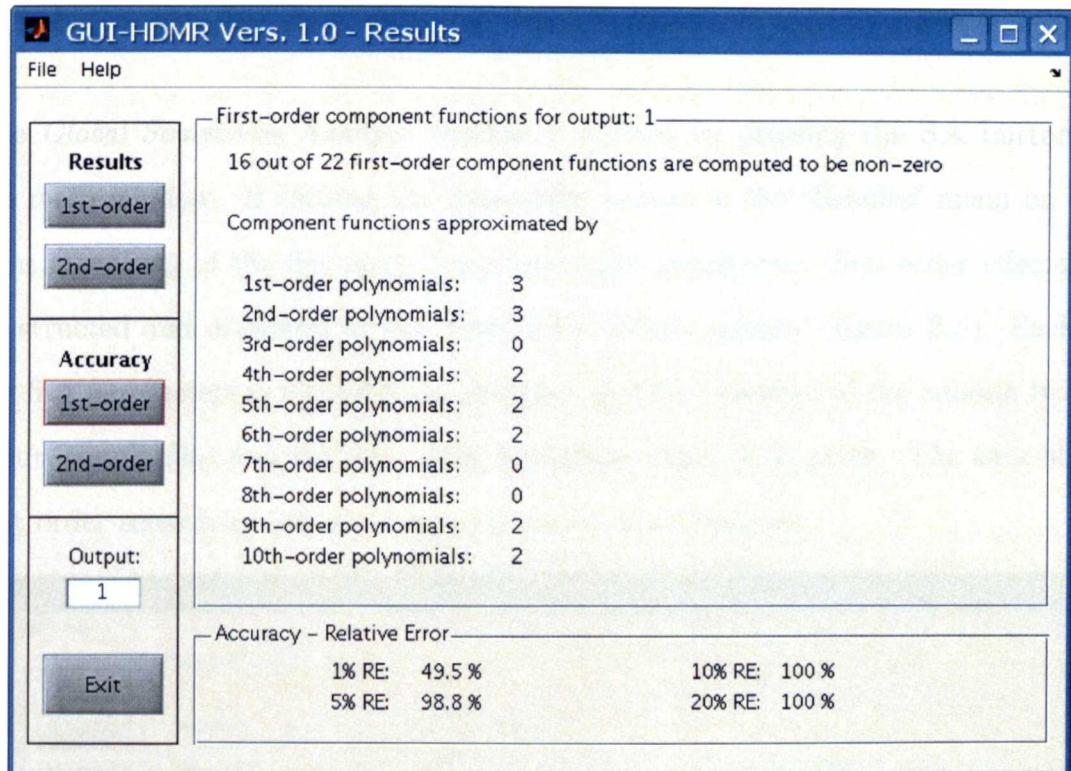


Figure B.4: GUI-HDMR results window. The optimal order is shown for the component functions which were approximated by polynomials. In order to assess the accuracy of the HDMR metamodel the relative error is also stated.

“Accuracy” menu provides the relative error for the second order HDMR metamodel. The relative error is given in the “Accuracy - Relative Error” section. It is stated how many of the tested samples (in %) are within a range of 1 %, 5 %, 10 % and 20 % of the relative error.

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 relative error are then shown for the selected output.

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

B.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 B.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.

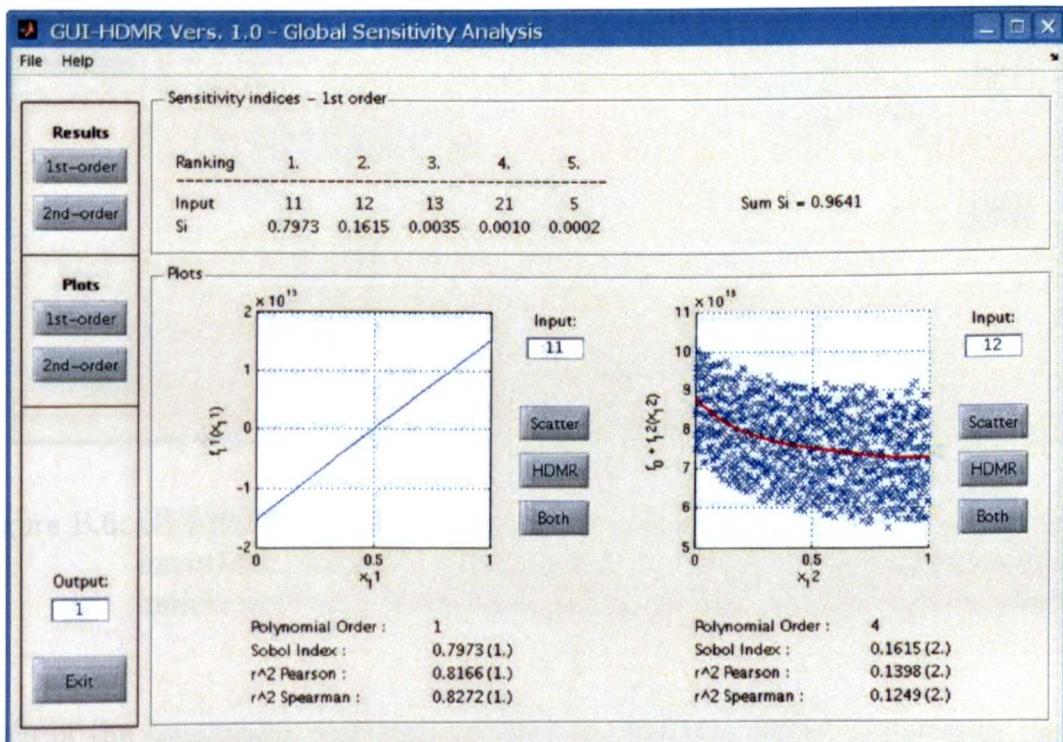


Figure B.5: GUI-HDMR sensitivity analysis window: The ranking of the five most important parameters based on the first order sensitivity indices is shown. Additionally all component functions can be plotted.

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 B.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.

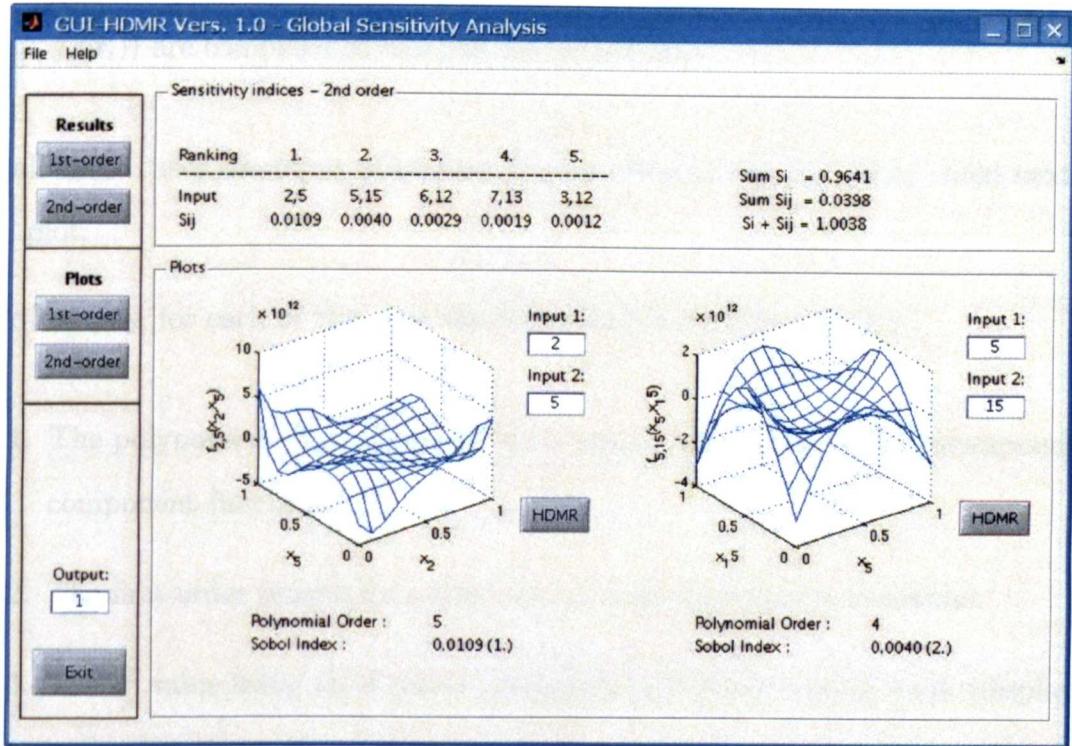


Figure B.6: GUI-HDMR sensitivity analysis window: The ranking of the five most important parameter interactions based on the second order sensitivity indices is shown. Additionally all component functions can be plotted.

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.

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 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 three different formats: Portable Network Graphics (png), Portable Document Format (pdf) and Encapsulated PostScript (eps). 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 three different versions with the endings .png, .pdf and .eps.

B.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.

B.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);
[approxist] = sub_comp_opt_1st(alpha,f0,opt_1st,sample_input,sample_output);
[opt_2nd]=sub_opt_order_2nd(beta,sample_input,sample_output,...  

    approxist,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.

B.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.

Bibliography

- Alzueta, M. U., Bilbao, R. & Glarborg, P. (2001), 'Inhibition and sensitization of fuel oxidation by SO₂', *Combustion and Flame* **127**(4), 2234–2251.
- Andria, G., D'Orazio, A., Lay Ekuakille, A. & Notarnicola, M. (2000), Modeling study in remote sensing of air pollution measured data, in 'Proceedings of the 17th IEEE Instrumentation and Measurement Technology Conference', Vol. 2, pp. 782–785.
- Atkinson, R., Baulch, D., Cox, R., Hampson, R., Kerr, J., Rossi, M. & Troe, J. (1997), 'Evaluated Kinetic and Photochemical Data for Atmospheric Chemistry. Supplement VI (Ox, HOx, NOx, SOx)', *Journal of Physical and Chemical Reference Data* **26**, 1329–1499.
- Atkinson, R., Baulch, D. L., Cox, R. A., Crowley, J. N., Hampson, R. F., Hynes, R. G., Jenkin, M. E., Rossi, M. J. & Troe, J. (2004), 'Evaluated kinetic and photochemical data for atmospheric chemistry: Volume I - gas phase reactions of Ox, HOx, NOx and SOx species.', *Atmospheric Chemistry and Physics* **4**, 1461–1738.
- Baker, J., Walker, H. L. & Cai, X. (2004), 'A study of the dispersion and transport of reactive pollutants in and above street canyons - a large eddy simulation', *Atmospheric Environment* **38**(39), 6883–6892.

- Balakrishnan, S., Roy, A., Ierapetritou, M. C., Flach, G. P. & Georgopoulos, P. G. (2005), 'A comparative assessment of efficient uncertainty analysis techniques for environmental fate and transport models: application to the FACT model', *Journal of Hydrology* **307**, 204–218.
- Baldwin, R. R. & Howarth, J. A. (1982), 'Surface termination in chain reactions and the interaction with homogeneous termination', *Journal of the Chemical Society, Faraday Transactions, 1* **78**, 451–464.
- Battin-Leclerc, F. (2008), 'Detailed chemical kinetic models for the low-temperature combustion of hydrocarbons with the application to gasoline and diesel fuel surrogates', *Progress in Energy and Combustion Science* **34**, 440–498.
- Bechmann, A., Sørensen, N. & Johansen, J. (2007), Atmospheric flow over terrain using hybrid RANS/LES, in 'Scientific proceedings, Ewec 2007'.
- Benson, J., Dixon, N. S., Ziehn, T. & Tomlin, A. S. (2007), The Sensitivity of a 3D Street Canyon CFD Model to Uncertainties in Input Parameters, in 'Proceedings of the 11th International Conference on Harmonisation within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO)', Vol. 1, pp. 63–65.
- Benson, J., Ziehn, T., Dixon, N. S. & Tomlin, A. S. (2008), 'Global sensitivity analysis of a 3D street canyon model - Part II: Application and physical insight using sensitivity analysis', *Atmospheric Environment* **42**, 1874–1891.
- Benson, S. W. (1976), *Thermochemical Kinetics. Methods for the Estimation of Thermochemical Data and Rate Parameters.*, John Wiley & Sons.
- Berkowicz, R. (2000), 'OSPM - A Parameterised Street Pollution Model', *Environmental Monitoring and Assessment* **65**(1-2), 323–331.
- Blitz, M. A., McKee, K. W. & Pilling, M. J. (2000), 'Temperature dependence of

- the reaction of OH with SO', *Proceedings of the Combustion Institute* **28**, 2491–2497.
- Boddy, J. W. D., Smalley, R. J., Goodman, P. S., Tate, J. E., Bell, M. C. & Tomlin, A. S. (2005), 'The spatial variability in concentrations of a traffic-related pollutant in two street canyons in York, UK Part I: The influence of background winds', *Atmospheric Environment* **39**(17), 3147–3161.
- Bonner, B. H. & Tipper, C. F. H. (1965), 'Cool flame combustion of hydrocarbons. I Cyclohexane', *Combustion and Flame* **19**, 317–327.
- Borgonovo, E. (2007), 'A new uncertainty importance measure', *Reliability Engineering & System Safety* **92**, 771–784.
- Box, G. E. P. (1954), 'The Exploration and Exploitation of Response Surfaces: Some General Considerations and Examples', *Biometrics* **10**(1), 16–60.
- Box, G. E. P. & Draper, N. R. (1987), *Empirical Model-Building and Response Surfaces*, John Wiley & Sons, Inc.
- Box, G. E. P. & Wilson, K. B. (1951), 'On the Experimental Attainment of Optimum Conditions', *Journal of the Royal Statistical Society. Series B (Methodological)* **13**(1), 1–45.
- Box, G. E. P., Hunter, W. G. & Hunter, J. S. (1978), *Statistics for Experimenters : An Introduction to Design, Data Analysis, and Model Building*, Wiley.
- Bratley, P., Fox, B. & Niederreiter, H. (1992), 'Implementation and tests of low-discrepancy sequences', *ACM Transactions on Modeling and Computer Simulation* **2**(3), 195–213.
- Brown, R. J. & Bilger, R. W. (1996), 'An experimental study of a reactive plume in grid turbulence', *Journal of Fluid Mechanics* **312**, 373–407.

Brunekreef, B. & Holgate, S. T. (2002), 'Air pollution and health', *The Lancet* **360**, 1233–1242.

Buda, F., Bounaceur, R., Warth, V., Glaude, P. A., Fournet, R. & Battin-Leclerc, F. (2005), 'Progress toward a unified detailed kinetic model for the autoignition of alkanes from C₄ to C₁₀ between 600 and 1200 K', *Combustion and Flame* **142**, 170–186.

Buda, F., Heyberger, B., Fournet, R., Glaude, P.-A., Warth, V. & Battin-Leclerc, F. (2006), 'Modeling of the Gas-Phase Oxidation of Cyclohexane', *Energy & Fuels* **20**(4), 1450 – 1459.

Burcat, A. (1984), Thermochemical Data for Combustion Calculations, in W. Gardiner, ed., 'Combustion Chemistry', Springer-Verlag, chapter 8.

Burcat, A. & McBride, B. (1997), 1997 Ideal Gas Thermodynamic Data for Combustion and Air-Pollution Use, Aerospace Engineering Report TAE 804, Technion Israel Institute of Technology.

Burcat, A. & Ruscic, B. (2005), Third Millennium Ideal Gas and Condensed Phase Thermochemical Database for Combustion with Updates from Active Thermochemical Tables, Technical Report ANL-05/20 TAE 960, Argonne National Laboratory.

Campolongo, F. & Braddock, R. (1999), 'The use of graph theory in the sensitivity analysis of the model output: a second order screening method', *Reliability Engineering & System Safety* **64**(1), 1–12.

Campolongo, F., Cariboni, J. & Saltelli, A. (2007), 'An effective screening design for sensitivity analysis of large models', *Environmental Modelling & Software* **22**(10), 1509–1518.

- Campolongo, F., Kleijnen, J. & Andres, T. (2000a), Screening Methods, in 'Sensitivity Analysis', John Wiley & Sons, Ltd.
- Campolongo, F., Saltelli, A., Sørenson, T. & Tarantola, S. (2000b), Hitchhiker's Guide to Sensitivity Analysis, in 'Sensitivity Analysis', John Wiley & Sons, Ltd.
- Campolongo, F., Tarantola, S. & Saltelli, A. (1999), 'Tackling quantitatively large dimensionality problems', *Computer Physics Communications* **117**, 75–85.
- Carslaw, D. C., Beevers, S. D. & Bell, M. C. (2007), 'Risks of exceeding the hourly EU limit value for nitrogen dioxide resulting from increased road transport emissions of primary nitrogen dioxide', *Atmospheric Environment* **41**, 2073–2082.
- Cassiani, M., Franzese, P. & Giostra, U. (2005), 'A PDF micromixing model of dispersion for atmospheric flow. Part I: development of the model, application to homogeneous turbulence and to neutral boundary layer', *Atmospheric Environment* **39**, 1457–1469.
- Cathonnet, M. (2003), Advances and challenges in the chemical kinetics of combustion, in 'Proceedings of the European Combustion Meeting'.
- Chan, K., Tarantola, S., Saltelli, A. & Sobol, I. M. (2000), Variance-Based Methods, in 'Sensitivity Analysis', John Wiley & Sons.
- Collett, R. S. & Oduyemi, K. (1997), 'Air quality modelling: a technical review of mathematical approaches', *Meteorological Applications* **4**, 235–246.
- Commission of the European Communities (2005), 'Proposal for a Directive of the European Parliament and of the Council on ambient air quality and cleaner air for Europe'.

- Cukier, R. I., Fortuin, C. M., Shuler, K. E., Petschek, A. G. & Schaibly, J. H. (1973), 'Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. I Theory', *Journal of Chemical Physics* 59(8), 3873–3878.
- Cukier, R. I., Schaibly, J. H. & Shuler, K. E. (1975), 'Study of the sensitivities of coupled reaction systems to uncertainties in rate coefficients. III. Analysis of the approximations', *The Journal of Chemical Physics* 63(3), 1140–1149.
- Daniel, C. (1973), 'One-at-a-Time Plans', *Journal of the American Statistical Association* 68(342), 353–360.
- Department for Environment, Food and Rural Affairs (DEFRA) (2007), 'The Air Quality Strategy for England, Scotland, Wales and Northern Ireland'.
- Dixon, N. S. & Tomlin, A. S. (2007), 'A Lagrangian stochastic model for predicting concentration fluctuations in urban areas', *Atmospheric Environment* 41, 8114–8127.
- Dixon, N. S., Boddy, J. W. D., Smalley, R. J. & Tomlin, A. S. (2006), 'Evaluation of a turbulent flow and dispersion model in a typical street canyon in York, UK', *Atmospheric Environment* 40(5), 958–972.
- Djurisic, Z. M., Langston, M. & Frenklach, M. (2007), Submission of Reaction Rates to PrIME, in 'Proceedings of the 5th U.S. National Combustion Meeting'.
- Downing, D. J., Gardner, R. H. & Hoffman, F. O. (1985), 'An Examination of Response-Surface Methodologies for Uncertainty Analysis in Assessment Models', *Technometrics* 27(2), 151–163.
- Du, S. (1997), 'Universality of the Lagrangian Velocity Structure Function Constant (c_0) Across Different Kinds of Turbulence', *Boundary-Layer Meteorology* 83, 207–219.

- Ehrhard, J., Khatib, I. A., Winkler, C., Kunz, R., Moussiopoulos, N. & Ernst, G. (2000), 'The microscale model MIMO: development and assessment', *Journal of Wind Engineering and Industrial Aerodynamics* **85**, 163–176.
- Eichhorn, J. (1996), Validation of a Microscale Pollution Dispersal Model, in 'Proceedings of the 21st International Meeting on Air Pollution Modeling and its Application'.
- Eichhorn, J. (2008), *MISKAM manual for version 5*, giese-eichhorn environmental meteorological software.
- El Bakali, A., Braun-Unkhoff, M., Dagaut, P., Frank, P. & Cathonnet, M. (2000), 'Detailed kinetic reaction mechanism for cyclohexane oxidation at pressure up to ten atmospheres', *Proceedings of the Combustion Institute* **28**(2), 1631–1638.
- Espenson, J. H. (1995), *Chemical Kinetics and Reaction Mechanisms*, second edn, McGraw-Hill.
- Fackrell, J. E. & Robins, A. G. (1982), 'Concentration fluctuations and fluxes in plumes from point sources in a turbulent boundary layer', *Journal of Fluid Mechanics* **117**, 1–26.
- Fairweather, M., Griffiths, J. F., Hughes, K. J., Pilling, M. J., Porter, R. & Tomlin, A. S. (2005), Reduced Kinetic Models for Different Classes of Problems, Technical Report Project No. EVG1-CT-2002-00072, Deliverable 38, SAFEKINEX: SAFe and Efficient hydrocarbon oxidation processes by KINETics and Explosion eXpertise EU 5th Framework Programme on Energy, Environment and Sustainable Development.
- Fang, S. and Gertner, G. Z., Shinkareva, S., Wang, G. & Anderson, A. (2003), 'Improved generalized Fourier amplitude sensitivity test (FAST) for model assessment', *Statistics and Computing* **13**, 221–226.

- Faure, H. (1992), 'Good permutations for extreme discrepancy', *Journal of Number Theory* **42**, 47–56.
- Firestone, M., Fenner-Crisp, P., Barry, T., Bennett, T., Chang, S. & Callhan, M. (1997), Guiding Principles for Monte Carlo Analysis, Technical Report EPA/630/R-97/001, U.S. Environmental Protection Agency.
- Franzese, P. (2003), 'Lagrangian stochastic modeling of a fluctuating plume in the convective boundary layer', *Atmospheric Environment* **37**, 1691–1701.
- Galanti, S. & Jung, A. (1997), 'Low-discrepancy sequences: Monte Carlo simulation of option prices', *Journal of Derivatives* **5**, 63–83.
- Glarborg, P., Kubel, D., Dam-Johansen, K., Chiang, H.-M. & Bozzelli, J. W. (1996), 'Impact of SO₂ and NO on CO oxidation under post-flame conditions', *International Journal of Chemical Kinetics* **28**(10), 773–790.
- Gordon, S. & McBride, B. J. (1971), Computer Program for Calculation of Complex Chemical Equilibrium Composition, Rocket Performance, Incident and Reflected Shocks and Chapman-Jouguet Detonations, Technical Report SP-273, NASA.
- Halton, J. H. (1960), 'On the efficiency of certain quasi-random sequences of points in evaluating multi-dimensional integrals', *Numerische Mathematik* **2**, 84–90.
- Hamby, D. M. (1994), 'A review of techniques for parameter sensitivity analysis of environmental models', *Environmental Monitoring and Assessment* **32**(2), 135 – 154.
- Hampartsoumian, E., Nimmo, W., Tomlin, A. S., Hughes, K. J. & Mahmud, T. (1998), 'Studies of sulphur-nitrogen interactions in turbulent-spray flames', *Journal of the Institute of Energy* **71**(488), 137–144.

- Helton, J. (1993), 'Uncertainty and sensitivity analysis techniques for use in performance assessment for radioactive waste disposal', *Reliability Engineering & System Safety* 42, 327–367.
- Helton, J. C. & Davis, F. J. (2000a), Sampling-Based Methods, in A. Saltelli, K. Chan & E. Scott, eds, 'Sensitivity Analysis', John Wiley & Sons, chapter 6, pp. 101–153.
- Helton, J. C. & Davis, F. J. (2000b), Sampling-Based Methods for Uncertainty and Sensitivity Analysis, Sandia report sand99-2240, Sandia National Laboratories.
- Helton, J. C. & Davis, F. J. (2002), 'Illustration of Sampling-Based Methods for Uncertainty and Sensitivity Analysis', *Risk Analysis* 22(3), 591–622.
- Helton, J. C. & Davis, F. J. (2003), 'Latin hypercube sampling and the propagation of uncertainty in analyses of complex systems', *Reliability Engineering & System Safety* 81(1), 23–69.
- Helton, J., Johnson, J., Sallaberry, C. & Stolie, C. (2006), 'Survey of sampling-based methods for uncertainty and sensitivity analysis', *Reliability Engineering & System Safety* 91, 1175–1209.
- Hochstrasser, U. W. (1972), *Handbook of Mathematical Functions*, 9th edn, Dover Publications, Inc., New York, chapter 22, pp. 771–802.
- Homma, T. & Saltelli, A. (1996), 'Importance measures in global sensitivity analysis of nonlinear models', *Reliability Engineering & System Safety* 52, 1–17.
- Horng, I. R. & Chou, J. H. (1986), 'Design of Optimal Observers with Specified Eigenvalues via Shifted Legendre Polynomials', *Journal of Optimization Theory and Applications* 51(1), 179–188.

- Hughes, K. J., Fairweather, M., Griffiths, J. F., Porter, R. & Tomlin, A. S. (2008), 'The application of the QSSA via reaction lumping for the reduction of complex hydrocarbon oxidation mechanisms', *Proceedings of the Combustion Institute* **28**, in press.
- Hughes, K. J., Griffiths, J. F., Fairweather, M. & Tomlin, A. S. (2006), 'Evaluation of models for the low temperature combustion of alkanes through interpretation of pressure-temperature ignition diagrams', *Physical Chemistry Chemical Physics* **8**, 3197 – 3210.
- Hughes, K. J., Tomlin, A. S., Dupont, V. & Pourkashanian, M. (2001a), 'Experimental and modelling study of sulfur and nitrogen doped premixed methane flames at low pressure', *Faraday Discussions* **119**, 337–352.
- Hughes, K. J., Tomlin, A. S., Hampartsoumian, E., Nimmo, W., Zsély, I. G., Ujvári, M., Turányi, T., Clague, A. R. & Pilling, M. J. (2001b), 'An investigation of important gas phase reactions of nitrogen species from the simulation of bulk experimental data in combustion systems', *Combustion and Flame* **124**, 573–589.
- Hughes, K. J., Turányi, T., Clague, A. R. & Pilling, M. J. (2001c), 'Development and Testing of a Comprehensive Chemical Mechanism for the Oxidation of Methane', *International Journal of Chemical Kinetics* **33**, 513–538.
- Hunter, L. J., Johnson, G. T. & Watson, I. D. (1992), 'An investigation of three-dimensional characteristics of flow regimes within the urban canyon', *Atmospheric Environment. Part B. Urban Atmosphere* **26**(4), 425–432.
- 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 Modeling and Analysis', pp. 398–403.

- Jackson, J. L. (1951), 'Spontaneous Ignition Temperatures', *Industrial & Engineering Chemistry* **43**(12), 2869–2870.
- Jakeman, A. J., Letcher, R. A. & Norton, J. P. (2006), 'Ten iterative steps in development and evaluation of environmental models', *Environmental Modelling & Software* **21**(5), 602–614.
- Kalos, M. H. & Whitlock, P. (1986), *Monte Carlo Methods*, John Wiley & Sons.
- Kee, R. J., Grcar, J. F., Smooke, M. D. & Miller, J. A. (1985), A Fortran Program for Modeling Steady Laminar One-Dimensional Premixed Flames, Technical Report SAND85-8240, Sandia National Laboratories.
- Kee, R., Rupley, F. & Miller, J. (1991), Chemkin-II: A Fortran Chemical Kinetics Package for the Analysis of Gas Phase Chemical Kinetics, Technical Report SAND89-8009B, Sandia National Laboratories.
- Khuri, A. I. & Cornell, J. A. (1996), *Response Surface: Design and Analysis*, Vol. 152 of *Statistics: textbooks and monographs*, Marcel Dekker, Inc.
- Klai, S. E. & Baronnet, F. (1993), 'Study of the homogeneous gas-phase oxidation of cyclohexane. I: Experimental study', *Journal de chimie physique* **90**(10), 1929–1950.
- Krzyzanowski, M., Kuna-Dibbert, B. & Schneider, J., eds (2005), *Health effects of transport-related air pollution*, WHO Regional Office for Europe.
- 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)', pp. 35–36.

- Lemaire, O., Ribaucour, M., Carlier, M. & Minetti, R. (2001), 'The production of benzene in the low-temperature oxidation of cyclohexane, cyclohexene, and cyclohexa-1,3-diene', *Combustion and Flame* **127**(1-2), 1971–1980.
- Li, G. & Rabitz, H. (2006), 'Ratio control variate method for efficiently determining high-dimensional model representations', *Journal of Computational Chemistry* **27**(10), 1112 – 1118.
- Li, G., Artamonov, M., Rabitz, H., Wang, S.-W., Georgopoulos, P. C. & Demiralp, M. (2003a), 'High-Dimensional Model Representations Generated from Low Order Terms - lp-RS-HDMR', *Journal of Computational Chemistry* **24**(5), 647 – 656.
- Li, G., Hu, J., Wang, S.-W., Georgopoulos, P. G., Schoendorf, J. & Rabitz, H. (2006), 'Random Sampling-High Dimensional Model Representation (RS-HDMR) and Orthogonality of Its Different Order Component Functions', *Journal of Physical Chemistry A* **110**(7), 2474 –2485.
- Li, G., Rabitz, H., Wang, S.-W. & Georgopoulos, P. G. (2003b), 'Correlation method for variance reduction of Monte Carlo integration in RS-HDMR', *Journal of Computational Chemistry* **24**(3), 277 – 283.
- Li, G., Rosenthal, C. & Rabitz, H. (2001a), 'High Dimensional Model Representations', *Journal of Physical Chemistry A* **105**(33), 7765 –7777.
- Li, G., Schoendorf, J., Ho, T.-S. & Rabitz, H. (2004), 'Multicut-HDMR with an application to an ionospheric model', *Journal of Computational Chemistry* **25**(9), 1149 – 1156.
- Li, G., Wang, S.-H. & Rabitz, H. (2000), 'High Dimensional Model Representations (HDMR): Concepts and Applications'. IMA Workshop: Atmospheric Modeling.

- Li, G., Wang, S.-W. & Rabitz, H. (2002a), 'Practical Approaches To Construct RS-HDMR Component Functions', *Journal of Physical Chemistry A* **106**(37), 8721–8733.
- Li, G., Wang, S.-W., Rabitz, H., Wang, S. & Jaffé, P. (2002b), 'Global uncertainty assessments by high dimensional model representations (HDMR)', *Chemical Engineering Science* **57**(21), 4445–4460.
- Li, G., Wang, S.-W., Rosenthal, C. & Rabitz, H. (2001b), 'High Dimensional Model Representations Generated from Low Dimensional Data Samples. I. mp-Cut-HDMR', *Journal of Mathematical Chemistry* **30**(1), 1 – 30.
- Lien, R.-C. & D'Asaro, E. A. (2002), 'The Kolmogorov constant for the Lagrangian velocity spectrum and structure function', *Physics of Fluids* **14**, 4456–4459.
- Lightfoot, P. D., Cox, R. A., Crowley, R. A., Destriau, M., Hayman, G. D., Jenkin, M. E., Moortgat, G. K. & Zabel, F. (1992), 'Organic peroxy radicals: Kinetics, spectroscopy and tropospheric chemistry', *Atmospheric Environment* **26**(10), 1805–1961.
- Lindstedt, R. P. & Maurice, L. Q. (1995), 'Detailed Kinetic Modelling of n-heptane combustion', *Combustion Science and Technology* **107**, 317 – 353.
- Liu, C.-H. & Barth, M. C. (2002), 'Large-Eddy Simulation of Flow and Scalar Transport in a Modeled Street Canyon', *Journal of Applied Meteorology* **41**(6), 660–673.
- Louka, P., Belcher, S. E. & Harrison, R. G. (2000), 'Coupling between air flow in streets and the well-developed boundary layer aloft', *Atmospheric Environment* **34**(16), 2613–2621.
- MacDougall, M., Smith, R. I. & Scott, E. M. (2005), 'Comprehensive sensitivity

- analysis of an SO₂ deposition model for three measurement sites: Consequences for SO₂ deposition fluxes', *Atmospheric Environment* **39**(28), 5025–5039.
- Manomaiphiboon, K. & Russell, A. G. (2004), 'Effects of uncertainties in parameters of a Lagrangian particle model on mean ground-level concentrations under stable conditions', *Atmospheric Environment* **38**(33), 5529–5543.
- Marinov, M. N. (1998), 'A detailed Chemical Kinetic Model for High Temperature Ethanol Oxidation', *International Journal of Chemical Kinetics* **31**, 183–220.
- McEnally, C. S. & Pfefferle, L. D. (2005), 'Fuel decomposition and hydrocarbon growth processes for substituted cyclohexanes and for alkenes in nonpremixed flames', *Proceedings of the Combustion Institute* **30**(1), 1425–1432.
- McKay, M. D., Conover, W. J. & Beckman, R. J. (1979), 'A Comparison of Three Methods for Selecting Values of Input Variables in the Analysis of Output from a Computer Code', *Technometrics* **21**, 239–245.
- Mendenhall, W., Beaver, R. J. & Beaver, B. M. (2003), *Introduction to Probability and Statistics*, Brooks/Cole.
- Moin, P. & Mahesh, K. (1998), 'DIRECT NUMERICAL SIMULATION: A Tool in Turbulence Research', *Annual Review of Fluid Mechanics* **30**, 539–578.
- Morokoff, W. J. & Caflisch, R. E. (1995), 'Quasi-Monte Carlo Integration', *Journal of Computational Physics* **122**, 218–230.
- Morris, M. D. (1991), 'Factorial Sampling Plans for Preliminary Computational Experiments', *Technometrics* **33**(2), 161–174.
- Myers, R. H., Khuri, A. I. & Carter, W. H. (1989), 'Response Surface Methodology: 1966 - 1988', *Technometrics* **31**, 137–157.

Nakamura, Y. & Oke, T. R. (1988), 'Wind, temperature and stability conditions in an east-west oriented urban canyon', *Atmospheric Environment* **22**(12), 2691–2700.

Niederreiter, H. (1987), 'Point sets and sequences with small discrepancy', *Monatshefte für Mathematik* **104**, 273–337.

Niederreiter, H. (1988), 'Low-discrepancy and low-dispersion sequences', *Journal of Number Theory* **30**, 51–70.

Niederreiter, H. (1992), *Random Number Generation and Quasi-Monte Carlo Methods*, Vol. 63 of *CBMS-NSF Regional Conference Series in Applied Mathematics*, Society for Industrial and Applied Mathematics.

Nimmo, W., Hampartsoumian, E., Hughes, K. J. & Tomlin, A. S. (1998), 'Experimental and kinetic studies on the effect of sulphur-nitrogen interactions on NO formation in flames', *Proceedings of the Combustion Institute* **27**, 1419–1426.

Oke, T. R. (1987), *Boundary Layer Climates*, Routledge.

Oke, T. R. (1988), 'Street design and urban canopy layer climate', *Energy and Buildings* **11**, 103–113.

Pavageau, M. & Schatzmann, M. (1999), 'Wind tunnel measurements of concentration fluctuations in an urban street canyon', *Atmospheric Environment* **33**, 3961–3971.

Peebles, L. R. & Marshall, P. (2002), 'A coupled-cluster study of the enthalpy of formation of nitrogen sulfide, NS', *Chemical Physics Letters* **366**, 520–524.

Phenix, B. D., Dinaro, J. L., Tatang, M. A., Tester, J. T., Howard, J. B. & McRae, G. J. (1998), 'Incorporation of parametric uncertainty into complex kinetic

- mechanisms: Application to hydrogen oxidation in supercritical water', *Combustion and Flame* **112**, 132–146.
- Pilling, M. J. (2008), 'From elementary reactions to evaluated chemical mechanisms for combustion models', *Proceedings of the Combustion Institute*.
- Pilling, M. J., ed. (1997), *Low-Temperature Combustion and Autoignition*, Elsevier.
- Porter, R. T. J. (2007), Kinetic Mechanism Recuction for Chemical Process Hazard Application, PhD thesis, University of Leeds.
- Rabitz, H. & Aliş, O. F. (1999), 'General foundations of high-dimensional model representations', *Journal of Mathematical Chemistry* **25**, 197 – 233.
- Rabitz, H. & Aliş, O. F. (2000), Managing the Tyranny of Parameters in Mathematical Modelling of Physical Systems, in A. Saltelli, K. Chan & E. Scott, eds, 'Sensitivity Analysis', John Wiley & Sons, chapter 9, pp. 199–224.
- Rabitz, H., Aliş, O. F., Shorter, J. & Shim, K. (1999), 'Efficient input–output model representations', *Computer Physics Communications* **117**(1-2), 11–20.
- Rabitz, H., Kramer, M. & Dacol, D. (1983), 'Sensitivity Analysis in Chemical Kinetics', *Annual Review of Physical Chemistry* **34**, 419–461.
- Ratto, M. (2008), 'Analysing DSGE Models with Global Sensitivity Analysis', *Computational Economics* **31**(2), 115–139.
- Ratto, M., Pagano, A. & Young, P. (2007), 'State Dependent Parameter meta-modelling and sensitivity analysis', *Computer Physics and Communications* **177**, 863–876.
- Reynolds, A. M. (1998), 'Comments on the 'Universality of the Lagrangian velocity structure function constant (c_0) across different kinds of turbulence'', *Boundary-Layer Meteorology* **89**, 161–170.

- Ries, K. & Eichhorn, J. (2001), 'Simulation of effects of vegetation on the dispersion of pollutants in street canyons', *Meteorologische Zeitschrift* **10**, 229–233.
- Rizza, U., Mangia, C., Carvalho, J. C. & Anfossi, D. (2006), 'Estimation of the Lagrangian Velocity Structure Function Constant C0 by Large-Eddy Simulation', *Boundary-Layer Meteorology* **120**, 25–37.
- Ro, K. S. & Hunt, P. G. (2007), 'Characteristic Wind Speed Distributions and Reliability of the Logarithmic Wind Profile', *Journal of Environmental Engineering* **133**(3), 313–318.
- Robert, C. P. & Casella, G. (2004), *Monte Carlo Statistical Methods*, second edn, Springer, chapter 3, pp. 79–122.
- Rosenbrock, H. H. (1963), 'Some General Implicit Processes for the Numerical Solution of Differential Equations', *The Computer Journal* **5**, 329–330.
- Rowley, D. M., Lesclaux, R., Lightfoot, P. D., Noziere, B., Wallington, T. J. & Hurley, M. . D. (1992), 'Kinetic and mechanistic studies of the reactions of cyclopentylperoxy and cyclohexylperoxy radicals with HO₂', *Journal of physical chemistry* **96**(12), 4889–4894.
- Sacks, J., Schiller, S. B. & Welch, W. J. (1989a), 'Designs for Computer Experiments', *Technometrics* **31**(1), 41–47.
- Sacks, J., Welch, W. J., Mitchell, T. J. & Wynn, H. P. (1989b), 'Design and Analysis of Computer Experiments', *Statistical Science* **4**(4), 409–423.
- Saltelli, A. (2000), What is Sensitivity Analysis?, in 'Sensitivity Analysis', John Wiley & Sons.

- Saltelli, A. & Sobol', I. M. (1995), 'About the use of rank transformation in sensitivity analysis of model output', *Reliability Engineering and System Safety* **50**, 225–239.
- Saltelli, A., Chan, K. & Scott, A. M., eds (2000), *Sensitivity Analysis*, John Wiley & Sons, LTD.
- Saltelli, A., Tarantola, S. & Chan, K. P.-S. (1999), 'A Quantitative Model-Independent Method for Global Sensitivity Analysis of Model Output', *Technometrics* **41**, 39–56.
- Saunders, A. M., Jenkin, M. E., Derwent, R. G. & Pilling, M. J. (2003), 'Protocol for the development of the Master Chemical Mechanism, MCM v3 (Part A): tropospheric degradation of non-aromatic volatile organic compounds', *Atmospheric Chemistry and Physics* **3**, 161–180.
- Sawford, B. L. (2004), 'Micro-Mixing Modelling of Scalar Fluctuations for Plumes in Homogeneous Turbulence', *Flow, Turbulence and Combustion* **72**, 133–160.
- Schaibly, J. H. & Shuler, K. E. (1973), 'Study of the sensitivity of coupled reaction systems to uncertainties in rate coefficients. II Applications', *Journal of Chemical Physics* **59**(8), 3879–3888.
- Seinfeld, J. H. & Pandis, S. N. (1998), *Atmospheric Chemistry and Physics*, John Wiley & Sons.
- Shetter, R. E., Cantrell, C. A., Lantz, K. O., Flocke, S. J., Orlando, J. J., Tyndall, G. S., Gilpin, T. M., Fischer, C. A., Madronich, S., Calvert, J. G. & Junkermann, W. (1996), 'Actinometric and radiometric measurement and modelling of the photolysis rate coefficient of ozone to O(1D) during Mauna Loa Observatory Photochemistry Experiment 2', *Journal of Geophysical Research* **101**, 14631–14641.

- Shetter, R., McDaniel, A., Cantrell, C., Madronich, S. & Calvert, J. (1992), 'Actinometer and Eppley Radiometer Measurement of NO₂ Photolysis Rate Coefficient During the Mauna Loa Observatory Photochemistry Experiment', *Journal of Geophysical Research* **97**, 10349–10359.
- Silke, E. J., Pitz, W. J., Westbrook, C. K. & Ribaucour, M. (2007), 'Detailed Chemical Kinetic Modeling of Cyclohexane Oxidation', *The Journal of Physical Chemistry A* **111**(19), 3761 – 3775.
- Sini, J.-F., anquetin, S. & Mestayer, P. G. (1996), 'Pollutant dispersion and thermal effects in urban street canyons', *Atmospheric Environment* **30**, 2659–2677.
- Snee, T. J. & Griffiths, J. F. (1989), 'Criteria for spontaneous ignition in exothermic, autocatalytic reactions: Chain branching and self-heating in the oxidation of cyclohexane in closed vessels', *Combustion and Flame* **75**(3-4), 381–395.
- Sobol', I. M. (1967), 'On the distribution of points in a cube and the approximate evaluation of integrals', *USSR Computational Mathematics and Mathematical Physics* **7**(4), 86–112.
- Sobol', I. M. (1993), 'Sensitivity analysis for non-linear mathematical models', *Mathematical Modeling & Computational Experiment* **1**, 407–414. Englisch translation of Russian original paper Sobol' (1990).
- Sobol', I. M. (2001), 'Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates', *Mathematics and Computers in Simulation* **55**, 271 – 280.
- Sobol', I. M. (2003), 'Theorems and examples on high dimensional model representation', *Reliability Engineering & System Safety* **79**(2), 187–193.
- Sobol', I. M. & Kucherenko, S. S. (2005), 'Global Sensitivity Indices for Nonlinear Mathematical Models. Review', *Wilmott magazine* **2**, 2–7.

- Stull, R. B. (1988), *An Introduction to Boundary Layer Meteorology*, Kluwer Academic Publishers.
- Thomson, D. J. (1987), 'Criteria for the selection of stochastic models of particle trajectories in turbulent flows', *Journal of Fluid Mechanics* **180**, 529–556.
- Tomlin, A. S. (2006), 'The use of global uncertainty methods for the evaluation of combustion mechanisms', *Reliability Engineering & System Safety* **91**(10-11), 1219–1231.
- Tomlin, A. S., Turányi, T. & Pilling, M. J. (1997), Mathematical tools for the construction, investigation and reduction of combustion mechanisms, in M. J. Pilling, ed., 'Low Temperature Combustion and Autoignition', Elsevier.
- Turányi, T. (1990), 'Sensitivity analysis of complex kinetic systems. Tools and applications', *Journal of Mathematical Chemistry* **5**(3), 1572–8897.
- Turányi, T. (1997), 'Applications of sensitivity analysis to combustion chemistry', *Reliability Engineering & System Safety* **57**(1), 41–48.
- Turányi, T. & Rabitz, H. (2000), Local methods, in A. Saltelli, K. Chan & E. M. Scott, eds, 'Sensitivity Analysis', Wiley, New York.
- Turns, S. R. (2000), *An introduction to combustion: Concepts and Applications*, McGraw-Hill.
- Upahyay, S. K. (2006), *Chemical Kinetics and Reaction Dynamics*, Springer.
- van Thuijl, E., Roos, C. J. & Beurskens, L. W. M. (2003), An Overview of Biofuel Technologies, Markets and Policies in Europe, Technical report, Energy research Centre of the Netherlands.

- Vardoulakis, S., Fisher, B. E. A., Gonzalez-Flesca, N. & Pericleous, K. (2002), 'Model sensitivity and uncertainty analysis using roadside air quality measurements', *Atmospheric Environment* **36**(13), 2121–2134.
- Vardoulakis, S., Fisher, B. E. A., Pericleous, K. & Gonzalez-Flesca, N. (2003), 'Modelling air quality in street canyons - a review', *Atmospheric Environment* **37**, 155–182.
- Wang, M. L. & Chang, R. Y. (1985), 'Optimal Control of Lumped Parameter Systems via Shifted Legendre Polynomial Approximation', *Journal of Optimization Theory and Applications* **45**(2), 313–324.
- Wang, S.-W. & Georgopoulos, P. G. (2003), 'Random Sampling-High Dimensional Model Representation (RS-HDMR) with Nonuniformly Distributed Variables: Application to an Integrated Multimedia/Multipathway Exposure and Dose Model for Trichloroethylene', *Journal of Physical Chemistry A* **107**, 4707 – 4716.
- Wang, S. W., Balakrishnan, S. & Georgopoulos, P. (2005), 'Fast equivalent operational model of tropospheric alkane photochemistry', *AIChE Journal* **51**, 1297–1303.
- Wang, S. W., Georgopoulos, P. G., Li, G. & Rabitz, H. (2001), 'Computationally Efficient Atmospheric Chemical Kinetic Modeling by Means of High Dimensional Model Representation (HDMR)', *Lecture Notes in Computer Science* **2179**, 326–333.
- Wilson, J. D. & Sawford, B. L. (1996), 'Review of Lagrangian stochastic models for trajectories in the turbulent atmosphere', *Boundary-Layer Meteorology (Historical Archive)* **78**, 191 – 210.

World Health Organization (2000), *Air Quality Guidelines for Europe*, number 91
in 'European Series', second edn, WHO Regional Publications.

Ziehn, T. & Tomlin, A. S. (2007a), A global sensitivity study of sulphur chemistry
in a premixed methane flame model using HDMR, *in 'Proceedings of the 5th
International Conference on Sensitivity Analysis of Model Output (SAMO)',*
pp. 69–70.

Ziehn, T. & Tomlin, A. S. (2007b), Efficient Methods for Assessing Uncertainties
and Sensitivities in Environmental Models, *in 'Proceedings of the 11th Interna-
tional Conference on Harmonisation within Atmospheric Dispersion Modelling
for Regulatory Purposes (HARMO)', Vol. 2, pp. 73–75.*

Ziehn, T. & Tomlin, A. S. (2008a), 'A global sensitivity study of sulphur chemistry
in a premixed methane flame model using HDMR', *International Journal of
Chemical Kinetics* **40**(11), 742–753.

Ziehn, T. & Tomlin, A. S. (2008b), 'Global sensitivity analysis of a 3D street canyon
model - Part I: The development of high dimensional model representations',
Atmospheric Environment **42**, 1857–1873.

Ziehn, T. & Tomlin, A. S. (2008c), 'GUI-HDMR - A Software Tool for Global
Sensitivity Analysis of Complex Models', *Environmental Modelling & Software*.
doi:10.1016/j.envsoft.2008.12.002.

Ziehn, T., Dixon, N. S. & Tomlin, A. S. (2008a), The effects of parametric uncer-
tainties in simulations of a reactive plume using a lagrangian stochastic model
layout, *in 'Proceedings of the 12th International Conference on Harmonisation
within Atmospheric Dispersion Modelling for Regulatory Purposes (HARMO)'*.

Ziehn, T., Hughes, K. J., Griffiths, J. F., Porter, R. & Tomlin, A. S. (2008b), 'A
global sensitivity study of cyclohexane oxidation under low temperature fuel

rich conditions using HDMR methods', Presented at the 32nd International Symposium on Combustion.

Ziehn, T., Hughes, K. J., Griffiths, J. F., Porter, R. & Tomlin, A. S. (2008c), 'A global sensitivity study of cyclohexane oxidation under low temperature fuel rich conditions using HDMR methods', *Combustion Theory and Modelling*. under review.

Zsély, I. G., Kramarics, A., Ruscic, B. & Turányi, T. (2008), 'Correlation of the enthalpies of formation: impact on the uncertainty of combustion models', Presented at the 32nd International Symposium on Combustion.